How the Spread of COVID-19 Has Been Lowered by Quarantining

Alec Barker

SIR models can map the spread of an infection through a population by using differential equations to track the number of susceptible, infected, and recovered individuals. A Python program was created to use these equations in an attempt to map the current worldwide spread of COVID-19. After obtaining values from the SIR model equations, those values were compared to actual COVID-19 data. Once the data was compared, it became clear that widespread social distancing and other safety precautions have slowed the virus' spread significantly. It also shows that the SIR model is not indicative of real life. With people around the globe taking action to combat the virus, the infection rate has lowered and will likely result in less people becoming infected overall. If the world had not acted the way it has, the peak number of infected individuals would have been significantly higher as well. This change in the peak helps researchers and medical staff combat the virus by giving them more time to study it. The information in this report shows that reducing face-to-face interactions during the COVID-19 pandemic has kept the world healthier and minimized the effect that that the virus has had on the worldwide population. It also shows that the SIR model doesn't fit perfectly to the real world where many factors can influence the spread of an infection.

1. Introduction

1A. SIR Model (1)

The rate that an infection can spread through a population can be roughly mapped using the SIR model as proposed by W. O. Kermack and A. G. McKendrick in 1927. Using this model, the population is split into three categories; **S**usceptible, **I**nfected, and **R**ecovered. The rates of each category can be mapped using the following differential equations:

$$\frac{dS}{dt}=\frac{-\beta IS}{N} \qquad \frac{dI}{dt}=\frac{\beta IS}{N}-\gamma I \qquad \frac{dR}{dt}=\gamma I$$

Where:

S = Number of susceptible individuals in a population

I = Number of infected individuals in a population

R = Number of individuals in a population that can no longer spread the disease (due to either recovering from the infection or dying)

N = Total population, equal to S + I + R

β = The rate of infection, equal to $\dfrac{1}{\text{number of infections per infected individial per time unit}}$

γ = The rate of recovery, equal to $\dfrac{1}{\text{duration of disease in time units}}$

The default model does not take extra real-world variables into account, such as birth and death rates, incubation periods, quarantining, etc. The basic SIR model has a constant population size and constant rates of infection and recovery. It also assumes that the population is well-mixed through interaction and that every individual has an equal chance of contracting the infection. These restraints on the model can never fully simulate the real world and the many factors that go into how likely it is that an individual will contract a certain infection. The SIR model is probably more accurate in smaller populations that have little change over the lifespan of the infection. Smaller groups live in similar environments and likely have similar immune systems. The infection can also live and die fairly quickly in a small population before it becomes widely known, which would make it less likely to be seriously combated by health organizations. For these reasons and others, the SIR model is better suited for small groups. However, the model does give a basic idea of how an infection will spread through a population, regardless of size. Larger populations will still follow the same trends, but they may not follow the SIR model as closely as smaller populations do.

1B. Basic Reproduction Number, R_0

R_0 is a number that represents the number of susceptible individuals that will become infected by a single infected individual *(2)*. It is worth noting that R_0 is not a rate, but rather a count of individuals within a population. R_0 is useful because it gives researchers an idea of how steep the exponential curve of infected individuals will be during the spread of an infection. It also determines if an infection will spread (if $R_0 > 1$) or die out (if $R_0 < 1$). R_0 can be calculated using the following formula (please note that Jones uses "v" to denote the rate of recovery, while this paper uses "γ") *(3)*:

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

Some R_0 values can be found in Table I below for well-known infectious diseases, as well as the current R_0 value for COVID-19.

Disease	**R_0**	**Reference**
Measles	12-18	*(4)*
Chickenpox	10-12	*(5)*
Polio	5-7	*(6)*
Mumps	4-7	*(6)*
Smallpox	3.5-6	*(7)*
HIV/AIDS	2-5	*(6)*
SARS	2-5	*(8)*
COVID-19	2.4-3.3	*(9)*
Seasonal Influenza	0.9-2.1	*(10)*

Table I. R_0 values for well-know infectious diseases.

The R_0 value for COVID-19 listed in this table (2.4-3.3) will be used in calculations for this report. This is particularly helpful since infection rates of COVID-19 are not as well publicized. Using the equation for R_0 and knowing the length that an individual is infected, the infection and recovery rates can be calculated and plugged into the SIR equations.

1C. Scope and Objectives

Given recent events of the COVID-19 worldwide pandemic, it seems informative to understand how the outbreak might play out. The purpose of this report is to show what might lie ahead and how social distancing and other safety precautions have helped reduce the impact of the outbreak thus far. This report also illustrates that the SIR model for mapping infections is not fully accurate to the real world and can be affected by many factors.

Using COVID-19 data found from various sources, this report will first attempt to find an SIR model that matches the current trend of COVID-19. The purpose of this is to show a comparison between real numbers from the European Centre for Disease Prevention and Control (ECDC) and the numbers estimated by the SIR equations. Some values will be manipulated in the SIR model to best fit the original spread of COVID-19 so that a comparison can be made. If the manipulated values in the SIR model are similar to the actual data, then the model can be considered fairly accurate to the real-world situation and can be used to discuss the spread of COVID-19. The SIR model will show how the virus might have spread thus far and how it would have spread in the future if people around the world hadn't tried to reduce its impact. The SIR model will be graphed alongside the ECDC data to show how quarantining has affected the spread of the virus. There is a large difference between the infection curves of the real and modeled data, and the gap between them should help display how effective the precautionary measures have been thus far.

The graphs and data found in this report should help some people understand the effects of avoiding others, washing hands, wearing gloves and masks, and other safety measures during an epidemic. Hopefully this report will inform readers that if they continue to follow the recommended safety measures, they will not only be keeping themselves healthy, but they will also help lower the worldwide spread of COVID-19.

2. Results

The real data values obtained from the ECDC list the number of confirmed cases of COVID-19 per day *(11)*. The values slowly grow in the beginning and then increase exponentially. Around late March and early April of 2020, the numbers begin to flatten and grow at a much slower rate. This can be seen in Figure 1 below. The flatter curve is likely due to massive quarantine measures put in place by world governments around that time. Please note that this paper uses January 13, 2020 as a starting date of COVID-19. There were cases before January 13, but they were either not recorded by health organizations or were lumped together as one value on December 31, 2019. To compare the data to an SIR model, January 13 was chosen because it has one confirmed case and continues to grow from there.

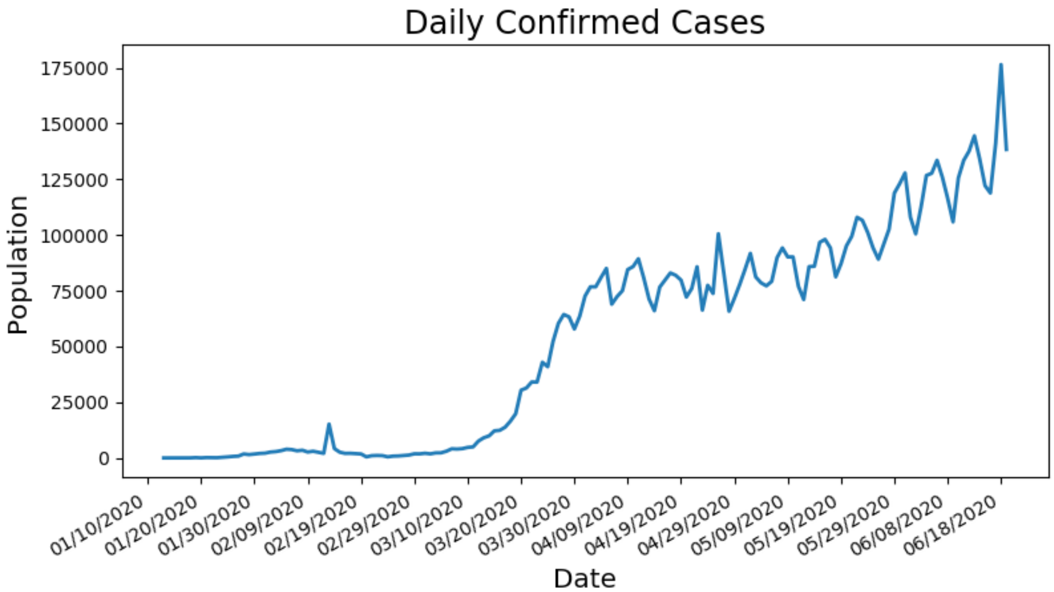


Figure 1. The number of daily confirmed cases worldwide of COVID-19 from January 13, 2020 to June 19, 2020.

From here, an SIR model was created to attempt to mimic the early parts of COVID-19's spread. To create the SIR model, the ranges of values for R_0 and the duration of the disease were each averaged and then slightly adjusted until they appeared to fit the initial curve of real data. The values that produce a graph that is most similar to the actual data provided by ECDC are 3.09 for R_0 and 14 days for the duration of the virus. Using these two values, the program was able to run through the SIR equations to create an estimation of the spread of COVID-19. Figure 2 shows the curves calculated using the SIR model equations.

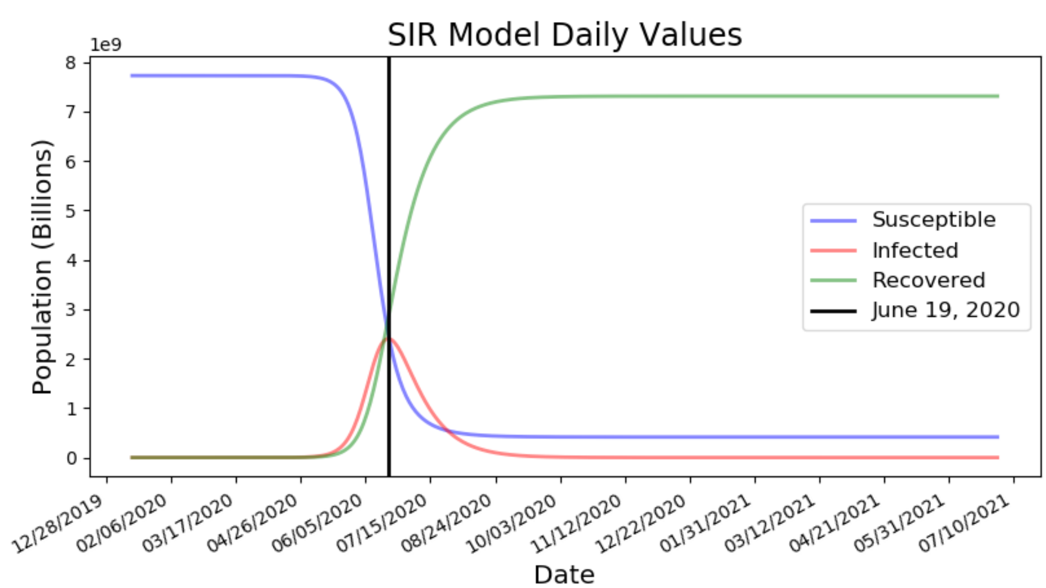


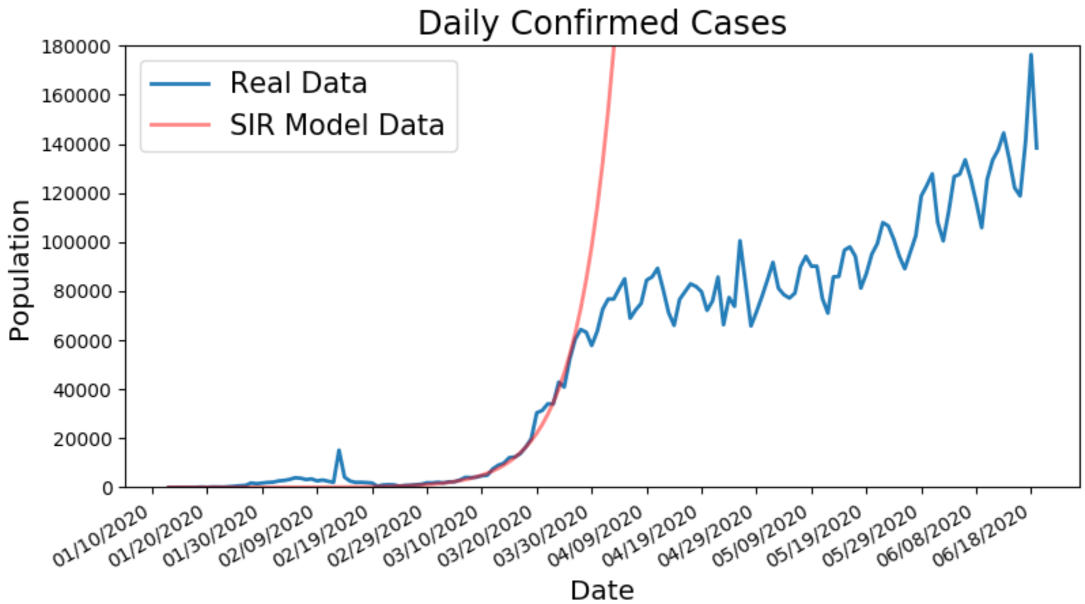
Figure 2. The estimated progress of COVID-19 over time between January 13, 2020 and June 30, 2021 as predicted by the optimized SIR equations.

The curves of this graph show the overall trends of COVID-19 if it had been left unchecked while it spread through the world population. As the number of infected individuals increases, the number of susceptible individuals decreases. At the same time, infected individuals begin to recover from the virus and the number of recovered individuals increases. Eventually, the number of individuals being infected each day levels out and then decreases. If COVID-19 had not been combated in any meaningful way to impact its infection rate, then the data obtained in the SIR equations can be used to predict some statistics about the way the virus might have spread. These statistics can be found in Table II.

Category	**Value**
Total Number of Individuals in the World	7,722,935,550
Total Number of Individuals Never Infected	414,858,541
Total Number of Infected/Recovered Individuals	7,308,077,009
Peak Number of Infected Individuals	2,403,926,457
Date of Peak Number of Infected Individuals	June 19, 2020
Total Duration of Pandemic	534 days
Last Date of Pandemic	June 30, 2021

Table II. Statistics about the optimized SIR model and the potential spread of COVID-19 if it had continued its course without intervention.

These statistics represent how COVID-19 might have spread through the world population if it had been left unchecked. The number of individuals infected by the virus would have risen until 2.4 million people would have been infected with COVID-19 in a single day, and the virus would have completely died out after a total of 534 days. However, these statistics were changed in the real world by reducing the infection rate. R_0 was lowered by people practicing social distancing and staying at home, and the infection rate went down as a result. SIR models don't capture all of the dynamics that may affect the spread of a disease, and this is clearly shown by the discrepancy between the real and modeled data. Figure 3 shows how much change there has been to the spread of the virus once more people began to protect themselves and others from the virus.



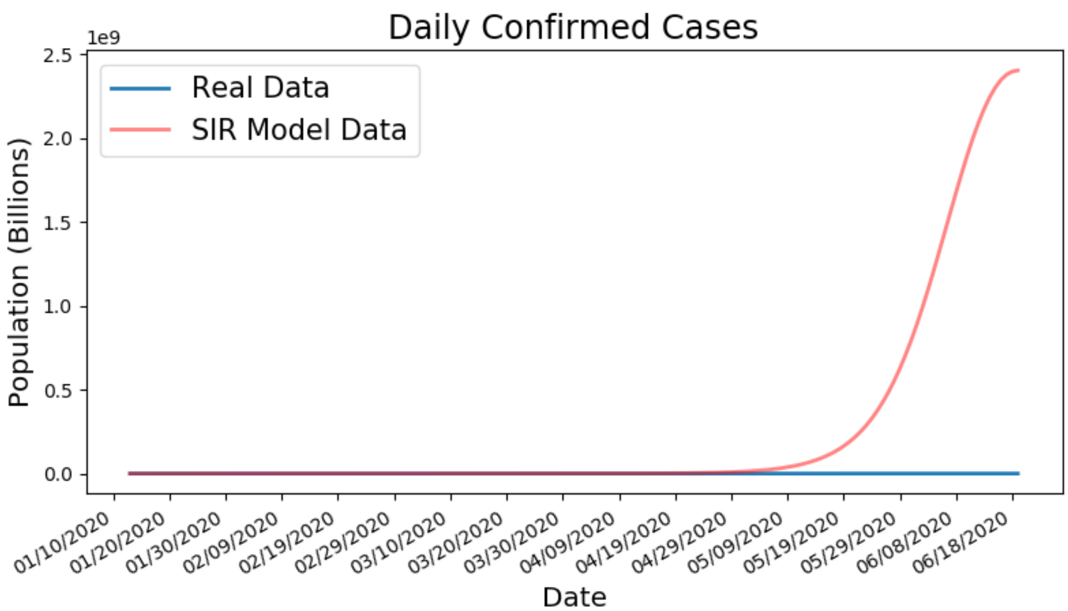


Figure 3. A comparison between the actual spread of COVID-19 and the spread that was predicted using the optimized SIR equations from January 13 to June 19. The top graph is zoomed in to make the real ECDC data visible. The bottom graph is zoomed out to show the full extend of the SIR model as it appears on June 19, 2020.

Figure 3 shows that as the number of daily confirmed cases began to flatten and grow at a much slower rate since late March and early April, the impact of the virus could have been much worse. The second graph shows that if no efforts had been put into stopping the virus, then the number of daily confirmed cases would have been much higher than they currently are. Table III shows some statistics related to these graphs.

Category	Value
Real Number of Confirmed Cases for June 19, 2020	138,360
Optimized SIR Model's Number of Cases for June 19, 2020	2,403,926,457
Difference Between the Real and Modeled Number of Cases for June 19, 2020	2,403,788,097

Table III. Statistic comparing the real number of confirmed cases of COVID-19 and the number obtained by the optimized SIR model for June 19, 2020.

The difference between the real number of confirmed cases and the optimized SIR model's number of cases for June 19 is rather large. Some may think that this indicates that the SIR model is incorrect and does not actually fit with the real data. This is partially true. As stated previously, the SIR model is not perfect and does not account for many external variables that affect the spread of an infection. A large reason that the model and the real data points do not match up currently is that governments around the world began to mass quarantine their citizens starting in March. With less people traveling outside of their homes and interacting with one another, there is a lesser chance that there will be new confirmed cases. Many people have also been taking some steps to protect themselves, such as washing their hands more often, wearing masks and gloves when they need to go outside, and more. The SIR model does not account for this type of behavior in a society. The model assumes that every individual in a population will do nothing to combat an infection and that every individual interacts greatly with other people. If quarantines hadn't been put into place and people hadn't practiced social distancing, the graph of actual data would be very different. It would probably be close to the modeled data as the number of confirmed cases continued to grow exponentially.

3. Methods

The first thing that needed done to test the SIR model was to get actual data to compare to. This was done by taking data reported by the ECDC on their website about the daily number of cases *(11)*. The CSV file containing this data can be found in Appendix 1. A Python script was made to pull data from the downloaded CSV file, as seen in Appendix 2. The script goes through the file, totaling the number of infected individuals for each country in a single day. The program intends to show data regarding the worldwide outbreak of COVID-19, so it does not distinguish between countries. The ECDC file contains death counts but no numbers on the number of people that recover from the virus, so this script is only worried about the number of confirmed cases each day. The script goes through each line and pulls the date and the number of confirmed cases. For each line, the program checks if the date has already been stored in a Python dictionary. If it has not, the date and number of confirmed cases is added to the dictionary. If the date already exists in the dictionary, the program will add the number of confirmed cases for that line of the file to the value that already exists in the dictionary. Once the file has been completely sorted, the data is stored in a CSV file, and it can be viewed in Appendix 3.

Next, an estimation was made about the values required for the SIR equations. As noted in Table I, the R_0 of COVID-19 is 2.4-3.3. Reports vary on the length that an individual is infected with COVID-19, but the Centers for Disease Control and Prevention (CDC) provide some information on their website. They state that the symptoms typically appear 2-14 days after contracting the virus *(12)*. They also say that for most people, symptoms will subside and the person will recover from the virus after a week *(13)*. This means that the overall duration of time that an infected individual may infect others is roughly 9-21 days. The averages of each value were initially used. The average of R_0 is 2.85, and the average for the duration is 15 days. Using these averages, the R_0 equation was rearranged in the following way:

$$\beta = R_0 \gamma = \frac{R_0}{\text{duration of disease in time units}}$$

Using the modified equation, both β and γ values can be found and plugged into the SIR equations. The total number of individuals in the population must also be known to complete the equations. Sources vary on the world's current population, so four values were taken from various websites at the time of writing, and the average was used *(14-17)*. This average value states that 7,722,935,550 people are currently alive in the world. The average values of R_0, the duration of the infection, and the global population were used to produce an SIR model, but the curve that they created did not match well with the actual data curve. The numbers were tweaked slightly until the curve looked like it more closely matched the curve of the real data. The values that created the optimized curve are 3.09 for R_0 and 14 days as the duration. Using these values, another SIR model was created and some statistics about the model were tracked and stored. This script can be viewed in Appendix 4.

4. Discussion

The SIR model can be used to understand the way that an infection may spread through a population. SIR models use sets of differential equations to keep track of a population of susceptible, infected, and recovered individuals. SIR models are often not completely accurate to the spread of a real disease throughout a population. They lack various variables that exist in the real world. This lack of external variables is shown clearly in this report. The SIR model did not account for people quarantining, social distancing, and attempting to stay healthy. This goes directly against the SIR model's assumption that the population is well mixed. After a short time, the real data diverged from the model that was created and created a large gap between reality and the model.

R_0 values represent the number of susceptible individuals that are infected by a single infected individual. This number is useful for determining how severe an outbreak of an infection will be. Higher numbers results in a quick spread of the infection, while lower numbers have a low but steady incline. If the R_0 is lower than 1, the infection will die out quickly. The value of R_0 can be used to determine the infection rate that plugs into the SIR equations.

Taking data supplied by ECDC about the spread COVID-19, this program attempts to create an SIR model that mimics the real number of infected individuals. It does so by using known facts about COVID-19 provided by CDC. Figures were generated from this model that display the way COVID-19 is currently spreading and how it might have spread in the future if humanity had not attempted to combat it. Tables were also created to show some statistics about the virus and how its effects have been minimized by quarantining.

It is worth noting that this data is not completely accurate to real life. Instead, it is meant to show a general trend of how the spread of COVID-19 may have played out. Every individual is different and has different odds of becoming infected. Some notable changes between individuals that change the likelihood of infection are different immune systems, ages, areas of residence, occupations, and many other factors. It is impossible to map out all of these differences between individuals. Therefore, a general population must be used. There are also some factors about the virus or overall population that are not included. For example, this study does not include the births and deaths that occur each day or the idea that recovering from the virus might not make an individual immune to COVID-19, thus allowing them to become infected again. There are various models that map these sorts of factors that may occur in the real world. In a future study, one or more of them may be used to gain a more accurate representation of the virus.

Some people may ask why anyone would want to flatten the curve. Looking at the graphs provided in this report, it become apparent that less people are being infected daily and that humanity is in a much better place than it could be. It is extremely beneficial to attempt to flatten the curve and reduce the infection rate of COVID-19. There are a few reasons that a shallower curve would help combat the virus. One reason is that it helps with hospital capacity. Hospitals currently have to deal with the normal amount of emergency-related incidents and long-term patients, and they now face an influx of new patients who have COVID-19. Hospitals only have a limited number of beds and staff, and if there is a steep curve of infected individuals, then hospitals will have a hard time dealing with the high number of patients all at once. Many hospitals will reach capacity and be forced to turn away people who need medical attention. Another benefit of flattening the curve is that it gives researchers more time to create a vaccine. Instead of everyone getting sick quickly and scientists trying to find a cure in a short period of time, a flattened curve means that there will only be a small infected population at any given time, and the researchers have a bit more time to develop a vaccine. If scientists are able to vaccinate people to be immune to the virus, then the remainder of the population can avoid to the effects of the virus. If researchers find a vaccine, then the overall number of infected individuals will be lowered, and they are more likely to do so over a longer period of time.

Thanks to everyone practicing social distancing and staying at home whenever possible, the infection rate of COVID-19 has go down. This has put less stress on hospitals and gives researchers more time develop a vaccine that can save lives. While there might not be much that an average person can do to combat the virus, avoiding contact with others can be done by anyone. If people distance themselves from one another, they will be less likely to contract the virus themselves, and they also help the world get rid of COVID-19.

Works Cited

1. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **115** (772), 700-772 (1918).
2. C. Fraser, C. A. Donnelly, S. Cauchemez, et al., Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* **324** (5934), 1557-1561 (2009).
3. J. H. Jones, Notes on R_0. *Califonia: Department of Anthropological Sciences*, 2 (2007).
4. F. M. Guerra, S. Bolotin, G. Lim, J. Heffernan, S. L. Deeks, Y. Li, N. S. Crowcroft, The basic reproduction number (R0) of measles: a systematic review. *The Lancet Infectious Diseases* **17** (12), e420-e428 (2017).
5. Ireland's Health Services, Chapter 23 – varicella-zoster. *Health Care Worker Information – Immunisation Guidelines*, 2 (2020).
6. Centers for Disease Control and Prevention (U.S.), World Health Organization., History and epidemiology of global smallpox eradication. *Smallpox: Disease, Prevention, and Intervention*, 17 (2001).
7. R. Gani, S. Leach, Transmission potential of smallpox in contemporary populations. *Nature* **414** (6865), 748-751 (2001).
8. J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* **160** (6), 509-516 (2004).
9. P. GT. Walker, C. Whittaker, O. Watson, et al., The global impact of COVID-19 and strategies for mitigation and suppression. *Imperial College London – COVID-19 Reports* **13**, 7 (2020).
10. B. J. Coburn, B. G. Wagner, S. Blower, Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Medicine* **7** (30), (2009).
11. Situation update worldwide. *European Centre for Disease Prevention and Control* (2020).
12. Coronavirus disease 2019 (COVID-19) – symptoms & testing – symptoms. *Centers for Disease Control and Prevention* (2020).
13. Coronavirus disease 2019 (COVID-19) – if you are sick – caring for someone. *Centers for Disease Control and Prevention* (2020).
14. World population. *countrymeters* (2020).
15. Live world population clock. *Population of the world* (2020).
16. Current world population. *worldometer* (2020).
17. U.S. and world population clock. *United States Census Bureau* (2020).

Appendices

Appendix 1: *covid19_data* (No File Extension)

Data provided in CSV format by ECDC regarding the number of people worldwide that are infected by COVID-19 daily. This file was obtained for this report on June 19, 2020.

Appendix 2: *data_scraper.py*

A script that takes the data from the *covid19_data* CSV file provided by the ECDC and condenses it into the total number of infected individuals worldwide for each day since the start of the outbreak.

Appendix 3: *covid19_condensed_data* (No File Extension)

The data taken from ECDC's CSV file that has been condensed by *data_scraper.py* that represents the total number of infected individuals each day since the beginning of the pandemic.

Appendix 4: *covid19_sir.py*

A script that takes data from the *covid_condensed_data* file and creates an SIR model that attempts to replicate the *covid_condensed_data* data.