6-2020

# Modeling the Spread of Measles

Alexandria Le Beau

MODELING THE SPREAD OF MEASLES

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

by

Alexandria Le Beau

June 2020

Modeling the Spread of Measles

_____

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

_____

by

Alexandria Le Beau

June 2020

Approved by:

Shawnee McMurran, Committee Chair

Min-Lin Lo, Committee Member

Lida Ahmadi, Committee Member

David Maynard, Chair, Department of Mathematics

Corey Dunn, Graduate Coordinator

Abstract

The measles virus has been around since the 9th century. Throughout the years measles have become less problematic in certain areas of the world due to research and the creation of vaccinations. Sadly, not all countries are fortunate enough to have adequate access to the vaccination, which leads to yearly outbreaks.

The goal of this project is to experiment with different mathematical growth models and examine their suitability for modeling outbreaks of measles. We will compare and contrast the exponential model, the logistic model, the SIR model, and the SEIR model. In addition, we will show how the epidemiological models are built and explain their parameters. We will then consider a model for the spread of measles and analyze its suitability using real data.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Measles have been a problem around the world since the 1900s. The measles vaccination was a huge breakthrough, which granted many people immunity from the disease and saved many lives. The Centers for Disease Control and Prevention (CDC) reports that in the year 2000 the disease was considered eliminated in the United States. However, in the last several years, measles have become a problem again. Between 2015 and 2019, the United States started experiencing epidemics again, with 2019 having the most outbreaks since 1992.

In Chad, Africa measles outbreaks happen yearly, but they usually die off around June when the rainy season begins. Researchers are still not sure why this happens. In 2018 the measles outbreak grew in intensity and continued into 2019. Not enough people in the country have been vaccinated and this has resulted in thousands of people losing their lives.

Mathematical models are extremely important in improving our understanding of population dynamics of infectious diseases. Models are a crucial tool

to help with controlling and preventing the spread of disease based on scientific evidence.

# Chapter 2

# History of Measles

## 2.1 Measles

### 2.1.1 History

The measles were first accounted for and the first written accounts were published in the 9th century by a Persian doctor, Muhammad ibn Zakariya ar-Razi[Man19]. In 1757, Scottish physician Francis Home demonstrated that the measles disease is caused by an infectious agent present in the blood of those who contract the measles[CDC19].

Measles became a notifiable disease in the United States in 1912, requiring all healthcare professionals and laboratories to report each diagnosed case of the measles to government officials. The reported information allows the authorities to monitor the disease and provide early warning of possible outbreaks. Within the first decade of mandatory reporting, the United States averaged 6000 deaths per year caused by measles.

In the 1950s, before a measles vaccine was available, almost all children contracted the measles before the age of fifteen. Each year in the United States, three to four million people contracted the measles, 400 to 500 people died, approximately 48 thousand were hospitalized, and about 1000 people suffered from brain swelling caused by the measles disease.

### 2.1.2 The Measles Virus

Measles is a very contagious airborne disease that is caused by a virus and may be spread via an infected person who sneezes or coughs. The measles virus can remain live in the air for up to two hours, so an infected person need not be present in order to infect others. The virus is extremely contagious; if only one person is infected, up to 90 percent of the people around that person may also become infected if they have not been vaccinated. Measles is very serious in the fact that it can create many complications and may result in death.

When infected with the measles, symptoms do not show for approximately seven to fourteen days. The first symptoms are a high fever (can be over 104 degrees), runny nose, cough, and eyes may become red and watery. Small white spots, called Koplik spots, will develop inside the mouth two to three days after the first symptoms appear, and a rash will appear within three to five days of the first symptoms. The rash commonly starts as flat, red spots on the face near the hairline, then advances to the neck, arms, legs and feet. Raised bumps might also develop on top of the Koplik spots, and as the rash spreads the spots may join together. When this is happening, one's fever could spike to 104 degrees or more. The total duration of the measles infection is about two to three weeks.

## 2.2   Vaccination

### 2.2.1   Vaccine Creation

In 1954, the research of creating a measles vaccination began. John F. Enders and Dr. Thomas C. Peebles began their research in Boston, Massachusetts during a measles outbreak. They started by collecting blood samples from a handful of ill students. Their goal was to isolate the virus in the blood, and in turn, create a vaccination. They were successful in doing so with 13 year old David Edmonston[CDC19].

Enders and colleagues were able to transform their Edmonston-B strain of the measles virus into a vaccination and license it in the United States in 1963. An improved and weaker vaccination, called the Edmonston-Enders (formerly "Moraten") strain, was developed and distributed in 1968 by Maurice Hilleman and his colleagues. The Edmonston-Enders strain has been the only measles vaccination to be used in the United States since it was developed in 1968[CDC19].

The mealses vaccination works by enabling the immune system to create antibodies that fight off the measles virus that is contained in the vaccination. The vaccination defends the body from every type of mealses genotype. If someone is exposed to the measles virus and they have been vaccinated, their body will remember how to fight off the virus because it had been trained to do so when the vaccination was administered.

### 2.2.2    Vaccine Effectiveness

According to the CDC, the measles vaccine is very effective. One dose of the vaccine is approximately 93% effective, and two doses of the vaccine are approximately 97% effective, at preventing the measles if one is to be exposed to the virus.

### 2.2.3    Measles in the United States

In 1978, The CDC set a goal to eliminate the measles in the United States by 1982. In order to do this, vaccinations needed to be distributed among a large number of people. Although the objective was not achieved, the vaccination campaigns greatly reduced the number of people contracting the measles. By 1981, measles cases had dropped over 80%. Unfortunately, in 1989 the United States experienced measles outbreaks in school-aged children that had previously been vaccinated. This inspired the Advisory Committee on Immunization Practices (ACIP), the American Academy of Pediatrics (AAP), and the American Academy of Family Physicians (AAFP) to suggest that children should receive a second dose of the measles vaccination. There was a decline in reported cases after the second dose suggestion and after the improvement of coverage of the first dose.

In the year 2000, there were no reported measles cases for more than twelve consecutive months. Consequently, the measles were considered eliminated from the United States. This achievement was brought about by a widespread vaccination program and better measles control. The United States was able to maintain this situation for 19 years, which was a huge public health accomplishment. Unfortunately, measles outbreaks began again in 2019. [WHO19]

### 2.2.4 Measles Around the World

Measles and a measles vaccination is present all over the world. Although the vaccination is affordable and safe, a significant number of people are still not vaccinated. There were an estimated 110,000 deaths in the world from the disease in 2017, of which most were children under five years old. However, due to the vaccination there was an 80% decline of deaths globally between the years 2000 and 2017. This amounts to about 21.1 million spared lives. The measles vaccination is one of the best vaccinations to have been created. By 2017, approximately 85% of children around the world had received at least one of two doses of the vaccination before their first birthday. This was a 13% increase from the year 2000. [WHO19]

### 2.2.5 Why Not Vaccinate?

Although the measles vaccination is affordable and available all over the world, many people refuse the vaccination for a variety of reasons. Some of these reasons include, but are not limited to: religious reasons, personal beliefs or philosophical reasons, safety concerns, and the need of additional information.

# Chapter 3

# Chad, Africa



Figure 3.1: *Chad, Africa.*

## 3.1 History

### 3.1.1 Lifestyle

The country of Chad is located in Central Africa and is bordered by Sudan, Cameroon, the Central African Republic, and Nigeria. The territory spreads

across 1,284,000 square kilometers of desert and rain forest. Chad is the fifth largest country in Africa and the 21st largest country in the world, with a population of 15.5 million.

Chad is an extremely poor country with one of the highest levels of hunger in the entire world. Approximately 66.2 percent of Chad's population is living in severe poverty. In Chad, children are required to attend school to receive an education, yet only about 68% of children attend. Half of the country's population is illiterate.

### 3.1.2 Vaccination

The healthcare system in Chad is extremely poor. The healthcare facilities are overpopulated, under supported, and are not spread equally throughout the country. The available healthcare facilities average only about three doctors, fifteen nurses, and two midwives per 100,000 people. Deficits in the healthcare system are primarily due to the political uncertainty that has been a concern for many years.

Before a measles vaccination was present in Africa, the measles virus mainly affected young children. More than a million cases were documented yearly. Between 1970 and 1980, a measles vaccination was introduced in the African countries by the World Health Organization (WHO). The measles vaccination helped the countries significantly, leading to longer times between epidemics. However, the vaccination still has not been distributed to enough children to stop annual outbreaks.

### 3.1.3   Outbreaks

Table 3.1 and Figure 3.2 display the number of newly infected individuals per week and the cumulative number of individuals that have been infected. Week 1 corresponds to the first week of January, 2018. Our chart begins week 20, May 12 - May 18, 2018, since this is when the epidemic was declared.

| Time (Weeks) | Newly Infected | Total Infected |
|:---:|:---:|:---:|
| 20 | 77 | 77 |
| 21 | 74 | 151 |
| 22 | 66 | 217 |
| 25 | 175 | 392 |
| 26 | 91 | 483 |
| 27 | 118 | 601 |
| 29 | 162 | 763 |
| 30 | 98 | 861 |
| 33 | 237 | 1098 |
| 35 | 122 | 1220 |
| 36 | 155 | 1375 |
| 48 | 259 | 1634 |
| 55 | 2000 | 3634 |
| 61 | 8333 | 11967 |
| 68 | 912 | 12879 |
| 69 | 846 | 13725 |

Table 3.1: *Chad Data from the 2018 Outbreak.*
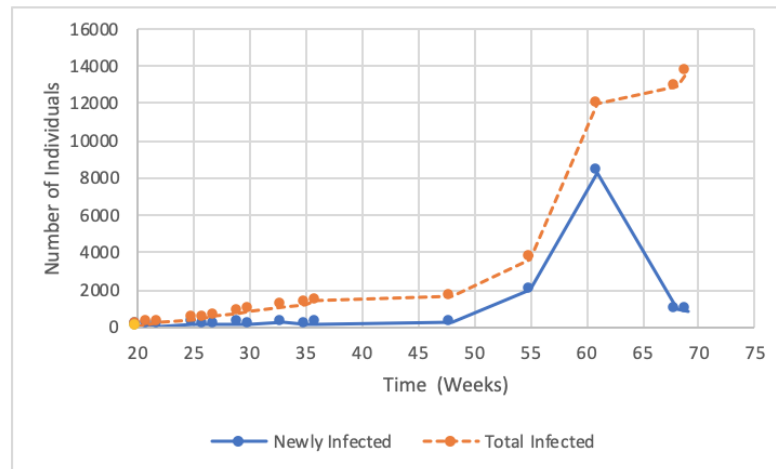
Figure 3.2: *Scatterplot of Chad Data from the 2018 Outbreak.*

Every year Chad experiences measles outbreaks. The outbreaks normally begin in the spring and die off around June, when the rainy season begins. In 2018, the measles outbreak did not end in June, like most years. Instead, it grew in intensity and continued into 2019, as depicted in Table 3.1 and Figure 3.2. This was a huge problem because Chad's pre-harvest season is from June to September, which is when hunger is at an all time high. When children are malnourished they are more susceptible to becoming ill and suffering from the measles. Those who are malnourished have a lessened chance of surviving the disease.

In the next chapter we will explore different mathematical models and evealuate their suitability to model the data shown in Table 3.1 and Figure 3.2.

# Chapter 4

# Mathematical Models

Mathematical models can be used as a mechanism or tool that help explain, describe, and predict behavior of physical phenomena. One area in which models are exceptionally useful is in the study of epidemiology. There are many epidemiology mathematical models, from simple to complex. As scientists come to better understand how disease spreads, they continue to work towards building models that will be appropriate for whatever problem may be at hand. In this thesis we begin our study with the analysis of a simple growth model and demonstrate how to increase the complexity of a model to better represent the scenario under investigation.

## 4.1   Exponential Model
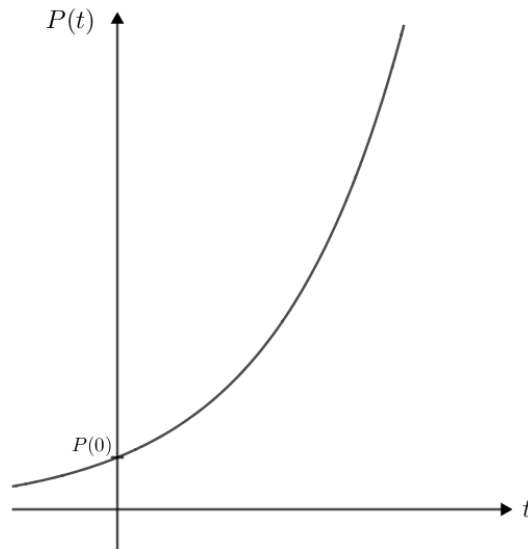


Figure 4.1: *Graph Depicting Exponential Growth.*

The exponential model was popularized by Thomas Robert Malthus (1766-1834)[Sha97]. An exponential model has an unconstrained pattern. The assumption that a growth rate will be proportional to the population leads to the differential equation $\frac{dP}{dt} = rP$, where the dependent variable $P$ represents a population as a function of time $t$ and $r$ is a parameter that represents the per capita growth rate. All solutions of this equation are of the form $P(t) = P_0 e^{rt}$, where $P_0$ denotes the initial population $P(0)$ and $e$ is the base rate of growth shared by all continuous growth processes. When $r < 0$, the population declines. When $r > 0$, the populations increases. If $r = 0$, the population remains fixed.

The exponential model may sometimes be used to model early stages in

the spread of disease, but it is too simplistic of a model to use long term. It is useful because it relates to predicting impact on resources and can be used to estimate parameters. One drawback to the model is that it does not predict when the number of ill individuals will peak. When working with this model, we assume an infinite population, unless the domain is restricted. In order to use the model appropriately, the domain over which behavior would be approximately exponential would need to be known ahead of time. We consider an example to help explain some of the key drawbacks and benefits of the exponential model.

### 4.1.1 Exponential Model Example

In a short article published in 1978 by the British Medical Journal, the authors present Figure 4.2 [BMJ78]. The graph depicts data of students confined to bed and convalescent students from the influenza outbreak at an English boarding school. By embedding the original graph in a *GeoGebra* worksheet and scaling appropriately (see Figure 4.3), we obtain the data estimates given in Table 3.1. The data estimates were then entered into *Excel* to create Figure 4.4.
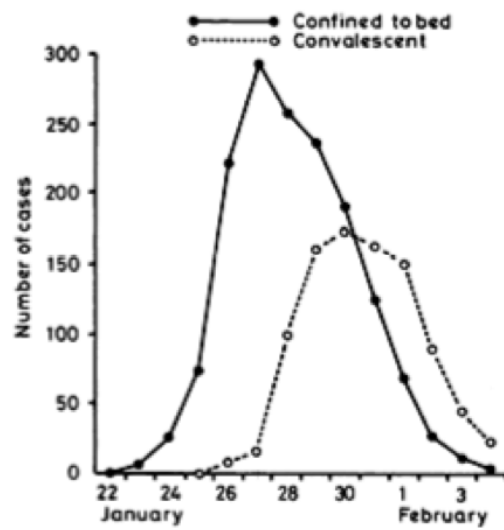
Figure 4.2: *Graph of 1978 England Boarding School Influenza Estimates.*
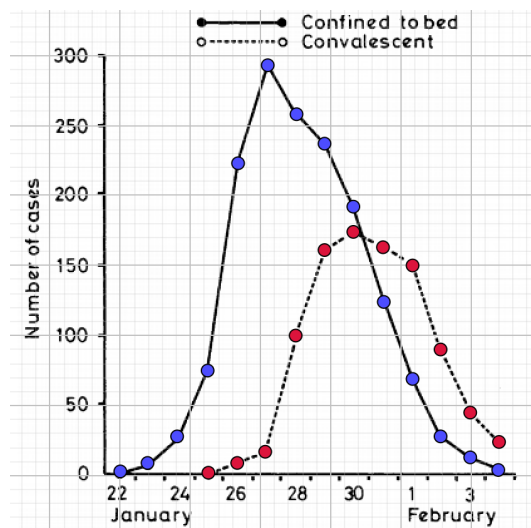


Figure 4.3: *GeoGebra Estimates of Number of Students Confined to Bed and Number of Convalescent Students.*

| Time (Days) | Confined to Bed | Convalescent |
|:-----------:|:---------------:|:------------:|
| 0 | 1 | 0 |
| 1 | 3 | 0 |
| 2 | 9 | 0 |
| 3 | 28 | 0 |
| 4 | 75 | 2 |
| 5 | 223 | 9 |
| 6 | 293 | 17 |
| 7 | 258 | 100 |
| 8 | 237 | 161 |
| 9 | 192 | 174 |
| 10 | 124 | 163 |
| 11 | 69 | 150 |
| 12 | 28 | 90 |
| 13 | 13 | 45 |
| 14 | 4 | 24 |

Table 4.1: *Estimates of 1978 England Boarding School Influenza Data.*
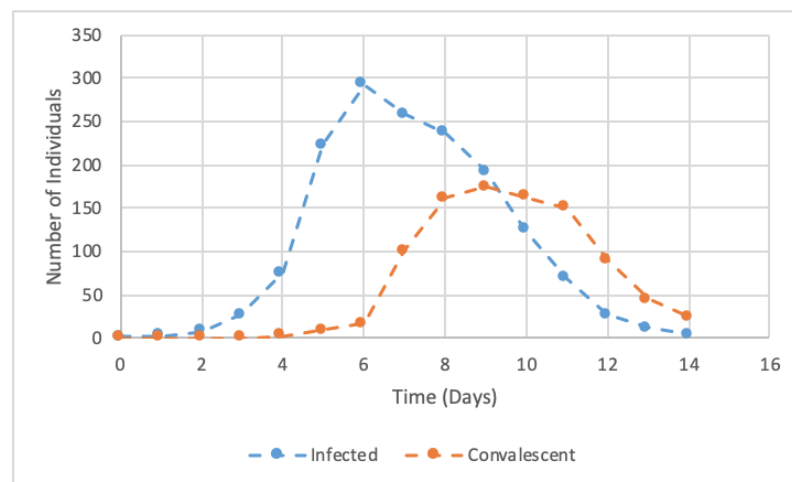


Figure 4.4: *Scatterplot of 1978 England Boarding School Influenza Data Estimates.*

When looking at Figure 4.4 it is clear that an exponential model does not fit our entire data set. However, we notice in Figure 4.4 that the first six data points on the graph of the "infected" students appear to follow an exponential growth pattern. Taking the first six data points and plotting them on a scatterplot gives us a closer look at the proposed exponential growth (see Figure 4.5). When an exponential best fit curve is added, we see that the fit appears close. With a growth pattern such as this one, and only a few data points, it can be difficult to tell visually whether the data fits an exponential model, or whether some other model such as a power function might provide a better fit. We can use a semi-logarithmic (semi-log) graph to help determine if the growth of our data is exponential or not. We want to see if the model $y = ab^t$ provides a good fit for the data for some $a$ and $b$, where $y$ represents the dependent variable. For our semi-log graph we will use a logarithmic scale for the dependent axis and a linear scale for the independent axis. By taking the natural log of both sides of the equation $y = ab^t$, we obtain the equation $\log y = \log a + t \log b$.[1] Hence the graph of the exponential equation will be linear on a semi-log graph.[2]

---

[1]In this paper we will let "log" denote the natural logarithm.

[2]A best fit curve is a curve that shows the overall idea of what the data is illustrating. It helps us to see if there is a correlation between the two factors being graphed.
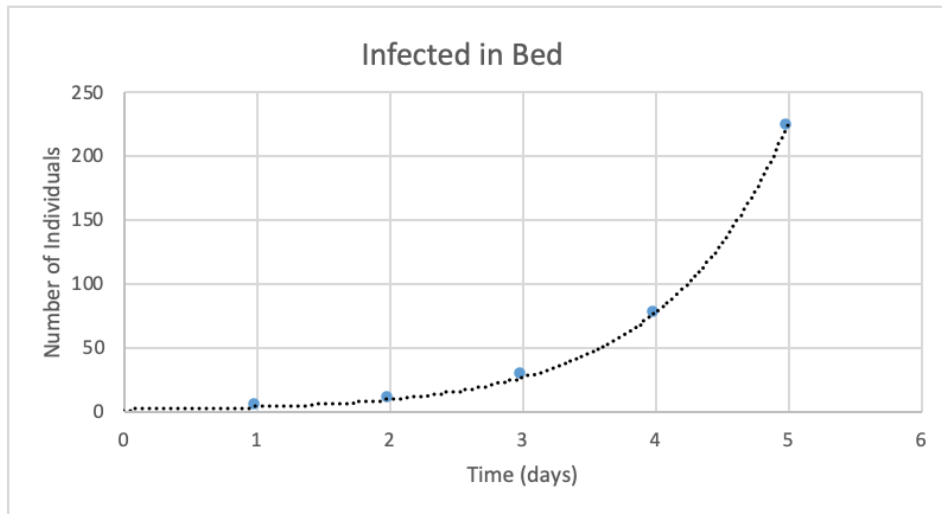
Figure 4.5: *First Six Data Points of the Boarding School Data.*

To test our hypothesis on the boarding school data, we began by taking the natural log of number of infected individuals, Table 4.2, and creating the corresponding semi-log scatterplot, Figure 4.6.

| Time (Days) | Confined to Bed | log(Infected) |
|:-----------:|:---------------:|:-------------:|
| 0 | 1 | 0 |
| 1 | 3 | 1.098612 |
| 2 | 9 | 2.197225 |
| 3 | 28 | 3.332205 |
| 4 | 75 | 4.317488 |
| 5 | 223 | 5.407172 |
| 6 | 293 | 5.680173 |
| 7 | 258 | 5.552960 |
| 8 | 237 | 5.468060 |
| 9 | 192 | 5.257495 |
| 10 | 124 | 4.820282 |
| 11 | 69 | 4.234107 |
| 12 | 28 | 3.332205 |
| 13 | 13 | 2.564949 |
| 14 | 4 | 1.386294 |

Table 4.2: *Boarding School Data with Natural Log of Number of Infected.*
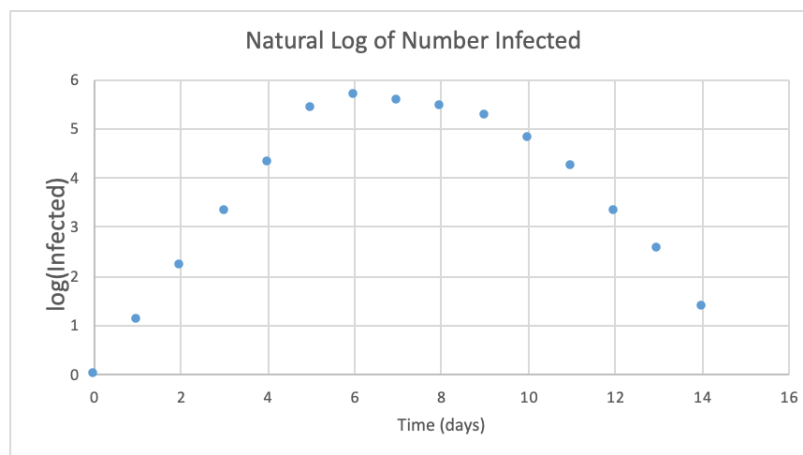


Figure 4.6: *Semi-Log Scatterplot of Boarding School Data.*

Notice that the first six data points seem to lie on a line, further con-firming our conjecture of exponential growth. Let us check the goodness-of-fit by finding a line of best fit for the first six points on the semi-log graph.
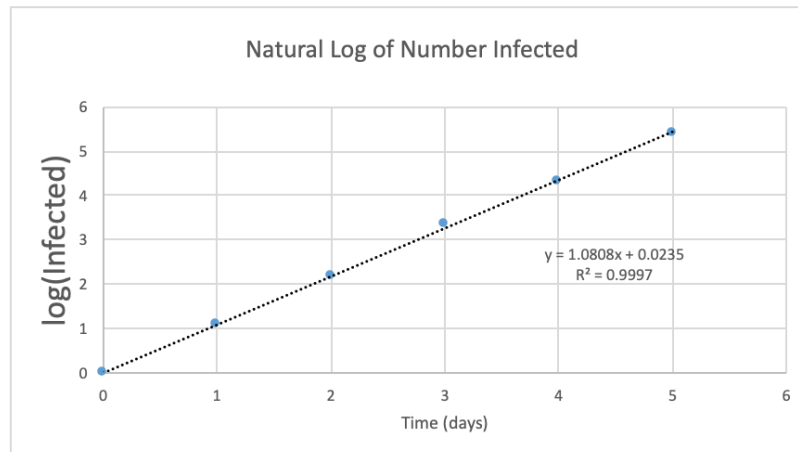


Figure 4.7: *Boarding School Data on a Semi-log Graph.*

In Figure 4.7 we find that the line of best fit appears to provide a sur-prisingly good fit for our data estimates. We find that the data given by $y = \log(\text{infected})$ seems to be modeled well by the line of best fit $y = 1.0808t + 0.0235$.

## 4.1.2    Diagnostics and Model Validity

The coefficient of determination for this model is given by $R^2 = 0.9997$. Since this value is very close to 1, we may be able to conclude that the linear model fits well. The coefficient of determination, R-squared, is a goodness-of-fit measure for linear regression models. This statistic indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. The

R-squared statistic measures the strength of the relationship between a model and the dependent variable on a convenient $0 - 100\%$ scale [Fro20]. We also examine the residual plot to see if our model is biased. An unbiased model has residuals that are randomly scattered around zero. If the model is unbiased, we can trust our result of R-squared.
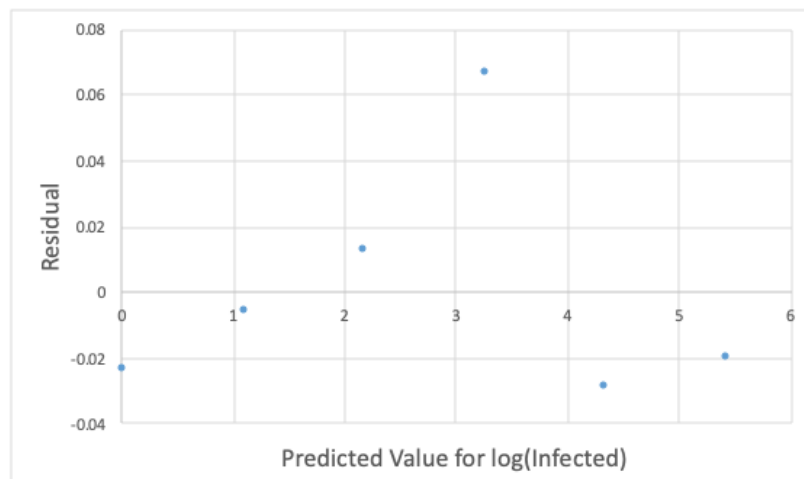


Figure 4.8: *Boarding School Data Residual Plot*

When examining our residual plot in Figure 4.8 it is difficult to conclude if the differences between the data values and the values predicted by the model are biased or not. Our residuals fall both above and below 0, therefore the model appears to be unbiased, but with so few data points we cannot confidently make that assumption. In order to strengthen our claim of goodness-of-fit, we can use *Excel Data Analysis* to run regression statistics for our data, see Table 4.3

| Regression Statistics | |
|---|---|
| R-squared | 0.9997 |
| Adjusted R-squared | 0.9996 |
| Standard Error | 0.03698 |
| Observations | 6 |

Table 4.3: *Regression Statistics of Semi-log Boarding School Data.*

Table 4.3 confirms our R-squared value of 0.9997 and provides us with a standard error value of 0.0368. Standard error tells us if our data is precise enough to use for prediction. Since our standard error is 0.0368, our 95% prediction interval would be $2 \times 0.0368 = \pm 0.0736$ units. Our data set lies between 0 and 6, therefore our standard error value of 0.0368 is small enough to confirm goodness-of-fit. Now that we have obtained our R-squared value, examined our residual plot, and obtained our standard error value, we may conclude that the linear model fits well.

The model is given by $y = 1.0808t + 0.0235$, which translates to the exponential model for the number of infected given by $I \approx 1.02e^{1.0808t} \approx 1.02(2.947)^t$. So the model predicts an approximate 195% increase in infected individuals per day for the first five days. Clearly such behavior cannot continue indefinitely for a finite population and we indeed see that it discontinues after day six. What our findings tell us is that the early data fits the exponential model well, but as time goes on, the spread slows down. This may be the case since the population was confined to the boarding school. The infections grew quickly the first few days and then slowed down as individuals were removed from the population the first day they showed symptoms.

## 4.2 Logistic Model
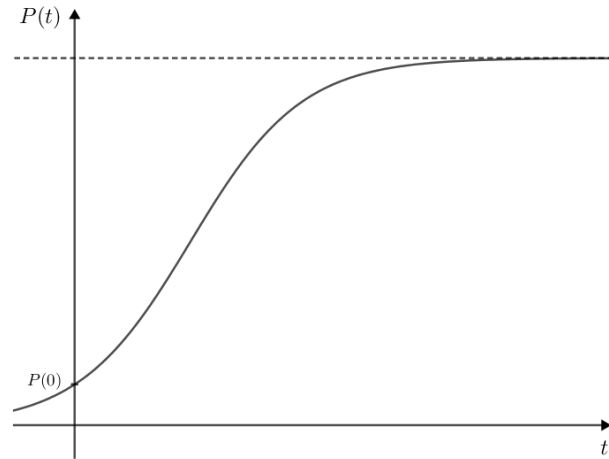


Figure 4.9: *Logistic Growth Model.*

    The logistic model was introduced by Pierre Verhulst in 1838, who for-
mulated it as a model of population growth by adjusting the exponential growth
model[Sha97]. The logistic model has a constrained pattern. The population tends
toward a fixed value $k$, usually referred to as the carrying capacity. If the popula-
tion is small relative to $k$, the pattern approximates the pattern of the exponential
model. As $P$, a population as a function of time $t$, becomes a significant fraction
of $k$, the curve diverges from the exponential curve, and as $P$ approaches $k$, the
growth drops to zero. We can model this behavior by modifying the exponential
model with the inclusion of a factor of $\frac{k-P}{k}$, which is close to 1 (has little effect)
when $P$ is much smaller than $k$, and is close to zero when $P$ is close to $k$. When $P$
is less than $k$, the expression $k - P$ tells us how many more people may be added
to the population before the carrying capacity is reached. The logistic model is

then defined by the initial value problem:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{k}\right), \quad P(0) = P_0.$$

The differential equation is separable. The solution to the initial value problem may be obtained analytically via a straightforward integration:

$$\int_{P_0}^{P} \frac{1}{\rho\left(1 - \frac{\rho}{k}\right)} d\rho = \int_0^t r d\tau.$$

Integrating both sides yields the equation

$$\ln\left|\frac{P}{k-P}\right| - \ln\left|\frac{P_0}{k-P_0}\right| = rt.$$

Solving for $P$ we obtain

$$P = \frac{k}{1 + Ae^{-rt}},$$

where $A = \frac{k}{P_0} - 1$.

Although the logistic model has a sigmoid shape characteristic of many models for the spread of disease, its parameters do not account for factors such as individuals who have recovered from the disease and have gained lifelong immunity or for randomness in infection or time. In particular, to use the model one would need to know ahead of time what percent of the population is expected to become infected.

## 4.2.1  Logistic Model Example

In section 4.1.1 we notice that the data follows an exponential pattern for the first 5 days. The data from the British Medical Journal just provides us with the number of boys confined to bed rest on any particular day. In this section we

wish to look at the number of infected boys in the population on any particular day. The flu in this situation is the 1977-78 H1N1 influenza, where symptoms usually present 24 hours after exposure. The boys were removed from the population and confined to bed rest once symptoms presented. If we can estimate how many new boys are added to the bed rest group each day, then we will have an estimate of how many infected boys were in the population on the previous day. According to the article, the boys were confined to bed for two to four days and then spent another one to two days convalescing [BMJ78]. We will use this information to break down the data and predict where in the recovery process each infected boy is on any particular day. A problem with our data is that we can not determine, after three to four days, how many new boys are added to the bed rest group each day since the boys leave the group and, after a day or two of convalescing, re-enter the population.

Using the data estimates from Table 4.1 and *Excel* we can estimate the cumulative number of infections and fit it to a logistic curve in *GeoGebra*. First we must estimate two transmission rates, $\beta_1$ and $\beta_2$. We estimate two transmission rates because when using a single vallue for $\beta$ it only allowed us to fit the first few days, and after the first few days the model predicted a faster transmission than the data showed, so we introduce a second $\beta$. We do this by taking the first five data points and using the differential equation:

$$\frac{dS}{dt} = -\beta SI$$
$$\Rightarrow \quad \delta S = -\beta SI$$
$$\Rightarrow \quad S_n - S_{n-1} = -\beta S_{n-1} I_{n-1}.$$

We average the five values of $\beta$ to get the a first approximation of $\beta_1 = 0.00265$. We will use the same value, 0.00265, for $\beta_2$'s first approximation. We then run *Excel Solver* to obtain a better approximation, giving us $\beta_1 = 0.00342$ and $\beta_2 = 0.00198$. We assume every infected individual needs at least two days of bed rest. So we also need to find the probability of needing a 3rd and 4th day of bed rest, $B_3$ and $B_4$. We will create a table, Table 4.4, and adjust the values of $B_3$ and $B_4$ until they give us values that are close to the real data values. We get $B_3 = 0.3854$ and $B_4 = 0.3$.

| Time (Days) | Bed Day 1 | Bed Day 2 | Bed Day 3 | Bed Day 4 | Conv. Day 1 | Conv. Day 2 |
|---|---|---|---|---|---|---|
| 1 | 3 | 0 | 0 | 0 | 0 | 0 |
| 2 | 7.8 | 3 | 0 | 0 | 0 | 0 |
| 3 | 19.9 | 7.8 | 1.2 | 0 | 3 | 0 |
| 4 | 49.7 | 19.9 | 3 | 0.3 | 5.6 | 3 |
| 5 | 115.6 | 49.7 | 7.7 | 0.9 | 14.7 | 5.6 |
| 6 | 129.1 | 115.6 | 19.2 | 2.3 | 36.9 | 14.7 |
| 7 | 111.2 | 129.1 | 44.6 | 5.8 | 86.8 | 36.9 |
| 8 | 71.3 | 111.2 | 49.8 | 13.4 | 116.3 | 86.8 |
| 9 | 35.7 | 71.3 | 42.9 | 14.9 | 116.6 | 116.3 |
| 10 | 15.3 | 35.7 | 27.5 | 12.9 | 88.8 | 116.6 |
| 11 | 6.1 | 15.3 | 13.8 | 8.2 | 54 | 88.8 |
| 12 | 2.4 | 6.1 | 5.9 | 4.1 | 27.3 | 54.0 |
| 13 | 0.9 | 2.4 | 2.4 | 1.8 | 12 | 27.3 |
| 14 | 0.3 | 0.9 | 0.9 | 0.7 | 4.9 | 12 |
| total sick | 568.6 | | | | | |

Table 4.4: *Predicted Data of the Boarding School Influenza: Breakdown by Day.*

| Time (Days) | Predicted New Infections | Cumulative Predicted Infections |
|:---:|:---:|:---:|
| 1 | 3 | 3 |
| 2 | 7.8 | 10.8 |
| 3 | 19.9 | 30.7 |
| 4 | 49.7 | 80.5 |
| 5 | 115.6 | 196.1 |
| 6 | 129.1 | 325.2 |
| 7 | 111.2 | 436.5 |
| 8 | 71.3 | 507.8 |
| 9 | 35.7 | 543.5 |
| 10 | 15.3 | 558.8 |
| 11 | 6.1 | 565 |
| 12 | 2.4 | 567.3 |
| 13 | 0.9 | 568.3 |
| 14 | 0.3 | 568.6 |

Table 4.5: *Predicted Data of the Boarding School Influenza: New and Cumulative.*

We can now fit a logistic curve in *GeoGebra* to the data from Table 4.5. We find that a logistic curve provides a surprisingly good fit for our data estimates.
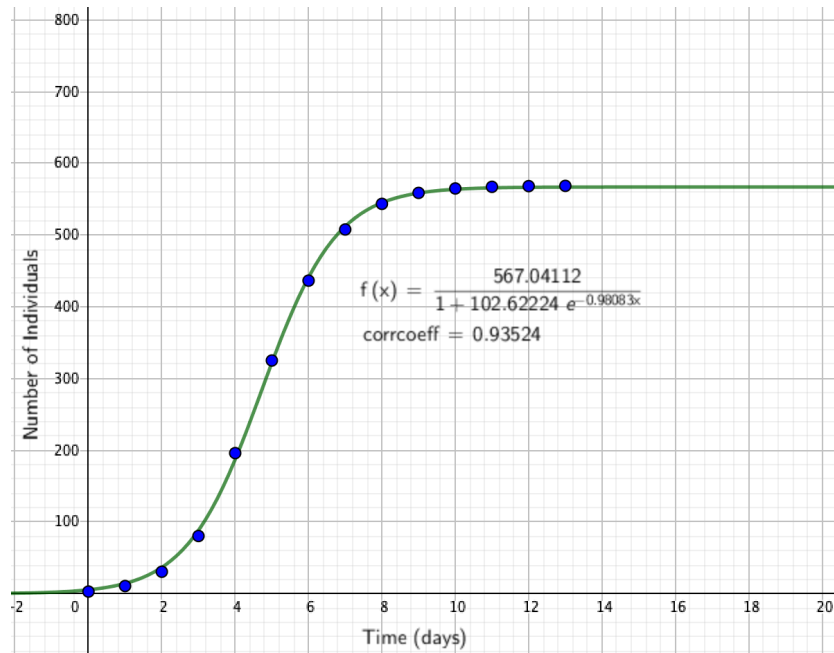
Figure 4.10: *Cumulative Predicted Infections Fit to a Logistic Curve.*

## 4.3   SIR Model

The SIR model, also called the Kermack-McKendrick model, was one of the early triumphs of mathematical epidemiology created by William Kermack and Anderson McKendrick (1927) in order to model the spread of disease in larger, closed populations. We assume the population, $N$, is classified into three distinct groups of individuals, which are represented by the letters S, I, and R. The groups are defined as follows:

**S(t):** Number of susceptible people. People belonging to the susceptible group have not been infected with the disease nor are they immune to it. They are

at risk of becoming infected with the disease in the future.

**I(t):** Number of infected or infectious people. These are people that are infected with the disease and can transmit the disease to susceptible people.

**R(t):** Number of recovered people. This represents the population who have recovered from the disease and are immune; they can no longer be infected with the disease. This group may also include members of the population who have been vaccinated.
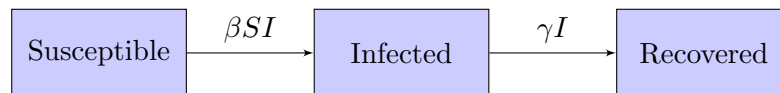


Figure 4.11: *SIR Model.*

Figure 4.11 illustrates how people move from one group to another as a function of time. The parameters $\beta$ and $\gamma$ describe the flow rates, leading to a system of differential equations.

New infections occur as a result of contact between infectives and susceptibles. In this simple model, the rate at which new infections occur is $\beta SI$ for some positive constant $\beta$ related to the probability of a new infection occurring. When a new infection occurs, the individual infected moves from the susceptible group to the infected group. In our simple model, there is no other way individuals can enter or leave the susceptible group, so we have our first differential equation:

$$\frac{dS}{dt} = -\beta SI.$$

The other process that may occur is the infective individuals may enter the recovered group. We assume that this happens at the rate $\gamma I$ for some positive constant $\gamma$. Thus we have our other two differential equations:

$$\frac{dI}{dt} = \beta SI - \gamma I,$$

$$\frac{dR}{dt} = \gamma I,$$

where $t$ represents time, $\beta$ represents the probability that there will be an interaction between an infected and susceptible that results in a new infection ($\beta$ = probability of transfer per contact $\times$ contacts per day), and $\gamma$ represents the per capita recovery rate $\left(\gamma = \frac{1}{\text{average length of illness}}\right)$. We are also interested in the reciprocal, $\frac{1}{\gamma}$, which describes the average infectious period.

In this model, $R$ is completely determined by $S$ and $I$ so we need not consider the $R$ equation in this model, leaving a system of two equations

$$\frac{dS}{dt} = -\beta SI,$$

$$\frac{dI}{dt} = (\beta S - \gamma)I, \tag{4.1}$$

together with initial conditions:

$$S(0) = S_0, \ I(0) = I_0, \ S_0 + I_0 = N. \tag{4.2}$$

The SIR model is only applicable if $S(t)$ and $I(t)$ remain positive. Population values must be non-negative, and whenever $S(t)$ or $I(t)$ reaches zero, the rates of change will be zero. This is because everyone would then be in the recovered population, $R$, and the disease would be terminated. It is recognized that $S'(t) \leq 0$ for all $t$, and $I'(t) > 0$ if and only if $S > \frac{\gamma}{\beta}$. Thus $I$ increases as long as $S > \frac{\gamma}{\beta}$. If $S'(t) < 0$, then $S$ decreases for all $t$, so $I$ decreases and approaches zero. This is because as susceptible individuals become infected, they leave $S$ and enter into $I$. They eventually move into $R$, so $I$ also decreases. If $S_0 < \frac{\gamma}{\beta}$, then $I$ decreases to zero and there will be no epidemic. If $S_0 > \frac{\gamma}{\beta}$, then $I$ first increases to a maximum when $S = \frac{\gamma}{\beta}$ and then decreases to zero, in which case there will be an epidemic.

The basic reproduction number determines if there will be an epidemic or not. It is the number of secondary infections that are caused by a single infected person during the course of their infection. The single infective has introduced the infection into an entirely susceptible population of size $N \approx S_0$. The basic reproduction number is denoted by $\mathfrak{R}_o$ and represented by the quantity $\frac{\beta S_0}{\gamma}$. If $\mathfrak{R}_o > 1$, there will be an epidemic, and if $\mathfrak{R}_o < 1$ the infection will die out and there will not be an epidemic. In this model, we assume an infected person makes $\beta N$ contacts per unit of time with susceptible individuals and hence new infections occur. The mean infective period is $\frac{1}{\gamma}$; thus the basic reproduction number is $\frac{\beta N}{\gamma}$. We assume $N \approx S_0$, then $\frac{\beta N}{\gamma} \approx \frac{\beta S_0}{\gamma}$.

The system 4.1 is a two-dimensional autonomous system of differential equations. We proceed to analyze the system.

The sum of the two equations of 4.1 is

$$(S + I)' = -\gamma I. \tag{4.3}$$

Thus $S + I$ is a nonnegative smooth decreasing function and therefore tends to a limit as $t$ tends to infinity. Also, the derivative of a nonnegative smooth decreasing function must tend to zero, so

$$\lim_{t \to \infty} (S + I)'(t) = 0.$$

It follows that

$$I_\infty = \lim_{t \to \infty} I(t) = 0.$$

Thus, there exists a finite value $S_\infty$ such that

$$\lim_{t \to \infty} S(t) + I(t) = \lim_{t \to \infty} S(t) = S_\infty.$$

Working with the above results, we can determine a relationship between the size of the epidemic and the basic reproduction number. Division of the first equation of 4.1 by $S$ and integration from 0 to $\infty$ gives the following.[3]

---

[3]Recall that in this thesis we use "log" to denote the natural logarithm.

$$S' = -\beta SI$$

$$\Rightarrow \quad \frac{S'}{S} = -\beta I$$

$$\Rightarrow \quad \int_0^\infty \frac{S'}{S} dt = -\beta \int_0^\infty I \, dt$$

$$\Rightarrow \quad -\int_\infty^0 \frac{S'}{S} dt = -\beta \int_0^\infty I \, dt$$

$$\Rightarrow \quad \int_\infty^0 \frac{S'}{S} dt = \beta \int_0^\infty I \, dt$$

$$\Rightarrow \quad \lim_{a \to \infty} \int_a^0 \frac{S'}{S} dt = \frac{\beta}{\gamma} \int_0^\infty \gamma I \, dt$$

$$\Rightarrow \quad \lim_{a \to \infty} [\log S(t)]_a^0 = \frac{\beta}{\gamma} \int_0^\infty R'(t) dt, \quad \frac{d}{dt} \log(S(t)) = \frac{S'(t)}{S(t)}$$

$$\Rightarrow \quad \log (S_0) - \log (S_\infty) = \frac{\beta}{\gamma} \lim_{b \to \infty} [R(b) - R(0)]$$

$$\Rightarrow \quad \log \frac{S_0}{S_\infty} = \frac{\beta}{\gamma} [N - S_\infty], \quad \text{assuming } R(0) = 0,$$

$$\Rightarrow \quad \log \frac{S_0}{S_\infty} = \frac{\beta N}{\gamma} \left[ 1 - \frac{S_\infty}{N} \right].$$

We can now rewrite this final result as

$$\log \frac{S_0}{S_\infty} = \mathfrak{R}_{\mathfrak{o}} \left[ 1 - \frac{S_\infty}{N} \right]. \tag{4.4}$$

Equation 4.4 is called the *final size relation*. The final size relation gives a connection between the size of the epidemic and the basic reproduction number. Assuming there are no deaths, the number of members of the population who are infected over the course of the epidemic, the final size of the epidemic, is $N - S_\infty$. This is sometimes described in terms of the attack ratio $\left(1 - \frac{S_\infty}{N}\right)$. The final size relation given by Equation (4.4) can be generalized to more complex epidemic models, like one that includes an exposed period.

We also have an SIR model with vaccination:
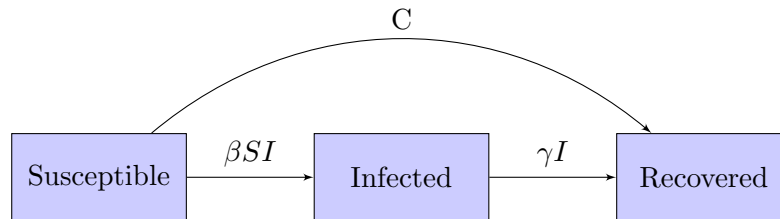


Figure 4.12: *SIR Model with Vaccination.*

We assume that the vaccinated population, C, are accounted for if they were vaccinated before the disease started to spread. Thus, the vaccinated population would start in the Recovered group. We assume no one is vaccinated during the time the disease is spread, and proceed as in the previous section.

The SIR model may work for modeling the spread of a disease such as the measles in a closed community, but we can refine and take the model a step

further by including an exposed period. We will explore this further in Section 4.4.

### 4.3.1  SIR Model Example

We will use the data from Table 4.1 to complete an SIR model. We complete our model by giving each differential equation an initial condition

$$S(0) = 762,$$

$$I(0) = 1,$$

$$R(0) = 0.$$

In terms of the scaled variables, these initial conditions are

$$s(0) = 1,$$

$$i(0) = 0.00131,$$

$$r(0) = 0.$$

We do not know the values for parameters $\beta$ and $\gamma$ yet, but we can estimate them and then adjust to better fit our data. The average time that boys were removed from the population due to the infection was five to six days. Let us propose a guess of $\gamma = \frac{1}{5} = 0.2$. If we guess that, on average, each infectious boy would infect one contact per day, then $\beta = 1$. We must remember that this is just a guess. The following model shows the solution curves for these choices of $\beta$ and $\gamma$.
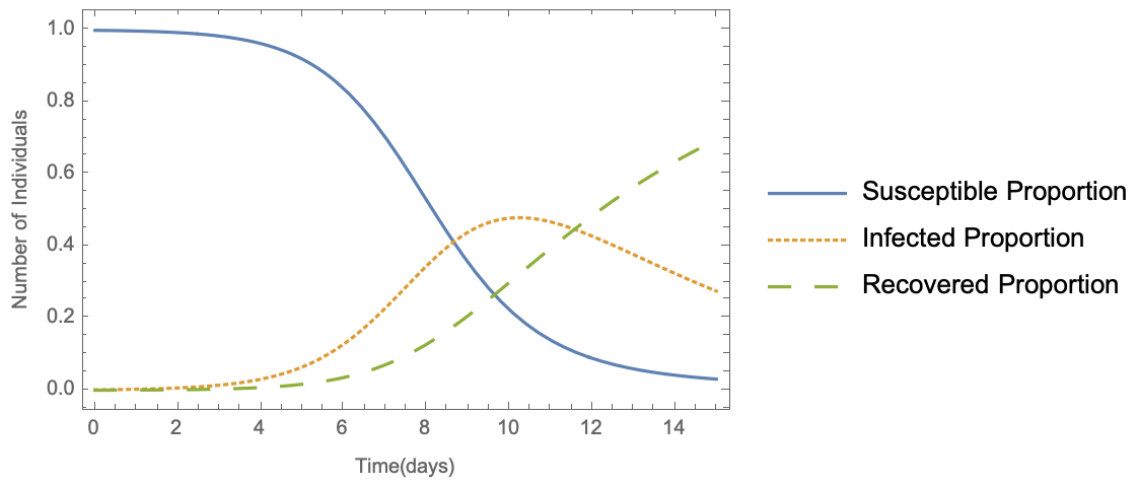
Figure 4.13: *SIR Model with $s(0) = 1, i(0) = 0.00131, r(0) = 0, \beta = 1$ and $\gamma = 0.2$.*

The average infectious period for influenza is known to be about five to seven days, so our $\gamma$ value is probably close, but our $\beta$ value was just a guess. We will focus our experiment on the infected solution, $i(t)$, since the function tells us about the progress of the infection. After experimenting with different parameter values in order to find values that fit our data, we conclude that $\beta = 1.6$ and $\gamma = 0.2$ provides a reasonably close fit for our infected data. Now let us compare our model with the data.
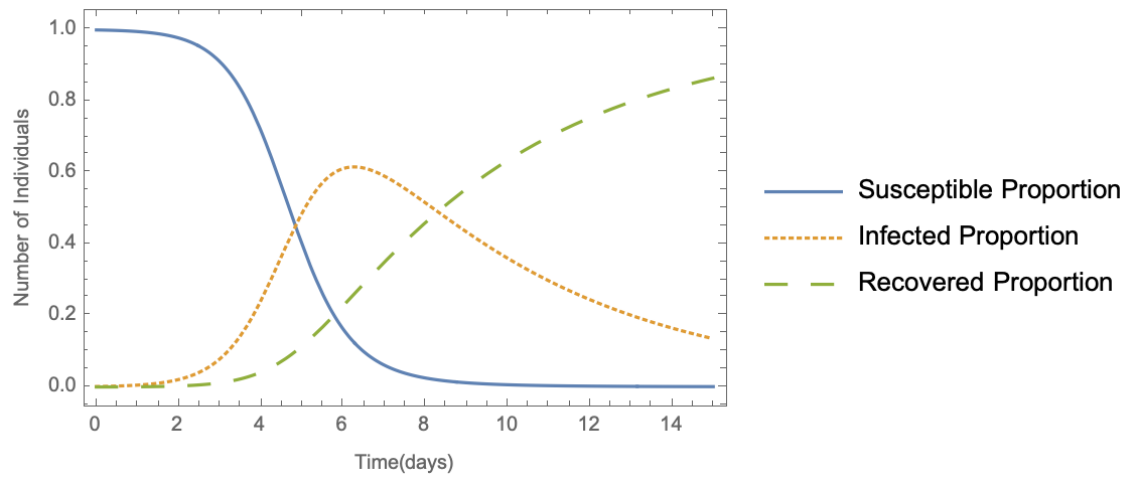
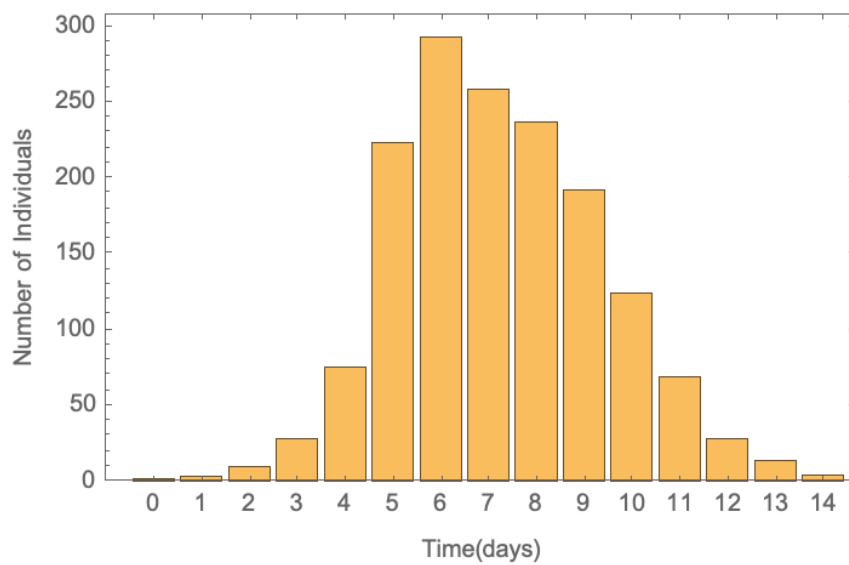Figure 4.14: *SIR Model with $s(0) = 1, i(0) = 0.00131, r(0) = 0, \beta = 1.6$ and $\gamma = 0.2$.*



Figure 4.15: *Bar Graph of the Number of Individuals in the Infected Population.*

When comparing the "Infected" curve in Figure 4.14 to Figure 4.15, we are able to see the resemblance. We see a jump in infections from day 4 to day 5, with a peak at day 6 and then a slow decline. On the other hand, a major flaw we see with this SIR model is how quickly the susceptible curve declines. Our susceptible population does decline quickly, because as soon as a student shows symptoms they are removed from the population, but this model predicts them to be removed too soon. When experimenting with the parameter values, it was not possible to get each curve to fit our real data. This is why we modified the parameter $\beta$ in section 4.2.1.

## 4.4 SEIR Model

### 4.4.1 About the Model

The SEIR model differs from the SIR model in the addition of a latency period. In many disease outbreaks, susceptible individuals go through an exposed period after they contract the infection, in which they are infected, but not yet experiencing symptoms and are not able to spread the infection to another susceptible individual. The population of exposed individuals, $E$, is incorporated by the mean exposed period, $\frac{1}{\gamma}$.

**S(t):** Number of susceptible people.

**E(t):** Number of exposed people. This represents the portion of the population who have been infected, but are not yet infectious.

**I(t):** Number of infected or infectious people.

**R(t):** Number of recovered people.
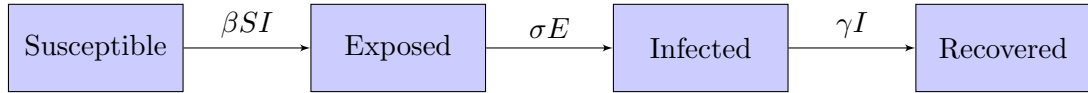


Figure 4.16: *SEIR Model.*

## 4.4.2 SEIR Without Vital Dynamics

When we assume no births or deaths with a closed population, the system of differential equations becomes:

$$\frac{dS}{dt} = -\beta SI,$$

$$\frac{dE}{dt} = \beta SI - \sigma E,$$

$$\tag{4.5}$$

$$\frac{dI}{dt} = \sigma E - \gamma I,$$

$$\frac{dR}{dt} = \gamma I,$$

where $N = S + E + I + R$ is the total population, the parameter $\beta$ represents the probability of an interaction between and infected and susceptible that results in a new infection, the parameter $\sigma$ represents the per capita rate of latent individuals becoming infectious, the parameter $\gamma$ represents the per capita recovery rate.

Our analysis of the SEIR model will be similar to that performed for the SIR model, but with $E + I$ replacing $I$. This means that the total number of

the infected population, if they are contagious or not, are considered instead of using the number of infectives as one of the variables. In some cases the exposed population may cause infectivity during the exposed period. This must be modeled by assuming infectivity is reduced by a factor which we will denote by $\varepsilon$, during the exposed period. A calculation of the rate of new infections per susceptible leads to the following model:

$$\frac{dS}{dt} = -\beta S(I + \varepsilon E),$$

$$\frac{dE}{dt} = \beta S(I + \varepsilon E) - \sigma E, \tag{4.6}$$

$$\frac{dI}{dt} = \sigma E - \gamma I,$$

with initial conditions:

$$S(0) = S_0, \; E(0) = E_0, \; I(0) = I_0.$$

For this model, the basic reproduction number is given by

$$\mathfrak{R}_{\mathfrak{o}} = \frac{\beta N}{\gamma} + \varepsilon \frac{\beta N}{\sigma} \tag{4.7}$$

because we must account for those in the exposed population that may be contagious.

In order to find our final size relation we first divide the first equation from 4.5 by $S$ and then integrate from 0 to $\infty$:

$$\log \frac{S_0}{S_\infty} = \beta \int_0^\infty [I(t) + \varepsilon E(t)] dt.$$

Using the fact that $I'(t) = \sigma E(t) - \gamma I(t)$ and substituting for $I(t)$ in the above integral, it follows that

$$\log \frac{S_0}{S_\infty} = \beta \int_0^\infty \left[ \frac{\sigma E(t) - I'(t)}{\gamma} + \varepsilon E(t) \right] dt$$

$$= \beta \int_0^\infty \left[ \frac{\sigma E(t)}{\gamma} + \varepsilon E(t) - \frac{I'(t)}{\gamma} \right] dt$$

$$= \beta \left( \frac{\sigma}{\gamma} + \varepsilon \right) \int_0^\infty E(t) dt - \frac{\beta}{\gamma} \int_0^\infty I'(t) dt$$

$$= \beta \left( \frac{\sigma}{\gamma} + \varepsilon \right) \int_0^\infty E(t) dt + \frac{\beta}{\gamma} I_0$$

Integrating the third equation from 4.5 yields

$$\sigma \int_0^\infty E(t) dt = \gamma \int_0^\infty I(t) dt - I_0.$$

It follows that

$$\log \frac{S_0}{S_\infty} = \beta \left( \frac{\sigma}{\gamma} + \varepsilon \right) \left[ \frac{\gamma}{\sigma} \int_0^\infty I(t) dt - \frac{I_0}{\sigma} \right] + \frac{\beta I_0}{\gamma}.$$

Now, integration of the sum of equations 4.5 from 0 to $\infty$ gives

$$N - S_\infty = \gamma \int_0^\infty I(t) dt.$$

Hence

$$\log \frac{S_0}{S_\infty} = \beta \left( \frac{\sigma}{\gamma} + \varepsilon \right) \frac{1}{\sigma} \left[ N - S_\infty \right] - \beta \left( \frac{\sigma}{\gamma} + \varepsilon \right) \frac{I_0}{\sigma} + \frac{\beta I_0}{\gamma}$$

$$= \frac{\beta N \sigma}{\gamma \sigma} + \frac{\beta N \varepsilon}{\sigma} \left[ 1 - \frac{S_\infty}{N} \right] - \frac{\beta I_0}{\gamma} - \frac{\varepsilon \beta I_0}{\sigma} + \frac{\beta I_0}{\gamma}.$$

Finally, we recall Equation 4.7 to arrive at our final size relation

$$\log \frac{S_0}{S_\infty} = \mathfrak{R}_o \left[ 1 - \frac{S_\infty}{N} \right] - \frac{\varepsilon \beta I_0}{\sigma}.$$

It is assumed that there are some individuals who were originally infected, that are now past the exposed stage of infectivity. Therefore, we obtain an initial term $\frac{\beta I_0}{\alpha}$ in the final size relation. It must be assumed that $I(0) = 0$ in order to obtain a final size relation without this initial term. Assuming $I(0) = 0$ is assuming that the initial infectives are in the first stage of disease, in which they can transmit the infection.

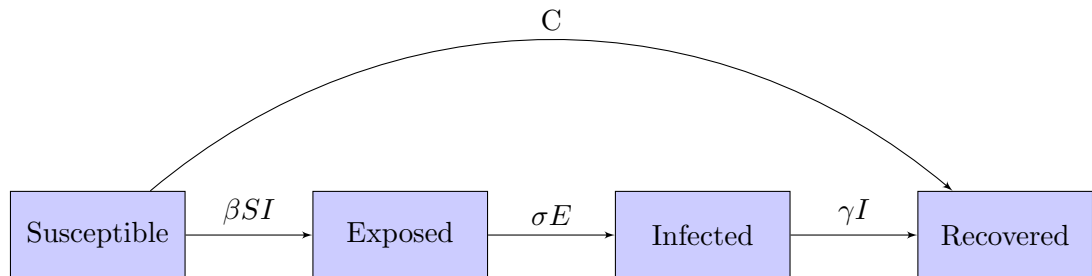We also have an SEIR model with vaccination, C:



Figure 4.17: *SEIR Model with Vaccination.*

As with the SIR model, if we wish to include vaccinated members of

the population, we will include them with the recovered group at the start of the spread.

### 4.4.3   SEIR with Birth and Death

Introducing births and deaths in the model for the spread of a disease may predict either a sustained epidemic or the further spread of the disease since new births create additional susceptible individuals. When we have a realistic population, including births and deaths, the dynamic of the disease will reach a steady state. Suppose we assume that the birth rate ($\mu$) and death rate ($\nu$) are equal. Then the population size remains constant, so the system of differential equations become:

$$\frac{dS}{dt} = \mu N - \nu S - \frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \nu E - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \nu I$$

$$\frac{dR}{dt} = \gamma I - \nu R,$$

where $N = S + E + I + R$ is the total population.[SEI19]

When the full course of an outbreak is observed, we can see that after the initial fast growth, the epidemic reduces the number of susceptible individuals in the population. This in turn reduces the chances of a susceptible individual

coming infected, so the virus will die off. Introducing an incubation period will not change the overall number of individuals who are infected.

Since the measles has a latency period, of of all the models studied so far, this type of model may be most appropriate for long term analyses.

### 4.4.4   SEIR Model Example

We will use the data from Table 4.1 and the example of the SIR model to complete the SEIR model. Again, we complete our model by giving each differential equation an initial condition

$$S(0) = 762,$$
$$E(0) = 1.6,$$
$$I(0) = 1,$$
$$R(0) = 0.$$

In terms of the scaled variables, these initial conditions are

$$s(0) = 1,$$
$$e(0) = 0.0021,$$
$$i(0) = 0.00131,$$
$$r(0) = 0.$$

We use the same values from Figure 4.14 along with an extra parameter, $\sigma$. We have $\beta = 1.6$ and $\gamma = 0.2$. The parameter $\sigma$ represents the per capita rate of latent individuals becoming infectious. The average per capita rate of latent individuals becoming infectious is 1 to 4 days. Since this is not included in our data, we will

compare different parameter values for $\sigma$ to illustrate how changes in $\sigma$ affect the model of an outbreak.



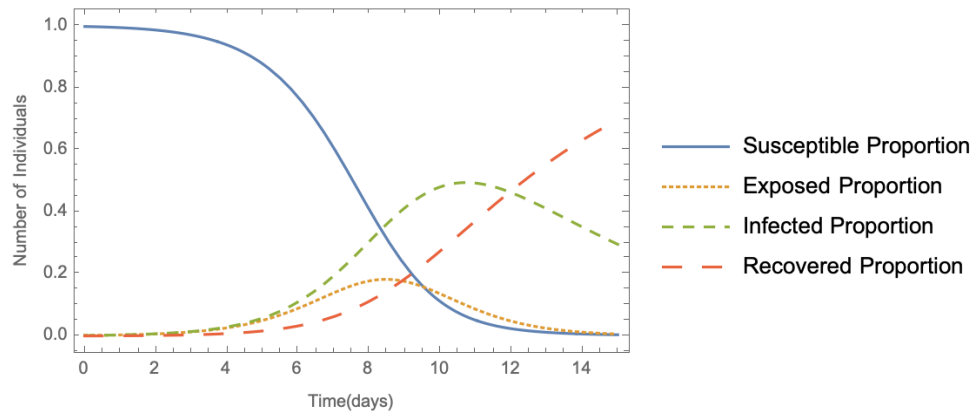Figure 4.18: *SEIR Model with $s(0) = 1, e(0) = 0.0021, i(0) = 0.00131, r(0) = 0,$ $\beta = 1.6, \sigma = 1,$ and $\gamma = 0.2$.*
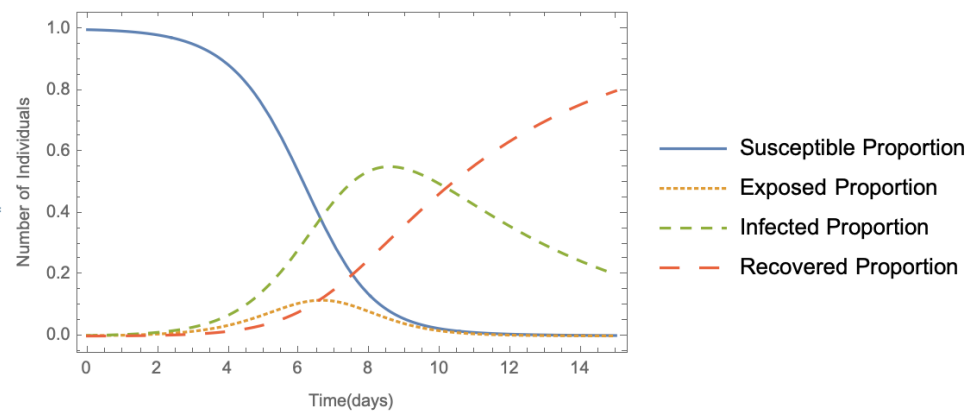


Figure 4.19: *SEIR Model with $s(0) = 1, e(0) = 0.0021, i(0) = 0.00131, r(0) = 0,$ $\beta = 1.6, \sigma = 2,$ and $\gamma = 0.2$.*
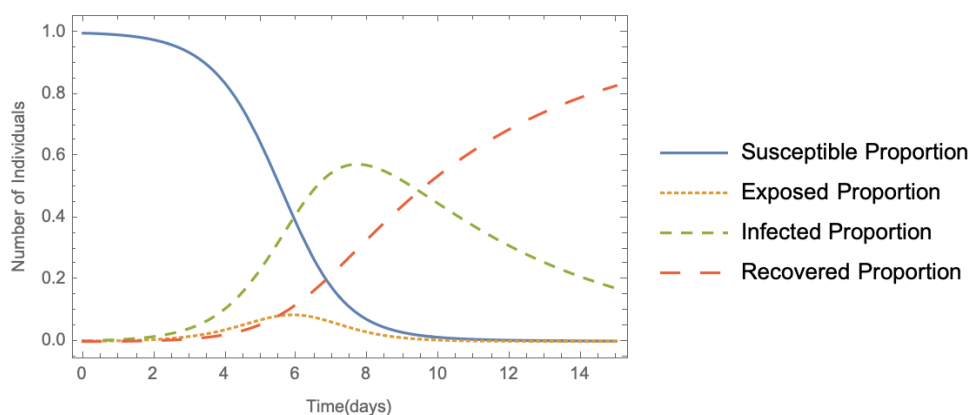
Figure 4.20: *SEIR Model with $s(0) = 1, e(0) = 0.0021, i(0) = 0.00131, r(0) = 0,$ $\beta = 1.6, \sigma = 3,$ and $\gamma = 0.2$.*
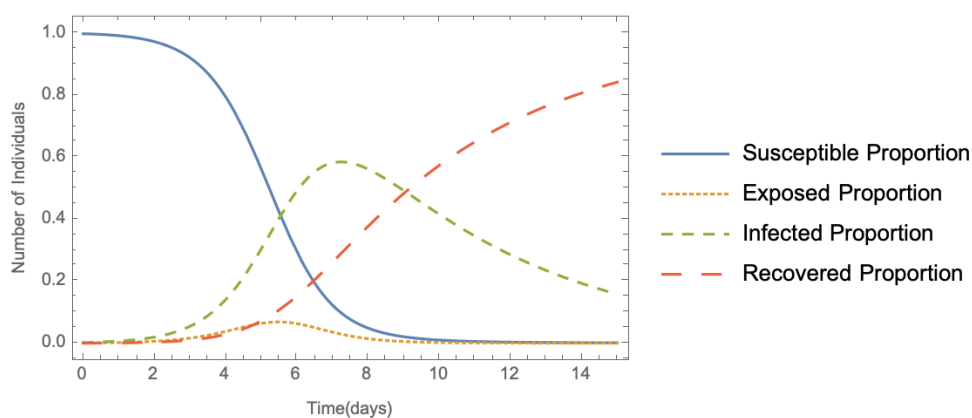


Figure 4.21: *SEIR Model with $s(0) = 1, e(0) = 0.0021, i(0) = 0.00131, r(0) = 0,$ $\beta = 1.6, \sigma = 4,$ and $\gamma = 0.2$.*

It is important to note the significant effect $\sigma$ has on our curves. The shorter the exposed period is, the slower the infection spreads. This is because symptoms show quickly, and we can remove the infected from the population before they are able to infect more individuals. What is also interesting is that when $\sigma = 1$

our "Susceptible" curve is similar to that given by our data, but our "Infected" curve does not provide a good fit. We note that when $\sigma = 4$, our "Infected" curve appears closest to fitting.

Again, we were not able to find parameter values that accurately depicted all of our data.

# Chapter 5

# Modeling Chad Outbreak

## 5.1   Outbreak of Study

In May of 2018, Chad declared a measles outbreak. This particular out-
break did not end in June when the rainy season began, like most years, and
continued into 2019. Data was collected and reported periodically. We will model
the outbreak using data from May 20, 2018 (week 20) to May 12, 2019 (week 69).
Earlier data may not have been collected because it was not clear if there was an
outbreak at that time. During the period from week 20 to week 69, it was reported
that 13,725 people were infected with measles and 123 people died. The average
age of those infected was 9.9 years old and the median age was six years old. About
39% of the affected people were four years old and below, 23% were between five
and nine years old, and 37% were 10 years old and older. Since measles out-
breaks affect Chad yearly and lifelong immunity is gained after recovering, we will
make the simplifying assumption that our population consists of children between

0 to 14 years old. The population of children in Chad aged 0 to 14 is 7,636,427. Approximately 37% of children are vaccinated, so they will be removed from the susceptible population and included in the recovered population. This gives us an initial susceptible population of 4,810,949.

Table 5.1 and Figure 5.1 show the number of newly infected individuals each week and the cumulative number of infected individuals each week. The data was reported and published to the World Health Organization. We have gaps in time because for certain weeks there were no changes reported. For example, week 23 and week 24 were not reported, although it is reasonable to assume changes took place. Therefore, our table does not include those weeks for which there was no report.

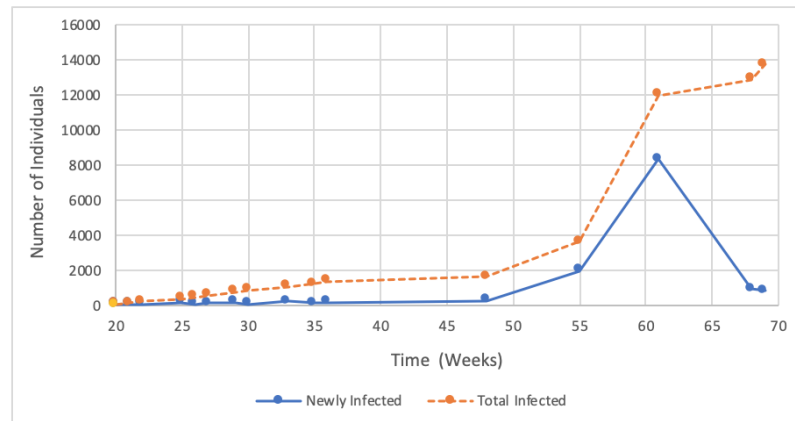| Time (Weeks) | Newly Infected | Total Infected |
|:---:|:---:|:---:|
| 20 | 77 | 77 |
| 21 | 74 | 151 |
| 22 | 66 | 217 |
| 25 | 175 | 392 |
| 26 | 91 | 483 |
| 27 | 118 | 601 |
| 29 | 162 | 763 |
| 30 | 98 | 861 |
| 33 | 237 | 1098 |
| 35 | 122 | 1220 |
| 36 | 155 | 1375 |
| 48 | 259 | 1634 |
| 55 | 2000 | 3634 |
| 61 | 8333 | 11967 |
| 68 | 912 | 12879 |
| 69 | 846 | 13725 |

Table 5.1: *Chad Measles Infections.*

Figure 5.1: *Scatterplot of Chad Measles Infections.*

### 5.1.1 Exponential Model of Chad Data

For our Chad measles data, we will analyze the total number of infected individuals each week. We begin by taking the log of the number of infected individuals, given in Table 5.2, and creating the corresponding log-linear scatterplot, Figure 5.2.

| Time (Weeks) | Total Infected | log(Total Infected) |
| --- | --- | --- |
| 20 | 77 | 4.343805422 |
| 21 | 151 | 5.017279837 |
| 22 | 217 | 5.379897354 |
| 25 | 392 | 5.97126184 |
| 26 | 483 | 6.180016654 |
| 27 | 601 | 6.398594935 |
| 29 | 763 | 6.637258031 |
| 30 | 861 | 6.7580945049 |
| 33 | 1098 | 7.001245622 |
| 35 | 1220 | 7.106606138 |
| 36 | 1375 | 7.226209017 |
| 48 | 1634 | 7.398786275 |
| 55 | 3634 | 8.198089249 |
| 61 | 11967 | 9.389908141 |
| 68 | 12879 | 9.463353357 |
| 69 | 13725 | 9.526974266 |

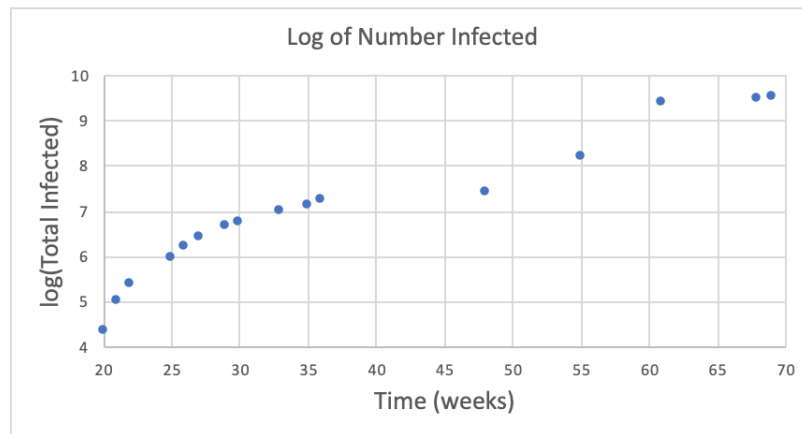Table 5.2: *Measles Data with Natural Log of Number of Infected.*



Figure 5.2: *Log-Linear Scatterplot of Measles Data.*

Notice that the trend of the data from week 20 to week 35 seems to align with that of a concave down curve, and the trend of the data from week 36 through week 61 seem to align with a concave up curve. We create Figure 5.3 and Figure 5.4 to take a closer look at this conjecture.
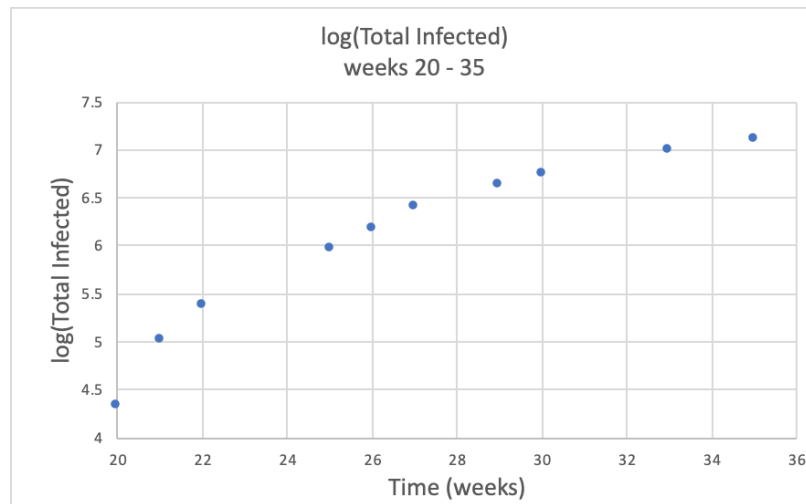


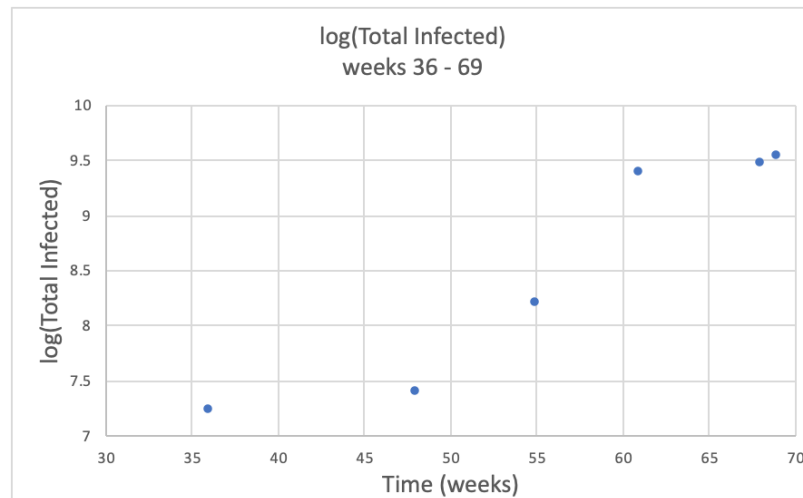Figure 5.3: *Measles Data on a Log-Linear Scatterplot: Weeks 20 to 35.*

Figure 5.4: *Measles Data on a Log-Linear Scatterplot: Weeks 36 to 69.*

It is clear that the data from weeks 20 to 35 follows a concave down and increasing trend. This tells us that the growth is not exponential, it is sub-exponential. The sub-exponential growth makes sense for growth that can be described by a sigmoid curve. Recall that in a sigmoid growth pattern the initial trend may be nearly exponential, but approaches a linear trend as the curve approaches its inflection point, at which point the rate of growth begins to decrease. Our graph may be suggesting that if the trend is sigmoid, the data being reported in weeks 20 to 35 may be near the inflection point, rather than at the beginning of the virus. It is also clear that the data from weeks 36 to 61 follow a concave up trend, which indicates faster than exponential growth. This suggests that something happened to cause the infection to spread more quickly. Until week 35, our data followed a pattern that was pretty much the shape we expected it to be if we assume a typical sigmoid growth pattern and assuming reports of data began only

after it was noticed that there was an outbreak. It makes sense that we are closer to the inflection point, rather than if we had data from the very beginning. Moreover, it may be the case that the data from weeks 20 to 35 lie along the steepest part of the sigmoid curve, which passes through the inflection point. With our conjecture that we are near an inflection point, we will fit a logistic curve to those weeks and examine the result.
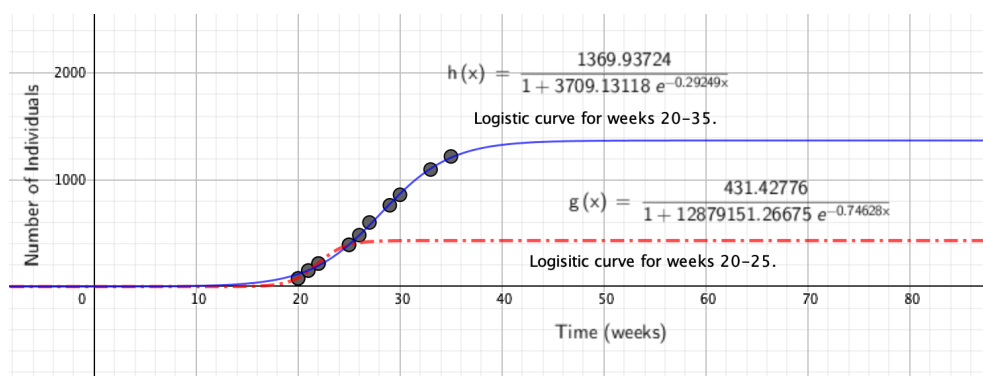
### 5.1.2   Logistic Model of Chad Data



Figure 5.5: *Measles Data on a Logistic Curve.*

If we hypothesize that we are near an inflection point and fit a logistic curve for weeks 20 to 35, then we observe that the curve fits fairly well. We get the confirmation that we might be near that inflection point and the logistic curve predicts a carrying capacity of about 1370 individuals, which is close to the mark before the unexpected continued growth in cases. It looks like the spread was slowing down as was normally the case at the start of the rainy season. Then, between weeks 36 and 48 the number of cases grew, so a lot more people than

previous years were sick, thereby sustaining the disease longer than usual. Looking at weeks 20 to 35 the model predicts an inflection point at week 26, which is mid-June and is usually when the measles outbreaks die off. In this outbreak, it appears a positive rate of increase continued longer than normal. Where the numbers of new cases should have started to decrease around week 26, a positive rate of growth is continuing, and perhaps even spiking. This may cause a larger than usual initial number of infected individuals to be brought into the new cycle. Our initial value for the next cycle is higher than it would have normally been and that is perhaps where the new jump comes from. The data suggests that perhaps there was something that caused the spread of the virus to behave more aggressively than in previous years.

We know the logistic model may not be an ideal model for the spread of measles. We are taking advantage of its sigmoid shape, characteristic of the pattern for cumulative infectives, to suggest what phase of the outbreak the data lies in. We found that the logistic curve may provide a good fit if we are near the inflection point.

### 5.1.3   SIR Model of Chad Data

We will use the data from Table 5.1 to complete an SIR model. We complete our model by giving each differential equation an initial condition

$$S(0) = 4{,}810{,}949,$$

$$I(0) = 77,$$

$$R(0) = 2{,}825{,}478.$$

In terms of the scaled variables, these initial conditions are

$$s(0) = 1,$$

$$i(0) = 0.000016,$$

$$r(0) = 0.00000035.$$

We do not know the values for parameters $\beta$ and $\gamma$ yet, but we can estimate them and then adjust them to fit our data. The average infectious period is 6 to 7 days. Let us propose a guess of $\gamma = \frac{1}{6} = 0.167$. If we guess that, on average, each infected would infect one contact per day, then $\beta = 1$. As before, we keep in mind that this is just a guess. The following model shows the solution curves for these choices of $\beta$ and $\gamma$.
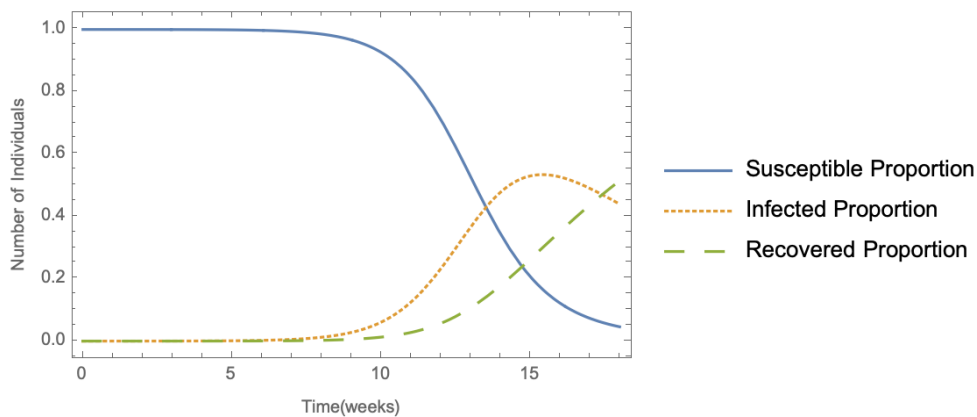


Figure 5.6: *SIR Model with $s(0) = 1, i(0) = 0.000016, r(0) = 0.00000035$, $\beta = 1$ and $\gamma = 0.167$.*

Since the average infectious period for the measles is known to be about 6 to 7 days, our $\gamma$ value is probably close, but our $\beta$ value was just a guess. We will focus our experiment on the infected solution, $i(t)$, since the function tells us

about the progress of the infection. After experimenting with different parameter values in order to find values that fit our data, we conclude that $\beta = 1.15$ and $\gamma = 0.154$ gives us an improved fit for our infected data. Now let us compare our model with the data.
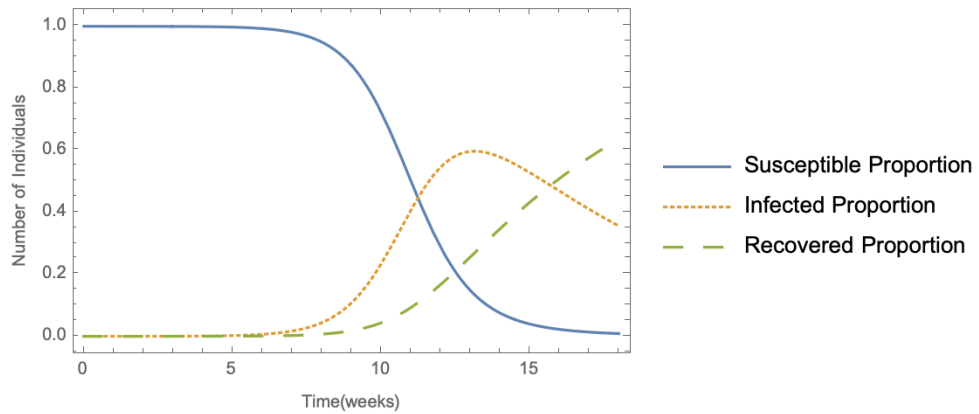


Figure 5.7: *SIR Model with* $s(0) = 1, i(0) = 0.000016, r(0) = 0.00000035, \beta = 1.15$ *and* $\gamma = 0.154$.
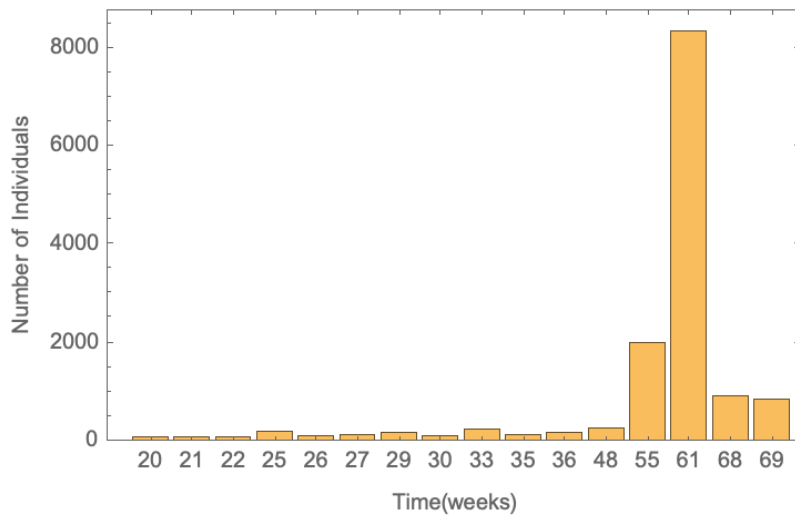
Figure 5.8: *Bar Graph.*

When comparing the "Infected" curve in Figure 5.7 to Figure 5.8, we see a slight resemblance only for the peak week. On our bar graph, Figure 5.8, we see a jump in infections from week 48 to week 55, with a peak at week 61 and then a sharp decline. A major flaw we see with this SIR model is the gradual infection growth and the gradual decline after the peak weak. When experimenting with the parameter values, we were not able to get each curve to fit our real data.

### 5.1.4 SEIR Model of Chad Data

We will use the data from Table 5.1 and the SIR model to complete an SEIR model. Again, we complete our model by giving each differential equation

an initial condition

$$S(0) = 4,810,949,$$

$$E(0) = 115.5,$$

$$I(0) = 77,$$

$$R(0) = 2,825,478.$$

In terms of the scaled variables, these initial conditions are

$$s(0) = 1,$$

$$e(0) = 0.000024,$$

$$i(0) = 0.000016,$$

$$r(0) = 0.00000035.$$

We will use the same values from Figure 4.14 along with an extra parameter, $\sigma$. We have $\beta = 1.15$ and $\gamma = 0.154$. The parameter $\sigma$ represents the per capita rate of latent individuals becoming infectious. The average per capita rate of latent individuals becoming infectious is 7 to 14. Since the exposed period is not included in our data, we will test different parameter values for $\sigma$ to show the effect on the oubreak model.

Figure 5.9: *SEIR Model with $s(0) = 1, e(0) = 0.000024, i(0) = 0.000016, r(0) = 0.00000035, \beta = 1.15, \sigma = 7,$ and $\gamma = 0.154$.*



Figure 5.10: *SEIR Model with $s(0) = 1, e(0) = 0.000024, i(0) = 0.000016, r(0) = 0.00000035, \beta = 1.15, \sigma = 8,$ and $\gamma = 0.154$.*

Figure 5.11: *SEIR Model with $s(0) = 1, e(0) = 0.000024, i(0) = 0.000016, r(0) = 0.00000035, \beta = 1.15, \sigma = 9,$ and $\gamma = 0.154$.*
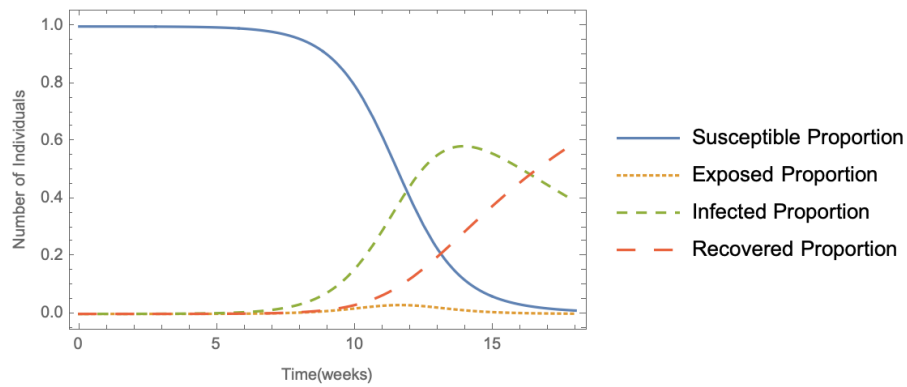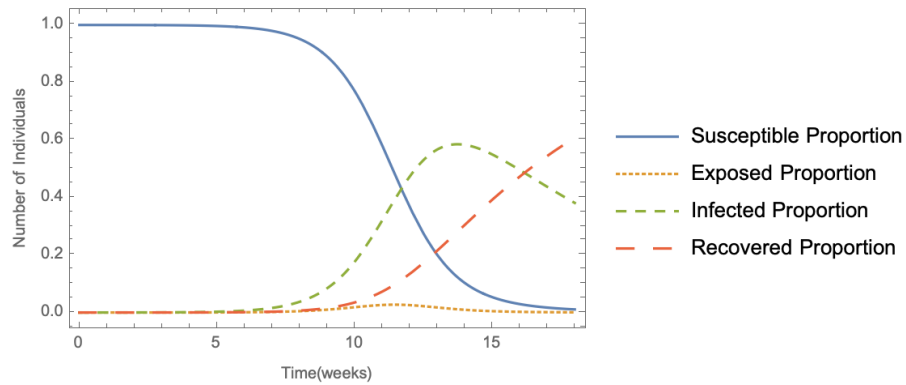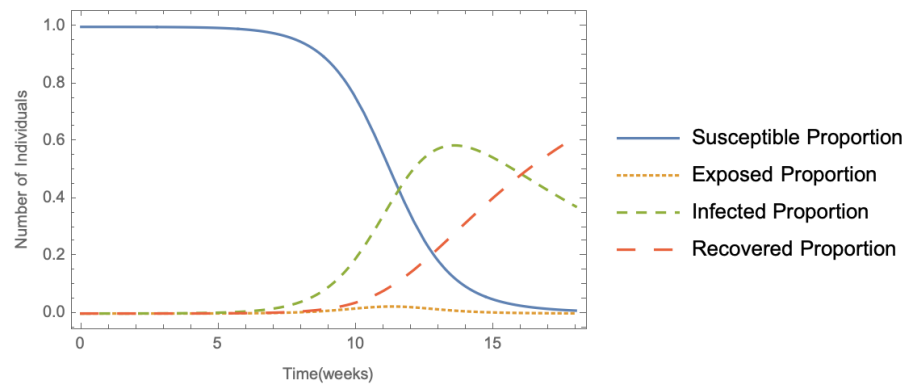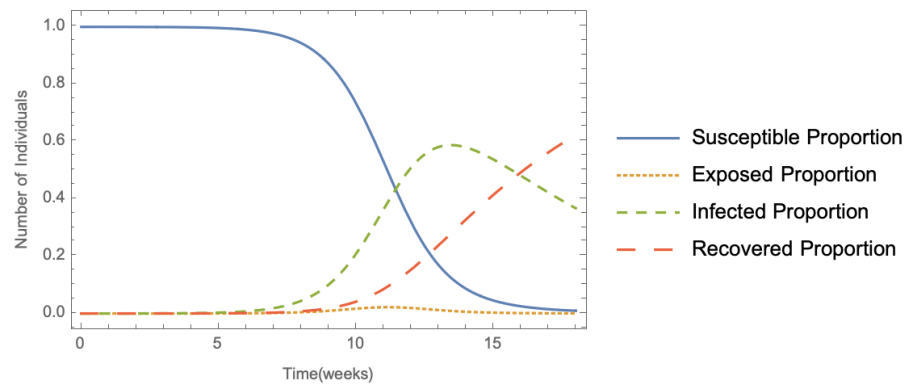


Figure 5.12: *SEIR Model with $s(0) = 1, e(0) = 0.000024, i(0) = 0.000016, r(0) = 0.00000035, \beta = 1.15, \sigma = 10,$ and $\gamma = 0.154$.*

We can see that the parameter, $\sigma$, does not have a significant impact on our SEIR models. Again, we were not able to find parameter values that accurately depicted our data. There is no way that the SIR and SEIR models could be a good fit because they predict too many infections in the population, and we know that was not the case.

If we take into account that the data from weeks 20 to 35 may be at the inflection point of the characteristic sigmoid shape, the latency period may be affecting how far the outbreak is extending into the rainy season. If we have more people getting infected than normal, and for longer into the cycle, and we also take into account that they have a latency period, then those individuals would be infecting people further into the rainy season. In this case, infections are carrying over and causing a higher initial value for the next cycle. So the latency period comes into effect when we compare the first and second outbreak cycle.

# Chapter 6

# Conclusion

The goal of this project was to experiment with different mathematical growth models and examine their suitability for modeling disease outbreaks. We compared and contrasted the exponential model, the logistic model, the SIR model, and the SEIR model. In addition, we showed how the epidemiological models are built and examined their parameters using the boarding school outbreak. We then considered each model for the spread of measles and evaluated its suitability using real data.

The exponential model may be used to say that our data is not exponential, and is actually sub-exponential. Data that follows an increasing concave down trend on a log-linear scale is exhibiting sub-exponential growth. We used the flattening out of the sub-exponential growth pattern to hypothesize that the data is at the inflection point of a sigmoid trend. We further supported this conjecture by testing the fit of a simplest sigmoid curve, the logistic model. This experiment supported our conjecture that the data could be near the inflection point. We

then noticed that where we expected our curve to flatten out, it did not. When we considered the first few weeks of the Chad data on the logistic curve, the trend appeared to flatten out where it normally would. When we extended the data longer, we saw that the data grew, instead of flattening out. This suggested that the disease was not progressing in the usual way. This led us to wonder what factors might lead to an outbreak that caused more people to get sick than usual, and further resulted in an increase in the initial number of infectious individuals moving into the next cycle. We are not sure what is causing this to happen, but we hypothesize that weather, travel, conflict, and people gathering to water sources during a drought may be possible factors.

We also found that neither the SIR nor SEIR models provided suitable fits for our data. A key problem encountered was that the total susceptible population was extremely large in relation to the amount of infections. With these models, we would also assume that the total population is equally distributed over a common area, which is not the case in Chad. The models may still be useful if modified to adapt to the situation we have. Perhaps there is a way to estimate a value to replace $S_0$ that might lead to a modified model that better represents the data. My initial hypothesis when beginning this project was that, in the end, the SEIR model would provide a good fit for the spread of measles in Chad, Africa. My results do not support this hypothesis.

We see with the boarding school data and Chad measles data, that it is very difficult to find a mathematical model that fits real data. Our models fit the boarding school data better because it was a completely closed and controlled population. With the Chad measles data, there are many factors that could be

affecting our data. There are many parameters and factors that play into the collection and analysis of real data, that create problems and complications in mathematical modeling. A question we are left with is, what do we have to do to adapt these models to create a model that makes sense?

# Bibliography

[AM82]    Roy M. Anderson and Robert M. May. Directly Transmitted Infectious Diseases: Control by Vaccination. 1982.

[BCC10]   Fred Brauer and Carlos Castillo-Chavez. *Mathematical Models in Population Biology and Epidemiology*. Springer, 2010.

[BMJ78]   Epidemiology. *British Medical Journal*, page 587, 1978.

[CDC19]   Measles (Rubeola), 2019.

[CH220]   Chad: Measles Outbreak - May 2018, 2020.

[Fro20]   Jim Frost. How To Interpret R-squared in Regression Analysis, 2020.

[JS05]    V.A.A. Jansen and N. Stollenwerk. 7.6 Modeling Measles Outbreaks. In *Branching Processes: Variation, Growth, and Extinction of Populations*, pages 236–249. Cambridge University Press, 2005.

[Man19]   Dr. Ananya Mandal. Measles history, 2019.

[POP19]   Chad Population, 2019.

[SEI19]   SEIR and SEIRS models, 2019.

[Sha97]    Alexei Sharov. Exponential model, 1997.

[WHO19]  Measles, 2019.