

Reza Vafaei

vafaei.r@northeastern.edu

1. Write down a SEIR model without births/deaths. Use values for the rates that correspond to the spread of COVID-19 in New Jersey, and use the population size of New Jersey. Note that you may find a range of values as you search online. That is ok, but justify why you are choosing the values you do choose, and be sure to cite the source.

A SEIR model without births/deaths is as the following:

$$\frac{dS}{dt} = -\lambda(I)S$$

$$\frac{dE}{dt} = \lambda(I)S - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

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

$$S + E + I + R = N$$

$$\text{And } \lambda(I) = \beta \frac{I}{N}$$

where S, E, I, and R are the number of susceptible, exposed, infected and recovered people respectively. λ , β , σ and γ are the force of infection, transmission rate, the rate at which individuals transition from the exposed class to the infectious class and the recovery rate respectively. And N is population size. σ and γ can be obtained from the following formula:

$$\sigma = \frac{1}{\text{latent/incubation period}}$$

$$\gamma = \frac{1}{\text{infectious period}}$$

Authors of [1] used the following values as the infectious period and latent period for coronavirus. I am using the same amounts because this paper has been cited frequently in a short period of time (it cited 68 times until now).

$$\text{incubation period} = 5.2 \text{ days}$$

$$\text{infectious period} = 18 \text{ days}$$

To find a good estimation for β in NJ, we first tried to find a good estimation for R_0 , reproductive number, in NJ. To estimate R_0 in a specific area, several ways were explained in [2]. I took advantage of the first method explained in [2], Estimating R_0 in an invasion.

In this method, a SIR model was assumed. In a SIR model, the differential equation for the number of infected individuals with considering births/deaths is as

$$\frac{dI}{dt} = \lambda S - \gamma I - \mu I$$

$$\lambda = \beta \frac{I}{N}$$

where μ is the natural mortality rate. In early stages, number of susceptible individuals is almost equal to the population size, $S \approx N$. Therefore,

$$\frac{dI}{dt} = \beta \frac{I}{N} N - (\gamma + \mu)I$$

After some simplifications,

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

We also know [3] the formula for R_0 in this model is

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

Therefore, we can rewrite the previous equation as

$$\frac{dI}{dt} = (R_0 - 1)(\gamma + \mu)I$$

So, the number of infectious individuals $I(t)$ is can be obtained as

$$I(t) \approx I_0 e^{(R_0 - 1)(\gamma + \mu)t}$$

By taking logarithm from both sides, we have

$$\log I(t) \approx \log I_0 + (R_0 - 1)(\gamma + \mu)t \quad (*)$$

We can use the data of the number of infected individuals in the early stages for the NJ to estimate R_0 . If we can fit a linear line to this data, the slope of the line can be used to find a good estimation of R_0 . A linear line has slope and y intercept as the following:

$$y = slope \times x + y_intercept$$

Let's assume we used the data and by using simple linear regression, we now obtained a line like above. By comparing the last two equations, we can write

$$(R_0 - 1)(\gamma + \mu) = slope$$

$$\text{So, } R_0 = \frac{slope}{(\gamma + \mu)} + 1$$

The only point we need to consider is we need to use logarithm version of our data instead of the raw data since the (*) equation is using log.

In addition, since in the first part of this project, we should ignore the births/deaths rate, this rate should be also ignored in obtaining the (*) equation. By ignoring the birth/death rate and repeating all of the steps, the star equation would be in the form of

$$\log I(t) \approx \log I_0 + (R_0 - 1) \times \gamma \times t \quad (**)$$

Therefore, R_0 this time is

$$R_0 = \frac{\text{slope}}{\gamma} + 1$$

The New York Times is collecting a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time. This dataset is available on github [4]. We used this data to find the linear line and also to initialize the system on differential equations in the next steps. Since it was asked to use March 15 as the starting point of our simulation, we used the data collected by New York Times for before March 15 to estimate R_0 in NJ. This data is extracted from this dataset and listed in the following table.

date	state	cases	deaths
3/4/2020	New Jersey	1	0
3/5/2020	New Jersey	2	0
3/6/2020	New Jersey	4	0
3/7/2020	New Jersey	4	0
3/8/2020	New Jersey	6	0
3/9/2020	New Jersey	11	0
3/10/2020	New Jersey	15	1
3/11/2020	New Jersey	23	1
3/12/2020	New Jersey	29	1
3/13/2020	New Jersey	50	1
3/14/2020	New Jersey	75	2
3/15/2020	New Jersey	98	2

The MATLAB code used to estimate R_0 and then β for NJ without considering birth and death rate is as the following

```
function [beta,R0] = beta_without(gamma)
% beta function without demography
t = 1:12; % From 03/04/2020 to 03/15/2020
t = t';
T = [ones(length(t),1) t]; % add a column of ones
I = [1; 2; 4; 4; 6; 11; 15; 23; 29; 50; 75; 98]; % Infected cases in NJ
from 03/04/2020 to 03/15/2020 based on [3]
% linear regression
b = T\log(I);
slope = b(2);
% estimating R0 based on [5]
R0 = (slope/gamma)+1;
% Beta for NJ
beta = R0 * gamma;
```

The β that was obtained by this method and without considering birth and death rate for New Jersey is 0.4576.

- (a) Using initial values based on numbers of S, E, I, and R in New Jersey on March 15, solve the model numerically using a Runge-Kutta solver (4th order is good). Note that both Matlab and Python have Runge-Kutta routines you can use. You should run your code for long enough to see the entire epidemic. I would like to see a plot showing number of infectious individuals versus time, as well as a plot showing S, E, I, and R together on the same plot as a function of time.

The next step is initializing S, E, I and R to solve the system of equations of a SEIR model. As mentioned before, we used the dataset collected by the New York Times to initialize I. Since the beginning time assumed to be March 15, the number of infected individuals reported for this date in the above table used as the initialization value for I ($I(t = 0) = I_0 = 98$). The initial exposed population was assumed to be $E(t = 0) = 20 \times I(t = 0)$ in accordance with [1, 5 and 6], and the number of recovered individuals was set to zero value $R(t = 0) \approx 0$. Number of susceptible individuals can be obtained simply as $S = N - E - I - R$. By looking several online sources on the internet, we assumed 9×10^6 as the population size of New Jersey.

The MATLAB code to obtain the graphs asked in part (a) is as the following:

```
% Parameters
latent_period = 5.2;           % latent period /Obtained from [2]
infectious_period = 18;       % infectious period /Obtained from [2]

delta = 1 / latent_period;
gamma = 1 / infectious_period;

[beta,~] = beta_without(gamma); % beta function without demography

% Initialization
N = 9e6; % Total population of NJ state
I_0 = 98; % infected/ Obtained form [3]
E_0 = 20 * I_0; % exposed / in accordance with [2]
R_0 = 0; % recovered / immune
S_0 = N - E_0 - I_0 - R_0; % susceptible
y0 = [S_0, E_0, I_0, R_0];

% days of simulation
days = 365; % one year
h = 1; % step (1 day)
tspan = 1:h:days;

[t,y] = ode45(@(t,y) without(t,y,N,beta,delta,gamma), tspan, y0);
% y = [S, E, I, R];

figure;
plot(t,y(:,1),t,y(:,2),t,y(:,3),t,y(:,4)); grid on
axis([0 days 0 10e6]);
title('Solution of SEIR model without demography');
xlabel('Time (days)');
ylabel('Individuals');
legend('S','E','I','R');
```

```

figure;
plot(t,y(:,3)); grid on
axis([0 days 0 5e6]);
title('Number of infectious individuals versus time');
xlabel('Time (days)');
ylabel('Individuals');

function dydt = without(t,y,N,beta,delta,gamma)

dydt = zeros(4,1);

dydt(1) = (-beta/N) * y(1) * y(3);
dydt(2) = ((beta/N) * y(1) * y(3)) - (delta * y(2));
dydt(3) = (delta * y(2)) - (gamma * y(3));
dydt(4) = (gamma * y(3));

% dS = (-beta/N) * S * I;
% dE = ((beta/N) * S * I) - (delta * E);
% dI = (delta * E) - (gamma * I);
% dR = (gamma * I);
% y = [S, E, I, R];

```

From the code, we can see that we used one-year period of time to graph the solution; in addition, ode45 function of MATLAB was employed to solve SEIR model's system of equations. ode45 is based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair. The method is a member of the Runge-Kutta family of ODE solvers. More specifically, it uses six function evaluations to calculate fourth- and fifth-order accurate solutions. The difference between these solutions is then taken to be the error of the (fourth-order) solution. The plot showing number of infectious individuals versus time is shown in the figure 1.

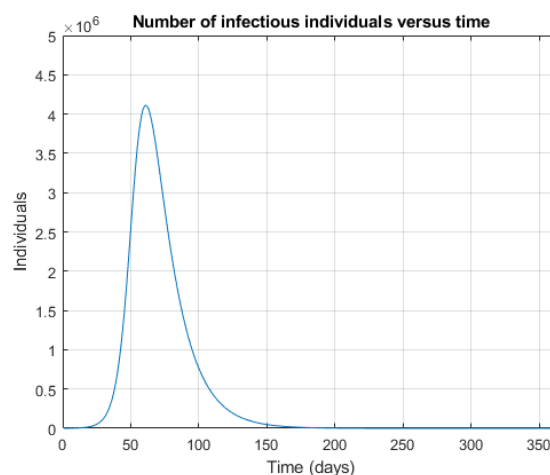


Figure 1. Number of infectious individuals versus time

In addition, the plot showing S, E, I, and R together on the same plot as a function of time is depicted in figure 2.

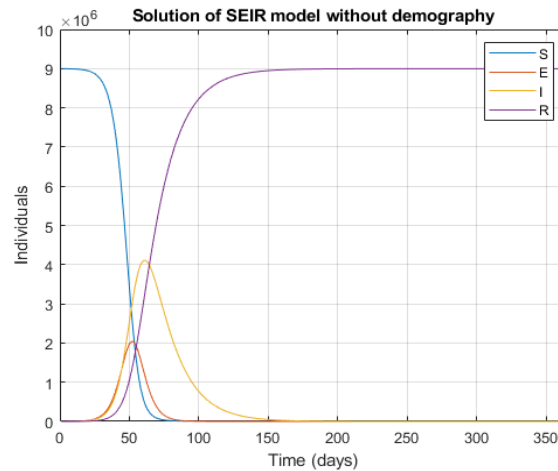


Figure 2. S, E, I, and R together as a function of time

(b) Now assume that NJ's control policies were good enough to reduce the transmission rate by 80%. Re-solve your model and plot infectious individuals versus time for both the case of part (a) and the reduced transmission case of part (b).

MATLAB code to obtain the solution of this part is

```
%% Reducing the transmission rate by 80%
days = 2 * days; % 2-year period
tspan = 1:h:days;
sol = zeros (length(tspan),2); % space to save number of infectious
individuals
[~,y] = ode45(@(t,y) without(t,y,N,beta,delta,gamma), tspan, y0);
sol(:,1) = y(:,3);

Reduced_beta = 0.2 * beta;
[t,y] = ode45(@(t,y) without(t,y,N,Reduced_beta,delta,gamma), tspan, y0);
sol(:,2) = y(:,3);

figure;
plot(t,sol(:,1),t,sol(:,2)); grid on
axis([0 days 0 5e6]);
title('Number of infectious individuals versus time');
xlabel('Time (days)');
ylabel('Individuals');
legend('Original solution','Reducing beta by 80%');
```

Figure 3 shows the requested graph. This graph is plotted for a 2-year period.

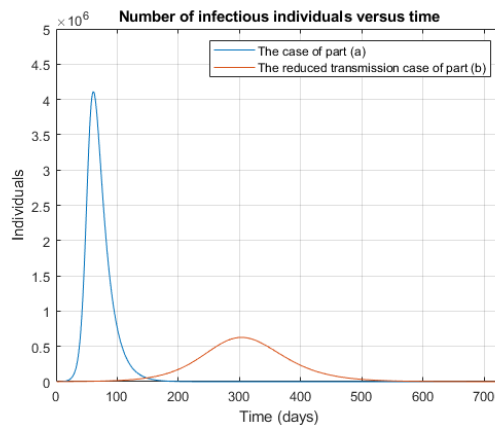


Figure 3. Infectious individuals versus time for both the case of part (a) and the reduced transmission case of part (b).

2. Add in births and deaths to the SEIR model and re-solve, showing 2 years worth of epidemic. Use the same value for both births and deaths, and choose it to be the average of the NJ birth rate and NJ death rate. Plot infectious individuals as a function of time.

After spending some time on the internet, I have decided to use birth & death rate collected from New Jersey State Health Assessment Data [7] for my simulation. Based on [7], birth and death rate in NJ in 2018 are 11.4 and 8.5 per 1000 persons respectively. As requested, we used the same value for both births and deaths, and chose it to be the average of the NJ birth rate and NJ death rate. Therefore, μ in our simulation is:

$$\mu = \frac{11.4 + 8.5}{2 \times 1000} = 0.00995$$

As we discussed before, β should recalculate this time by considering birth/death rate. The star equation must be used this time instead of double star equation. The MATLAB function for calculating beta by considering birth/death rate of New Jersey is

```
function [bet,R0] = beta_with(gamma,mio)
% beta function with demography
t = 1:12; % From 03/04/2020 to 03/15/2020
t = t';
T = [ones(length(t),1) t]; % add a column of ones
I = [1; 2; 4; 4; 6; 11; 15; 23; 29; 50; 75; 98]; % Infected cases in NJ
from 03/04/2020 to 03/15/2020 based on [3]
% linear regression
b = T\log(I);
slope = b(2);
% estimating R0 based on [5]
R0 = (slope/(gamma+mio))+1;
% Beta for NJ
bet = R0 * (gamma+mio);
```

Beta value this time was calculated as 0.4675.

To obtain the requested graph for this part, I used the following MATLAB code

```

%%
                %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
                % SEIR model with births/deaths%
                %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% dS/dt = mio*N - lambda(I)*S - mio*S
% dE/dt = lambda(I)*S - (delta + mio)*E
% dI/dt = delta*E - (gamma + mio)*I
% dR/dt = gamma*I - mio*R

% mio-----
% Birth rate of NJ in 2018 = 11.4 per 1,000
% Death rate of NJ in 2018 = 8.5 per 1,000
% Birth&Death rate collected from New Jersey State Health Assessment Data
[1]
mio = (11.4+8.5)/(2*1000); % per 1 person

[beta,~] = beta_with(gamma,mio); % beta function with demography
% days = 2 * days; % 2-year period
% tspan = 1:h:days;
[t_w,y_w] = ode45(@(t,y) with(t,y,N,beta,delta,gamma,mio), tspan, y0);

figure;
plot(t_w,y_w(:,1),t_w,y_w(:,2),t_w,y_w(:,3),t_w,y_w(:,4)); grid on
axis([0 days 0 10e6]);
title('Solution of SEIR model with demography');
xlabel('Time (days)');
ylabel('Individuals');
legend('S','E','I','R');

figure;
plot(t_w,y_w(:,3)); grid on
axis([0 days 0 4e6]);
title('Number of infectious individuals versus time with demography');
xlabel('Time (days)');
ylabel('Individuals')

function dydt = with(t,y,N,beta,delta,gamma,mio)

dydt = zeros(4,1);

dydt(1) = (mio * N) - ((beta/N) * y(1) * y(3)) - (mio * y(1));
dydt(2) = ((beta/N) * y(1) * y(3)) - ((delta + mio) * y(2));
dydt(3) = (delta * y(2)) - ((gamma + mio) * y(3));
dydt(4) = (gamma * y(3)) - (mio * y(4));

% y = [S, E, I, R];

```

Figure 4 & figure 5 show the output graphs for this piece of code.

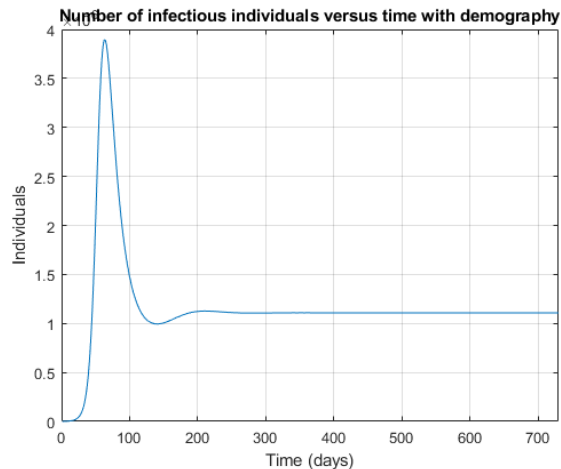


Figure 4. Number of infectious individuals versus time with demography

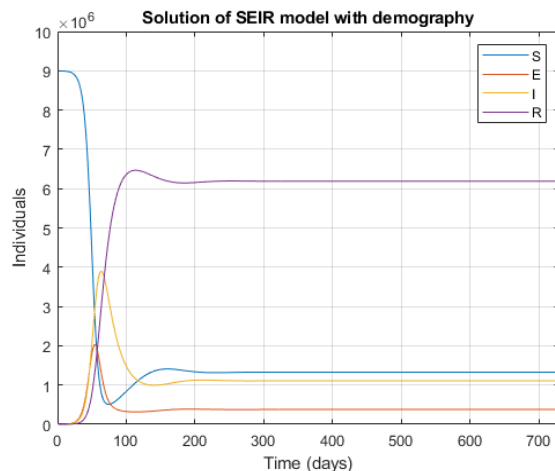


Figure 5. S, E, I, and R together as a function of time with demography

3. Two models that many governments are using are the CHIME model from the Univ. of Pennsylvania, and the IHME model from the Univ. of Washington. Investigate both models and describe what they are and how they work. What are the pros and cons to both models? Compare your COVID-19 predictions from question 1 to the predictions of CHIME and IHME models, and discuss any similarities or differences there may be. Take screen shots of these model outputs to include in your results.

COVID-19 Hospital Impact Model for Epidemics (CHIME)

CHIME is a modified version of SIR model. This model was established for a short term forecasting. CHIME is only practical and applicable for the time before a region's peak infections; in addition, it considers only a single significant distancing policy. This online tool was built by predictive healthcare at Penn Medicine to assist hospitals and public health official with hospital capacity planning. This application can be used by anyone in the world and is available online.

CHIME can give us an estimation of total daily and running totals of inpatient hospitalizations, ICU admissions, and patients requiring ventilation. It specifically allows users to forecast the

expected number of in-hospital patients and daily admissions so that they can estimate the impact in their location and when they will need to have mitigation strategies in place. This app was released on 03/15/2020 [10].

The advantages of this model are:

- Providing a good estimation of the total number of inpatient hospitalizations, ICU admissions, and mechanical ventilation.
- Using a simple and understandable model, SIR.
- Being able to modify assumptions and parameters to adapt it for use by any hospital system to reflect local contexts.
- Purposing default values for the parameters.
- Incorporates hospitalizations factors.

The disadvantages of this model are:

- It is limited to short term forecasting (Although Penn Medicine is actively working on a new version of CHIME with the name of BayesCHIME, the current version (v1.1.5) is still limited to short term forecasting).
- It can take into account only one single significant social distancing policy.
- Some of the parameters such as “Hospital Market Share (%)” need to collect data (data collection is needed).
- The number of days to project in this model is limited to 30 days.
- There is no clear instruction to adopt the Severity Parameters (They referred us to the paper of Verity et al [9]).
- Being applicable for the time before a region’s peak infections.

To be able to get output from this app, the following parameters were set on the browser

Row	Description	Value	Reason for choosing this value
Hospital Parameters			
1	Regional population	9,000,000	Population size of NJ
2	Hospital market share (%)	15	Left as default
3	Currently hospitalized COVID-19 patients	98	Number of infectious individuals on 03/15/2020 [4]
Spread and Contact Parameters			
4	I know the date of the first hospitalized case	03/04/2020	Based on [4]
5	Date of social distancing measures effect	03/15/2020	To be comparable with the predications of question 1
6	Social distancing (% reduction in social contact going forward)	0 & 80	Since we want to compare these outputs with COVID-19 predictions from question 1
Severity Parameters			
7	Hospitalization %(total infections)	2.5	Left as default
8	ICU %(total infections)	0.75	Left as default
9	Ventilated %(total infections)	0.5	Left as default
10	Infectious days	18	Our γ , Based on [1]
11	Average hospital length of stay (in days)	7	Left as default
12	Average days in ICU	9	Left as default
13	Average days on ventilator	10	Left as default

Display Parameters			
14	Number of days to project	30	The max number of days is set to 30 for v1.1.5
15	Current date	03/15/2020	As requested in this project

When social distancing parameter set to be 0, the output curve was as the figure 6.



Figure 6. Projected number of daily COVID-19 admissions obtained by CHIME model

By comparing this figure with figure 1 which was the number of infectious individuals for SEIR model without demography, we can see that the number of hospitalized admissions (blue line) takes about 15 days to meet its peak in this picture; however, this period for figure 1 is around 60 days. Existing exposed compartment in our simulation can explain this delay in arriving to the peak point. Furthermore, the amount of peak here is about 4,000 person; in contrast, figure 1 has a peak with amount of 4,000,000. Since in this figure (Figure 6) only admitted cases to hospitals are plotted, the value of peak in this graph is a portion of the value of peak in figure 1, so the differences are reasonable.

If we reduce the transmission rate by 80% the figure 6 will change to the form figure 7.



Figure 7. Projected number of daily COVID-19 admissions obtained by CHIME model with reduced transmission rate

The model does not allow me to import the same date for “Date of social distancing measures effect” and “Date of first hospitalized case”. Therefore, I imported 03/15/2020 as “Date of social distancing measures effect” which is different from 03/04/2020, “Date of first hospitalized case”. Thus, in the graph demonstrated in figure 7, there is a sharp upward trend at the beginning of the graph for the number of hospitalized admissions (blue line). However, this upward trend changes to a moderate one at date 03/15/2020. Since we reduced the transmission rate after 03/15/2020, the sharpness of the blue line is different before and after this date.

In addition, since the max number of days is set to 30 in v1.1.5 of CHIME model, the graph could not continue enough to see the new peak. Also, as we reduced the transmission rate, this graph increases very slowly. We knew same thing happened when we reduce the transmission rate in part b of question 1 (see figure 3).

IHME's COVID-19 projections

It seems that two main versions of IHME have been available online. The first one released on May 4, 2020 and the later one released on March 26, 2020. The following descriptions and information are about the later version.

The model released on March 26, 2020 is not a disease transmission model. It is a hybrid model that combines both a statistical modeling approach (Data driven modeling approach) and a disease transmission approach. This can help to take advantage of the strengths of both types of models as well as to scale the results of the disease transmission model to the results of the statistical model. More specifically, this model was developed based on a mixed effects non-linear regression framework to estimate the trajectory of the cumulative and daily death rate as a function of the implementation of social distancing measures, supported by additional evidence from mobile phone data [8].

IHME model forecast not only the number of infected individuals but also the number of deaths in states. The model takes into account mobility, testing, temperature, and population density. Similar to CHIME model, it is also designed to help hospital administrators and government officials figure out when demand on health system resources will be greatest [11].

The advantages of this model are:

- It considers data of deaths, hospitalizations, and testing.
- It takes into account factors in mobility (decreased mobility reduces COVID-19 transmission and deaths).
- By combining a disease transmission approach by a statistical modeling approach, this model relies on not only assumptions about how the disease will spread but also real-world data.
- Takes into account variation in transmission across locations and over time.

And disadvantages of this model are:

- It relies on data, so data collection and data processing will be needed.
- Data collection might be a time-consuming process and even expensive.
- As the number of deaths and testing equipment/kits would be needed, there might be some security embargos about sharing data at the early stages.

- Since a curve-fitting tool should be developed to fit a nonlinear mixed effects model to the available data, this process might impose a huge computational load to our hardware in the training step.
- A fast and probably expensive hardware might be needed to carry the computational load.

The output of this model for New Jersey for the same period as CHIME is demonstrated in figure 8.

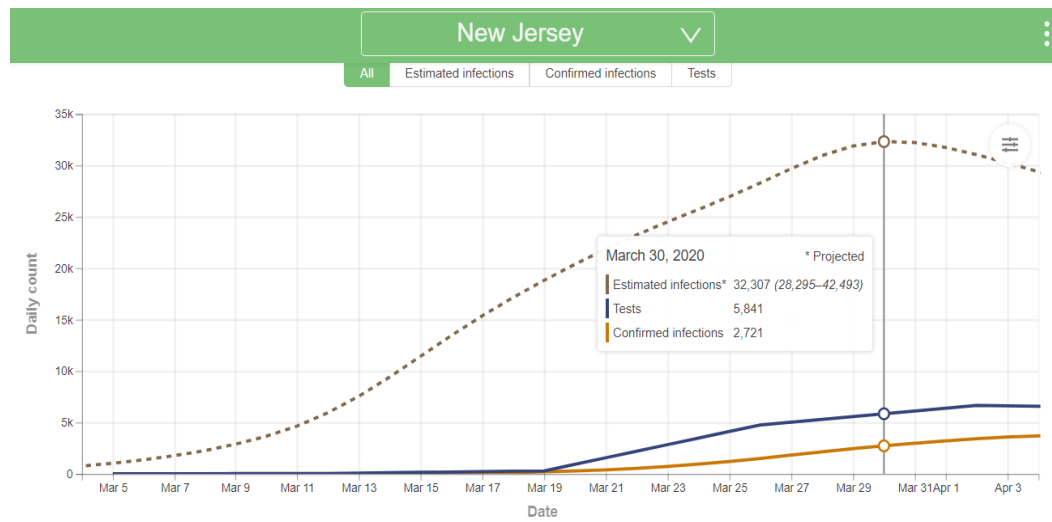


Figure 8. Output of IHME model for NJ

From this figure, we can see that the estimated infections arrives to its peak after around 15 days. This period is same as the other model (CHIME model).

4. For this last problem, think of yourself as an expert mathematical epidemiologist who can use modelling to make predictions and provide guidance to the Governor of NJ. What problem do you think is important to model now as the state tries to emerge from a near complete shut-down. Provide a brief outline of the model you would have to develop to obtain the answer.

Since New Jersey passed the first peak of COVID-19 deaths, attention is now on how best to prevent and manage a resurgence of the disease while safely enabling people to get back to work and school. In locations where the pandemic has yet to be curbed, and sufficient testing and contact tracing is not in place, easing social distancing could prolong the pandemic and lead to a greater number of deaths. It also could lead to a higher risk of resurgence compared to locations that have maintained social distancing.

Therefore, in the model, I am suggesting to develop the following factors will be incorporated:

- The impact of easing social distancing.
- Similar to both discussed models, hospitalizations factors (hospitalization, ICU beds, and mechanical ventilation).
- Data on deaths and testing
- Mobility
- Considering variation in transmission across locations and over time

- Considering several social distancing policies.
- If the data were simply available for New Jersey, I would prefer to use a hybrid model (Traditional model + Data driven modeling approach (Machine learning)), so my model will rely on real data.

References:

- [1] Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., ... & Wang, Y. (2020). Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell discovery*, 6(1), 1-8.
- [2] <https://kingaa.github.io/clim-dis/parest/parest.html>
- [3] Blackwood, J. C., & Childs, L. M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5(1), 195-221.
- [4] <https://github.com/nytimes/covid-19-data>
- [5] Dandekar, R., & Barbastathis, G. (2020). Neural Network aided quarantine control model estimation of global Covid-19 spread. *arXiv preprint arXiv:2004.02752*.
- [6] Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A., & Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *MedRxiv*.
- [7] <https://www-doh.state.nj.us/doh-shad/>
- [8] COVID, I., & Murray, C. J. (2020). Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. *medRxiv*.
- [9] Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., ... & Dighe, A. (2020). Estimates of the severity of COVID-19 disease. *MedRxiv*.
- [10] <https://code-for-philly.gitbook.io/chime/>
- [11] <http://www.healthdata.org/covid/faqs#differences%20in%20modeling>