Title: Mathematical Modeling Study of the 2020 CoVID-19 outbreak in the United States

Durgesh Nandini Sinha[1]*, Nicholas Klahn[2]

Durgesh N Sinha (Adjunct Assistant Professor, Temple University, Philadelphia/ Strayer University (online)/Rowan College at Burlington County, New Jersey/ Community College of Philadelphia/ Mercer County Community College, New Jersey)

Nicholas Klahn (Senior student of Medical track, Rowan College at Burlington County)
Corresponding author: Durgesh N Sinha (durgesh.sinha4@gmail.com)

Abstract
  A mathematical model was developed for the currently evolving COVID-19 outbreak. Data analysis and model fitting using Latin Hypercube Sampling partial Correlation Coefficient Method was used to determine the model's parameters and basic reproduction numbers. The infectivity values from symptomatic infectious people was 0.118461389 (95% CI [0.1136278, 0.12329497]), asymptomatic transmission was 0.100111427 (95% CI [0.1000297, 0.10019314]), and quarantined transmission was 0.057337278 (95% CI [0.0504738, 0.0642008]). The United states reached its peak basic reproduction number on March 10th where $R_0=58$, but it has since lowered to 1.47 as of April 5th. Also, those in quarantine had contributed the most to the basic reproduction number, with asymptomatic people being second, and regular symptomatic people contributing the least. Our simulations showed that the United States has reached its peak occurred on April 11, 2020 with a total 461,700 number of cases and it will reach on June 12, 2020 where the confirmed case count would reach 1.439 million. As for the longevity of the virus, our prediction shows that it could be under preventive measure within two years by February 10, 2022, would be 14,130.

Keywords: COVID-19, US population, Mathematical Modeling, SARS-CoV-2, 2019-nCoV
Introduction
  In 2020, a new strain of severe acute respiratory syndrome (SARS) virus was discovered when cases of pneumonia erupted in the Wuhan district of China [1,2]. This new strain is called SARS-CoV-2, due to its similarities with SARS, but it also goes by COVID-19, and previously, 2019-nCoV [2]. Since then, it has spread to many countries around the world with it breaking over 1 million cases by April 4, 2020 [3]. Of those cases, the United States holds the highest case count with 241,703 cases on the same day, but the numbers have risen to 333,811 cases in just three days [3,4].
  The goal of this paper is to analyze the spread of COVID-19 within the United States, and to make predictions about the outbreak timeline and case count based on the available data.
Methods
  Data Sources
  In order to obtain the necessary parameter values for our COVID-19 transmission model we used multiple outside sources as well as calculations based on those sources as seen in table 1. In order to find the number of cases by day for the United States and the most up to date numbers we used CDC, WHO, and worldometer data. As for the other parameters, we used data

from other papers that were published earlier in the outbreak in order to estimate the viral dynamics.

Procedures

With our $SEIAQI_M$ model we divided the human population into six classes in order to analyze the spread of COVID-19. The schematic flow of this model is shown in figure 1 and the state variables, associated parameters, and the equations of this model are given below:

$S(t)$: Susceptible humans in time t,

$E(t)$: Exposed humans in time t,

$I(t)$: Infectious humans in time t,

$Im(t)$: Recovered humans with Immunity in time t,

$A(t)$: Asymptotic infectious human in time t,

$Q(t)$: Quarantine human in time,

$N(t)$: Total human population in time t,

B : Birth rate of humans,

βi : Infectivity rate of COVID-19 from Symptomatic infectious class,

β_A : Infectivity rate of COVID-19 from Asymptotic population class,

β_θ : Infectivity rate of COVID-19 from Quarantine population class,

η : Rate of transition from exposed to infected humans,

α : Rate of transition from Quarantine to recovered humans,

σ: Rate of Immune humans from COVID-19,

γ: Rate of transition from Infected class to Quarantine class,

γ_m : Rate of transition from Infected class to recovered class,

ξ: Rate of transition from exposed to Asymptotic infected class,

λ: Rate of transition from Asymptotic class to Quarantine class,

Θ: Rate of transition from Asymptotic class to Recovered class with Immunity,

μ : Natural death rate of humans, and

δ : Death rate of humans due to COVID-19.

The susceptible, S, class contains all people that are able to be infected by COVID-19. They can be exposed to the virus from both infected, I, and asymptomatic people, A [5]. Both I and A are not yet in quarantine, but can be tested for the virus and put into quarantine, Q, if they test positive. Those in quarantine are still able to infect others that may be around them, which gives the virus another transmission route through β_θ. Those in quarantine will remain there for up to 3-6 weeks, before moving to the recovered class with immunity against COVID-19, I_M [6]. All infectious classes I, A, and Q are able to move directly to the recovered class. The asymptomatic class may not realize they are infected and will recover before receiving any tests, θ, the infected class may never go out to get tested, and will recover without knowing about their condition through γ_m, and the quarantine group will recover some time during their quarantine, α.

Based on the flow of transmission of COVID-19 in the human population as depicted in figure 1, we have the following system of equations:

$$\frac{dS}{dt} = BN - \beta i SI - \beta_A SA - \beta_\theta SQ - \mu S$$

$$\frac{dE}{dt} = \beta i SI + \beta_A SA + \beta_\theta SQ - (\mu + \eta + \xi + \sigma)E$$

$$\frac{dA}{dt} = \xi E - (\lambda + \theta + \mu)A$$

$$\frac{dI}{dt} = \eta E - (\mu + \gamma + \gamma_m)I \tag{1}$$

$$\frac{dQ}{dt} = \gamma I + \lambda A - (\mu + \delta + \alpha)Q$$

$$\frac{dIm}{dt} = \alpha Q + \theta A + \sigma E + \gamma_m I - \mu Im$$

And $N(t) = S(t) + E(t) + A(t) + I(t) + Im(t)$ (2)

All our parameters are positive or non-negative, therefore all parameters must remain positive or non-negative for initial conditions for t≥0. From our model we have $\frac{dN}{dt} = B - \mu N - \delta Q \le B - \mu N$. The closed set D= $\{(S, E, A, I, Q, Im) \in R+^6 ; N \le \frac{B}{\mu} \}$ is a feasible region for the graph.

Some model parameters were calculated using the Latin Hypercube Sampling partial Correlation Coefficient Method (LHS PRCC). We used this method to calculate all of the infectivity values, β, as well as the transitions between E, I, A, and Q.

We have also proved that our model is locally and globally stable for both diseases-free-equilibrium and endemic equilibrium points through block matrix and Lyapunov function methods that can be provided if needed upon request.

Since all of our model parameters are positive or non-negative, it is important to show that all state variables remain positive or non-negative for all positive initial conditions for $t≥0$.

As for the basic reproduction number, it is the average number of secondary infectious cases produced by a single infection in the total susceptible population. The basic reproduction number is calculated by $R_0 = \mathscr{L}(FV^{-1})$, where $\mathscr{L}$ is the spectral radius of the matrix FV^{-1}, and F & V are the matrices of new infection terms and the remaining transmission terms respectively. By solving equation (1) and (2), we get F and V of the human population as given below:

$$F = \begin{bmatrix} 0 & \beta_A & \beta_I & \beta_\theta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V^{-1} = \begin{bmatrix} \frac{1}{x_1} & 0 & 0 & 0 \\ \frac{\xi}{x_1 x_2} & \frac{1}{x_2} & 0 & 0 \\ \frac{\eta}{x_1 x_3} & 0 & \frac{1}{x_3} & 0 \end{bmatrix}$$

$$\left| \frac{1}{x_1 x_4} - \left[\frac{\xi\lambda}{x_2} + \frac{\eta\gamma}{x_3}\right] \qquad \frac{\xi}{x_2 x_4} \qquad \frac{\gamma}{x_3 x_4} \qquad \frac{1}{x_4} \right.$$

Let $x_1 = (\mu + \eta + \xi + \gamma)$, $x_2 = (\lambda + \theta + \mu)$, $x_3 = (\mu + \gamma + \gamma_m)$, $x_4 = (\mu + \delta + \alpha)$.

$R_0 = Trace\ (FV^{-1})$ give us the equation:

$$R_0 = \frac{\beta_A \xi}{(\mu+\eta+\xi+\gamma)(\lambda+\theta+\mu)} + \frac{\beta_I \eta}{(\mu+\eta+\xi+\gamma)(\mu+\gamma+\gamma_m)} + \frac{\beta_\theta}{(\mu+\eta+\xi+\gamma)(\mu+\delta+\alpha)}\left[\frac{\xi\lambda}{(\lambda+\theta+\mu)} + \frac{\eta\gamma}{(\mu+\gamma+\gamma_m)}\right].$$

In order to interpret R_o in terms of the dynamics of the spread of COVID-19, R_o is broken down into three components. Those components are the new infections induced by susceptible contacts with infectious individuals, R_I, contact with asymptomatic survivors, R_A, and contact with those in quarantine R_θ.

For the MATLAB simulations we used the parameter values in table 1. All of the data required to reproduce the analysis is available online.
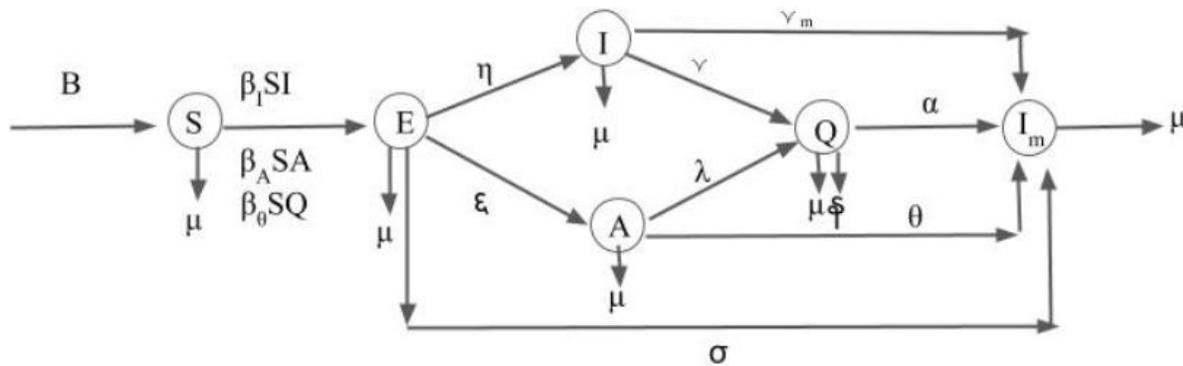
Figure 1. The COVID-19 transmission model.

Table 1. Values and sources of the model parameters.

Parameter	Value	Source		Parameter	Value	Source
βi	[0.1136278, 0.12329497]	fitted		η	0.00003056924157	7,8
βa	[0.1000297, 0.10019314]	fitted		α	0.0008012198307	6,8
βq	[0.0504738, 0.0642008]	fitted		δ	0.01736643891	8
θ	[0.0073,0.08871677]	fitted		μ	0.00002328767123	9
ξ	[0.000149625, 0.00019222]	fitted		B	0.00003397260274	9
γ	[0.00040403, 0.0004857174]	fitted		N	331002651	10
λ	[0.815424,0.836734]	fitted		R0	[9.631140619, 18.16764997]	calculated
γm	[0.9995727,0.999491]	calculated		Ri	[0.00502, 0.0059]	calculated
Rθ	[0.62042, 0.95161]	calculated		Ra	[0.02487, 0.03552]	calculated

3.2) Discussion of results.

Figure 2. The model parameters and their 95% confidence intervals.

Box Plot of Infectivity parameters, and transitivity parameters by LHS PRCC methods
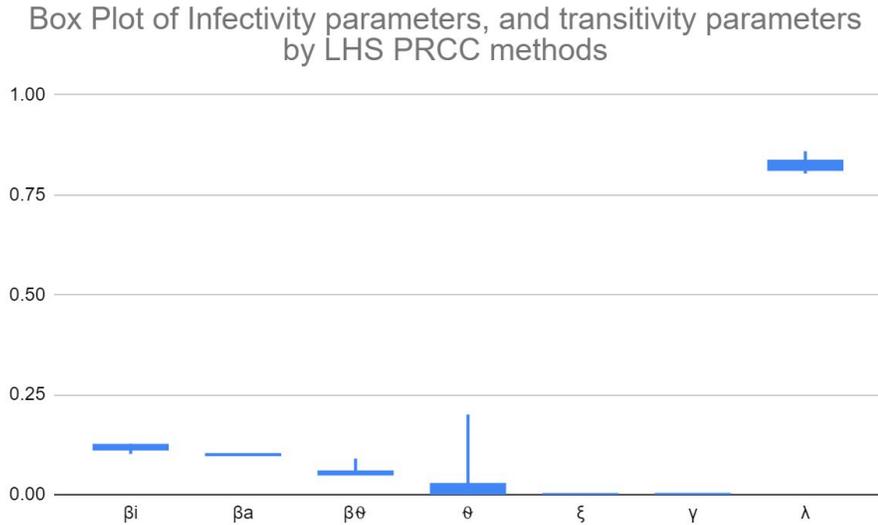
In figure 2 we have the values that were calculated using the LHS PRCC as well as their 95% confidence intervals based on the reported case data. The values are as follows $\beta_i = 0.118461389$ (95% CI [0.1136278, 0.12329497]), $\beta_a = 0.100111427$ (95% CI [0.1000297, 0.10019314]), $\beta_\theta = 0.057337278$ (95% CI [0.0504738, 0.0642008]), $\theta = 0.04070604$ (95% CI [00073, 0.08871677]), $\xi = 0.00017092$ (95% CI [0.000149625, 0.00019222]), $\gamma = 0.00044499991$ (95% CI [0.00040403, 0.0004857174]), and $\lambda = 0.8260789$ (95% CI [0.815424, 0.836734]).

Figure 3. Basic reproduction numbers and their 95% confidence intervals for different outbreak regions.
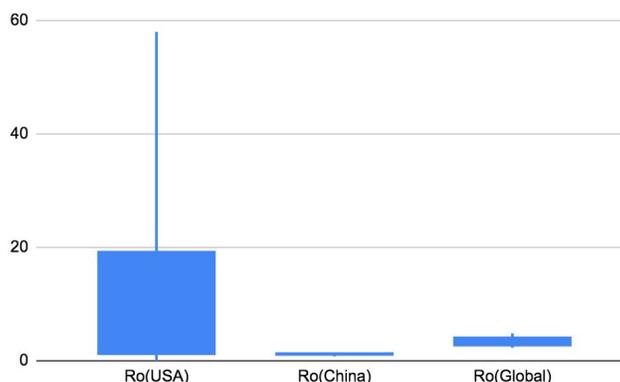
As for the basic reproduction numbers (figure 3), we found that R0=13.89939529 (95% CI [9.631140619,18.16764997]) for the United States based on a 9 day incubation period, 1.185729498 (95% CI [1.071562815, 1.29989618]) for China based on and incubation period of 7-14 days, and 3.46761042 (95% CI [2.655971597,4.059512574]) for the world also based on an incubation period of 7 to 14 days [11].

Upon analysis of the individual components, we found that R_θ contributed the most to the basic reproduction number as seen in table 1. This could be due to others (family members or healthcare workers) being in close proximity to the infectious person, resulting in constant repeated exposures to the virus rather than a one-time exposure when coming into contact with an infectious or asymptomatic person. That being said, asymptomatic people contributed the second most. Since they don't know that they're infected, they likely won't take as many precautions to prevent others from getting sick like a symptomatic infectious person would (not care about distancing themselves or sanitizing what they touch) since they feel fine. Not taking these precautions result in more possible exposure events.

Figure 4. The changing basic reproduction numbers from day to day in the United states population and the control line.

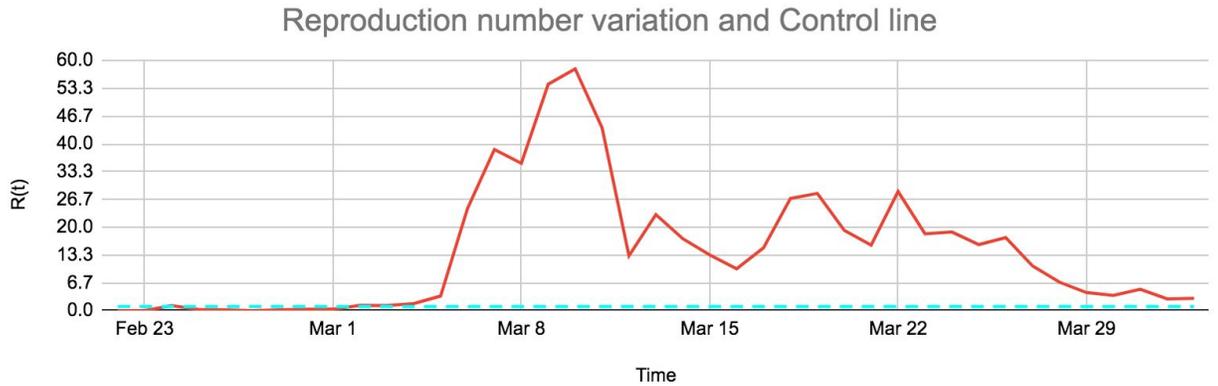Reproduction number variation and Control line

Figure 4 shows how the basic reproduction number has changed in the United States since the start of the outbreak. The control line for the outbreak is set at R0=1. We see that the reproduction number remains under the control line from Feb 24 to March 1. It then began to grow until March 10, where it reached its peak at March 10 where R(t)= 58. With the help of preventative measures such as quarantine and the shutdown of nonessential businesses lowering the amount of contacts that people can have, the basic reproduction number has dropped significantly in recent days, reaching a low of R(t) = 1.46 on April 5. Despite the measures taken, we see that the current number is above the control line, so the outbreak is not fully under control yet.

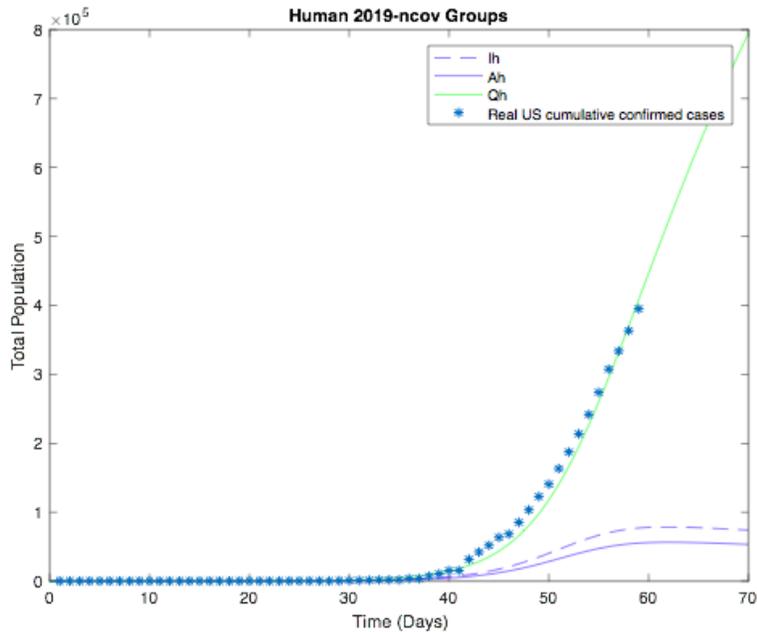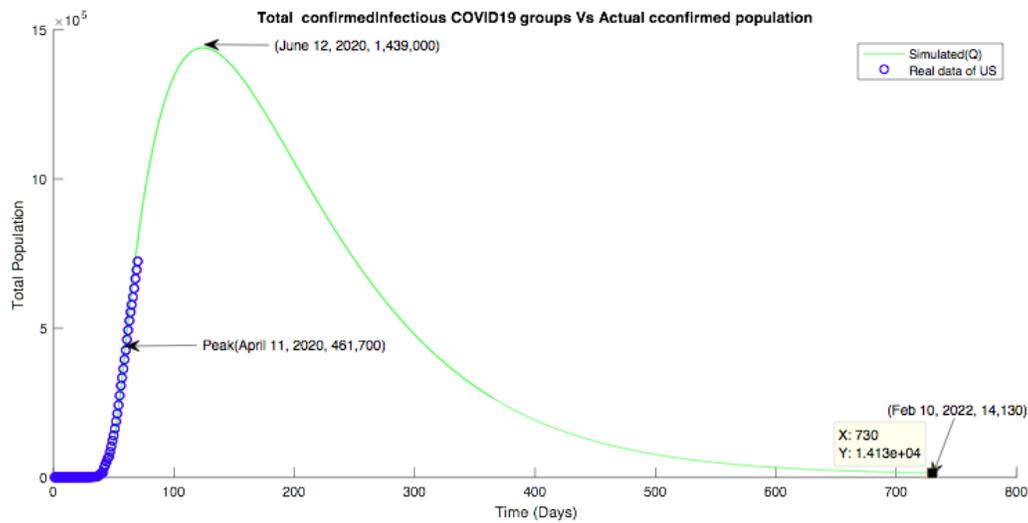**Figure 5a.** Fitting the simulated outbreak to the actual cumulative case numbers.
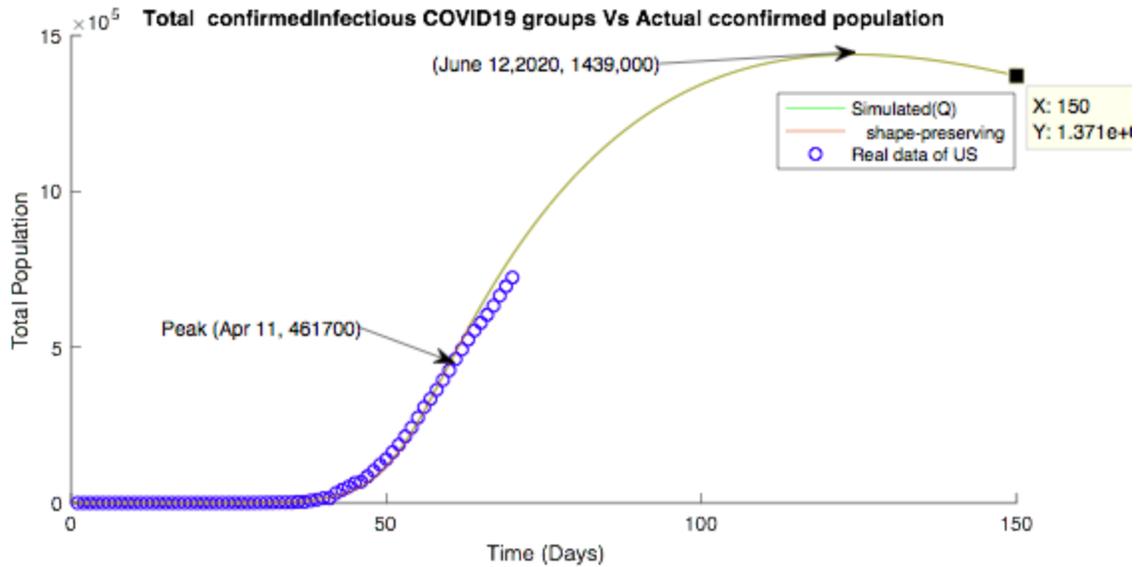
Figure 5a shows the comparison of our model's case count compared to the actual case count for 55 days. Our simulation is counting confirmed cases class as Q(t). There are still some non-quarantine infectious and asymptomatic people which could be attributed to the lack of available and affordable testing, motivation to get tested, or unawareness of their condition. This model shows that despite the majority of people being quarantined and isolating themselves from the general population, there are still people that are spreading the virus unknowingly.

Figure 5b, 5c. Predictions for the total number of COVID-19 cases.

Total confirmedInfectious COVID19 groups Vs Actual cconfirmed population

Total confirmedInfectious COVID19 groups Vs Actual cconfirmed population

Using the same simulation as we used in Figure 5a, Figure 5b and Figure 5c is run for a longer time frame in order to predict the number of future COVID-19 Infections. With our model we see that cases of COVID-19 will continue to rise for approximately 123 days where it will reach the cumulative number of infections at 1.439 million cumulative infectious cases. After that, the number of cases will start dropping for another 606 days after the maximum and will persist over the course of 2 years by Feb. 10, 2022. Judging by this graph, we are still in the early stages of the outbreak and have about 71 more days of quickly rising newly reported cases before the outbreak begins to die out. The extended trajectory of the graph shows that COVID-19 will continue to persist in the world for some time and may or may not fully disappear due to the infectious population being so close to zero.

5. Conclusion:

$SEIAQI_M$ (susceptible-exposed-infectious-asymptomatic-quarantine-recovered) model is what we use to analyze COVID-19's transmission dynamics in the United states population. We found that the United States reached its peaked reproduction number on March 10th (R0=58), but since then, the basic reproduction number has lowered to 1.46 (April 5th). However, that is still above the control line which sits at 1, so the outbreak will continue until $R0 \leq 1$. We found the peak of COVID-19 already occured on April 11, 2020 with a cumulative number of cases 461700. We also calculated the average basic reproduction numbers for the Chinese and Global outbreaks. These numbers were 1.185729498 (95% CI [1.071562815, 1.29989618]) and 3.46761042 (95% CI [2.655971597, 4.059512574]) respectively, with the United States being the highest of the three with R0=13.89939529 (95% CI [9.631140619,18.16764997]).

We fit our simulation to the actual cumulative cases and ran the simulation for an extended amount of time (730 days) to predict the life of the outbreak. We found that we would not reach the maximum number of cases until June 17, 2020 where there would be 1.439 million cases in the United States. However, after that, the number of new cases would lower over the course of 606 days, and could fully disappear by Feb 10, 2022.

So, despite the basic reproduction number dropping significantly since March 10 (R(t)= 58, to April 5 where R(t) = 1.47), due to it being above the control line, our simulations show that COVID-19 will continue to spread easily through the United States population well into June of 2020.

Competing Interest: The authors declare no competing interest.

Designed the project (D. S.); performed simulations (D. S.), performed analysis of the data (D. S., N. K.), wrote the paper (D. S., N. K.)

6: References:

1) del Rio, C., & Malani, P. N. (2020). 2019 Novel Coronavirus—Important Information for Clinicians. *JAMA*. doi: 10.1001/jama.2020.1490.
2) Gorbalenya, A. E. (n.d.). Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. *Biorxiv*. doi: 10.1101/2020.02.07.937862.
3) Coronavirus disease 2019 (COVID-19) Situation Report – 75. (2020). *World Health Organization*.
4) Coronavirus disease 2019 (COVID-19) Situation Report – 78. (2020). *World Health Organization*.
5) Coronavirus disease 2019 (COVID-19) Situation Report – 73. (2020). *World Health Organization*.
6) Aylward, B., et al. (2020, February 24). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). *World Health Organization*.

7) Wang, W., Tang, J., & Wei, F. (2020). Updated understanding of the outbreak of 2019 novel coronavirus (2019‑nCoV) in Wuhan, China. *Journal of Medical Virology*, *92*(4), 441–447. doi: https://doi.org/10.1002/jmv.25689.

8) COVID-19 CORONAVIRUS PANDEMIC. (2020). *Worldometer*.

9) Central Intelligence Agency. (2020, March 31). The World Factbook.

10) Countries in the world by population. (2020). *Worldometer*.

11) Centers for Disease Control. (2020). Symptoms of Coronavirus.

12) Centers for Disease Control (2020, April 10). Cases in U.S.