# Variable Factors and the Epidemiology of COVID-19

Using Differential Equation-based Modified SIR Models to Simulate the Impacts of Seasonal Fluctuations, Vaccinations, and Social Distancing on the Transmission of COVID-19

**By John Lee, Sahil Buddharaju, Sahit Kavukuntla, and Eric Zhang**

Mr. Linahres
Calculus 3
Period 5
19 May 2020

# 1 Introduction

COVID-19 is the official name for the disease transmitted by the SARS-CoV-2 coronavirus. The disease affects the respiratory tract in humans, its most common symptoms being fever, dry cough, and fatigue.

Over 80% of patients with COVID-19 recover without needing hospital treatment, and the disease has shown to be fatal primarily only to patients with previous medical conditions or older patients. Despite this, COVID-19 has spread rapidly across the globe, being declared as an epidemic on March 11th, 2020 by the WHO.[1] There is currently no vaccine that has been approved by any national agency, and the spread of COVID only looks to continue in the forseeable future.

Thus, our goal is to analyze the spread patterns of COVID-19, mapping out key distinctions and being able to model the effects of key measures. This paper specifically looks at the epidemiology of COVID-19 in Santa Clara County. We will be using a SIR model based on a set of differential equations in FreeMat in order to reach clear conclusions about the impact of seasonal transmissions, a fully deployed vaccine, and social distancing.

# 2 Derivation of Base Model

The base model is derived from a simple SIR model.[2] In such modeling, the population at hand is split into three groups:

$S$, the susceptible group, refers to all people who have not been in contact with the virus, and thus have the ability to become infectious

$I$, the infectious group, refers to all people at a given time who have the virus and the ability to spread it to others.

$R$, the recovered group, refers to all people who have previously carried the disease and now no longer have the ability to infect others, either due to full recovery or death from the disease.

In a differential equation-based model, we can simulate the expected change in each population per timestep as the output of a differential equation. In order to make the conclusions more intuitive to understand, we use one day for each timestep.

In each timestep, each person in the $S$ group can either not become infected and remain susceptible to the disease or contract the disease and move to the

[1] WHO. "Coronavirus Disease (COVID-19) - Events as They Happen." World Health Organization, World Health Organization, May 2020, www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.

[2] Sasaki, Kai, "COVID-19 dynamics with SIR model", First Cry of Atom, Accessed from https://www.lewuathe.com/covid-19-dynamics-with-sir-model.html

$I$ group. They cannot move directly to the $R$ group, as one must recover fully from or die as a result of the disease to be considered "recovered."

The $I$ group will receive an influx of people who become infected from the $S$ group each timestep. It will also lose a contingency of people who either die or recover from the diease, who will move to the $R$ group.

The change in the $R$ group each timestep is simple: there will simply be an addition of people from the $I$ group.

There are two parameters that control the spread and lethality of the disease. The first is $\beta$, or the transmission rate, which controls how many people in the susceptible population will get infected, and the second is $\gamma$, or the recovery rate, which controls how many people in the infected population will recover or die. At the time of writing this paper, for COVID-19, the rates are 0.93 and 0.0426, respectively.[3] It must be mentioned, however, that these rates are constantly in flux and might not necessarily reflect the current situation, but rather the global average over the course of the epidemic.

It is worth noting that the $R_0$ figure, often cited as the average number of people an infectious person will infect, is derived from the $\beta$ and $\gamma$ parameters, as the number is obtained by $R_0 = \frac{\beta}{\gamma}$.

With the necessary parameters, we can now build our set of ODEs. For a simple SIR model, the differential equations involved are defined as:

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

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

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

## 2.1 Initial Modifications

The simple SIR model that we have derived carries two flaws, which we can attempt to fix.

First, it doesn't take into account the natural fluctuations of the population, namely natural births and deaths. Though this may seem insignificant in the short term, incorporating these factors in a long term simulation can help make the model much more realistic.

[3] Huang, Yubei, et al. "Epidemic Situation and Forecasting of COVID-19 in and Outside China." WHO Bulletin, 16 Mar. 2020, doi:10.2471/blt.20.255158.

Thus, in order to reflect the natural birth and death rate as best as possible, we can add in two more modifiers to $dS$. The natural birth rate of the population, or $\Omega S$, and the natural death rate of the population, or $-\alpha S$. The figures for our model are 0.0000461 and 0.00002142, respectively.[4]

Second, and more importantly, the baseline simple model treats each recovery and death as the same. While this simplifying assumption may be acceptable for looking at simply the infection data, it doesn't allow for an accurate measurement of the true impact of the disease. In order to do so, we must separate actual recoveries from deaths caused by the disease.

To do so, we can simply create a new subset of the population to model: $D$, or the dead group, which consists of everyone who has died from the disease. In order to reflect the influx of people into this population per timestep, we can split up our old recovery rate into two. $\gamma$, or the actual rate of infectious people who have recovered, is now 0.359, and $\mu$, or the rate of people who have died from the disease, is now 0.067. Obviously, there will be no people in the $D$ group who will move to a different group, since by definition every person in the group is dead.

We now have the new baseline set of ODEs to model from:

$$\frac{dS}{dt} = -\beta SI + \Omega S - \alpha S$$

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

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

$$\frac{dD}{dt} = \mu I.$$

### 2.1.1 Preliminary Results

Using the new set of ODEs, we can do a preliminary run modeling the spread of COVID-19 in Santa Clara County. We can use the actual population of Santa Clara County, $1,928,000$ people, and set the initial populations to $1,927,999$, 1, 0, and 0 for $S$, $I$, $R$, and $D$, respectively.

Running the model for 80 days gives a pretty standard plot showing the progression of an epidemic, as can be seen in Figure 1. The epidemic doesn't really start creating noticeable levels of infections until around day 20, and from there it takes until around day 40 for the epidemic to stabilize.

[4] "World Health Organization, Natural Population Growth Rate." World Health Organization, 13 July 2017, origin.searo.who.int.

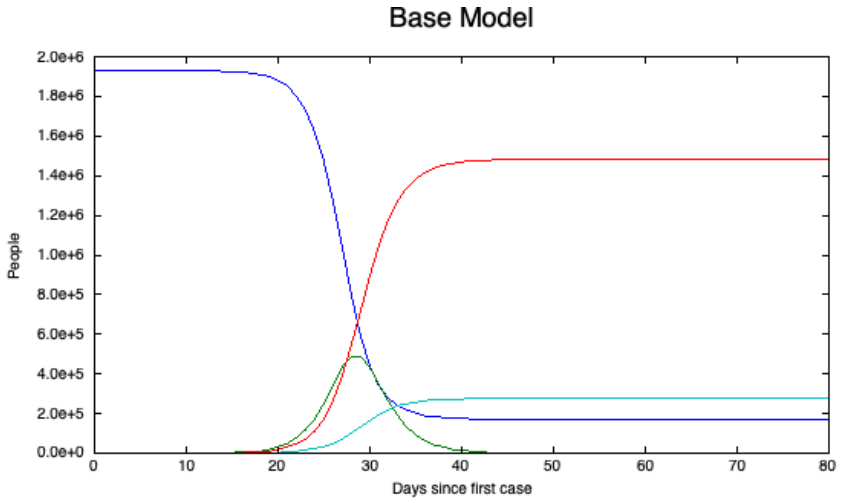



Figure 1: The Base Model

Some key figures show that with the current parameters, the epidemic will have a devastating effect on Santa Clara County. As it stands, the epidemic length, or the number of days from the first infection to zero infectious people, is 59 days. During those 59 days, a total of 276,500 people are being modeled to die. The peak of the infection, at Day 27, brings 89,600 infections in a single day. Meanwhile, 1.48 million people are estimated to catch, then recover from the disease; only 171,000 people in our model will stay completely free of the infection.

Obviously, those numbers are quite bleak. They do, however, reflect the reality and gravity of the situation as it stands right now. Without any countermeasures taken, COVID-19 will affect most of the population greatly.

# 3 Variable Factors

From the baseline model that we set up, we can now find ways to mathematically incorporate potential variable factors. We will attempt to address three factors that have been relevant in academic and policy-making discussions related to COVID-19: seasonal transmissions, the deployment of a vaccine, and social distancing.

## 3.1 Seasonal Transmissions

It has been debated that like many infections coming from influenza viruses, the spread of COVID-19 may be impacted by seasonal fluctuations. It is entirely plausible that the infection may decrease in the summer, where people are gen-

erally more active in outdoor spaces as opposed to indoor.

In order to model seasonal fluctuations in transmission rate, we must make our transmission rate itself a function that depends on the date of the year as opposed to a flat constant. Using the widely accepted weight for the influenza A virus[5], we can make $\beta$ a function,

$$\beta(T) = \beta_0(1 + \epsilon(2\pi T/365),$$

where $\beta_0$ is our stock rate of transmission, $\epsilon$ is the weight of transmission (defined as 0.35 in this case), $T$ is the day in the year with 1 corresponding January 1st, and $\beta(T)$ is the transmission rate for that specific day $i$. Since our function is based on a cos curve with a period of 365 days, it will show a higher rate of transmission at the bookends, or the winter days, and a lower rate of transmission at the center, or the summer days.

It is worth noting that we must differentiate $T$ from $t$, as the former is used in the $\beta$ function to refer to the date of the year while the latter is simply a measure of the number of days after the first infection.
Thus, our modified set of equations is now:

$$\frac{dS}{dt} = -\beta(T)SI + \Omega S - \alpha S$$

$$\frac{dI}{dt} = \beta(T)SI - \gamma I - \mu I$$

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

$$\frac{dD}{dt} = \mu I.$$

### 3.1.1 Results

Using the new set of ODEs incorporating seasonal transmission, we can model a hypothetical scenario in which all other factors are set the same but the disease starts in midsummer.

In modeling this scenario, we set the initial value of $T$ as 182, or July 1st. Meanwhile, we keep all other initial figures and parameters the same; it is only the changed transmission rate that will affect any changes from the initial base model.

Figure 2 clearly shows a significant change from the initial base model in the number of days it takes the epidemic to develop. In the first 45 days or so after

[5]Oraby, T., Vasilyeva, O., Krewski, D., Lutscher, F. (2014). Modeling seasonal behavior changes and disease transmission with application to chronic wasting disease. Journal of Theoretical Biology, 340, 50-59

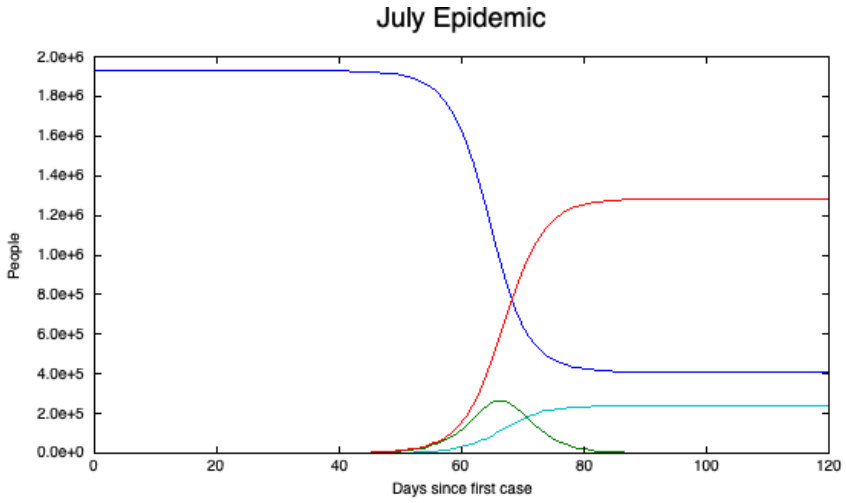



Figure 2: Summer Epidemic

the first infection, or in other words, during summer, there is no notable sign of an outbreak.

It may seem like besides the slight horizontal shift, there is no real effect on the epidemiology of the disease, especially when examining the final death toll: 276,500 for the base model and 239,000 for the summer epidemic.

However, the big change comes in the amplitude of the infectious population curve, colored in Figure 2 as green. The peak number of daily for the summer epidemic in 29,800, a far smaller number in comparison to the base model. While not reflected in our model, a lower number of daily infections would mean more healthcare resources become freed up, and ultimately, a lower death rate would be observed. In addition, the actual number of total people who contract the disease is also far lower; this model shows that at the end of the 119-day epidemic, 409,000 people would remain free of any contact with COVID-19.

## 3.2 Vaccinations

The impact of a potential vaccine is, perhaps, the single most important factor when it comes to stopping diseases in the modern era. It seems crucial to model the potential impact of a vaccine on the spread of COVID-19, especially as promising breakthroughs have posed the question of whether the deployment of a vaccine would mean an immediate return to normalcy.

We will be making the simplifying assumption that whatever vaccine gets approved and released to the general public will be significantly effective in preventing the disease. Thus, the main factors we wish to model are how many

people get vaccinated and when the vaccine gets deployed.

The first factor, how many people get vaccinated, can be modeled by incorporating a constant weight factor onto the vaccine, showing the daily vaccination rate derived from actual figures. The second factor of when the vaccine gets deployed can be solved using a simple if then statement, so that the weight of the vaccine factor stays at 0, essentially rendering the vaccine null, until the day we set as the parameter for the deployment day of the vaccine.

Since a vaccine is a preventative measure, it seems to be logical that it would only be administered to those in the susceptible population[6] Now, under the assumption that all those who receive the vaccine will no longer be susceptible to the disease, we can treat the vaccinated population as if they were recovered, and the new set of equations becomes:

$$\frac{dS}{dt} = -\beta SI + \Omega S - \alpha S - \text{PVD}S$$

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

$$\frac{dR}{dt} = \gamma I + \text{PVD}S$$

$$\frac{dD}{dt} = \mu I,$$

where PVD is the weight representing daily vaccinations.

### 3.2.1 Results

To set up our parameters, we take the average rate of vaccination for schoolchildren, 0.967.[7] We assume the deployment of the vaccine to be day 20, around the time when the outbreak starts in our base model.

Once again, we use the same parameters and initial figures for every other measurement for control. This way, we can be sure that any changes from the base model are the result of the added factor of a vaccine.

Figure 3 shows the clear effects that an early-response vaccine can have on the spread of a disease. The infectious population curve can now be clearly seen as being flattened; the number of peak daily infections, 11,700, is just over an eighth of what it was with the base model. The deaths are lower as well, at 116,500, but this is still probably an overestimation. Even if infections were

[6] Feng, Zhilan, et al. "Modeling the Effects of Vaccination and Treatment on Pandemic Influenza." The AAPS Journal, Springer US, Sept. 2011.

[7] Wadsworth, Jennifer. "As California's Vaccination Rates Fall, Medical Exemptions Rise." San Jose Inside, 12 June 2019, www.sanjoseinside.com/news/as-californias-vaccination-rates-fall-medical-exemptions-rise/.

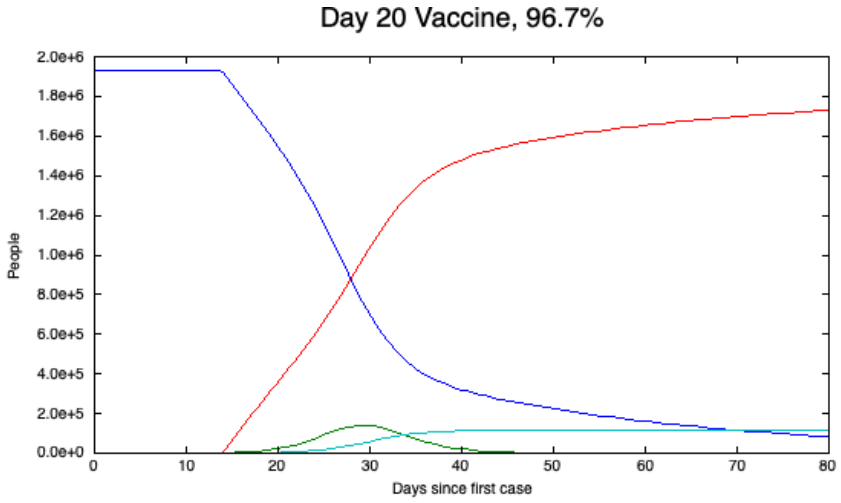



Figure 3: Vaccine

to be high, our model clearly shows that a vaccine can help meet the carrying capacity of hospitals and healthcare services.

Unfortunately, the way our model is set up makes it impossible to determine the full figures for infection; the steady outflow of people from the susceptible to the recovered population means that there isn't an exact figure we're able to compare.

It is worth noting, however, that a vaccine is only truly effective if both factors of an early enough deployment day and a high enough vaccination rate in the population are satisfied.

Figure 4 shows a graph in which the vaccine is developed early enough, but only 80% of the population actually gets vaccinated. Here, it is clear that the flattening effects on the infectious curve are greatly diminished in this case. The numbers reflect that with peak daily infections back up to 35,000 and deaths almost doubled from the earlier vaccination scenario with 176,000.

Figure 5 demonstrates that a late vaccine is even worse. In this case, the vaccine gets deployed on day 30, past the peak of the infections. The model demonstrates that there is almost no change in the number of peak daily infections in this case, at 85,000 compared to 90,000. The final death count also remains pretty close to the original death count, with a day 30 vaccine leading to 235,000 deaths.

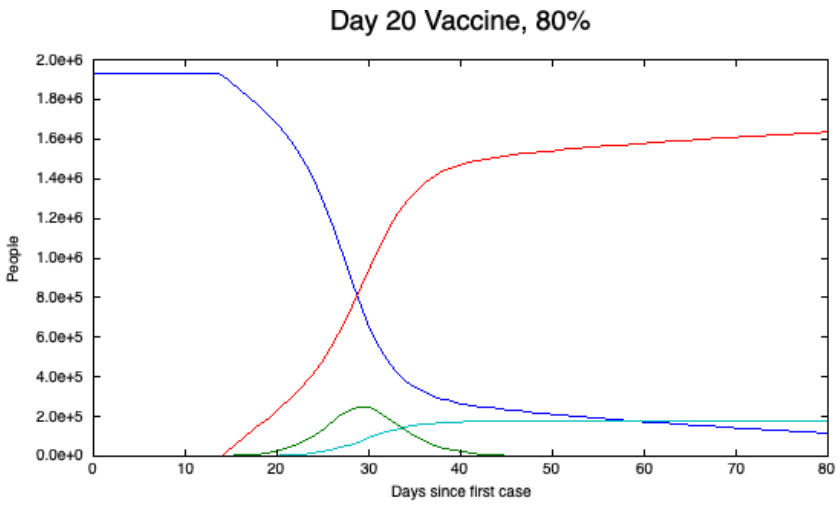
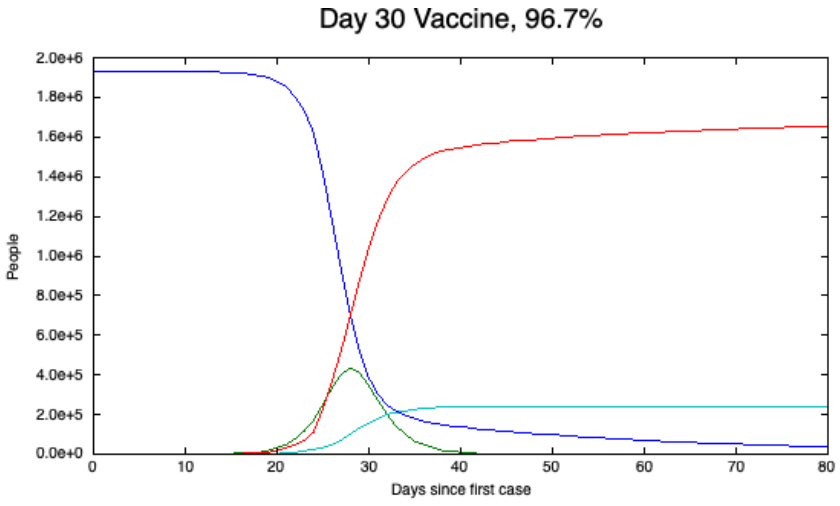



Figure 4: The Anti-Vaxxers Strike Back




,

Figure 5: It's Too Late

## 3.3 Social Distancing

The final factor which we can model is social distancing. For this factor, we don't necessarily have to change the model itself. Instead, we can change the parameters to adjust the $R_0$ value, thus simulating the effects of social distancing. In this case, we make a simplifying assumption that all changes to the $R_0$ factor come from the result of human activity.

### 3.3.1 Results

For our parameters, we use California's latest estimate for its $R_0$ figure, 0.87.[8] Since $R_0$ is derived from, and thus affected by, both $\beta$ and $\gamma$, we assume that the disease itself hasn't changed and that the change in $R_0$ comes mainly from the change in transmission rate.

This time, we cannot use the same initial figures as the other models due to the fact that the model will refuse to run with an initial infected population of 1. Thus, we must start with an initial population of 100. Even with that, the results are extreme.

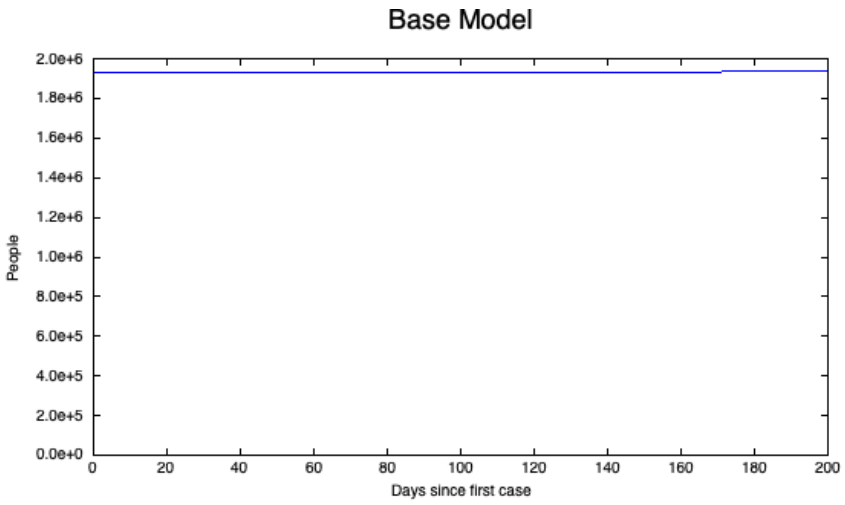



Figure 6: Social Distancing

Figure 6 shows a model in which Santa Clara County starts social distancing from the first infection, an admittedly impossible scenario. However, it is still notable that with an $R_0$ of 0.87, the epidemic is nowhere to be seen. The peak

[8] Krieger, Lisa M. "See California's Shrinking COVID-19 Outbreak, Thanks to Instagram Founders." The Mercury News, The Mercury News, 29 Apr. 2020, www.mercurynews.com/2020/04/28/watch-californias-shrinking-covid-19-outbreak-thanks-to-instagram-founders/.

daily infections maxes out at just 87, and by the end 99.99% of the population never comes into contact with COVID-19.

# 4 Discussion

The biggest conclusion that can be reached from the model is that the only thing outside factors are able to control significantly is the transmission of the disease itself. This seems to be reflected in real-world discussions of epidemics; while it's extremely hard to influence a disease's lethality directly, it is possible to impact the way that it spreads within a demographic.

Our model is a stark improvement from the simple SIR model that exists commonly across academia. The addition of a dead subgroup allows us to track total deaths, a pretty important figure in tracking an epidemic, very easily. While we have chosen to model three specific factors, it is also possible using the base modified model that we have developed to further model other factors.

There are two main flaws with the model that we discovered in the process of deriving results. First is the inconsistency of the data for the parameters. In every single one of our models, it feels as if the numbers of those who aren't affected by COVID-19, or the final susceptible numbers, are much too low. As of now, over half of all confirmed COVID-19 cases are active, which makes it extremely difficult to believe that transmission rates, recovery rates, and death rates are stabilized. Related to this is the fact that data isn't widely available for individual counties or even countries. Thus, we have to model a relatively specific area using very general data.

The other and more significant flaw is that of the fact that our death rates remain constant within the process of the entire model. The most widely discussed issue of treatment has to do with flattening the infectious curve and lowering the amount of daily new cases in order to ease the burden on medical workers. With a lower amount of people needing immediate care, it's obvious that the death rate should decrease; however, our model using a constant death rate means that the true effects of flattening the curve cannot be seen. To resolve this issue, we would need to model the death rate as a function dependent on the previous day's daily infections.

There are other, less significant issues specifically with our vaccination model that are more easily fixable. The ambiguity of the susceptible population in our vaccination model can be resolved by creating a fifth population subgroup, vaccinated in order to accurately track the number of people that have not come into contact with the disease. Meanwhile, the vaccination constant can be made more precise by locating data for the pattern of vaccinations after the deployment day and incorporating the figures to make the weight a function.

However, despite all of these flaws, our model should ultimately be seen as a successful one. It shows, very clearly, the effects of certain measures on the spread of COVID-19, such as how certain factors can actually help flatten the infectious curve. As previously mentioned, even without the results, the improvement on the current SIR model is an achievement in itself.

The next step for this model would be to incorporate it with a similar modified SIR model, but one that is agent-based. The question of spacial modeling is a difficult one with differential equation-based modeling, since the factor we are looking for is the change in each population. However, combining our results with a visual diagram for an agent-based model could improve the accuracy of our results in simulating the physical environment we are trying to model.

# 5 Freemat Code

```
1
2  totalpeople = 1928000
3  initial.S = 1927999;    % set the initial value of 'S'
4  initial.I = 1;  % set the initial value of 'I'
5  initial.R = 0;  % set the initial value of 'R'
6  initial.D = 0
7
8  param.gamma = 0.359;    % set the parameter 'r' of the
       model
9  param.beta = (2.2 * param.gamma)/(initial.S + initial.I +
        initial.R) % set the parameter 'beta' of the model
10 param.mu = 0.067
11
12 %days from January 1st, used for seasonal transmission
       for param.beta
13 d = 1;
14
15
16 end_time = 80;
17 time_interval = 1;
18
19 %percent vaccinated by end_time
20 pve = 0.967
21
22 %percent vaccinated per day
23 pvd = pve / 30
24
25 %day when vaccine is released
26 dayofvaccine = 25
27
```

```
28
29
30 N = initial.S + initial.R + initial.I + initial.D;
31 R_0 = param.beta * N / param.gamma;
32
33 x = [initial.S; initial.I; initial.R; initial.D];
34
35
36 initial_values = [];
37 variable_names = fieldnames(initial);
38 for i=1:length(variable_names)
39     initial_values = [initial_values; initial.(
          variable_names{i})];
40 end
41
42 % integrate the ODE system
43 %[t, y] = ode45(@(t, x) ode_system(0, x, param),[0
       end_time],initial_values,[]);
44
45 % prepare legend texts
46 legend_texts = cell(length(variable_names), 1);
47
48
49 pS = 0;
50 pI = 0;
51 pR = 0;
52 pD = 0;
53 oldS = initial.S;
54 oldI = initial.I;
55 oldR = initial.R;
56 oldD = initial.D;
57
58
59 seasonalbeta = 0;
60
61
62 smatrix = [];
63 imatrix = [];
64 rmatrix = [];
65 dmatrix = [];
66 ymatrix = [];
67
68
69 for i=0:end_time
70     %y = ode_system(i, x, param)
71
```

```
72      seasonalbeta = param.beta * (1 + (0.35 * cos((2 * pi
            * d)/365)))
73      param.beta
74
75
76      if (i >= dayofvaccine)
77          dS = -seasonalbeta * x(1) * x(2) - (0.00002142 *
                x(1)) + (0.0000461 * x(1)) - (pvd * x(1));
78          dI = +seasonalbeta * x(1) * x(2) - param.gamma *
                x(2) - (x(2) * param.mu);
79          dR = (param.gamma * x(2)) + (pvd * x(1));
80          dD = (param.mu* x(2))
81      else
82          dS = -seasonalbeta * x(1) * x(2) - (0.00002142 *
                x(1)) + (0.0000461 * x(1));
83          dI = +seasonalbeta * x(1) * x(2) - param.gamma *
                x(2) - (x(2) * param.mu);
84          dR = (param.gamma * x(2));
85          dD = param.mu*(x(2))
86      end
87
88      pS = oldS + dS;
89      pI = oldI + dI;
90      pR = oldR + dR;
91      pD = oldD + dD
92
93
94      %plot(i, pS)
95      %plot(i, pI)
96      %plot(i, pR)
97      pS
98      pI
99      pR
100     pD
101     smatrix = horzcat(smatrix, pS)
102     imatrix = horzcat(imatrix, pI)
103     rmatrix = horzcat(rmatrix, pR)
104     dmatrix = horzcat(dmatrix, pD)
105     ymatrix = horzcat(ymatrix, i)
106
107
108     oldS = pS;
109     oldI = pI;
110     oldR = pR;
111     oldD = pD;
112
```

```
113         x(1) = oldS;
114         x(2) = oldI;
115         x(3) = oldR;
116         x(4) = oldD;
117         d = d + 1;
118
119     end
120
121
122     plot(ymatrix, smatrix, ymatrix, imatrix, ymatrix, rmatrix
        , ymatrix, dmatrix);
123     clear xlabel;
124     clear ylabel;
125     clear title;
126     title('Day 20 Vaccine, 80%', 'fontsize', 18);
127     xlabel('Days since first case');
128     ylabel('People');
```