Project Number: MA-RYL-ACDR

Modeling Covid-19 in Massachusetts, Texas and Iowa

A Major Qualifying Project

submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

by

Justin Allen

Alexander Clopper

Javier Salido

March 19, 2021

Approved

Professor Roger Lui
Major Advisor

Contents

List of Figures

List of Tables

Abstract

CoViD-19 has rapidly spread around the world, and throughout the United States. The first case in the US was reported on January 19, in Washington State. It has already killed over 500,000 people in the US alone. Due to its rapid spread, it is difficult to research its epidemiology. New information comes to light at a speed that cannot be matched. Despite the thousands of papers written on its epidemiology and treatment, society is still struggling to control CoViD-19's spread.

In order to contribute to the research, we have chosen to use Tang et al.'s Mathematical Model on CoViD-19 spread, created using Markov Chain Monte Carlo Simulation. In proving the system of equations, we determined the Basic Reproduction Number formula. With this knowledge, we choose which parameters to research, based upon their importance in this calculation.

Although Tang et al. provided parameter values based on Wuhan, this information required examination. For some parameters, our research matched up with the values found by Tang et al. We conducted this research through two avenues: online research, for parameters with consistency between states; and, data analysis. This required us to use statistics from Massachusetts, Texas, and Iowa, which we cleaned and smoothed.

After we prepared the data, we performed Linear Regression Analysis and Constrained Non-Linear Optimization for Parameter Tuning. Our focus was on four of the parameters: $c, \beta, q,$ and θ. In our data analysis, we developed alternative parameter sets, and determined which parameters have the most impact on the model. These new parameter sets were tested in the Basic Reproduction Number formula twice, once with Tang et al.'s original parameters, and once with the updated values. This tested their accuracy and gave insight into what R_0 might be for CoViD-19.

We hope that this information is helpful for future analysis of CoViD-19, especially in determining a more accurate Basic Reproduction Number. Doing so will help understand Herd Immunity, which, with the development of the Moderna, Pfizer, and Johnson & Johnson vaccines, is becoming especially important.

Acknowledgement

There are many people who deserve appreciation for this project. Firstly, our families - John and AnnMarie Allen, Jeffrey and Carol Clopper, Ida Clayes, and Francisco and Elsa Salido have our humble gratitude for supporting us, both throughout college and during this ongoing pandemic. Without their guidance and advice, this project would not be the same.

Additionally, the professors who have taught us during our time at Worcester Polytechnic Institute deserve recognition, especially Professors Andrea Arnold, Balgobin Nandram, Ali Yousefi, and Andrew Trapp, as well as Professor Emeritus William Farr. The knowledge that these teachers passed on to us formed the foundation for the work done over the past six months.

Thirdly, thank you to our peers and our friends. To list all of you out is an impossible task, but we want you to understand that your input and camaraderie has been instrumental to our success. Your advice and support helped push this paper to new heights.

Furthermore, the countless people who have contributed to the body of research on CoViD-19, whether by compiling data on its spread or by studying the virus, must be recognized. We have stood on the shoulders of giants as we conducted our research, and their impact is inarguable.

Last, but certainly not least, thank you to Professor Lui. The guidance you provided has given all of us a chance to become better prepared for joining the workforce after graduation, and has shaped the way that we understand complex mathematical issues. Without your weekly meetings and your epidemiology crash-courses, there would be no paper. The opportunities that you provided for personal growth are innumerable, and our gratitude has no upper bound.

1 Introduction

CoViD-19 is a deadly virus that has already infected more than 120 million people worldwide [1]. The first positive case of CoViD-19 was confirmed on December 31, 2019, where a cluster of people in Wuhan, China started exhibiting symptoms of pneumonia. In January, 2020, Chinese biologists and officials announced the genetic sequence of the virus [2]. The sequence was very similar to the sequence of the Severe Acute Respiratory Syndrome (SARS) virus that mainly impacted Southern China, Vietnam, and Hong Kong in 2003. As a result, the new virus was given the name SARS-CoV-2. The virus was later renamed CoViD-19, an abbreviation of Coronavirus Disease 2019.

CoViD-19 quickly spread across the world, and has now infected 216 countries and territories [1]. The first cases in the United States were confirmed in February, 2020. Cases started spreading, and the World Health Organization (WHO) declared a pandemic on March 11, 2020 [3].

The term coronavirus is derived from the shape of the virus. Their external spikes create an appearance of a crown, which translates to "corona" in Latin [4]. The shape of coronaviruses can be seen in Figure 1:

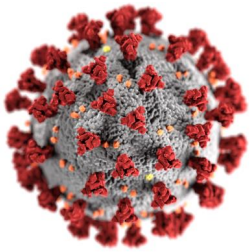

Figure 1: A picture of a coronavirus taken from [5]

Coronavirus can vary in severity. They can either be the cause of a common cold or cause more serious diseases like CoViD-19 or SARS [6]. The symptoms of CoViD-19 vary widely from person to person [7]. According to the Centers for Disease Control and Prevention (CDC), symptoms can include, but are not limited to, the following:

- Flu-like symptoms, such as running a temperature, a sore throat, nausea, vomiting, congestion, a runny nose or having chills;
- Respiratory difficulties, such as coughing, difficulty breathing, or gasping for air;
- Exhaustion, both physical and muscular;
- Loss of the senses of taste and smell; and,
- Diarrhea.

However, despite all these symptoms that can clue in doctors to the presence of CoViD-19, it has proven difficult to diagnose. Some people, while carriers of the virus, do not show any symptoms [8]. This makes tracking the virus difficult, and allows it to spread from person to person more efficiently, since asymptomatic people tend to continue to interact with others. As a result, testing for CoViD-19 is an integral step in controlling the spread of the disease, including for those who are asymptomatic.

There are two main types of tests people take to confirm if they have the virus: antigen and Polymerase Chain Reaction (PCR). Antigen tests look for antigens, or specific proteins, on the virus [9]. Antigen tests are cheaper and results can be verified within an hour, but antigen tests are prone to false negatives. PCR tests check for the RNA of the virus and duplicate it to amplify and confirm results. PCR tests are much more accurate than antigen tests, but the Polymerase Chain Reaction process can take hours or days to complete.

For the beginning of the CoViD-19 pandemic, the United States relied exclusively on human intervention to prevent the spread of CoViD-19. The current CDC guidelines recommend people wear masks whenever in public to prevent droplets potentially containing the virus to be spread throughout the air, keep a distance of at least six feet between people if possible, limiting public gatherings and avoiding crowded indoor areas, and keeping up personal hygiene such as washing hands and avoiding touching their face.

At the beginning of the pandemic, scientists and health experts were unaware of the fact that asymptomatic carriers were capable of spreading the virus to others. There was no strong emphasis on mask wearing at the time. Once this fact was discovered, the CDC updated their recommendations to include mask wearing in public.

Coronavirus has spread throughout the United States in waves. The first waves impacted major coastal cities, including New York City, Portland, Seattle, Boston, and Washington D.C. This is likely due to these cities being major places of international travel, which helped CoViD-19 spread across different countries, as well as the much larger population density of the cities. Afterward, cases started spiking in southern states and midwestern states.

The United States federal government has taken a largely hands-off approach to fighting CoViD-19. They have largely given responsibilities to the state governments to put up their own measures. Responses between states varied differently. Some states, like New York and Massachusetts, enforced stricter laws and mandates in order to fight the spread. On the other hand, some states were much less restrictive and did not enforce a statewide mask mandate when in public [10].

On October 22, the Food and Drug Administration (FDA) officially approved using Remdesivir intravenously for fighting CoViD-19, the first drug approved for such use. While human intervention should be used and promoted, Remdesivir has shown to lower the virus count in laboratory studies and animals [11].

In order to track the spread of CoViD-19, public health officials started implementing contact tracing. Contact tracing involves tracking the close contacts of people who either are probable or confirmed cases of CoViD-19. These people who are close contacts are considered "probable"

cases. CDC defines a close contact as "someone who was within 6 feet of an infected person for a cumulative total of 15 minutes or more over a 24-hour period starting from 2 days before illness onset (or, for asymptomatic patients, 2 days prior to test specimen collection) until the time the patient is isolated." The diagram below describes how contract tracing is performed:

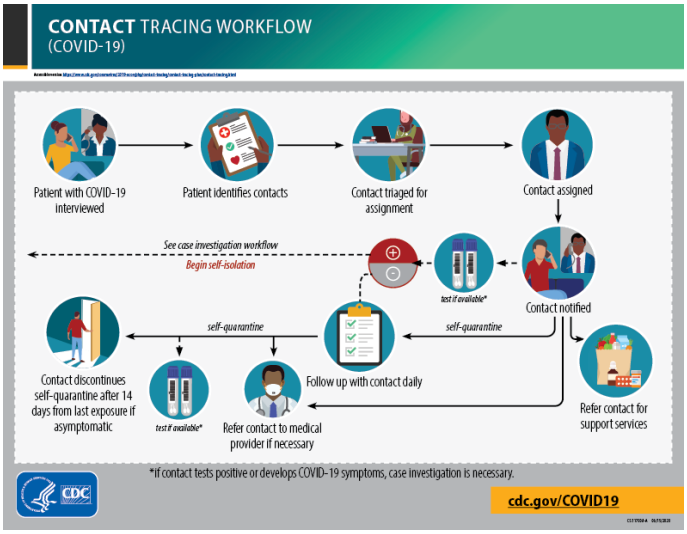

Figure 2: A chart on how contact tracing is conducted from [12]

People who are in close contact with a positive case are expected to self-quarantine for 14 days, even if a test comes back negative. CoViD-19 does not show symptoms immediately in those who become infected; there is a latency period between where people contract the virus and people become infectious. The median incubation period, time from contact to show of symptoms, for CoViD-19 is 5 days, but can vary from 1 day to 12 days [14]. This supports using 14 days as a safe length of self-quarantine.

People who do test positive are expected to self-quarantine for 14 days. The quarantine length of 14 days comes from the period in which someone is symptomatic. Most people who test positive are contagious for about 10 days [9]. After this period, if a person is no longer showing symptoms, they are very likely no longer contagious. Experts recommend the length of 14 days to confirm that a person is safe to leave quarantine.

Due to the high infection rate and the severity of some of the symptoms, it is important to add to our understanding of CoViD-19 as quickly as possible. We hope to develop further research on the spread of CoViD-19 from asymptomatic individuals based upon the level of mask mandate that is enforced.

The organization of this MQP is as follows:

- Introducing and Analyzing the Tang et al. Markov Chain Monte Carlo Simulation Model
- Selecting States for Analysis

- Processing Data
- Calculating the Results of Linear Regression Analysis and Constrained Non-Linear Optimization for Parameter Tuning, and Applying Said Results to the Basic Reproduction Number Formula
- Discussing the Methods Used, Examining the Results Obtained, and Recommending Future Research

2 Mathematical Model

The transmission diagram below is taken from Tang et al.'s Paper.

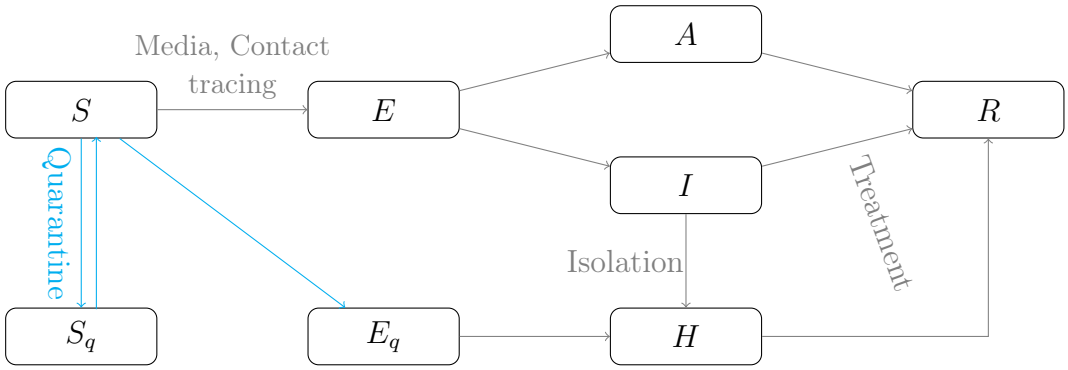

Figure 3: Transmission diagram for SARS-CoV-2.

In Figure (3),

- S - people susceptible to the disease.
- E - people exposed to someone with the disease.
- I - people who are infected and are symptomatic.
- A - people who are infected but are asymptomatic.
- S_q - people susceptible that have self-quarantined.
- E_q - people exposed that have self-quarantined .
- H - people that are hospitalized.
- R - people that have recovered from the disease.
- D - deaths due to the disease.

In order to model the pandemic, we use a version of the SEIRAH compartmental model. This model tracks the number of people who fall into each compartment and the changes in the amount of people in each compartment over time. The compartments are $S, E, I, A, S_q, E_q, H, R, D$. We assume homogeneous mixing so we can only need to consider ordinary differential equations. Based on the above transmission diagram, one can write down the model equations:

Our model is as follows:

$$
\begin{aligned}
\frac{dS}{dt} &= -(\beta c + cq(1-\beta))S(I+\theta A) + \lambda S_q \\
\frac{dE}{dt} &= \beta c(1-q)S(I+\theta A) - \sigma E \\
\frac{dI}{dt} &= \sigma\rho E - (\delta_I + \alpha + \gamma_I)I \\
\frac{dA}{dt} &= \sigma(1-\rho)E - \gamma_A A \\
\frac{dS_q}{dt} &= cq(1-\beta)S(I+\theta A) - \lambda S_q \\
\frac{dE_q}{dt} &= \beta cqS(I+\theta A) - \delta_q E_q \\
\frac{dH}{dt} &= \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \\
\frac{dR}{dt} &= \gamma_I I + \gamma_A A + \gamma_H H \\
\frac{dD}{dt} &= \alpha I + \alpha H
\end{aligned}
$$

The original set of values for the parameters is as follows:

Parameter	Est. Mean	Std	Meaning
c	14.78	0.904	Contact rate
β	2.1×10^{-8}	1.19×10^{-9}	Probability of transmission upon contact
q	1.89×10^{-7}	6.37×10^{-8}	Quarantine fraction of exposed class
σ	$1/7$	-	Transition rate from exposed class to infected class
λ	$1/14$	-	Rate of release of quarantined uninfected
ρ	0.87	4.923×10^{-2}	Probability of having symptoms
δ_I	0.13	2.132×10^{-2}	Transition rate symptomatic to quarantine infected
δ_q	0.13	5.203×10^{-2}	Transition rate of quarantined exposed to quarantine infected
θ	Unknown	Unknown	Reduced transmission of asymptomatic class of this standard deviation
γ_I	0.330	5.214×10^{-2}	Recovery rate of symptomatic infected
γ_A	0.140	3.462×10^{-2}	Recovery rate of asymptomatic infected class
γ_H	0.116	3.873×10^{-2}	Recovery rate of quarantine infected
α	1.782×10^{-5}	6.833×10^{-6}	Covid-19 induced death rate

Table 1: **Original Parameter Values and Meanings.** Taken from Table 1 of [13]. Unit of time is days.

We began the model with the values for the parameters from Table 1 from [13]. We then decided to find updated values for the parameters. A study of CoViD-19 cases in Singapore found that the mean latency period, or period lasting from when a person is infected to when they become contagious, is 5.5 days [14]. The transition rate from exposed to infected is the reciprocal of the latency period. A team of researchers from Bond University in Australia did a large meta-analysis of studies of CoViD-19 cases. In this study, 663 people tested positive and 111 people (16.7%) were asymptomatic; 552 people (83.3%) developed symptoms. The same study also found that the relative risk of transmission from asymptomatic people is 42% lower than from symptomatic people. These results are reflected in the following table.

Parameter	Est. Mean	Std	Meaning
c	14.78	0.904	Contact rate [13]
β	2.1×10^{-8}	1.19×10^{-9}	Probability of transmission upon contact [13]
q	1.89×10^{-7}	6.37×10^{-8}	Quarantine fraction of exposed class [13]
$\boldsymbol{\sigma}$	**1/5.5**	**0.13**	**Transition rate from exposed to infected [14]**
λ	1/14	-	Rate of release of quarantined uninfected
$\boldsymbol{\rho}$	**0.83**	**0.015**	**Probability of having symptoms [15]**
δ_I	0.13	2.132×10^{-2}	Transition rate symptomatic to quarantine infected [13]
δ_q	0.13	5.203×10^{-2}	Transition rate of quarantined exposed to quarantine infected [13]
$\boldsymbol{\theta}$	**0.58**	**0.21**	**Reduced transmission of asymptomatic [15]**
$\boldsymbol{\gamma_I}$	**1/10**	**-**	**Recovery rate of symptomatic infected [9]**
$\boldsymbol{\gamma_A}$	**1/6**	**-**	**Recovery rate of asymptomatic infected**
$\boldsymbol{\gamma_H}$	**1/11**	**-**	**Recovery rate of quarantine infected [17]**
$\boldsymbol{\alpha}$	**0.0221**	**-**	**Covid-19 induced death rate [1]**

Table 2: **Updated Parameter Values and Meanings.** Unit of time is days.
Updated parameters are in bold.

Symptomatic people are contagious for an average of approximately 10 days [9], and asymptomatic people are contagious for an average of 6 days. A study of in Belgium found the mean hospital stay due to CoViD-19 is 11.0 days [17]. Recovery rates are the reciprocal of these periods of time. The death rate from CoViD-19 is found by taking the fraction of total cases and total deaths [1].

For the parameters γ_I, γ_A, and γ_H, we assumed that the distributions of the length of time in which a person is contagious (or hospitalized in the case of γ_H) are normal with means of 10, 6, and 11 days respectively, due to the sufficiently large sample size. As a result, the parameters themselves each follow a reciprocal normal distribution. The reciprocal normal distribution does not have a mean, and as a consequence, does not have a standard deviation. The values in the table are instead the median values, since the reciprocal normal distribution is symmetric about the median.

3 Analysis of the Model

To have confidence in the model used, some basic proofs must be completed to show that the system of equations used is appropriate. This includes proofs that: show that the overall population is constant; a proof of the steady-states of the system of equations; and a proof of the Basic Reproduction Number of the system of equations.

Lemma 3.1. *Let* $N := S + E + I + A + S_q + E_q + H + R + D$*, then* $\dot{N} = 0$*, which implies that* $N(t) = N_0 := N(0)$ *for* $t > 0$*. This means the population size remains unchanged for all time.*

Proof.

$$
\begin{aligned}
\dot{N} &= -(\beta c + cq(1-\beta))S(w) + \lambda S_q + \beta c(1-q)S(w) - \sigma E + \sigma\rho E - \delta_I + \alpha + \gamma_I)I \\
&\quad + \sigma(1-\rho)E - \gamma_A A + cq(1-\beta)S(w) - \lambda S_q + \beta cqS(w) - \delta_q E_q + \delta_I I + \delta_q E_q \\
&\quad - (\alpha + \gamma_H)H + \gamma_I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= -\cancel{\beta cSw} - cqSw + cq\beta Sw + \lambda S_q + \cancel{\beta cSw} - \beta cqSw - \sigma E + \sigma\rho E - \delta_I I - \alpha I - \gamma_I I \\
&\quad + \sigma E - \sigma\rho E - \gamma_A A + cqSw - cq\beta Sw - \lambda S_q + \beta cqSw - \delta_q E_q + \delta_I I + \delta_q E_q - \alpha H - \gamma_H H \\
&\quad + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= -\cancel{cqSw} + cq\beta Sw + \lambda S_q - \beta cqSw - \sigma E + \sigma\rho E - \delta_I I - \alpha I - \gamma_I I + \sigma E - \sigma\rho E \\
&\quad - \gamma_A A + \cancel{cqSw} - cq\beta Sw - \lambda S_q + \beta cqSw - \delta_q E_q + \delta_I I + \delta_q E_q - \alpha H - \gamma_H H \\
&\quad + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= \cancel{cq\beta Sw} + \lambda S_q - \beta cqSw - \sigma E + \sigma\rho E - \delta_I I - \alpha I - \gamma_I I + \sigma E - \sigma\rho E - \gamma_A A - \cancel{cq\beta Sw} \\
&\quad - \lambda S_q + \beta cqSw - \delta_q E_q + \delta_I I + \delta_q E_q - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= \cancel{\lambda S_q} - \beta cqSw - \sigma E + \sigma\rho E - \delta_I I - \alpha I - \gamma_I I + \sigma E - \sigma\rho E - \gamma_A A - \cancel{\lambda S_q} + \beta cqSw \\
&\quad - \delta_q E_q + \delta_I I + \delta_q E_q - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= -\cancel{\beta cqSw} - \sigma E + \sigma\rho E - \delta_I I - \alpha I - \gamma_I I + \sigma E - \sigma\rho E - \gamma_A A + \cancel{\beta cqSw} - \delta_q E_q \\
&\quad + \delta_I I + \delta_q E_q - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= -\cancel{\sigma E} + \sigma\rho E - \delta_I I - \alpha I - \gamma_I I + \cancel{\sigma E} - \sigma\rho E - \gamma_A A - \delta_q E_q + \delta_I I + \delta_q E_q \\
&\quad - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= \cancel{\sigma\rho E} - \delta_I I - \alpha I - \gamma_I I - \cancel{\sigma\rho E} - \gamma_A A - \delta_q E_q \\
&\quad + \delta_I I + \delta_q E_q - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= -\cancel{\delta_I I} - \alpha I - \gamma_I I - \gamma_A A - \delta_q E_q + \cancel{\delta_I I} + \delta_q E_q - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \alpha I + \alpha H \\
&= -\cancel{\alpha I} - \gamma_I I - \gamma_A A - \delta_q E_q + \delta_q E_q - \alpha H - \gamma_H H + \gamma_I I + \gamma_A A + \gamma_H H + \cancel{\alpha I} + \alpha H \\
&= -\cancel{\gamma_I I} - \gamma_A A - \delta_q E_q + \delta_q E_q - \alpha H - \gamma_H H + \cancel{\gamma_I I} + \gamma_A A + \gamma_H H + \alpha H \\
&= -\cancel{\gamma_A A} - \delta_q E_q + \delta_q E_q - \alpha H - \gamma_H H + \cancel{\gamma_A A} + \gamma_H H + \alpha H \\
&= -\cancel{\delta_q E_q} + \cancel{\delta_q E_q} - \alpha H - \gamma_H H + \gamma_H H + \alpha H \\
&= -\cancel{\alpha H} - \gamma_H H + \gamma_H H + \cancel{\alpha H} \\
&= -\cancel{\gamma_H H} + \cancel{\gamma_H H} \\
&= 0\,.
\end{aligned}
$$

The proof of the lemma is complete. □

Lemma 3.2. *The only steady-state of the model is* $(S, 0, 0, 0, 0, 0, 0, R, D)$*, where* $S + R + D = N_0$*. Thus, in the long-run, part of the susceptible population remain uninfected (susceptible), while the rest of the population become infected and at the end either recovered or died.*

Proof. Steady states satisfy the equations:

$$
\begin{aligned}
&-(\beta c + cq(1-\beta))S(I+\theta A) + \lambda S_q = 0 \\
&\beta c(1-q)S(I+\theta A) - \sigma E = 0 \\
&\sigma\rho E - (\delta_I + \alpha + \gamma_I)I = 0 \qquad (3.1)\\
&\sigma(1-\rho)E - \gamma_A A = 0 \\
&cq(1-\beta)S(I+\theta A) - \lambda S_q = 0 \qquad (3.2)\\
&\beta cqS(I+\theta A) - \delta_q E_q = 0 \\
&\delta_I I + \delta_q E_q - (\alpha + \gamma_H)H = 0 \qquad (3.3)\\
&\gamma_I I + \gamma_A A + \gamma_H H = 0 \qquad (3.4)\\
&\alpha I + \alpha H = 0\,.
\end{aligned}
$$

Since $S, E, I, A, S_q, E_q, H, R, D \geq 0$ and the parameters $c, \beta, q, \sigma, \lambda, \rho, \delta_I, \delta_q, \theta, \gamma_I, \gamma_A, \gamma_H, \alpha > 0$. From equation (3.4), $I, A, H = 0$ and from (3.3), $E_q = 0$. Additionally, from (3.1), $E = 0$ and from (3.2), $S_q = 0$. S, R, D cannot be shown to be 0 in all steady states. If $R, D = 0$, The CoViD-19 free steady state is found, and $S = N_0$; that is $(N_0, 0, 0, 0, 0, 0, 0, 0, 0)$. Otherwise, the steady state is $(N, 0, 0, 0, 0, 0, 0, R, D)$. Since $N = S + E + I + A + S_q + E_q + H + R + D$, we have $N_0 = S + R + D$ The proof of the lemma is complete. □

Lemma 3.3. *The basic reproduction number , R_0, of this model is given by (2.10) of the [13].*

Proof. The next-generation method is used to determine the basic reproduction number. Details of this method is given in Appendix A.

Equations $\dot{E}$, $\dot{I}$, $\dot{A}$, $\dot{A}$, and $\dot{H}$ are the only relevant elements of $\dot{N}$ for the basic reproduction number. After expansion, the relevant equations are as follows:

$$
\begin{aligned}
\dot{E} &= \beta c(1-q)SI + \beta c(1-q)S\theta A - \sigma E \\
\dot{I} &= \sigma\rho E - (\delta_I + \alpha + \gamma_I)I \\
\dot{A} &= \sigma(1-\rho)E - \gamma_A A \\
\dot{E}_q &= \beta cqSI + \beta cqS\theta A - \delta_q E_q \\
\dot{H} &= \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H
\end{aligned}
$$

These equations can be rewritten:

$$
\begin{aligned}
\dot{E} &= F_1 - V_1 \\
\dot{I} &= F_2 - V_2 \\
\dot{A} &= F_3 - V_3 \\
\dot{E}_q &= F_4 - V_4 \\
\dot{H} &= F_5 - V_5,
\end{aligned}
$$

where

$$
\begin{aligned}
F_1 &= \beta c(1-q)S(I+\theta A) \\
F_2 &= 0 \\
F_3 &= 0 \\
F_4 &= \beta cqS(I+\theta A) \\
F_5 &= \delta_I I + \delta_q E_q
\end{aligned}
$$

and

$$
\begin{aligned}
V_1 &= \sigma E \\
V_2 &= -\sigma\rho E + (\delta_I + \alpha + \gamma_I)I \\
V_3 &= -(\sigma - \sigma\rho)E + \gamma_A A \\
V_4 &= \delta_q E_q \\
V_5 &= (\alpha + \gamma_H)H\,.
\end{aligned}
$$

Thus, the Jacobian Matrices for $F = (F_1, F_2, F_3, F_4, F_5)$ with respect to (E, I, A, E_q, H) is

$$
\mathcal{F} = \begin{bmatrix}
0 & \beta cS - \beta cqS & \beta cS\theta - \beta cqS\theta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & \beta cqS & \beta cqS\theta & 0 & 0 \\
0 & \delta_I & 0 & \delta_q & 0
\end{bmatrix}
$$

Similarly, the Jacobian matrix of $V = (V_1, V_2, V_3, V_4, V_5)$ with respect to the same variables is

$$
\mathcal{V} = \begin{bmatrix}
\sigma & 0 & 0 & 0 & 0 \\
-\sigma\rho & \delta_I + \alpha + \gamma_I & 0 & 0 & 0 \\
\sigma\rho - \sigma & 0 & \gamma & 0 & 0 \\
0 & 0 & 0 & \delta_q & 0 \\
0 & 0 & 0 & 0 & \alpha + \gamma_H
\end{bmatrix}
$$

Therefore,

$$
\mathcal{V}^{-1} = \begin{bmatrix}
\frac{1}{\sigma} & 0 & 0 & 0 & 0 \\
\frac{\rho}{\delta_I + \alpha + \gamma_I} & \frac{1}{\delta_I + \alpha + \gamma_I} & 0 & 0 & 0 \\
\frac{1-\rho}{\gamma_A} & 0 & \frac{1}{\gamma_A} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\delta_q} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\alpha + \gamma_H}
\end{bmatrix}
$$

The next generation matrix is defined as

$$\mathcal{K} = \mathcal{F}\mathcal{V}^{-1} = \begin{bmatrix} \frac{S\beta c\theta(q-1)(\rho-1)}{\gamma_A} & -\frac{S\beta c\rho(q-1)}{\delta_I+\alpha+\gamma_I} & \frac{-S\beta c(q-1)}{\delta_I+\alpha+\gamma_I} & \frac{-S\beta c\theta(q-1)}{\gamma_A} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{S\beta cq\rho}{\delta_I+\alpha+\gamma_I} & -\frac{S\beta cq\theta(\rho-1)}{\gamma_A} & \frac{S\beta cq}{\delta_I+\alpha+\gamma_I} & \frac{S\beta cq\theta}{\gamma_A} & 0 & 0 \\ \frac{\delta_I\rho}{\delta_I+\alpha+\gamma_I} & & \frac{\delta_I}{\delta_I+\alpha+\gamma_I} & 0 & 1 & 0 \end{bmatrix}$$

The basic reproduction number, R_0, is the eigenvalue of $\mathcal{K}$ furthest from the origin. The characteristic equation is $det(\xi\mathcal{I} - \mathcal{K}) = 0$, where $\mathcal{I}$ represents the identity matrix. The eigenvalues are: $\xi = [0, 0, 0, \xi_4]$, where ξ_4 equals

$$\xi_4 = \frac{S\alpha\beta c\theta + S\beta c\delta_I\theta + S\beta c\gamma_A\rho + S\beta c\gamma_I\theta - S\alpha\beta cq\theta - S\alpha\beta c\rho\theta - S\beta c\delta_I q\theta - S\beta c\gamma_A q\rho}{\delta_I\gamma_A + \alpha\gamma_A + \gamma_A\gamma_I}$$

$$+\frac{-S\beta c\delta_I\rho\theta - S\beta c\gamma_I q\theta - S\beta c\gamma_I\rho\theta + S\alpha\beta cq\rho\theta + S\beta c\delta_I q\rho\theta + S\beta c\gamma_I q\rho\theta}{\delta_I\gamma_A + \alpha\gamma_A + \gamma_A\gamma_I}$$

Upon simplification,

$$R_0 = -\frac{S\beta c\gamma_A\rho - S\beta c\gamma_A q\rho}{(\delta_I\gamma_A + \alpha\gamma_A + \gamma_A\gamma_I}$$

$$+\frac{\theta(S\alpha\beta c + S\beta c\delta_I + S\beta c\gamma_I - S\alpha\beta cq - S\alpha\beta c\rho - S\beta c\delta_I q - S\beta c\delta_I\rho}{\delta_I\gamma_A + \alpha\gamma_A + \gamma_A\gamma_I}$$

$$+\frac{-S\beta c\gamma_I q - S\beta c\gamma_I\rho + S\alpha\beta cq\rho + S\beta c\delta_I q\rho + S\beta c\gamma_I q\rho}{\delta_I\gamma_A + \alpha\gamma_A + \gamma_A\gamma_I}$$

$$R_0 = -\frac{S\beta c\gamma_A\rho - S\beta c\gamma_A q\rho}{(\delta_I\gamma_A + \alpha\gamma_A + \gamma_A\gamma_I} + \frac{S\beta c\theta(q-1)(\rho-1)}{\gamma_A}$$

$$R_0 = -\frac{S\beta c\rho(q-1)}{\delta_I+\alpha+\gamma_I} + \frac{S\beta c\theta(q-1)(\rho-1)}{\gamma_A}$$

$$R_0 = \frac{S\beta c\rho(1-q)}{\delta_I+\alpha+\gamma_I} + \frac{S\beta c\theta(1-q)(1-\rho)}{\gamma_A}$$

$$R_0 = S\left[\frac{\beta c\rho(1-q)}{\delta_I+\alpha+\gamma_I} + \frac{\beta c\theta(1-q)(1-\rho)}{\gamma_A}\right].$$

The proof of the lemma is complete. □

4 Covid-19 in Massachusetts, Texas and Iowa

We downloaded our data from The COVID Tracking project, which collected data from all fifty states on cases, hospitalizations, tests, test results, deaths, recoveries, and how many people were put in intensive care units. We wanted to choose three states that took different approaches to fighting the spread of the virus and got affected at different times. We also wanted states with a lot of mixing between people within the state because it best fits our model. We considered using metropolitan areas instead of states because metropolitan areas are the most interconnected. The metropolitan areas were split up into counties. However, we had difficulty finding all of the data we needed from counties, so we decided to use states.

We chose Massachusetts, Iowa, and Texas as a result of where their people are located. We wanted to avoid people travelling in and out of state on a constant basis. Since New York City's metropolitan statistical area (MSA) includes parts of northern New Jersey and eastern Pennsylvania, and New York City's combined statistical area (CSA) also includes parts of southwestern Connecticut, we decided to not use New York [19]. We also did not use New Jersey or Pennsylvania because Philadelphia's MSA includes parts of New Jersey, Delaware and Maryland [20].

When we attempted to find accurate 2020 Census data, it was not available. Instead, we utilized predictions of 2019 populations in each state and metropolitan area, based on previous census records.

4.1 Massachusetts

Massachusetts is a small, densely populated state in New England. The virus hit Massachusetts especially hard in the early stages of the pandemic. Massachusetts has a large percentage of their population living in these three metropolitan areas:

- Boston-Cambridge-Newton (population: 4,873,019 [21])
- Worcester (population: 947,404 [21])
- Springfield (population: 697,382 [21])

A significantly smaller percentage of these metropolitan areas include people from other states than New York City and Philadelphia. This means that fewer people will be traveling in and out of the state everyday, and allows us to have more accurate data with fewer outside influences; however, it is important to note that the Boston-Cambridge-Newton population includes parts of New Hampshire, and the Worcester population includes parts of Connecticut, and that a portion of Massachusetts's population living in the Providence, Rhode Island metropolitan area. These metropolitan areas make up 94.56368% [22] of Massachusetts's overall population (6,892,503), but this is inaccurate due to the aforementioned issues.

The first case of the virus in Massachusetts was identified on March 2, [23]. Governor Charlie Baker announced a state of emergency for Massachusetts on March 10 [24]. Baker

issued a stay-at-home order starting March 23. This was to last until April 7, but was extended to May 4 on March 31 [25]. The stay-at-home order was then extended to May 18 on April 28. Starting on May 18, Massachusetts started a phase-by-phase approach to reopening businesses across the state.

At 9:30 on May 1, Baker issued an order that mandated wearing face masks when social distancing was not possible, and required masks in retail settings and in all public transportation. This took effect on May 6 and stayed in effect until it was revised by a later order on November 2. The November 2 order mandated wearing face masks in all public places, even if social distancing is possible, and allows for businesses and schools to require proof for exemption of wearing a mask. The revised order took effect November 6.

On July 24, Governor Baker issued an order introducing a travel form that all adults and unaccompanied minors must fill out prior to entering unless they fall into one of the following exceptions:

- Coming from a lower risk state. This list changes based on rolling averages of positive test rates and cases per capita. As of November 18, only Maine, New Hampshire, Vermont, and Hawaii are considered low-risk states.
- Transitory travel. This means people traveling into Massachusetts so they can travel to another state. Examples include traveling along Interstate 95 from Rhode Island to New Hampshire, and someone from New Hampshire traveling to Boston's Logan Airport to catch a flight.
- Commuting for work or school. People must be traveling to a fixed location at least once a week and this only applies to work or school.
- Patients receiving or seeking medical treatment. Those who are required to accompany the patient are exempt from the travel form as well.
- People in the military.
- Workers providing critical infrastructure services.

Visitors and residents who do not meet at least one of the exemptions above must either self-quarantine for 14 days or produce a negative test result that was administered within 72 hours of entering the state [26]. Those who fail to comply are required to pay a fine of $500 per day.

4.2 Texas

Texas is a large state in southern United States with a large population. Texas was chosen for similar reasons as Massachusetts. We wanted a state that did not take as many measures as Massachusetts to fight CoViD-19 but did enact some rules and regulations. Texas has a large portion of its population in its metropolitan areas:

- Dallas-Fort Worth-Arlington (population: 7,573,136 [21])
- Houston-The Woodlands-Sugarland (population: 7,066,141 [21])
- San Antonio-New Braunfels (population: 2,550,960 [21])
- Austin-Round Rock-Georgetown (population: 2,227,083 [21])
- McAllen-Edinburg-Mission (population: 868,707 [21])
- El Paso (population: 844,124 [21])

These six metropolitan areas do not spread into any other state and make up about 72.872940% [22] of Texas's population (28,995,881).

Texas first declared a disaster proclamation on March 13 [27]. On March 31, governor Greg Abbott issued what is effectively a stay-at-home order that would take effect on April 2 and last until April 30 [28]. Abbott signed executive orders on April 27 to start reopening; Texas then started a fast reopening plan on April 30.

Governor Abbott issued an order that mandates people wear masks in most public places that took effect on June 3 [29]. On June 26, Abbott signed an executive order that lowered the capacity of food services [30]. Bars were shut down from 50% capacity effective that day, and restaurant capacity was lowered from 75% to 50% effective June 29. Abbott then issued a statewide executive order stating masks must be worn indoors and when maintaining social distancing is infeasible on July 2. People under the age of 10 and those in counties with fewer than 20 active cases are exempt.

On November 11, Texas became the first state to surpass one million cases [31]. However, Abbott has stated he opposes enforcing another lockdown [32].

4.3 Iowa

Iowa is a landlocked, less densely populated state in the Midwest. We wanted a state that took few to no actions in combating the virus. Iowa is one of two states that have not enforced statewide mask mandates for any measure by early November [33], the other state being South Dakota [34]. We chose Iowa over South Dakota for our third state because Iowa has about quadruple the population than South Dakota. Both states have a comparable proportion of people in metropolitan areas. Iowa's major metropolitan area, Des Moines-West Des Moines, has a population of about 699,292 [21], 22.1641% [22] of Iowa's population (3,155,070). There is a little bit of mixing with Nebraska in the Omaha-Council Bluffs MSA and with Illinois in the Davenport-Moline-Rock Island MSA, but we decided the portion of the population in this category was small enough to where we can use Iowa.

Governor Kim Reynolds first issued a disaster proclamation on March 9 [35]. This disaster proclamation has been renewed continually throughout the pandemic. Interestingly, in the

April 27 disaster proclamation renewal, Reynolds loosened social distancing measures in 77 of Iowa's 99 counties that took effect on May 1 [36].

Iowa was one of five states to not issue any sort of stay-at-home order through April [37]. However, certain cities in the state have issued city-wide mask mandates. On July 21, Iowa City put out a mask mandate that took effect immediately and lasted until September 15, [38]. On September 14, Iowa City Mayor Bruce Teague extended the mask mandate until November 13 [39]. Des Moines enforced their own mask mandate on August 26, effective immediately [40]. Waukee, a suburb of Des Moines located 15 miles to the west, issued their own mask mandate on September 15, which took effect on September 18 [41]. This mask mandate applied to anyone three years old or older and did not have any exceptions. Governor Reynolds put out the first statewide mask mandate for Iowa on November 16, which took effect the next day [42].

5 Cleaning and Smoothing of Downloaded Data

COVID Tracking did a great job aggregating raw data from across the United States and centralizing all of the information on their site. Throughout the course of this project, they continuously updated their definitions and made changes to the data in an effort to make it more accessible and more uniform. At the time of writing, we pulled the data from `https://covidtracking.com/data`. However, not all states reported their data uniformly, and there were features in the data to suggest that certain data was gathered once a week or following another pattern. Since our goal was to use this data to inform our model, a handful of preparations needed to be made. The starting date for the data collection was April 30, since by then most sources were reporting data for not just CoViD-19 deaths and cases, but hospitalizations as well. The end date of November 30 was chosen both because more data was not available yet and because the trends were likely to change in response to vaccines and the significant differences in social behavior precipitated by the holiday season.

Since the model seeks to simulate data for each day, the data needed to be cleaned and smoothed into a comparable set. This means that mitigating the impact of errors in reporting and patterns is needed to attain a format more accurate to the true daily trends. The cleaning involved understanding the origin of outliers and removing or modifying them as needed. In the case of holes or missing values in the data, a regression was fit to the data using a subset of the previous 7 days as predictors for the last day and then the missing day was predicted using this regression. That is:

$$\mathcal{D}_k = \beta_1\mathcal{D}_{k-1} + \beta_2\mathcal{D}_{k-2} + \beta_3\mathcal{D}_{k-3} + \beta_4\mathcal{D}_{k-4} + \beta_5\mathcal{D}_{k-5} + \beta_6\mathcal{D}_{k-6} + \beta_7\mathcal{D}_{k-7} + \beta_0 + \epsilon$$

Where $\mathcal{D}_k$ is a day, $\mathcal{D}_{k-1}$ is the day before $\mathcal{D}_k$, $\mathcal{D}_{k-2}$ is the day before c_{k-1}, etc. For each state and each column, a specific subset of previous days was selected to predict $\mathcal{D}_k$ using a backwards step-wise approach. This was done in response to the varying nature of the data sources, as information such as hospital cases are collected differently than CoViD-19 test results. These regressions were used solely for filling in "holes"- missing values not immediately preceded or followed by more missing values. We used Python for the cleaning and smoothing of the data, including the Statsmodels package for regressions and Pandas for the data frames. After backwards-stepwise model selection for the auto-regressions, the holes were filled in with the following code:

```
# Data table for the regression
arf = pd.DataFrame(nf.loc[(cIndex+7):k-1, newcol].values, columns=['y'])
arf.reset_index(drop=True, inplace=True)
X7 = arf.iloc[:,1:len(var)+1]
y = arf['y']
X7 = sm.add_constant(X7)

# Gather data for the point we want to fill in
for pred in var:
    predS = 'x{0}'.format(pred)
    arf[predS] = nf.loc[(cIndex+7-pred):(k-1-pred),newcol].values
    kpred.append(nf.at[k-pred,newcol])

# Generate regression model
model1 = sm.OLS(y, X7).fit()

#Fill in the missing data point with the prediction
prediction = round(model1.predict(kpred)[0])
if prediction < 0:
    prediction = 0
nf.at[k,newcol] = prediction
```

Figure 4: Python Code for Autoregressions

To smooth the data, a moving average of 7 was used. The smoothing needed to be a moving average of around 7 or more to handle the noise as well as the cyclical nature of patterns present in the data which occasionally led to consistent changes in the data (such as an uptick in hospitalizations on Mondays after a dip on Sundays). The smooth value s_k of a point $\mathcal{D}_k$ was computed as:

$$s_k = \text{Average}\left(\mathcal{D}_{k-3}, \mathcal{D}_{k-2}, \mathcal{D}_{k-1}, \mathcal{D}_k, \mathcal{D}_{k+1}, \mathcal{D}_{k+2}, \mathcal{D}_{k+3}\right).$$

Once the holes were removed, the tail ends of the data were prepared and the rest of the data was easily smoothed according to the above equation using some code:

```
#Generate moving avg
for i in range(cIndex+3,n-2,1):
    nf.at[i,mavg7]= round(np.average(nf.loc[i-3:i+3, newcol].values))
```

Figure 5: Python Code for 7-Day Average Smoothing

Below are graphs of the data from the three states: Massachusetts, Texas and Iowa, before and after the smoothing process of taking the moving 7-day average.

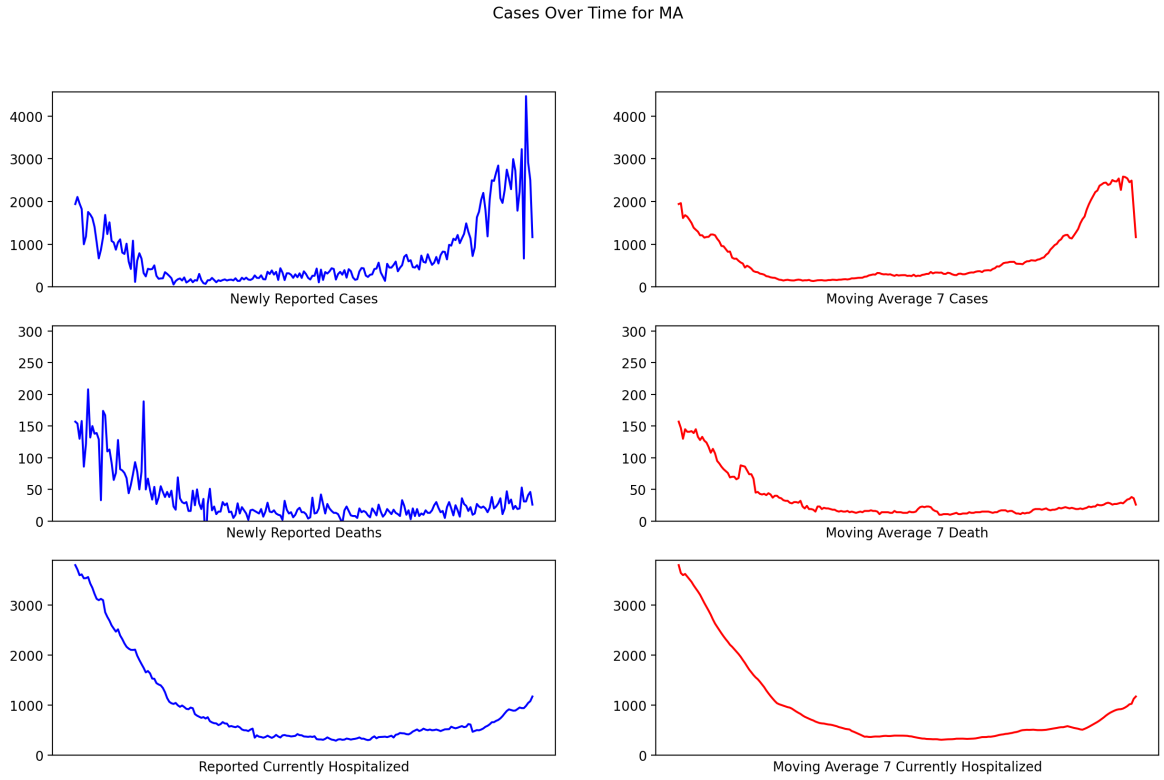

Figure 6: Massachusetts Data Smoothing

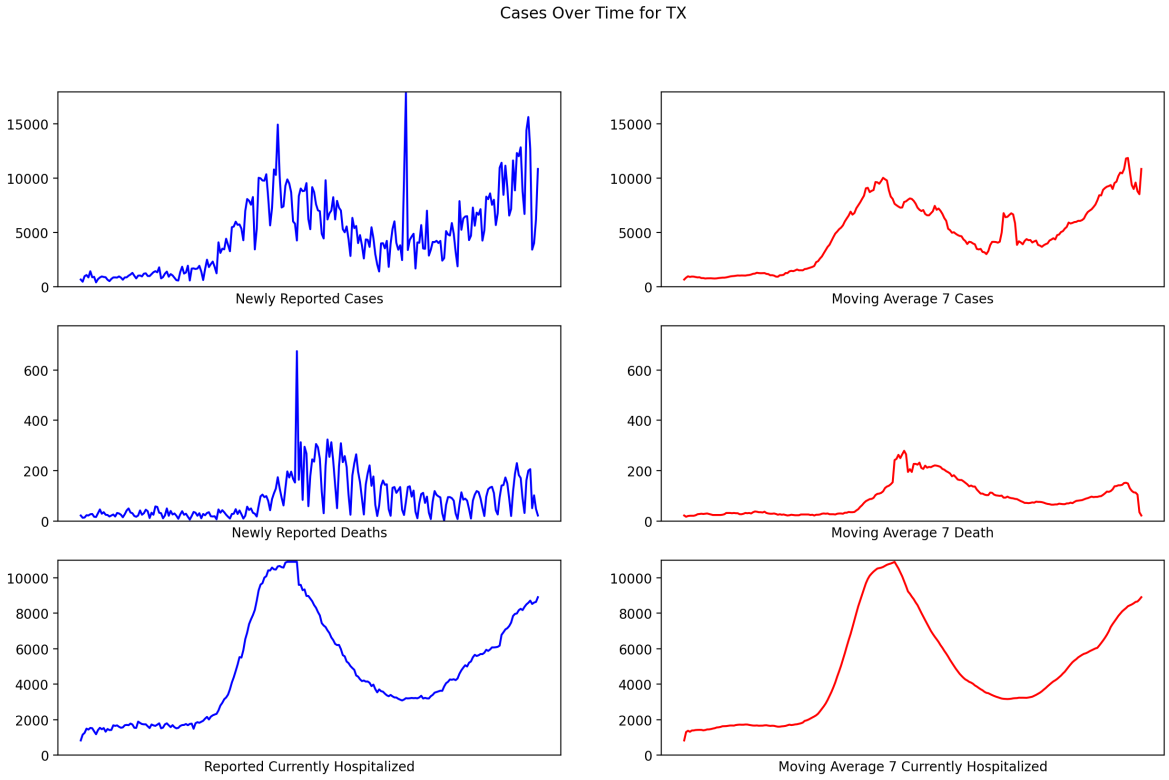

Figure 7: Texas Data Smoothing

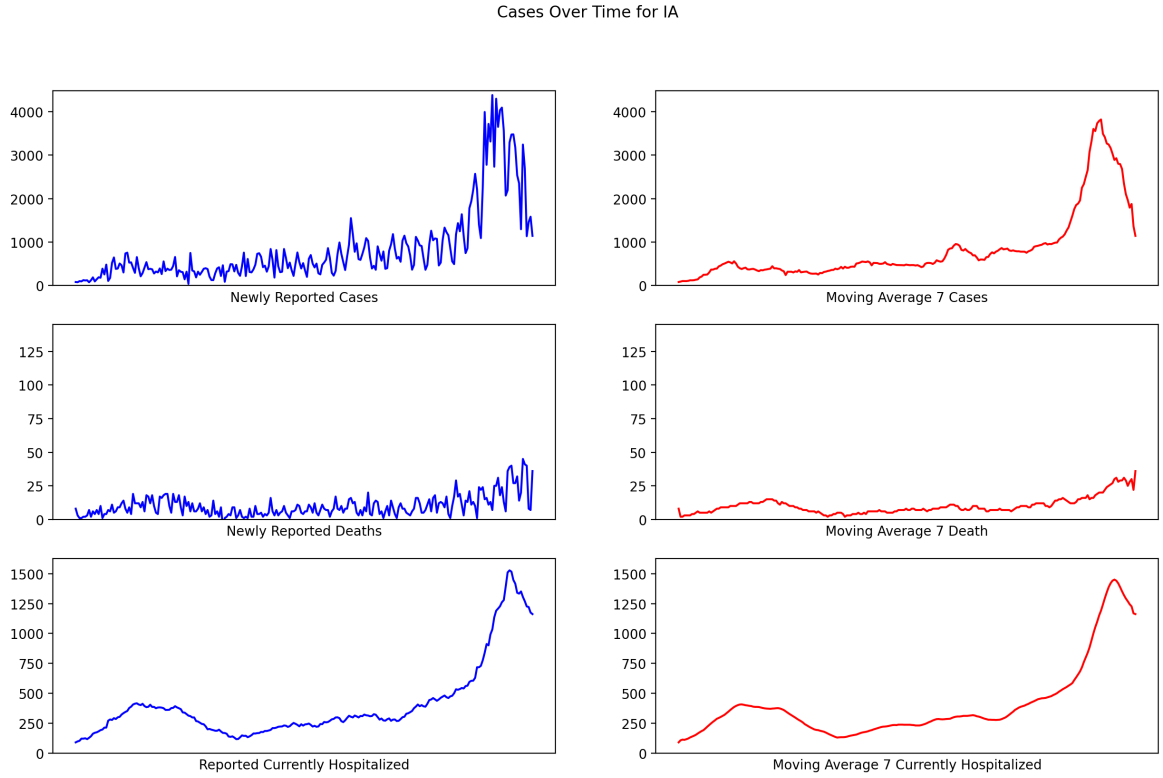

Figure 8: Iowa Data Smoothing

6 Results

After the data was prepared, two analyses of the model were performed. This was done to gain a better understanding of four model parameters: the contact rate c, chance of transmission upon contact β, the fraction of the exposed class in quarantine q, and the ratio of the infectiousness from the asymptomatic class to the infectiousness of the symptomatic class θ. First we approached the parameters with Linear Regression Analysis and then Constrained Non-Linear Optimization for Parameter Tuning. Each was important for different reasons.

Linear Regression Analysis shows the strength of the model which parameters have the greatest impact on the results. This is important for determining which combination of parameters best fits the model without adding excessive predictors. Determining parameters can help inform how to slow the spread of CoViD-19, as targeting some of these parameters in real life involves public action and legislation. This would also be useful for future research, including other approaches to parameter tuning, where these parameters can be targeted and adjusted to according to various criteria. This analysis was conducted using a MATLAB program to generate a Latin Hypercube Sample, and R to perform the linear regression.

Constrained Non-Linear Optimization for Parameter Tuning adjusts values for the target parameters to better fit the data we gathered. By observing different sets of parameters, researchers can have a better idea of how adjusting one parameter might have an impact on the others, as well as the overall model accuracy. Having improved estimates of these parameters can further our understanding of CoViD-19's transmission, the efficacy of public policies, and differences in symptomatic and asymptomatic transmission. Again, this was determined using MATLAB.

The results for both analysis methods are explained in greater detail, below.

6.1 Linear Regression Analysis

In order to figure out how to best combat CoViD-19, we wanted to determine which parameters were the most impactful and which parameters had little to no impact on the results. We generated a Latin Hypercube Sample 2000 times for the four parameters we were testing, as well as the state variables S, I, A, and H. We then ran a linear regression, modeling the state variables using the predictors c, β, q, θ, and all second-order cross terms.

The linear regression is presented by a table of estimated coefficients, the standard error, the normalized t-statistic value, and the probability of exceeding the absolute value of t, also called a p-value. The t-statistic follows a distribution very close to the standard normal distribution, but has more variance to take into account the size of the sample. Along with the tables, the linear regression gives us the degrees of freedom, the coefficient of determination R^2, F-statistic, and the p-value associated with the F-statistic. The coefficient of determination, bounded by zero and one, is a measure of how much of the variance in the data is explained by the model. R^2 equal to 0.500 means that 50% of the variance in the data is captured by the model in the

regression. The adjusted R^2, denoted by $\bar{R}^2$, takes into account the degrees of freedom in the model, as adding predictors to the model will always increase R^2 but not always increase $\bar{R}^2$. The F-statistic gives an idea of how well the model fits the data. A higher F-statistic indicates the model is a good fit for the data.

We looked at the individual p-values to determine which predictors impacted the data the most. The predictors that have the strongest impact have the lowest p-values and are considered significant. The null hypothesis was that a predictor is considered insignificant to the model and could have a coefficient of zero. If the p-value is below a certain threshold, then the null hypothesis is rejected and the predictor is considered significant. We used a p-value threshold of 0.05 to determine if predictors were considered significant.

```
Call:
lm(formula = S ~ c + q + beta + theta + cc + cq + cb + ct + qq +
    qb + qt + bb + bt + tt, data = lhs)

Residuals:
    Min      1Q  Median      3Q     Max
-6809.4 -2050.4  -172.6  1709.8  9919.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.559e+06  5.087e+04  69.965  < 2e-16 ***
c           -1.897e+05  3.765e+03 -50.392  < 2e-16 ***
q            6.625e+11  1.778e+11   3.726   0.0002 ***
beta        -6.482e+13  1.676e+12 -38.673  < 2e-16 ***
theta       -1.944e+05  9.316e+04  -2.086   0.0371 *
cc           3.228e+03  9.872e+01  32.698  < 2e-16 ***
cq           6.182e+09  6.645e+09   0.930   0.3524
cb           1.731e+12  5.996e+10  28.874  < 2e-16 ***
ct           1.384e+04  2.245e+03   6.165 8.52e-10 ***
qq           1.061e+18  4.736e+17   2.240   0.0252 *
qb          -5.169e+19  4.502e+18 -11.482  < 2e-16 ***
qt           3.349e+11  1.685e+11   1.988   0.0470 *
bb           4.694e+20  3.914e+19  11.993  < 2e-16 ***
bt          -1.243e+13  1.510e+12  -8.229 3.38e-16 ***
tt           1.254e+04  6.433e+04   0.195   0.8455
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2869 on 1985 degrees of freedom
Multiple R-squared:  0.9978,    Adjusted R-squared:  0.9977
F-statistic: 6.329e+04 on 14 and 1985 DF,  p-value: < 2.2e-16
```

Figure 9: Regression on S.

```
Call:
lm(formula = I ~ c + q + beta + theta + cc + cq + cb + ct + qq +
    qb + qt + bb + bt + tt, data = lhs)

Residuals:
    Min      1Q  Median      3Q     Max
-135.61  -27.82   -0.30   29.19  135.13

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.388e+04  7.106e+02  33.604  < 2e-16 ***
c            7.707e+02  5.258e+01  14.656  < 2e-16 ***
q            5.141e+10  2.484e+09  20.698  < 2e-16 ***
beta        -2.967e+11  2.341e+10 -12.676  < 2e-16 ***
theta        8.756e+02  1.301e+03   0.673    0.501
cc          -1.385e+01  1.379e+00 -10.041  < 2e-16 ***
cq           1.674e+09  9.282e+07  18.031  < 2e-16 ***
cb          -2.326e+10  8.376e+08 -27.768  < 2e-16 ***
ct           1.969e+00  3.136e+01   0.063    0.950
qq          -5.357e+17  6.615e+15 -80.982  < 2e-16 ***
qb           9.457e+18  6.288e+16 150.385  < 2e-16 ***
qt           1.906e+10  2.354e+09   8.099 9.59e-16 ***
bb          -3.915e+19  5.467e+17 -71.603  < 2e-16 ***
bt          -1.993e+11  2.110e+10  -9.447  < 2e-16 ***
tt           2.435e+02  8.986e+02   0.271    0.786
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.07 on 1985 degrees of freedom
Multiple R-squared:  0.9989,    Adjusted R-squared:  0.9989
F-statistic: 1.283e+05 on 14 and 1985 DF,  p-value: < 2.2e-16
```

Figure 10: Regression on I.

```
Call:
lm(formula = A ~ c + q + beta + theta + cc + cq + cb + ct + qq +
    qb + qt + bb + bt + tt, data = lhs)

Residuals:
    Min      1Q  Median      3Q     Max
-66.343  -7.926   0.518   8.398  58.840

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.986e+03  3.380e+02  20.669  < 2e-16 ***
c            1.843e+02  2.501e+01   7.367 2.54e-13 ***
q            1.490e+10  1.181e+09  12.617  < 2e-16 ***
beta        -8.191e+10  1.113e+10  -7.356 2.76e-13 ***
theta        4.316e+02  6.189e+02   0.697    0.486
cc          -3.172e+00  6.558e-01  -4.837 1.42e-06 ***
cq           5.194e+08  4.415e+07  11.765  < 2e-16 ***
cb          -6.338e+09  3.984e+08 -15.910  < 2e-16 ***
ct          -1.474e+00  1.491e+01  -0.099    0.921
qq          -1.552e+17  3.146e+15 -49.336  < 2e-16 ***
qb           2.669e+18  2.991e+16  89.227  < 2e-16 ***
qt           5.720e+09  1.119e+09   5.110 3.53e-07 ***
bb          -1.107e+19  2.600e+17 -42.588  < 2e-16 ***
bt          -5.753e+10  1.003e+10  -5.733 1.14e-08 ***
tt          -8.231e+01  4.274e+02  -0.193    0.847
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.06 on 1985 degrees of freedom
Multiple R-squared:  0.9968,    Adjusted R-squared:  0.9968
F-statistic: 4.404e+04 on 14 and 1985 DF,  p-value: < 2.2e-16
```

Figure 11: Regression on A.

```
Call:
lm(formula = H ~ c + q + beta + theta + cc + cq + cb + ct + qq +
    qb + qt + bb + bt + tt, data = lhs)

Residuals:
     Min       1Q   Median       3Q      Max
-132.998  -28.899    0.608   31.257  122.983

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  3.679e+04  7.756e+02   47.442   <2e-16 ***
c            1.146e+03  5.739e+01   19.965   <2e-16 ***
q            6.482e+10  2.711e+09   23.915   <2e-16 ***
beta        -2.954e+11  2.555e+10  -11.561   <2e-16 ***
theta        2.289e+03  1.420e+03    1.612    0.107
cc          -2.039e+01  1.505e+00  -13.550   <2e-16 ***
cq           2.464e+09  1.013e+08   24.320   <2e-16 ***
cb          -3.324e+10  9.141e+08  -36.362   <2e-16 ***
ct          -2.586e+01  3.422e+01   -0.756    0.450
qq          -8.265e+17  7.220e+15 -114.476   <2e-16 ***
qb           1.451e+19  6.863e+16  211.430   <2e-16 ***
qt           2.427e+10  2.569e+09    9.450   <2e-16 ***
bb          -6.089e+19  5.967e+17 -102.055   <2e-16 ***
bt          -2.636e+11  2.303e+10  -11.448   <2e-16 ***
tt          -1.652e+01  9.807e+02   -0.017    0.987
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.73 on 1985 degrees of freedom
Multiple R-squared:  0.9992,    Adjusted R-squared:  0.9992
F-statistic: 1.765e+05 on 14 and 1985 DF,  p-value: < 2.2e-16
```

Figure 12: Regression on H.

Our regression resulted in $\bar{R}^2$ values of 0.9968 or greater, meaning almost all of the variance in the data is captured by the model in the linear regression. As a result, the model is an excellent fit for the data.

Our regression on S, I, A, and H shows most of the predictors are significant in some way. Regression on I, A, and H resulted in the same variables being considered significant. The predictor θ^2 is not significant for any of the four state variables in the regression. θ and $c\theta$ were only significant on S. θ had a p-value of 0.0371, indicating that it is not as significant to our model as most of the other predictors, though $c\theta$ had a p-value of 8.52×10^{-10} when regressing on S.

On the other hand, c, q, and β were all very significant in impacting S, I, A, and H. Every second order cross term involving these three parameters were all significant in each regression, with the exception of cq having a p-value of 0.3524 in the regression on S. From this, we can conclude that quarantine and mask wearing habits will have a significant impact on the pandemic.

6.2 Constrained Non-Linear Optimization for Parameter Tuning

As part of this investigation, we used the data gathered to tune parameters which were likely to vary with the culture and environment, or not enumerated in the literature. When comparing our model to the collected data, it becomes unclear what the relationship between features of the collected data and most of the compartments in our model may be. In the case of the infected class, it is not immediately obvious whether this should follow the new positive CoViD-19 tests, some transformation of that data, or if a calculated column from multiple

compartments in the model might be a better fit. Out of the different compartments in our model, only H, or the number of individuals currently hospitalized, is directly represented in the collected data from different states. Thus, we tuned the four parameters to minimize the mean-squared error between the model and the hospitalization data for Massachusetts. We chose Massachusetts because we were more confident in the integrity of their hospitalization data, and because their hospital system was not overwhelmed between the months represented in the simulated data.

The bounds for the parameters were set using their 95% confidence intervals established through Table 2, or through common sense since c, β, and q must be greater than 0, and β must be less than 1. We relaxed the bounds for β because it was likely to change with environment and strain of virus. Similarly, the bounds for q were relaxed because the previous estimate seemed too restrictive for the stage in the pandemic which we analyzed, because people were more informed and therefore willing to quarantine. The problem is as follows:

$$\begin{aligned} \underset{c,\beta,q,\theta}{\text{minimize}} \quad & MSE = \frac{1}{215}\sum_{k=1}^{215}(H_k - \mathcal{H}_k)^2 \\ \text{subject to} \quad & 0 < c < 17 \\ & 0 < \beta < 1 \\ & 0 \leq q < 0.12 \\ & 0.168 \leq \theta \leq 0.992 \end{aligned}$$

Where H_k is the number of hospitalized patients predicted by the model on day k and $\mathcal{H}_k$ is the corresponding number of hospitalized patients in Massachusetts according to the data. The model reevaluates each time there is a change in the parameter values. With MATLAB's *fmincon* function, local minima were found according to the interior point method. The initial MSE (Mean Squared Error) before tuning was extremely high 1.186×10^{11}.

	c	β	q	θ	MSE
Original Values	14.78	2.1×10^{-8}	1.89×10^{-7}	0.58	1.186×10^{11}
First Minimum	2.84	3.15×10^{-7}	1.51×10^{-3}	0.991	3.188×10^{5}
Second Minimum	4.83	5.68×10^{-6}	2.88×10^{-2}	0.168	3.177×10^{5}
Third Minimum	16.72	8.42×10^{-8}	4.21×10^{-4}	0.726	3.181×10^{5}

Table 3: **Tuned Parameter Values.** In no particular order, these minima represent possible combinations of values for these parameters in the state of Massachusetts.

Based on these results, it is most apparent that the tuning favors increasing the value of the quarantine class by a significant margin, well outside of the 95% confidence interval for the

Wuhan data. However, this idea requires further development, and is beyond the scope of this analysis.

There exists the possibility that perhaps the value of c is much lower than 14.78 in Massachusetts, though this reasonably means β might also be greater than the initial estimates. If c and β follow these pattern, then the value of the quarantine class, q, is likely higher than the value in Wuhan during the time when the initial data was gathered. The values for θ vary wildly from minimum to minimum, which is supported in the research since the 95% confidence interval is so large. This could point to many things, including differences across strains of CoViD-19, but it is difficult to be certain. Without further research and more concrete estimates of these parameters, it is difficult to draw more conclusions without simply speculating. Out of the minima obtained, it is important to note that the first minimum was found at the maximum value for θ, making it less favorable.

Below are two graphs showing the difference in comparing the original model to the data using the parameters from Wuhan to the tuned parameters values for the second minimum in Table 3.

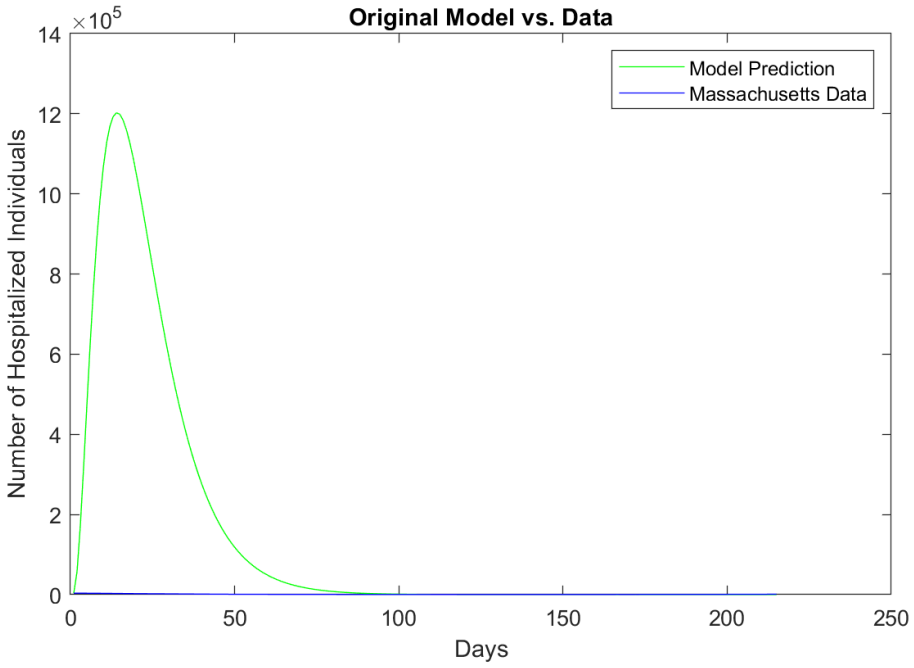

Figure 13: Simulated versus actual MA hospitalization data with Wuhan parameters. Graph corresponds to the original values of c, q, β, and θ.

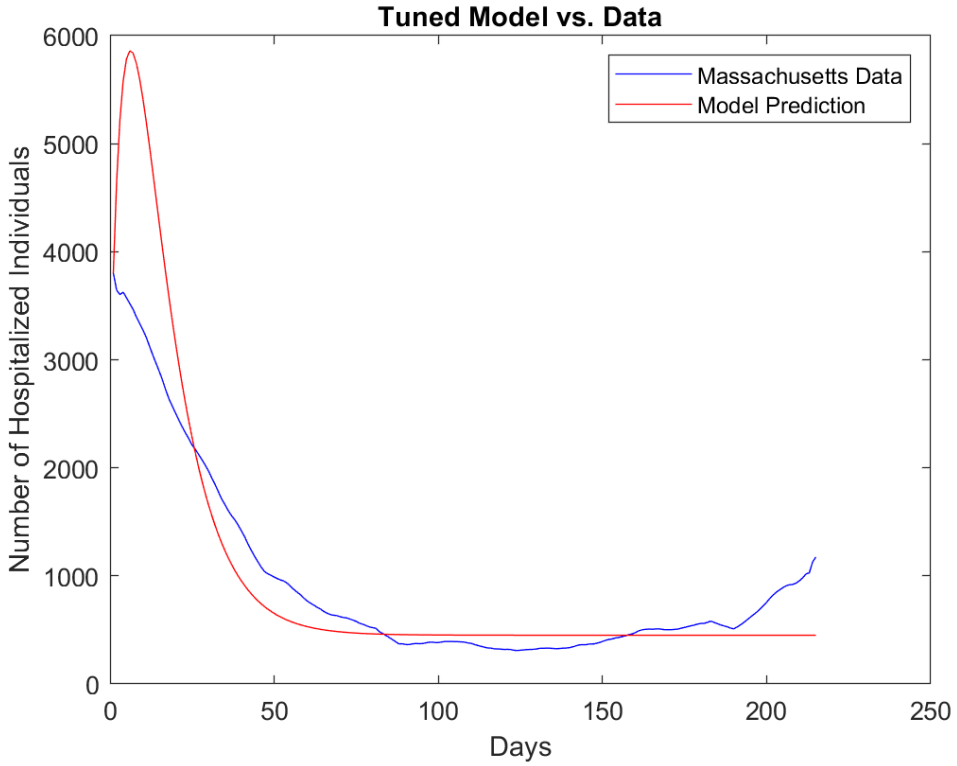

Figure 14: Simulated versus actual MA hospitalization data with Tuned parameters. Graph corresponds to the original values of c, q, β, and θ.

6.3 Analyzing the Basic Reproduction Number

Having determined three alternative sets of parameters, the next step was to try these values in the Basic Reproduction Number formula, as found in Chapter 3, Lemma 3.

The values were tested twice: once with the original parameters, as found by Tang et al., and again using the updated parameters. $\rho, \gamma_I, \gamma_A,$ and α were updated, but δ_I was the same for both models. The S in the Basic Reproduction Number formula is representative of the initial population in question. Because the parameter tuning analyzed Massachusetts, their population, as discussed in Section 4.1, was used.

The formula for herd immunity,

$$HI = 1 - \frac{1}{R_0},$$

Can be applied to each set of parameters to determine how much of the population needs to have recovered from, or have been vaccinated against, CoViD-19.

The results can be found in the following table:

	c	β	q	θ	OrigBRN	NewBRN	OrigHI	NewHI
Tang	14.78	2.1×10^{-8}	1.89×10^{-7}	0.58	5.20	8.42	80.76%	88.12%
1^{st} Min	2.84	3.15×10^{-7}	1.51×10^{-3}	0.991	17.31	26.81	94.22%	96.27%
2^{nd} Min	4.83	5.68×10^{-6}	2.88×10^{-2}	0.168	375.97	645.60	99.73%	99.85%
3^{rd} Min	16.72	8.42×10^{-8}	4.21×10^{-4}	0.726	24.88	39.62	95.98%	97.48%

Table 4: **Basic Reproduction Number Calculations.**

This shows that the parameters were better approximated by Tang et al. using Markov Chain Monte Carlo Simulation than by Constrained Non-Linear Optimization for Parameter Tuning. The researched parameters also seem to be less accurate than Tang et al.'s values. An R value cannot be expected to be of the size found using the Tuned Parameters, or using the updated values of $\rho, \gamma_I, \gamma_A,$ and α. As it is, 5.20 is surprisingly large.

This Basic Reproduction Number means that every person who contracts CoViD-19, more than five other people get infected. The maximum R_0 value, created using the First Minimum set and the Updated parameters, would mean that about 646 people are infected by a single individual with CoViD-19.

A range for Herd Immunity can be found above: $HI = [80.76, 99.85]$. Thus, it can be determined that somewhere between 80.8% and 99.85% of the population must have recovered from CoViD-19 or have the vaccine in order to control the spread of CoViD-19.

It should be clear that a common cause of the unreasonable values found here is β, given that the value of Tang et al. was of order of magnitude -8, and that two of values found using Parameter Tuning were at least ten times larger.

7 Discussion

CoViD-19 urgently requires research. Without a standardized response, supported by scientific articles, it has proven to be difficult to contain the virus. Nowhere is this more clear than in the more than 500,000 Americans who have died as a result of CoViD-19; a number that is always increasing. Providing scientific research that unequivocally proves whether or not mask mandates and lockdown protocol reduces the spread of CoViD-19 is necessary to prevent the pandemic from becoming worse, lasting longer, or both.

This study utilized Tang et al.'s Mathematical Model. This was created using a Markov Chain Monte Carlo simulation, based on data from the Wuhan Region. Its data was limited in important ways, necessitating research without the restrictions faced by Tang et al. The first problem was in scope. As Wuhan was the epicenter at the time, and there was neither a standardized response nor a body of research on how to reduce spread and how to treat patients. By focusing on data from at least three months after the advent of the novel coronavirus, and analyzing data from states with varying responses, this same criticism was avoided in this paper. The other major issue was that of time - Tang et al. were only able to research patterns for a month before publishing their study. To resolve this limitation, six months of data was analyzed.

In using Tang et al.'s Model, this study utilized an $SEIAS_qE_qHRD$ Model. This accounts for those in quarantine, both uninfected and exposed, asymptomatic infected individuals, hospitalized patients, and those who died. The quarantine subclasses S_q and E_q clearly are divisions of S and E, respectively, creating an $SEIAHRD$ Model. This in turn can be simplified, with E, A, and H as subclasses of I. Thus, the Model is reduced to a $SIRD$ format. Again, this can be simplified: D is a subclass of R. It is clear that the model use is an expansion of the basic SIR model. These expansions provided more insight into the epidemiology of CoViD-19 than a basic model could have.

The $SEIAS_qE_qHRD$ Model utilized in this research includes 13 parameters, each of which helps map the changes between groups. Each one can explained as follows:

- c represents either the average number of people that one interacts with during the course of a day, or the average number of infectious people that one interacts with during a day.
- β is either the chance of getting CoViD-19 from any singular person in c, or the chance of getting CoViD-19 from an infectious individual, respective to what c represents.
- q is the likelihood that someone who is exposed will start quarantining on any given day
- σ is the probability that someone who was exposed will become infected on any given day
- λ is the release of uninfected individuals from quarantine. This length of quarantine has since decreased, but during the study, 14 days was standard
- ρ is the likelihood that someone who has CoViD-19 will display symptoms

- δ_I represents the transfer between the Infected Class and the Hospitalized Class
- δ_q similarly indicates the transfer from the Exposed Quarantine Class and the Hospitalized Class
- θ is the reduced efficacy of transmission of the Asymptomatic Class
- γ_I is the rate of transition from the Infected Class to the Recovered Class
- γ_A is the rate of transition from the Asymptomatic Class to the Recovered Class
- γ_H is the rate of transition from the Hospitalized Class to the Recovered Class
- α is the rate of death caused by CoViD-19

It is worth noting that there are two possible interpretations of both c and β. The first versions of each go together, as does the second set. Either way, the results are still the same. What the parameters represent does not impact their value, or their use in the Basic Reproduction Number formula.

In addition to carefully choosing the Model, the data was similarly carefully chosen. By choosing three states that do not share as much of their metropolitan areas with other states, the data examined has less external variation. With this caveat in mind, states were selected with differing levels of mask mandates, different ratios of metropolitan population, and from different areas of the United States.

The data inevitably had to be cut off at a certain point. The decision was made to do so at the beginning of November. This served a few purposes: firstly, it eliminated the need to create a vaccinated class. Secondly, it prevented the data from being skewed. The holiday season resulted in the spread of CoViD-19, in a way that was inconsistent with the time that was analyzed. Finally, with the advent of co-mutations, it would have become impossible to properly track all necessary data given the resources available during this research.

This information has flaws due to the bureaucracy that all information suffers from at the beginning of documentation. Some information was irrelevant to this study, and was removed during the cleaning process. Additionally, it was clear that some days had less or no CoViD-19 test results reported, and additionally that more tests were conducted on some days of the week than others. To make the data more useful, smoothing, which utilized seven-day moving averages, was applied.

This study was not just limited by the data used. Without an understanding of Markov Chain Monte Carlo Simulations, the same accuracy of data as Tang et al.'s paper could not be expected. Therefore, Linear Regression Analysis and Constrained Non-Linear Optimization were applied to determine each parameter's significance and to tune the parameters.

The Linear Regression Analysis focused on the S, I, A, and H equations. In this research, it was determined that θ is the least significant parameter. Thus, future research should focus on c, q, and β.

Constrained Non-Linear Optimization provided an opportunity to examine the parameters used. This provided new possible values for c, β, q, and θ, by running a program based on MATLAB's $fmincon$ function. This created three new potential values for each parameter. These will require testing in future research, and a Markov Chain Monte Carlo Simulation will likely provide more accurate values if utilized.

Using these new parameter sets and the Basic Reproduction Number formula, estimations of R_0 value can be made. In doing so, it becomes clear that the new parameters result in values that are unreasonably high. This can be primarily attributed to the β value, which is at least one orders of magnitude larger than those found by Tang et al. in two of the Tuned Parameter sets. It will take future research to verify that the original values are correct, but at the very least, the β values found should be heavily examined.

Although this research should not be viewed as an end-all, be-all, it should be part of the larger conversation regarding CoViD-19 protocol. As information changes and develops, so too does the response to CoViD-19, both in the United States, and abroad. Study of the novel coronavirus should similarly evolve with these developments.

Some notable events that will be important for future studies include changes in mask mandates. With many states moving away from punishing those who choose not to wear masks, the spread of CoViD-19 will be very important to note. In the states that this study analyzed, changes occurred. Iowa put in place and removed a mask mandate in between the period where data was gathered and our results were published [44]. Additionally, Texas repealed their mask mandate [45]. This is representative of the events occurring throughout the United States - 15 states were without mask mandates by March 3, 2021, not including Texas [46].

It should be noted that these changes are occurring, at least in part, because of the advent of vaccines for CoViD-19. With Moderna, Pfizer, and Johnson & Johnson providing doses across the United States, and to an extent, abroad, the landscape regarding CoViD-19 is changing dramatically [47]. Future studies may wish to create a Mathematical Model, in which an additional subclass is added to the Removed Class, for those who have been vaccinated. The results of this hypothetical study would help determine the necessity of mask mandates in a society where vaccines are common.

The final major change is the development of mutated CoViD-19 strains [48]. Although this should not come as a surprise, it does add an additional wrinkle to future studies. The vaccines available have been proven to have reduced efficiency against at least one strain each [49]. This will make vaccinations harder to have complete confidence in, and necessitates future study.

The research in this paper lays out an effective starting point for all this research. With confidence in the parameters for the above mathematical model, and knowledge on which ones are most important to understand, future researchers might be able to focus their research to a greater precision. It is the hope of the authors that this is the case.

8 Appendices

8.1 Appendix A: The Basic Reproduction Number

The model used in this report is an expansion of the SIR model, which is given by

$$\begin{aligned}\dot{S} &= -\frac{\beta SI}{N} \\ \dot{I} &= \frac{\beta SI}{N} - \gamma I \\ \dot{R} &= \gamma I\,.\end{aligned}$$

This simpler model only has compartments for Susceptible, Infected, and Removed.

Common expansions of this model include: the SIRD model, where Removed is separated into Recovered and Deceased; the SEIR model, where Infected is subcategorized as Exposed and Infectious; and, the SIRC model, where Removed has the subclass of Carrier.

The SIRD model is given by:

$$\begin{aligned}\dot{S} &= -\frac{\beta IS}{N} \\ \dot{I} &= \frac{\beta IS}{N} - \gamma I - \mu I \\ \dot{R} &= \gamma I \\ \dot{D} &= \mu I\,.\end{aligned}$$

The SEIR model is the following:

$$\begin{aligned}\dot{S} &= \lambda N - \mu S - \frac{\beta IS}{N} \\ \dot{E} &= \frac{\beta IS}{N} - (\mu + \alpha)E \\ \dot{I} &= \alpha E - (\gamma + \mu)I \\ \dot{R} &= \gamma I - \mu R\,.\end{aligned}$$

The model used in this paper combines the elements of all of these models. By using an $SEIAS_qE_qHRD$ model, the carrier class is effectively replaced by asymptomatic carriers. It is important to note that these asymptomatic carriers are not a subgroup of the Removed compartment; rather, it is an element of the Infected compartment. In essence: S and S_q are the Susceptible, $E, I, A, E_q,$ and H are Infected, and R and D are Removed.

For all compartmental models, there exist steady states. These exist where there is no transfer between classes; ie. when

$$\dot{S}, \dot{I}, \dot{R} = 0 \text{ (for an SIR model)}$$

.

For the SIR model, the steady states exist at

$$(S, 0, R)$$

.

It follows that the steady states for the SIRD model are of the form

$$(S, 0, R, D),$$

and that of the SEIR model are of the form

$$(S, 0, 0, 0)$$

For these steady states, $S, R,$ and D are all variable, but must add up to N. Therefore, there exists a steady state wherein

$$S = N$$

This is known as the disease-free steady state, where no one has ever been infected.

In addition to providing information to determine steady states, the equations can help determine the Basic Reproduction Number, $\mathcal{R}_0$, of a disease. This is determined by the number of exposures to those in the Susceptible class that an Infected individual has before they become Removed.

As more categories are added, a matrix can be formed of the exposures an individual has between classes ($\mathcal{F}$), and another can be made of the removal from the infectious classes ($\mathcal{V}$). To determine the Basic Reproduction Number the exposure matrix can be divided by the removal matrix, $\mathcal{K} = \mathcal{F}\mathcal{V}^{-1}$, and the eigenvalue with the largest modulus is the Basic Reproduction Number. The Basic Reproduction Model is represented by R_0.

By this definition, $\mathcal{F}$ is β, β, and $\alpha\beta$, for SIR, SIRD, and SEIR models, respectively. Similarly, $\mathcal{V}$ is γ, $\mu + \gamma$, and $(\mu + \alpha)(\mu + \gamma)$.

Thus, for the SIR model, the Basic Reproduction Number is

$$R_0 = \frac{\beta}{\gamma},$$

For the SIRD model, the Basic Reproduction Number is

$$R_0 = \frac{\beta}{\mu + \gamma},$$

And for the SEIR model, the Basic Reproduction Number is

$$R_0 = \frac{\alpha}{\mu + \alpha} \frac{\beta}{\mu + \gamma}.$$

The Basic Reproduction Number is important in epidemiology, as it helps determine how rapidly a virus spreads. If the R_o value is less than or equal to one, then the virus will only infect one person on average for every person infected, and the virus will be in an endemic state. This means that the growth is linear or logarithmic, as opposed to the exponential growth that occurs under R_0 values greater than one, also known as a pandemic state.

This leads to the idea of herd immunity. If $\left[100 \times (1 - \frac{1}{R_0})\right]$ % of the population has been vaccinated or has otherwise gained immunity from a virus, then the spread will not occur at pandemic levels.

Thus, from the equations that provide a model, the steady states, Basic Reproduction Value, and herd immunity percentage can all be determined.

8.2 Appendix B: MATLAB Code for Parameter Tuning

```
%Optimization function presets
global k0;
k0 = [c, beta, q, theta];
Aeq = [];
beq = [];
lb = [1e-19, 1e-19, 1e-19, theta-1.96*sdtheta];
ub = [17, 0.999, 0.12, theta+1.96*sdtheta];
A = [];
b = [];

%ODE function presets
global w tspan initState options;
w = currHospitalized;
tspan = days;
initState = [6890000,5000,10267,6666,0,0,3803,0,0];  %initial values for MA
options = odeset('NonNegative',1);

%Optimization problem
k = fmincon(@myfun, k0, A, b, Aeq, beq, lb, ub);

%Parameters for ODE and ODE
para = [k(1), k(2), k(3), sigma, lambda, rho, deltaI, deltaq,...
    k(4), gammaI, gammaA, gammaH, alpha];
[t,x]=ode15s(@(t,y)covid19model(y, para), tspan, initState, options);
mse = immse(w,x(:,7));
```

Figure 15: MATLAB Code for Constrained Non-Linear Optimization for Parameter Tuning. Not shown here: parameter values and the equation of the ODE model.

8.3 Appendix C: R Code for Linear Regression

```
# NOTE: To run codes in R, select the lines of code you want to run and click "Run".

# Importing the csv file (location will be different for everybody)
# lhs <- read.csv("C:\\Users\\alexa\\OneDrive\\Documents\\MQP
# \\LHS Roger Dec282020\\lhsresults.csv", header = TRUE, sep = ",", fileEncoding="UTF-8-BOM")

# Defining all of the columns in the lhs dataframe to use for regression
# Predictors
c <- lhs[,1]
q <- lhs[,2]
beta <- lhs[,3]
theta <- lhs[,4]
cc <- lhs[,5]
cq <- lhs[,6]
cb <- lhs[,7]
ct <- lhs[,8]
qq <- lhs[,9]
qb <- lhs[,10]
qt <- lhs[,11]
bb <- lhs[,12]
bt <- lhs[,13]
tt <- lhs[,14]

# State Variables
S <- lhs[,15]
E <- lhs[,16]
I <- lhs[,17]
A <- lhs[,18]
Sq <- lhs[,19]
Eq <- lhs[,20]
R <- lhs[,21]
H <- lhs[,22]

#Linear Regression on S, I, A, and H

Smodel <- lm(formula = S ~ c + q + beta + theta + cc + cq + cb + ct + qq + qb + qt + bb + bt + tt)
Imodel <- lm(formula = I ~ c + q + beta + theta + cc + cq + cb + ct + qq + qb + qt + bb + bt + tt)
Amodel <- lm(formula = A ~ c + q + beta + theta + cc + cq + cb + ct + qq + qb + qt + bb + bt + tt)
Hmodel <- lm(formula = H ~ c + q + beta + theta + cc + cq + cb + ct + qq + qb + qt + bb + bt + tt)

# Uncomment these to get a summary of the regression.

# summary(Smodel)
# summary(Imodel)
# summary(Amodel)
# summary(Hmodel)
```

Figure 16: R Code for Linear Regression Analysis.

References

[1] Coronavirus cases. (n.d.). Retrieved March 18, 2021, from https://www.worldometers.info/coronavirus/

[2] Spiteri, G., Fielding, J., Diercke, M., Campese, C., Enouf, V., Gaymard, A., ... Ciancio, B. (2020, March 5). First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7068164/

[3] Dos Santos, W. G. (2020, July 3). Natural history of COVID-19 and current knowledge on treatment therapeutic options. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7332915/

[4] Velavan, T. P., & Meyer, C. G. (2020, February 16). The COVID-19 epidemic. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1111/tmi.13383

[5] Details - Public Health Image Library(PHIL). (n.d.). Retrieved from https://phil.cdc.gov/Details.aspx?pid=23312

[6] Coronavirus. (2020, February 15). Retrieved from https://www.cdc.gov/coronavirus/types.html

[7] Symptoms of Coronavirus. (2020, May 13). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

[8] Van Beusekom, M. (2020, September 22). Studies explore asymptomatic COVID-19 cases, racial inequities. Retrieved from https://www.cidrap.umn.edu/news-perspective/2020/09/studies-explore-asymptomatic-covid-19-cases-racial-inequities

[9] Harvard Health Publishing. (2020, October 16). If you've been exposed to the coronavirus. Retrieved from https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus

[10] Markowitz, A. (2020, October 22). Does Your State Have a Mask Mandate Due to Coronavirus? Retrieved from https://www.aarp.org/health/healthy-living/info-2020/states-mask-mandates-coronavirus.html

[11] Edwards, E. (2020, October 22). FDA approves first drug for Covid-19: Remdesivir. Retrieved from https://www.nbcnews.com/health/health-news/fda-approves-first-drug-covid-19-remdesivir-n1244351

[12] Contact Tracing for COVID-19. (n.d.). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/contact-tracing.html

[13] Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu, *Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions.* J. Clin. Med. 2020, 9, 462; doi:10.3390/jcm9020462

[14] Tan, W. Y., Wong, L. Y., Leo, Y. S., & Toh, M. P. (2020, September 2). Does incubation period of COVID-19 vary with age? A study of epidemiologically linked cases in Singapore. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7484300/

[15] Byambasuren, O., Cardona, M., Bell, K., Clark, J., McLaws, M., & Glazsiou, P. (2020, December 11). Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis. Retrieved from https://jammi.utpjournals.press/doi/10.3138/jammi-2020-0030

[16] J. Dehning, et. al. *Inferring change point in the spread of COVID-19 reveals the effectiveness of interventions.* Science, May 2020. doi:10.1126/science.abb9789.

[17] Faes, C., Abrams, S., Van Beckhoven, D., Meyfroidt, G., Vlieghe, E., & Hens, N. (2020, October 17). Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7589278/

[18] Joseph T. Wu, Kathy Leung, and Gabriel M. Leung, *Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study.* The Lancet, vol. 395, issue 10225, p. 689-697, Feb. 2020, doi.org/10.1016/S0140-6736(20)30260-9

[19] Census profile: New York-Newark-Jersey City, NY-NJ-PA Metro Area. (n.d.). Retrieved from https://censusreporter.org/profiles/31000US35620-new-york-newark-jersey-city-ny-nj-pa-metro-area/

[20] Census profile: Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area. (n.d.). Retrieved from https://censusreporter.org/profiles/31000US37980-philadelphia-camden-wilmington-pa-nj-de-md-metro-area/

[21] Annual Estimates of the Resident Population for Metropolitan Statistical Areas in the United States and Puerto Rico: April 1, 2010 to July 1, 2019 (CBSA-MET-EST2019-ANNRES)

[22] Table 1. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019 (NST-EST2019-01)

[23] First Presumptive Positive Case of COVID-19 Identified by Massachusetts State Public Health Laboratory. (2020, March 2). Retrieved from https://www.mass.gov/news/first-presumptive-positive-case-of-covid-19-identified-by-massachusetts-state-public-health

[24] COVID-19 State of Emergency. (n.d.). Retrieved from https://www.mass.gov/info-details/covid-19-state-of-emergency

[25] Coronavirus In Massachusetts: Baker Extends Business Closures, Stay At Home Advisory To May 18. (2020, April 28). Retrieved from https://boston.cbslocal.com/2020/04/28/coronavirus-latest-massachusetts-news-gov-charlie-baker/

[26] COVID-19 Travel Order. (n.d.). Retrieved November 18, 2020, from https://www.mass.gov/info-details/covid-19-travel-order

[27] Governor Abbott Declares State of Disaster In Texas Due To COVID-19. (2020, March 13). Retrieved from https://gov.texas.gov/news/post/governor-abbott-declares-state-of-disaster-in-texas-due-to-covid-19

[28] Svitek, P. (2020, March 31). Gov. Greg Abbott tells Texans to stay home except for essential activity in April. Retrieved from https://www.texastribune.org/2020/03/31/greg-abbott-texas-executive-order-closures/

[29] Texas State Law Library. (n.d.). Guides: COVID-19 & Texas Law: Mask Laws. Retrieved from https://guides.sll.texas.gov/covid-19/mask-laws

[30] Svitek, P. (2020, June 26). Gov. Greg Abbott orders Texas bars to close again and restaurants to reduce to 50% occupancy as coronavirus spreads. Retrieved from https://www.texastribune.org/2020/06/26/texas-bars-restaurants-coronavirus-greg-abbott/

[31] Weber, P. J., &; Vertuno, J. (2020, November 11). Texas becomes 1st state to surpass 1 million COVID-19 cases. Retrieved from https://apnews.com/article/texas-1-million-covid-cases-first-state-097550b8324b0f54f06bf4d2a1fa4fbd

[32] Svitek, P., & Rosenzweig-Ziff, D. (2020, November 18). Coronavirus cases in Texas are soaring again. But this time Gov. Greg Abbott says no lockdown is coming. Retrieved from https://www.texastribune.org/2020/11/18/texas-coronavirus-lockdown/

[33] Iowa's watered-down mask mandate won't extinguish COVID inferno. (2020, November 11). Retrieved from https://www.thegazette.com/subject/opinion/staff-editorial/iowas-watered-down-mask-mandate-wont-extinguish-covid-inferno-20201111

[34] As deaths spiral, South Dakota Gov. Kristi Noem opposes mask rules. (2020, November 16). Retrieved from https://triblive.com/news/as-deaths-spiral-south-dakota-gov-kristi-noem-opposes-mask-rules/

[35] Proclamations. (n.d.). Retrieved November 24, 2020, from https://coronavirus.iowa.gov/pages/proclamations

[36] Gov. Reynolds signs new proclamation continuing the State Public Health Emergency Declaration. (2020, April 27). Retrieved from https://governor.iowa.gov/press-release/gov-reynolds-signs-new-proclamation-continuing-the-state-public-health-emergency-0

[37] Wu, J., Smith, S., Khurana, M., Siemaszko, C., & DeJesus-Banos, B. (2020, April 29). Coronavirus lockdowns and stay-at-home orders across the U.S. Retrieved from https://www.nbcnews.com/health/health-news/here-are-stay-home-orders-across-country-n1168736

[38] Bridgeman, M., & Leibowitz, B. (2020, July 21). Iowa City's mayor issued a mask mandate. Here's what you need to know:. Retrieved from https://www.press-citizen.com/story/news/2020/07/21/covid-iowa-city-mask-mandate-requirements-rules-mayor-face-cover-order/5482551002/

[39] Pearce, A. (2020, September 14). Iowa City face mask mandate extended to Nov. 13, new fines added for violations. Retrieved from https://kwwl.com/2020/09/14/iowa-city-face-mask-mandate-extended-to-nov-13-new-fines-added-for-offenses/

[40] Rossi, M. (2020, August 26). Des Moines Mayor Frank Cownie issues city-wide face mask mandate. Retrieved from https://www.weareiowa.com/article/news/health/coronavirus/city-of-des-moines-iowa-coronavirus-mask-mandate-mayor-frank-cownie/524-37a8c482-ea57-4dcd-bfbb-5e5849fbc24f

[41] Schlesselman, H. (2020, September 15). Waukee mask mandate goes into effect Friday. Retrieved from https://www.weareiowa.com/article/news/health/coronavirus/waukee-mask-mandate-effective-september-18-iowa/524-d1669e30-6bfd-4fb7-8b36-7121eda22a63

[42] Foley, R. J. (2020, November 17). Changing course, Iowa governor enacts limited mask mandate. Retrieved from https://apnews.com/article/iowa-city-iowa-coronavirus-pandemic-prisons-3cab3e58f26c0e42cafe8001ca2e06c6

[43] Fred Brauer, Pauline van den Driessche, and Jianhong Wu. *Mathematical Epidemiology.* Springer-Verlag, Berlin, Heidelberg, 2008.

[44] Nick Coltrain and Tony Leys (2021, February 10). Gov. Kim Reynolds defends decision to lift mask mandate, other COVID-19 restrictions. Retrieved from https://www.desmoinesregister.com/story/news/politics/2021/02/10/kim-reynolds-covid-live-stream-mask-rules-iowa-governor-poor-vaccine-distribution/4449134001/

[45] Governor Abbott Lifts Mask Mandate, Opens Texas 100 Percent. (2021, March 2). Retrieved from https://gov.texas.gov/news/post/governor-abbott-lifts-mask-mandate-opens-texas-100-percent

[46] Leah Asmelash, Nakia McNabb and Amanda Watts (2021, March 9). Texas ends its face mask rule on Wednesday. Here are the other states that don't require them. Retrieved from https://www.cnn.com/2020/11/09/us/biden-mask-mandate-nationwide-trnd/index.html

[47] Different Vaccines. (2021, March 4). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html

[48] Emerging SARS-CoV-2 Variants. (2021, January 28). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html

[49] Rita Rubin (2021, March 17). COVID-19 Vaccines vs Variants—Determining How Much Immunity Is Enough. Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2777785