

# UC Riverside

## UCR Honors Capstones 2017-2018

### Title

A Personal Take on "Outbreak! Modeling the Ebola Epidemic"

### Permalink

<https://escholarship.org/uc/item/59v8j60t>

### Author

Toomey, Alysha

### Publication Date

2018-04-01

By

A capstone project submitted for  
Graduation with University Honors

University Honors  
University of California, Riverside

APPROVED

---

Dr.  
Department of

---

Dr. Richard Cardullo, Howard H Hays Jr. Chair and Faculty Director, University Honors  
Interim Vice Provost, Undergraduate Education

## **Abstract**

## Acknowledgments

## Table of Contents

|                      |     |
|----------------------|-----|
| Abstract.....        | ii  |
| Acknowledgments..... | iii |

## **The 2014 Ebola Epidemic in West Africa**

The journey and symptoms of Ebola are a nightmare. According to the World Health Organization, Ebola is a deadly virus that causes severe bleeding and organ failure in a matter of a couple weeks. This nightmare became a reality for two year old Emile Ouamouno, and eventually most of West Africa, in December 2013. After likely contracting the disease from a fruit bat, young Emile was declared as Patient Zero of what became the 2014 Ebola epidemic in West Africa. Emile spread Ebola to many members of his family, which in turn spread across West Africa (Gholipour). The Centers for Disease Control and Prevention (CDC) states the Ebola outbreak was declared on March 23, 2014, with the outbreak reaching epidemic proportions a few months later. The epidemic lasted from August 8, 2014 until March 2016. According to the CDC, during this time there were approximately 28,600 total cases, in which 11,325 of these cases resulted in the death of the individual.

There are many aspects that contributed to the outbreak becoming an epidemic. West Africa had never experienced an Ebola outbreak before so they were not prepared. For example, there was a severe shortage of healthcare workers. Prior to the outbreak there were only one to two doctors per 100,000 people. Also, West Africa's cultural beliefs and behavioral practices led to the spread of Ebola. In West Africa, compassion is a cultural trait so people wanted to care for the sick. Since Ebola is transferred by direct contact with bodily fluids and objects that have been contaminated with bodily fluids, this compassion contributed to the spread of Ebola. Once the infected people died, they were still able to infect the living until they were buried. Because of this, funerals were the place of origin for a large number of new cases. ("Origins of the 2014 Ebola epidemic")

Due to the severity of the Ebola virus, there have been many attempts to predict how a future epidemic may be prevented once an outbreak takes place. Also, if an outbreak were to become an epidemic, how to predict the impact of the virus on the population is another important project. One method in dealing with these unknowns is mathematical modeling. More specifically, epidemiological models are systems of equations that can be manipulated in order to model the particular circumstances involved. With undergraduate students Jennifer Rangel Ambriz, Andrew Whitaker, Jaime Madrigal, and Emily Ortego, and under the mentorship of Jolene Britton, Samuel Britton, Dr. Kevin Costello, and Dr. Vyjayanthi Chari, our group worked to create one of these models for the area of Sierra Leone, one of the countries where the epidemic hit the hardest.

### **Introduction to Mathematical Modeling**

After researching the Ebola virus and gathering the information needed to understand it from a mathematical standpoint, we transitioned into learning about examples of mathematical models. By utilizing the book *Mathematical Models in Population Biology and Epidemiology* by Fred Brauer and Carlos Castillo-Chavez, we discussed many types of models. Each week every member of the research group was assigned to read, learn, and teach a section of the book to the other group members. Since there was too much material for each of us to cover individually, we relied on our group members' abilities to teach the difficult material. The first models our group discussed were the Exponential Growth model and the Logistic Population model. These examples were a nice introduction to modeling due to their simplicity. However, these models are not complex enough to be applied to the Ebola epidemic. From there, we began studying the Predator-Prey model and the Competing Species model. These models gave us insight into how

two species can impact each other in a closed environment. Although the two interacting species models are more complex, they remained as only a stepping stone to understanding models used for epidemics.

After briefly covering the above models, we shifted our focus to epidemiological models. Jennifer started our study of these models by discussing, in great detail, the Kermack-McKendrick Epidemic model, also known as the SIR model. She explained the history, necessary conditions, observations, and equations of the SIR model. With this foundation, I researched, learned and taught the group members about expanding upon this model to make it more realistic. The model I discussed was the SEIR model. This model divides a population into four class labels:  $S(t)$ ,  $E(t)$ ,  $I(t)$ , and  $R(t)$ . Each of these class labels are with respect to time.  $S(t)$  represents the number of individuals in the population who are susceptible to the disease but are not yet infected. In other words,  $S(t)$  is the number of individuals in the population that have the chance to come into contact with the virus through other infected individuals or equipment.  $E(t)$  is the number of individuals that have been exposed to the virus. People in this category are infected with the virus but are not yet contagious. Individuals in this category move into  $I(t)$  once they become contagious with the virus.  $I(t)$  represents the number of individuals that have become infected with the virus and are able to infect others in  $S(t)$ . Finally,  $R(t)$  is the number of individuals that have been infected and then removed from  $S(t)$  and  $I(t)$  by means of isolation, immunization, recovery with full immunity against reinfection, or death (Brauer & Castillo-Chavez, 350). Individuals in the population of study move through the categories starting in  $S(t)$  and then possibly moving to  $E(t)$  then  $I(t)$  then finally  $R(t)$ . This flow is represented in the following system of equations:



$$\begin{aligned}
S' &= -\beta SI \\
E' &= \beta SI - \zeta E \\
I' &= \zeta E - \alpha I \\
R' &= \alpha I
\end{aligned}$$

The parameters involved in these equations are as follows:

$\beta$ : transmission rate of the disease  
 $\zeta$ : rate of incubation of the disease  
 $\alpha$ : average rate of infectiousness of the disease

These equations represent the rates at which the population of each category changes. In other words, the equations show how the individuals move from one class label to the next one.  $\beta SI$ , as seen in the  $S'$  and  $E'$  equations, represents the rate of new infections in the population. The  $\zeta E$  term is the rate at which the exposed individuals become infective/contagious with the virus. Lastly,  $\alpha I$  is the rate at which individuals leave the infective stage by one of the means listed above. To summarize:

$S'$  = the rate at which susceptibles become exposed  
 $E'$  = the rate that susceptibles become exposed minus the rate at which the exposed individuals become infective  
 $I'$  = the rate at which the exposed individuals become infective minus the rate at which individuals leave the infective stage  
 $R'$  = the rate at which individuals leave the infective stage

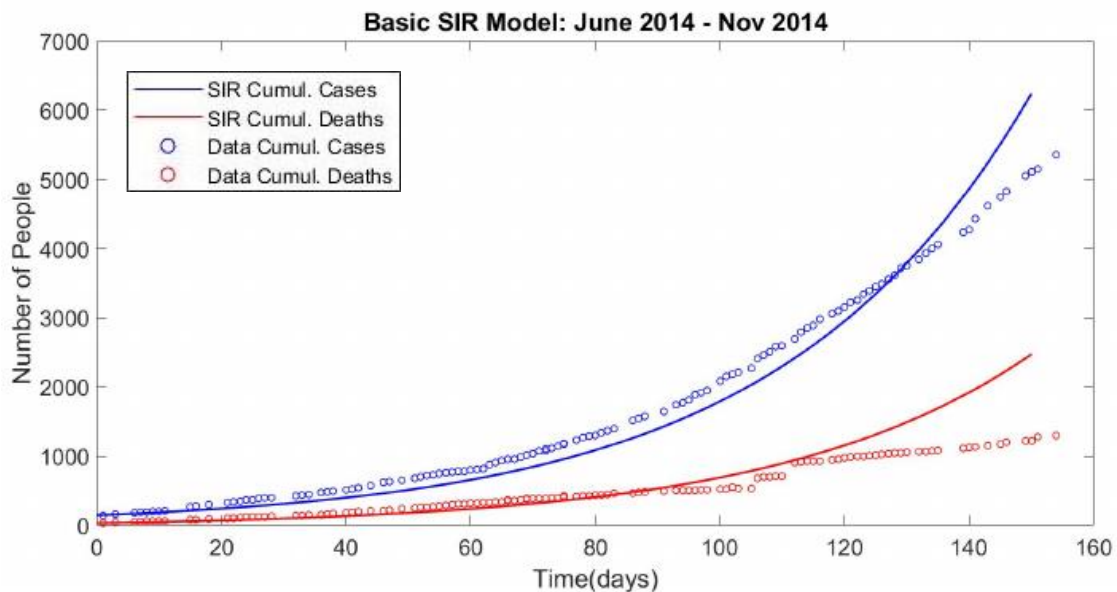
An important component of this SEIR model is the notion of the basic reproductive number  $R_0$ . Brauer and Castillo-Chavez define  $R_0$  as “the number of secondary infections caused by a single infective introduced into a wholly susceptible population of size  $N$  over the course of the infection of this single infective” (353). In other words,  $R_0$  is the number of uninfected individuals that one infected individual infects. For example, if someone had chickenpox and infected three other people with chickenpox, the reproductive number would be three. This basic reproductive number is important because it determines the outcome of an outbreak. If  $R_0$  is less than one, the infection eventually dies out and an epidemic does not occur. However, if  $R_0$  is

greater than one, an epidemic can take place. This basic reproductive number is a strong and useful component of the SEIR model.

The SEIR model can be used for many different types of diseases. The SIR and SEIR models, along with the related variations and expansions of these models, are used to model diseases that have immunity against reinfection, such as most viruses. It is for this reason that we studied these epidemiological models in our attempt to model the 2014 Ebola epidemic in Sierra Leone.

### Our Model

After an in-depth study of the SIR and SEIR models, we created our own SIR model. In order to do this, we used data from the Sierra Leone Ministry of Health and Sanitation as a reference. This source provided us with the number of reported deaths and the number of reported cumulative cases, which is the total number of people infected whether they died from the virus or not. Using these data from June 2014 through November 2014, we were able to create a basic SIR model. The graph of our model is shown below.



In this graph, the dots represent the data from the Sierra Leone Ministry of Health and Sanitation and the lines represent the results of the SIR model we created. The blue dots are the actual number of cumulative cases and the blue line is the SIR model's approximation of the number of cumulative cases. The red dots are the actual number of deaths from the virus and the red line is the SIR model's approximation of the number of deaths. The SIR model's approximation was fairly accurate until around day 120. As time approached November 2014, the model and the data began to diverge significantly. This remained true if we extended the model and data beyond November 2014. Thus, we came to the conclusion that the basic SIR model was not complex enough to accurately model the Ebola epidemic and it was time to move onto a more sophisticated model.

Since Ebola is a complex virus, our model needed to take some characteristics of Ebola into account. First, since the data we used was the number of deaths instead of the number of "removed" individuals as in the SIR model, we changed the class label  $R(t)$  from the number of removed individuals to the number of recovered individuals. We also added an additional class label of  $D(t)$ , which represents the number of deaths from the virus. Second, since many newly infected individuals were not yet infective, we created a class label of  $E(t)$  which represents the number of exposed individuals, similar to the SEIR model. This  $E(t)$  is the number of people in the incubation period for the virus, which is about six to seven days. To summarize:

$S(t)$ : Susceptibles  
 $E(t)$ : Exposed to Ebola but not yet infective  
 $I(t)$ : Infectives/contagious  
 $R(t)$ : Recovered  
 $D(t)$ : Dead

This model does not have a strictly linear flow as in the SIR and SEIR models. An individual in the SEIRD model can flow from  $S(t)$  to  $E(t)$  to  $I(t)$  and then either to  $R(t)$  or  $D(t)$ . This makes sense because an individual can either be recovered or dead, but not both.

Once we established the class labels we wanted to work with, it was time to create the system of ordinary differential equations. There have been many other models proposed that utilize these same class labels, such as in the article “Modeling Post-death Transmission of Ebola: Challenges for Inference and Opportunities for Control” by Joshua S. Weitz and Jonathan Dushoff. However, we did not create the same model that Weitz and Dushoff explained. Instead, we used the equations for the SIR and SEIR models as reference to create the following system:

$$\begin{aligned} S' &= -\beta SI/N \\ E' &= \beta SI/N - \alpha E \\ I' &= \alpha E - \kappa I \\ R' &= (1 - f) \kappa I \\ D' &= f \kappa I \end{aligned}$$

where  $N$  is the total population. The SEIRD model has similar parameters as explained earlier in the SEIR model, but some differences remain. The parameters for the SEIRD model are as follows:

- $\beta$ : transmission rate of the Ebola virus
- $\alpha$ : rate of incubation of the virus
- $\kappa$ : average rate of infectiousness
- $f$ : fatality rate
- $(1-f)$ : survival rate

We also added an additional equation labeled  $C(t)$  for cumulative cases. The  $C(t)$  equation is:

$$C' = \alpha E$$

The  $C(t)$  equation is not involved in the flow of individuals throughout their illness. Rather, it is used as a tool to help us compare our results with the data from the epidemic. As explained in the SEIR model, the system of equations represents the rates at which the population moves from

one class label to the next. Using these definitions of the parameters, the system of equations can be explicitly explained.

The  $\beta SI$  term, as in the  $S'$  and  $E'$  equations, is the transmission rate multiplied by the number of susceptibles and number of infectives. When it is divided by  $N$ , this then gives the rate at which individuals are infected with Ebola. These individuals are infected with the virus but are not yet contagious so they are categorized as exposed. Thus,  $S'$  represents the rate at which susceptibles move from  $S(t)$  to  $E(t)$  to become categorized as exposed.

The  $\alpha E$  term, as in  $E'$ ,  $I'$ , and  $C'$ , is the rate of incubation of the virus multiplied by the number of exposed individuals. This gives the rate at which the exposed individuals become infective due to reaching the end of the incubation period of the virus. Thus,  $E'$  is the rate at which susceptibles become exposed individuals minus the rate at which the exposed individuals become infective. This can also be used to find the total number of cumulative cases, as in  $C'$ .

The  $\kappa I$  term in the  $I'$ ,  $R'$  and  $D'$  equations is the average rate of infectiousness multiplied by the number of individuals in  $I(t)$ . This represents the rate at which people either die or recover from the disease. Thus,  $I'$  is the rate at which individuals become infective minus the rate they either die or recover.  $R'$  is the rate at which those leaving  $I(t)$  recover from the virus while  $D'$  is the rate at which those leaving  $I(t)$  die.

With this system of ordinary differential equations defined, we then had to find values for the defined parameters. Since many of these values are directly related to facts about Ebola, we used our knowledge of the virus to determine an interval to try for each parameter. Each person in the group was in charge of testing different values for the parameters to find out which values produced a result similar to the data. After many trials, the following values were determined as optimal:

$\beta$  = contact rate for one individual per unit of time multiplied by the probability that infection occurs given contact between a susceptible and an infected individual

$\alpha = \frac{1}{6}$  since exposure period/incubation is about 6 days

$\kappa = \frac{1}{9}$  since infective period is 9 days

f = number of deaths divided by total number of people infected

$R_0 = 1.5$

Notice  $R_0$  is greater than one. As explained previously, this means that an epidemic will most likely occur. In the case of the Ebola outbreak in West Africa in 2014, the outbreak most certainly reached epidemic proportions.

The last component needed to complete our model was the initial conditions. Like any system of ordinary differential equations, initial conditions were a key component in finding the best approximation of the data. Using the epidemic data from the Sierra Leone Ministry of Health and Sanitation, we found the SEIRD model initial conditions. The initial conditions are nonzero because the SIR, SEIR, and SEIRD models only work once an outbreak has been established. Thus, using June 22, 2014 as Day 0, the initial conditions for the model are:

$$N = 6,348,350 \text{ people}$$

$$E_0 = 10$$

$$D_0 = 34$$

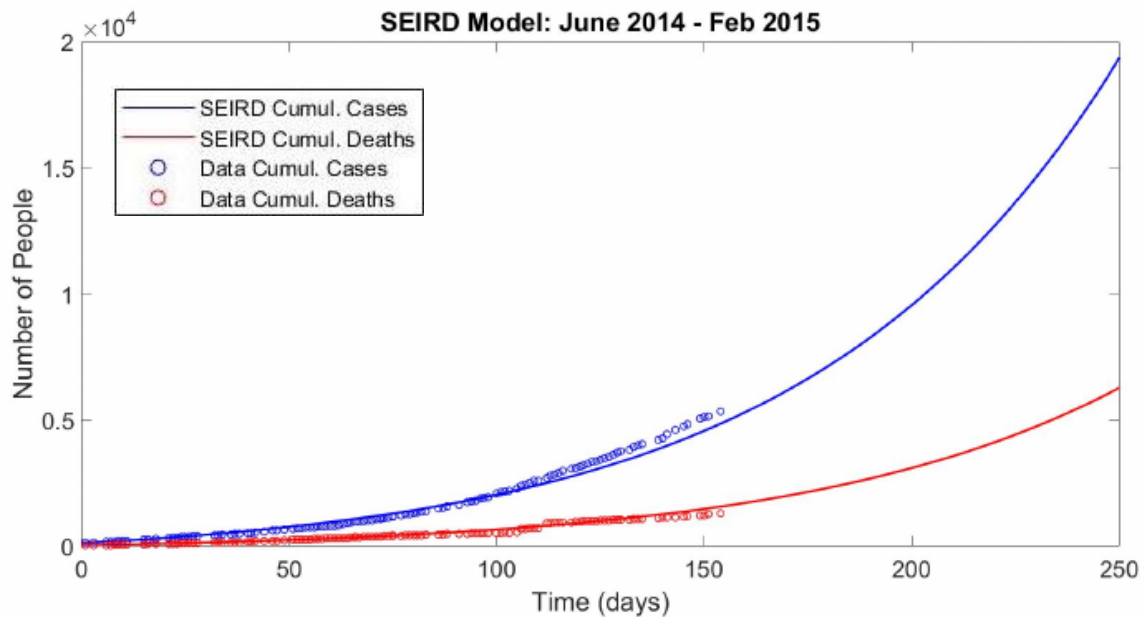
$$C_0 = 147$$

$$I_0 = C_0 - D_0$$

$$\text{Rec}_0 = 0$$

$$S_0 = N - E_0 - C_0 - \text{Rec}_0$$

where  $\text{Rec}_0$  is the initial number of recovered individuals. Combining these initial conditions with the previous information, we created the SEIRD model and graphed the approximations against the data. Thus, we obtained the following results:



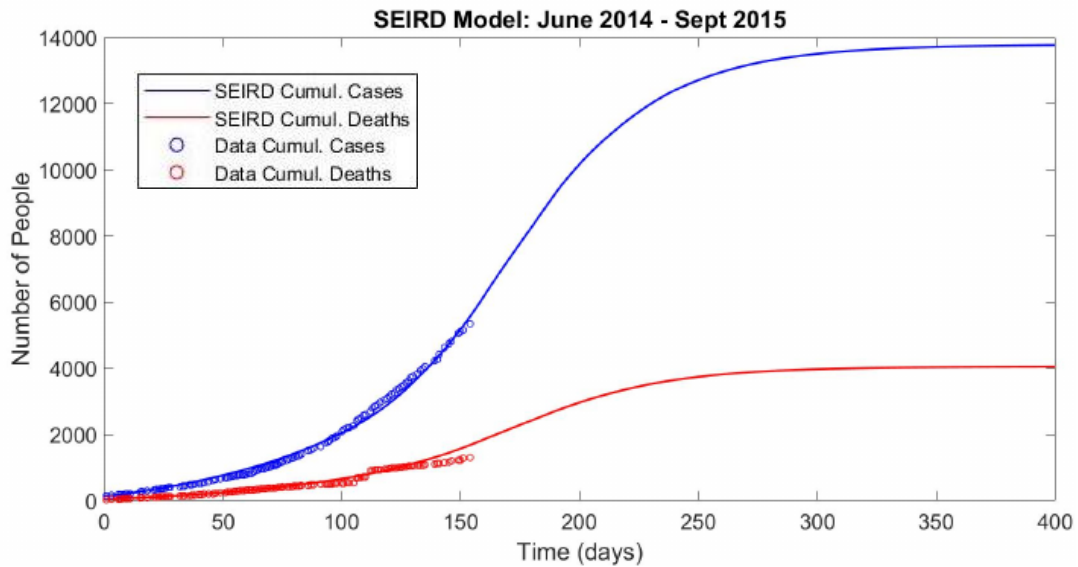
Similar to the graph of the basic SIR model, the lines represent the SEIRD model's approximations and the dots represent the reported data. The blue is the cumulative number of cases and the red is the number of deaths. For this graph we extended the data until February 2015 instead of November 2014. The model's approximation followed the data much closer than the SIR model. However, the epidemic eventually came to an end but the SEIRD model did not show this, as seen by the ever increasing solid lines.

The epidemic did not persist for an infinite period of time due to human intervention. The SEIRD model did not take human intervention into account. Therefore, the model did not show how the epidemic slowed and eventually ended over time. Thus, we needed to manipulate the model to account for this. The two types of intervention that had the biggest impact on the epidemic were quarantine methods and hospitalization. There are ways to add quarantine and isolation equations to the system of equations but due to time constraints and the level of complexity needed to incorporate these, we did not do so. Instead, we thought about the impact

of the types of human intervention on our SEIRD equations. Using our background knowledge of Ebola, we decided that quarantine methods will lower the number of new infections. The people in quarantine will not be able to transfer the virus to as many others. In other words, the transmission rate  $\beta$  will decrease. Additionally, hospitalization will decrease the fatality rate over time but may also increase  $K$ , the average rate of infectiousness. The fatality rate will decrease due to advances in medical preparedness over time. The duration of infectiousness may increase due to this medical preparedness because infected people may be kept alive for a longer period of time. However, this medical preparedness usually includes better sanitation methods which may decrease the amount of medical staff that become infected.

In terms of implementing this in the model, we first determined where in the data the epidemic began to slow down. Around day 225, the number of cases began to decrease. Similarly, the number of deaths began to decrease around day 200. Since the impact of human intervention is not instantaneous, we decided human intervention probably began around day 150. Thus, this is where we implemented the decrease in the value of  $\beta$ , increase in the value of  $K$ , and decrease in the value of  $f$ . By combining all of these factors, we reached a final model that produced the following results:





Note that the lines, dots, and colors represent the same data as the previous two graphs. As seen in the graph, implementing human intervention into the model significantly altered the results of the model. A very close correlation was produced between the SEIRD model's approximation with human intervention and the data of Sierra Leone from June 2014 to September 2015. The model predicted that the epidemic would end with around 14,000 cumulative cases and about 4,000 deaths in Sierra Leone. This is indeed the case with the 2014 Ebola epidemic in Sierra Leone. With human intervention involved, the SEIRD model we created finally produced results that reflected the true nature of the epidemic.

### Further Directions

Although our project ended here, there are many ways in which the model can be expanded upon and modified to create a more realistic and accurate model. First, one of the assumptions of the SIR, SEIR, and SEIRD models is that the population size stays constant. This is unrealistic for a number of reasons. To start, it is very rare for a population to stay constant.

Due to tourism, trade, and travel in general, people come and go from different populations. In the circumstance of an epidemic, there are many doctors, nurses, and other healthcare workers that come to assist with the medical care of infected individuals. These people are still at risk of infection and death so they should be added to the population over time. For these reasons, the model should be expanded to account for a fluctuating population size.

Second, instead of simply manipulating the SEIRD equations to produce the impact of human intervention, equations for quarantine and isolation should be added. This requires defining two new class labels  $Q(t)$  and  $J(t)$  for quarantine and isolation, respectively. These equations will cause the model to become far more complex. The flow of individuals through the class labels will become more intricate as well. However, the results will be closer to the nature of the epidemic.

Third, those who die from Ebola are still infectious until they are buried. The model we created did not account for this. In order to model this, the class label  $D(t)$  will represent those who have died but have not yet been buried. Individuals in this category can infect susceptibles. However, once they are buried the individuals will move to a new class label. Once individuals reach this category, they are no longer able to infect others with Ebola.

A final advancement that can be made to the SEIRD model is related to population flux. Since Sierra Leone is a third world country, modern and efficient transportation is in short supply. Because of this, if there is rough terrain between two populations, there will most likely be less travel between these two populations. On the other hand, if it is easy to move from one population to another, there will be more travel between these two cities. This shows that terrain can have an impact on how and where the virus spreads. Terrain may play a part in the spread of diseases in countries like Sierra Leone, which makes it an important aspect in the study of

epidemics. With these modifications, the model we developed could become a great tool if a future outbreak were to occur.

### **Impact of Project on My Future as a Mathematician**

Math has been my passion since I was young. However, I did not know the true power and need for math until I worked on this project. Witnessing how math can help people around the world deal with the worst imaginable circumstances truly made me appreciate the power of applied mathematics. Spending my time as an undergraduate student in the area of pure math, I quickly changed my area of focus to applied math. This experience also reinforced my decision to attend graduate school in order to attain a PhD in applied math in hope of starting a career as an applied mathematician. I hope to one day be able to help others while working on the subject that I love most.

### **Conclusion**

Ebola is a terrible virus that has killed thousands. After reaching epidemic proportions in West Africa during 2014, modeling the spread and impact of the virus became a priority for many. After studying different models and researching the necessary information about Ebola, our group was able to develop a basic SIR model. Using data from the Sierra Leone Ministry of Health and Sanitation, we were able to expand upon this model to create a more accurate SEIRD model. After further manipulation and implementing the impact of human intervention into the system of ordinary differential equations, we produced a model that approximates the accurate number of deaths and cumulative cases in Sierra Leone. Although the model's predictions are accurate, there are many ways in which this system of equations can be modified in order to

create a more realistic model. With these modifications, more complex models can be used by the government or research groups if another outbreak were to occur. I hope to someday be one of the individuals that can utilize mathematical models to help the world fight and recover from terrible circumstances, such as the 2014 Ebola epidemic. It is projects such as the one I had the opportunity to take part in that demonstrate the true power of mathematics.

## Resources

Brauer, and Carlos Castillo Chaves, *Mathematical Models in Population Biology and Epidemiology*, Springer, London, 2010.

"Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections — NEJM." *New England Journal of Medicine*. Massachusetts Medical Society, 23 Sept. 2014. Web. 20 May 2017.

<http://www.nejm.org/doi/full/10.1056/NEJMoa1411100t=article>.

Gholipour, Bahar. "Ebola 'Patient Zero': How Outbreak Started from Single Child." *LiveScience*. Purch, 30 Oct. 2014. Web. 20 May 2017. <http://www.livescience.com/48527-ebola-toddler-patient-zero.html>.

"Origins of the 2014 Ebola epidemic." WHO. World Health Organization, Jan. 2015. Web. 20 May 2017. <http://www.who.int/csr/disease/ebola/one-year-report/virus-origin/en/>.

Rivers, Caitlin. *Data for the 2014 Ebola Outbreak in West Africa*. GitHub cmrivers. 2014. <https://github.com/cmrivers/ebola>.

U.S. Department of Health and Human Services. "Ebola (Ebola Virus Disease)." Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 15 May 2017. Web. 20 May 2017. <https://www.cdc.gov/vhf/ebola/>.

Weitz, Joshua S., and Jonathan Dushoff. "Modeling Post-death Transmission of Ebola: Challenges for Inference and Opportunities for Control." *Nature News*, Nature Publishing Group, 4 Mar. 2015. <https://www.nature.com/articles/srep08751#abstract>.

"Why Is Logistic Growth More Realistic than Exponential Growth? | Socratic." Socratic.org. N.p., 11 May 2015. Web. 21 May 2017. <https://socratic.org/questions/why-is-logistic-growth-more-realistic-than-exponential-growth>.