Group members:

Oi Ting Cheung, Jamie Davis, Liqi Zhao, Jonathan Martinez, Vicky Kong

Professor: Longxiu Huang

Date: 5/14/21

A Model of the Effect of Vaccinations on the Spread of the COVID-19 Virus

ABSTRACT

The purpose of this report is to provide an insight into understanding the spreading of the

COVID-19 virus in the US. We developed a modified Susceptible-Infected-Recovered (SIR) model to study the contagion in the US in three different scenarios (1. Neglect vital dynamics and vaccination rate, 2. Neglect vaccination rate only, 3. Consider all parameters). Knowing whether the infection can be controlled assists our understanding of COVID-19 contagion. Thus, we studied the equilibrium and stability of the system. The addition of factors such as vital dynamics and the vaccination rate affects the equilibrium of the system. In case 1, the number of susceptible individuals or the number of infected individuals must equal zero in order to achieve equilibrium. For cases 2 and 3, the number of infected people must equal zero, and birth rates must equal death rates in each population in order to reach equilibrium. Case 3 achieves equilibrium fastest due to the introduction of vaccines.

Then, based on the recent official data, we simulate the model and create predictions by

training this model. For the simulation part, the number of COVID deaths increases if there is a high infection rate, infection death rate, or reinfection rate. A high recovery rate or vaccination

rate results in reducing the population of individuals who die from the virus. The graph of the prediction result obtained by training the model with a linear regression in an iterative manner is similar to the baseline graphs of simulations based on infection rate, reinfection rate, and infection death rate. The simulations show that vaccination can effectively control the spreading of COVID-19 in the long term.

INTRODUCTION

Over the past year, COVID-19 has caused many deaths. We are interested in how the spreading of COVID-19 affects the US population. Knowing how COVID-19 is spread may enable a more efficient distribution of vaccines and provide insight on controlling the virus. In this report, we modified a mathematical model (Susceptible-Infected-Recovered (SIR) model) and analyzed official statistics from the Centers for Disease Control and Prevention (CDC) to understand the COVID-19 contagion in the United States. A mathematical model is a description of a system using mathematical concepts and language. It can “offer a precious tool to public health authorities for the control of epidemics, potentially contributing to significant reductions in the number of infected people and deaths” (Giuseppe, Carlo, and Corrado, 2020). The model parameters will provide an overview of its effects on the US population. We modified the SIR model to SIRDVS (Susceptible-Infected-Recovered-Deceased- Vaccinated-Susceptible) model. We will discuss our model in three scenarios:

1) Neglecting the effect of natural birth rate and natural death rate (vital dynamics), and rate of vaccinated individuals on the US population
2) Neglecting the effect of the rate of vaccinated individuals on the US population
3) All model parameters take into account

Additionally, as COVID-19 vaccines were recently invented, we will also predict their long-term effect on the US population. We chose these three scenarios because we want to know how they change the proportion of different groups of individuals in the whole population. We will further explain the model in the section of Model Explanation. After the model explanation, we will investigate the model theoretically (through studying the equilibrium of the system) and simulate the model in the long term (through training the SIRDVS model based on current COVID-19 data) because we would like to know whether the mathematical model is able to provide information about the spreading of COVID-19 and the effects of different factors on the US population.

MODEL EXPLANATION

Let's start by defining the following components for our model:

- $S(t)$: the number of individuals susceptible to getting infected at time t;
- $I(t)$: the number of infected individuals at time t;
- $R(t)$: the number of individuals that have recovered from the disease at time t;
- $D(t)$: the cumulative number of individuals that are deceased due to the disease at time t;
- $V(t)$: the cumulative number of individuals that got a full vaccine at time t;

with the assumptions that:

- The US population is isolated from other regions with a total population $N = S + I + R$.
- Recovered individuals can still be susceptible to getting reinfected.
- The unit time t is expressed in days.
- Units S, I, R are expressed in persons.
- Vaccinated individuals can not be infected nor reinfected.

- All rates are constant.
- The initial conditions are $S_0 > 0, I_0 > 0, R_0 \geq 0$, $D_0 \geq 0$, and$V_0 \geq 0$.

We have decided to do three separate scenarios using a modified SIRS model, from Kermack and McKendrick, into SIRDVS (Susceptible-Infected-Recovered-Diseased-Vaccinated-Susceptible) model. Here are the following three cases we considered:

Case 1: Neglecting Natural Birth Rate, Death Rate, and Vaccination Rate

In this first case, we are neglecting the birth rate, general death rate, and the vaccine rate, where we define general death to mean deaths that are not caused by COVID-19. Using our SIRDVS model, we have:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} + \xi R$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \delta I - v'I$$

$$\frac{dR}{dt} = \delta I - \xi R$$

$$\frac{dD}{dt} = v'I$$

where $\beta$is the infection rate, $\delta$is the recovery rate, $v'$is the mortality rate from the disease, and $\xi$is the reinfection rate.

Case 2: Considering Vital Dynamics and Neglecting Vaccination Rate Only

In Case 1, the population only decreases caused by the disease. However in practice, our population fluctuates depending on the number of births and deaths every day, therefore, in the second case, we included the birth rate and natural death rate:

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} + \xi R - vS$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \delta I - v'I - vI$$

$$\frac{dR}{dt} = \delta I - \xi R - vR$$

$$\frac{dD}{dt} = v'I$$

where μ is the birth rate, and$v$is the natural death rate. We have two different death rates, which are the number of natural deaths that affect our population N(denoted as $\nu S, \nu I$,and νR), and the number of deaths caused by the disease (denoted as $\nu' I$).

Case 3: Accounting for Natural Birth Rate, Death Rate, and Vaccination Rate

With the recent development of vaccination shots, we have added the parameter of the vaccination rate. Thus, we have:

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} + \xi R - vS - \alpha S$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \delta I - v'I - vI$$

$$\frac{dR}{dt} = \delta I - \xi R - vR$$

$$\frac{dD}{dt} = v'I$$

$$\frac{dV}{dt} = \alpha S$$

where $\alpha$is the vaccination rate. Recovered individuals and non-infected individuals that took the vaccine shot are no longer susceptible. Therefore, they are not accounted for in the susceptible equation and are represented as αS.

EQUILIBRIUM AND STABILITY

In order to identify whether the infection is containable, we are interested in studying the equilibrium of the system. Equilibrium occurs when the number of people in each population stays constant, or when $\frac{dS}{dt}, \frac{dI}{dt}, \frac{dR}{dt}, \frac{dD}{dt}$ all equal zero. If there is no equilibrium, then it would be necessary to add one or more elements in order to mitigate the spreading of the disease. For the COVID-19 virus, governments legislated stay-at-home orders and mask policies. For our purposes, we assume that the rate of transmission, β, is determined with these policies in mind. As the entire system stems from the population of susceptible individuals, the simplest way to achieve equilibrium is to minimize the number of susceptible individuals in the population. In case 1, a net change in the populations of each category equal to zero is only achievable when

$$\xi R = \frac{\beta SI}{N}$$

$$\frac{\beta SI}{N} = \delta I + v' I$$

$$\delta I = \xi R$$

$$v' = 0 \ or \ I = 0$$

$$\Rightarrow \quad \xi R = \frac{\beta SI}{N} = \delta I$$

or, when the rate at which recovered individuals can become susceptible again, the proportion of the total population being infected, and the recovery rate are all equivalent. In order for the virus's death rate to reach equilibrium, the mortality rate must be equal to zero, or the number of infected individuals must be equal to zero. The data on the COVID-19 pandemic suggests that the mortality rate is not zero, therefore it is unreasonable to evaluate the systems of differential equations with this value. Then, the population of infected individuals must be zero. As the number of infected individuals stems from those that are susceptible, it would also be sufficient for that population to decrease to zero. That is, equilibrium can be reached when either the number of susceptible individuals or the number of infected individuals is zero. If the proportion of those recovered who can get the virus a second time is high, then, without mitigating factors, the spread of the virus would not reach equilibrium until the susceptible population has died. During the COVID-19 pandemic, governments legislated lockdowns in order to reach this equilibrium by isolating the infected individuals, and thus making the infection rate, β, very small and allowing the population of infected people to trend toward zero. However, imperfect lockdowns did not sufficiently lower the infection rate, meaning the equilibrium was not reached as quickly as intended.

Accounting for vital dynamics changes the considerations of the equilibrium.

In case 2, the conditions for equilibrium are as follows:

$$\mu N + \xi R = \frac{\beta SI}{N} + vS$$

$$\frac{\beta SI}{N} = \delta I + v'I + vI$$

$$\delta I = \xi R + vR$$

$$v' = 0$$

$$\Rightarrow \mu N = vR + vI + vS + v' I$$

$$or\ I = 0$$

$$\Rightarrow \mu N = vS$$

The equation for the death rate does not change, as we are not interested in deaths unrelated to the virus. Similar to case 1, v' still cannot be equal to zero. So again, in order to achieve equilibrium, the population of infected people must be equal to zero, and the total number of births in the population must be equal to the total deaths in each population. If the infected population is zero, then the recovered population will trend toward zero. So then to achieve equilibrium for the entire system, the birth rate and the death rate must be equal in order to maintain a stable population. This system will reach equilibrium more quickly than case 1, as the death rate decreases the infected population and susceptible population. However, the addition of mitigating factors will allow this system to reach equilibrium more quickly.

The conditions for case 3 are as follows:

$$\mu N + \xi R = \frac{\beta SI}{N} + vS + \alpha S$$

$$\frac{\beta SI}{N} = \delta I + v' I + vI$$

$$\delta I = \xi R + vR$$

$$v' = 0$$

$$\Rightarrow \mu N = vR + vI + vS + v' I + \alpha S$$

$$or\ I = 0$$

$$\Rightarrow \mu N = vS + \alpha S$$

This model achieves equilibrium more quickly than the previous two because the introduction of vaccines decreases the susceptible population to zero, and the infected population quickly follows, meaning the number of recovered individuals and individuals killed by the virus will remain constant.

It is also important to analyze the point at which $\frac{dI}{dt}$changes from positive to negative, as this means that the number of infected people is decreasing, and therefore trending toward stability.

In the following figures, we see that with the addition of vaccines, the number of susceptible individuals decreases more quickly than the other simulations, with a more dramatic slope as the vaccination rate increases. In figure 5, we see that when the vaccination rate is 0, the maximum number of individuals that are infected is 350,000 and this peak occurs around day 150. In the simulations where the vaccination rate is greater than 0, we see that the maximum number of individuals that are infected is less than 200,000 and this peak occurs before day 100. As the vaccination rate goes up, this maximum number decreases and the number of infected individuals begins to decrease in fewer days. This is sufficient evidence to show that the addition of vaccines leads to a shorter pandemic with fewer total infections.

SIMULATION AND PREDICTION

In this part, we will first simulate the distribution of population in each of the categories (susceptible, infected, recovered, deceased, and vaccinated) over a period of 2 years with different inferred parameters (infection rate, recovery rate, reinfection rate, covid death rate, and vaccination rate). Notice that instead of investigating the influence of the natural birth rate and natural death rate, we choose to use default data from online sources for better model convergence. After we understand the impact of individual parameters, we will train our SIRDVS model on the data we have for COVID deaths, cases, and recoveries over a span of 10 months with plain linear regression and gradient boosting linear regression methods. After we obtain the parameters for the SIRDVS model, we will compare the future trends predicted by both methods and discuss their differences.

To assess the influence of different parameters on the severity of the pandemic, we isolate each of them and simulate with the same initial conditions. Here we choose $S_0 = 10^6$, $I_0 = 10^2$, $R_0 = V_0 = D_0 = 0$ as the distribution at day 0. The baseline parameters are listed below in Table 1,

Parameter Name	Rate (per day)
Natural Birth Rate (μ)	$3.288*10^{-5*}$
Natural Death Rate (v)	$2.459*10^{-5*}$
Infection Rate (β)	$3*10^{-2}$
Reinfection Rate (δ)	$2*10^{-3}$
Infection Death Rate (v')	$1*10^{-3}$
Recovery Rate (δ)	$3*10^{-3}$
Vaccination Rate (α)	0

Table 1. Baseline values for parameters in SIRVDS model, data with * are from sources (The World Bank 2021)

a. Infection Rate (β)

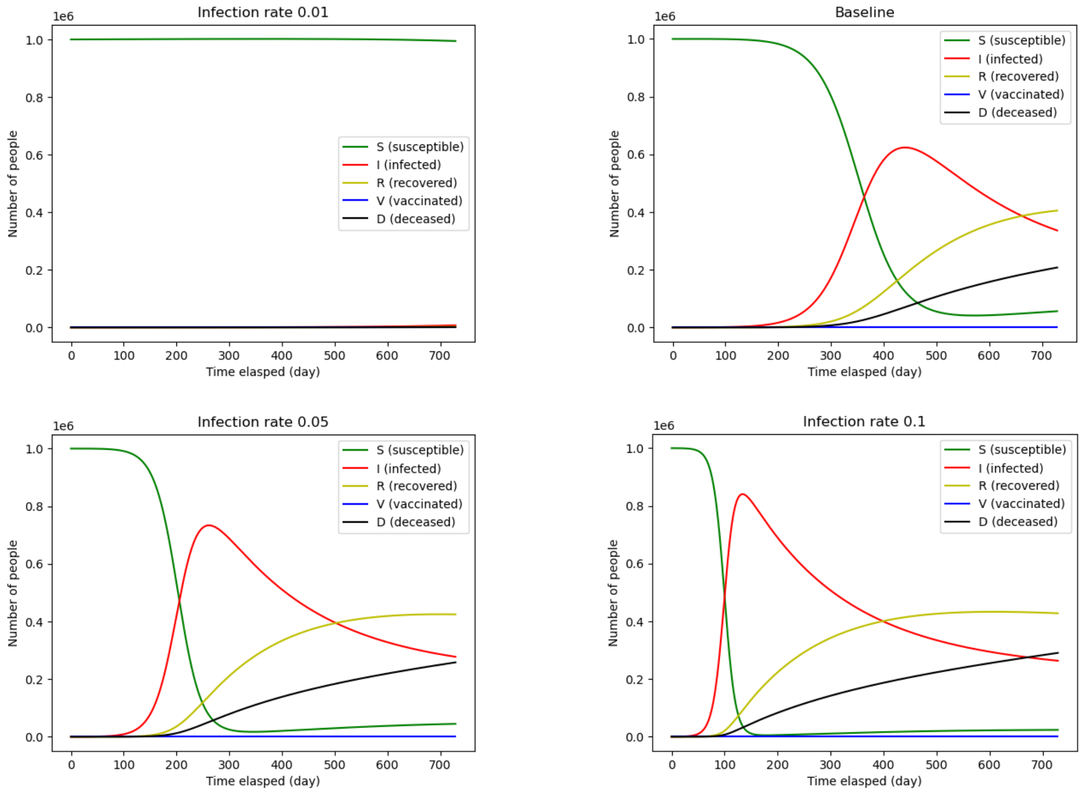

Figure 1. Simulations with different infection rates

From Figure 1, we could observe that as the infection rate increases, the curve for infected cases is steeper in the ascending phase and the peak comes earlier. Also, the peak will be higher and thinner when the infection rate is higher. Correspondingly, the deceased population will start to climb at a higher speed around the peak of infection.

b. Reinfection Rate (δ)

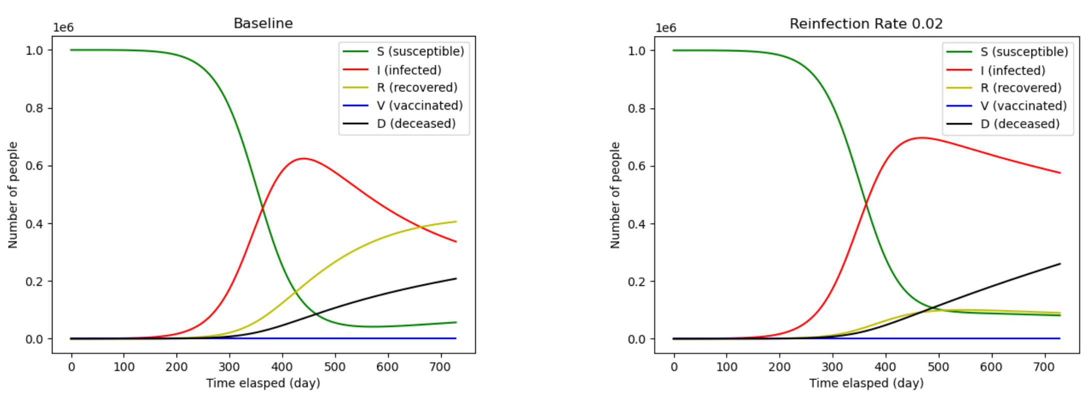

Figure 2. Simulations with different reinfection rates

From Figure 2, we could observe that a high reinfection rate will decrease the descending rate of the infection curve and increase the growth rate of deaths. The recovery curve will be flattened and the infected population will stay at a high level in the long term.

c. Infection Death Rate (v')

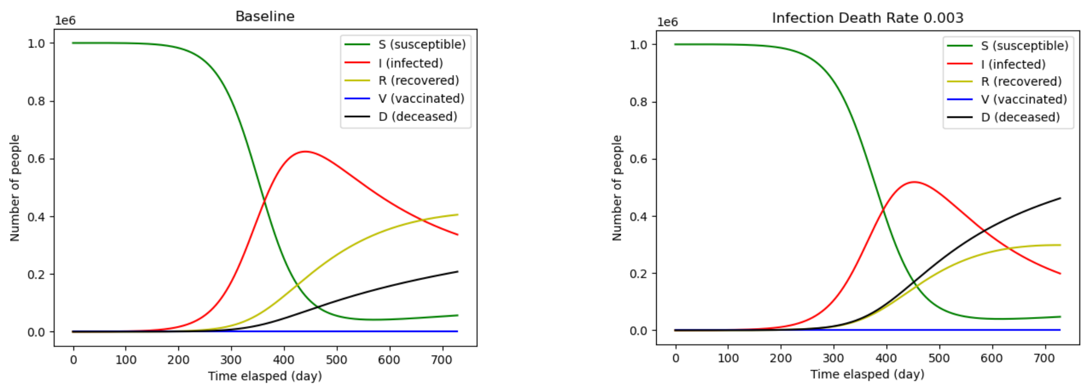

Figure 3. Simulations with different infection death rates

From Figure 3, we could observe that with a higher death rate during infection, the infection curve will be flattened while the death curve will increase faster. The deceased will surpass the infected at a very early stage. The pandemic will be shorter but more damaging.

d. Recovery Rate (δ)

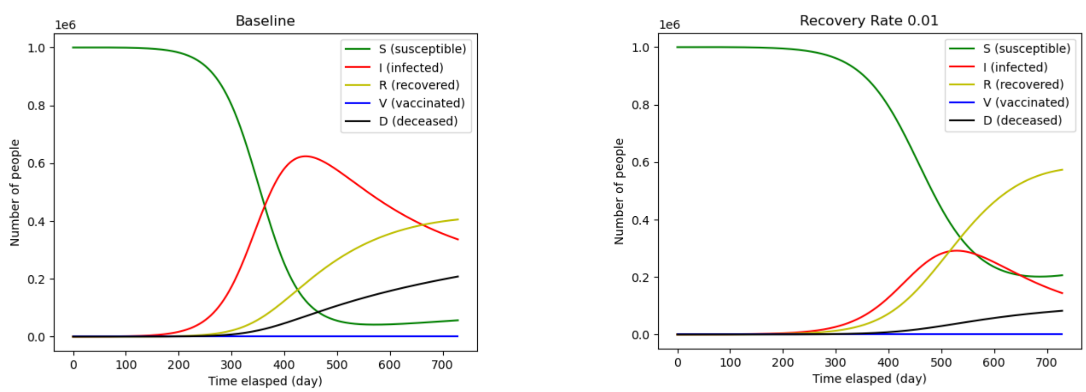

Figure 4. Simulations with different infection recovery rates

From Figure 4, we could observe that a higher recovery rate will effectively flatten the infection curve both at the ascending stage and descending stage. The death curve is greatly flattened with more healthy people staying healthy as well.

e. Vaccination Rate (α)

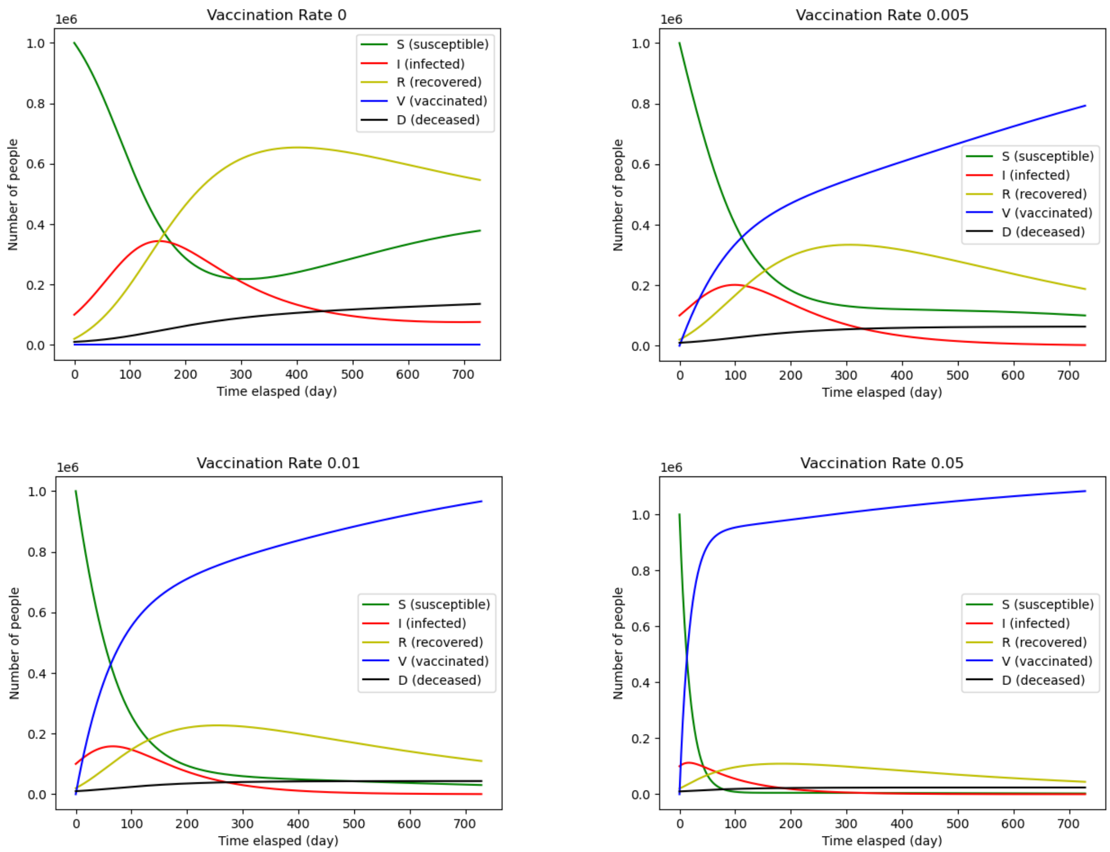

Figure 5. Simulations with different vaccination rates

To investigate the influence of vaccination, we will simulate the model in the middle of the pandemic. Hence, we will alter the initial conditions as follows: $S_0 = 10^6$, $I_0 = 10^5$, $R_0 = 2*10^4$, $D_0 = 1*10^4$, $V_0 = 0$. The results are shown below in Figure 5. We could see that vaccination in the middle of the pandemic could reduce the number of infections and deaths fairly effectively. The faster the vaccination rate is, the lower the peak of infections and deaths will be. Also, the peak of infection will occur around 40% to 60% vaccination rate, which indicates that when roughly half of the population are vaccinated, we might achieve herd immunity for the population.

After the simulation, we choose to train the SIRDVS model on the data we acquired online, including the number of COVID cases and deaths (New York Times 2021) and the number of recoveries (Rajkumar 2021) from January 2020 to December 2020. We derive the parameters from 2 linear models and simulate the future trend with generated results. The first model is the linear regression model with feature space being the aggregation of the matrix derived from the differential equations. We concatenate the linear equations for each day and construct one matrix for all the data as our feature space. The labels are concatenated from the ground truth for each date. To account for the diminishing effects of data over time, we apply a decay weighting parameter to reduce the weight of data from the distant past (Calafiore, Novara, and Possieri 2020). Below is the prediction generated by the linear regression model.

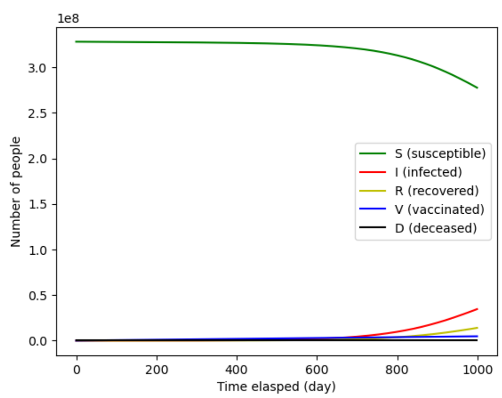

Figure 6. Prediction with plain linear regression

In Figure 6, the linear regression model predicts that the number of cases will increase continuously and it might reach equilibrium very far from now. The prediction indicates that the infection rate will be low but the recovery rate will be low as well, which might lead to bigger outbursts in the distant future. The model is slightly overfitted with 86.3% accuracy on the training set and 78.5% accuracy on the testing set.

The second model we consider is also using linear regression but in an iterative manner. We try to generate the estimations for parameters on each day individually and try to update the parameters with a certain learning rate over time. Instead of training on data from all dates at once, we train on one day and adjust our estimation on each of the following days. The learning rate is not fixed but with a decaying multiplier to prevent overfitting. The prediction result is shown in Figure 7. Compared with the first model, the boosted linear regression estimates the exponential outburst and peak of the pandemic earlier, which is more congruent with our observation. The prediction indicates that without vaccination, we might expect the peak of infection at around half of the total population.

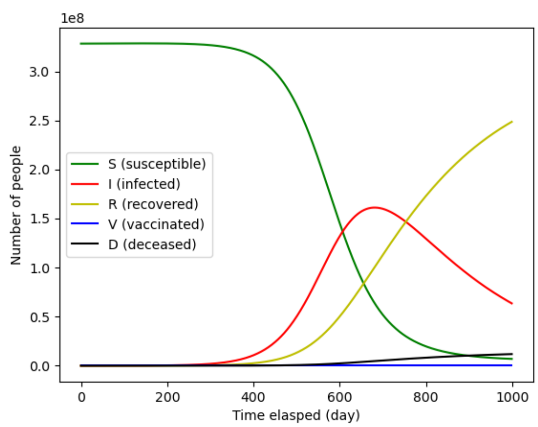

Figure 7. Prediction with gradient boosting for linear regression

CONCLUSION

COVID-19 is still a prevailing issue in the United States and our modified SIR model serves as a simple, yet effective tool in understanding the spread of the virus. To do so we considered three different scenarios for our model where we first neglected vital dynamics and vaccination rates, and then only neglected vaccination rates, and finally considered both rates. In our first case, we found that our model reached equilibrium when infected people are zero. In our second case, since we considered vital dynamics, our model also reaches equilibrium when the infected population is zero, but the natural birth and death rates must be equal. This was also true for the last case. However, we found that the last model reaches equilibrium more quickly than the previous two because of the additional vaccine parameter. Additionally, when simulating our model, we found that increasing infection rate, virus death rate, or reinfection rate increases accelerate the number of deaths from COVID whereas increasing recovery rate, slows down the number of deaths from COVID. When considering vaccinations our simulations suggest that it effectively decreases the numbers of infection and death, even in the middle of a pandemic.

In future manipulations of this model, given additional data on the effect of the vaccine in the population and its differences among age groups, it would be beneficial to consider factors such as the point of herd immunity, at which enough of the population is vaccinated to greatly decrease the infection rate and the differences in infection and death rates among different age groups. There is evidence to suggest that these variables were significantly higher in adults over 70, and accounting for these changes could change the trajectory of the spread, while also introducing the possibility of multiple solutions, such as quarantines for different age groups rather than the entire population. Overall, the SIRDVS model reflects the importance of mitigation factors such as lowering infection rates through mask-wearing, lockdowns, and vaccinations.

REFERENCES

Calafiore, Giuseppe C., Carlo Novara, and Corrado Possieri. 2020. "A time-varying SIRD model for the COVID-19 contagion in Italy." *Annual Reviews in Control* 50 (Mar): 361-372. https://doi.org/10.1016/j.arcontrol.2020.10.005.

New York Times. 2021. Data for covid cases and deaths. In *us.csv*. https://github.com/nytimes/covid-19-data.

Rajkumar, Sudalai. 2021. "Novel Corona Virus 2019 Dataset," Day level information on covid-19 affected cases. In *time_series_covid_19_recovered.csv*. https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset.

The World Bank. 2021. Birth rate, crude (per 1000 people) - United States. https://data.worldbank.org/indicator/SP.DYN.CBRT.IN?locations=US.