Sensitivity analysis of key variables of the SEIR model

Lars Bliss-Miller
Western Washington University
Bellingham, Washington
blissml@wwu.edu

Natthapong Choeypant
Western Washington University
Bellingham, Washington
choeypn@wwu.edu

Tyson Letham
Western Washington University
Bellingham, Washington
lethamt@wwu.edu

Noah Sinnes
Western Washington University
Bellingham, Washington
sinnesn@wwu.edu

ABSTRACT

This report overviews the results of variable-sensitivity analysis performed on the SEIR disease model. This SEIR model was implemented in the context of COVID-19, which is a pandemic caused by the coronavirus SARS-CoV-2. R_0 was found to be the most important factor in altering the behavior of the model, with the incubation and infectious periods following. This report suggests that public policy aimed at reducing R0 would be effective in reducing the impact of COVID-19.

KEYWORDS

R_0, I_0, CFR, P-severe, Hospital lag, TTD

ACM Reference Format:
Lars Bliss-Miller, Natthapong Choeypant, Tyson Letham, and Noah Sinnes. 2018. Sensitivity analysis of key variables of the SEIR model. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Due to the COVID-19 crisis, many people, but especially governments, are interested in how the virus will affect the population. To predict how that will happen, we turn to disease models. This paper focuses on the SEIR model of disease. The SEIR model divides populations into four categories, called Susceptible, Exposed, Infectious, and Recovered. The susceptible population are those people who can be infected but have not yet been. Exposed is the population where the people have the virus but are not able to transmit it to others. The infected population are those with the virus and capable of spreading it to the susceptible population, and the recovered are those that had the disease but no longer have it.

The government and medical industries only have specific factors that they can influence to control the outcome of the virus outbreak. Controlling the outcome is important to minimize the number of deaths due to the pandemic. Furthermore, governments and hospitals would be interested in knowing which factors they can control, and how much of an impact do those factors have on the outcome. The most important outcomes that a government would be interested in is the effect on the number of people who have the disease and can spread it, the number of people in the hospital, and the total number of dead. The number of people who have the disease and can spread it is an important metric as if there are more people who can spread the disease, then the impact on hospitals and the number of dead will increase. In order to minimize the number of dead, reducing the number of people with the virus is important. The most important metric however is the number of people in the hospital because having this number be too large can overwhelm the hospital and increase the death rate.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06.
https://doi.org/10.1145/1122445.1122456

2 MOTIVATION

Amid the COVID-19 epidemic, any accurate information on what factors have an impact on the spread of the disease can help in understanding more about the virus. Fortunately, we can study the pandemic effect with software that simulates the SEIR model. The results can help to provide estimations on sensitivity of the model to different variables. Identifying which factor(s) are crucial to handling the COVID-19 epidemic helps in forming policies to reduce the spread of the virus.

Reducing the spread of the virus is critical to minimize the number of deaths that occur. A model can help officials predict the spread and hospitalization rates and allow them to test possible intervention strategies and see how they affect the model. Therefore, knowing how sensitive the model is to changes in inputs can help officials decide which parameters they want to target, as well as which interventions may be the most effective. Measuring sensitivity is also of interest to researchers so that they can know how accurate the model is.

3 RELATED WORK

Disease model sensitivity analysis can be found in Okaïs et al. (2010), Powell et al. (2005), and Fang et al. (2020). This paper most closely follows Fang et al. but uses a novel methodology by the authors. Okaïs et al. (2010) and Powell et al. (2005) both concern theoretical analyses of disease models, whereas Fang et al. (2020) is a sensitivity analysis of the SEIR model as it relates to COVID-19, so it is the most similar to this paper. Okaïs et al. (2010) lays out a general methodology, where researchers must identify parameters of interest that they plan to analyze. Generally, the parameters fall into three categories: demographic, biological, and interventional. This paper is mostly concerned with biological parameters, especially those that relate to the virus. Powell et al. (2005) shows a methodology for running a sensitivity analysis, presented as steps.

4 METHOD

The equations for the four SEIR populations are

$$\frac{dS}{dt} = -\beta * I * S$$

$$\frac{dE}{dt} = \beta * I * S - (\alpha * E)$$

$$\frac{dR}{dt} = \frac{Mild}{D_{recovery_{mild}}} + \frac{Hospital}{D_{recovet_{severe}}}$$

where, $\beta = \frac{R_0}{D_{infectious}}$ $\alpha = \frac{1}{D_{incubation}}$ $\gamma = \frac{1}{D_{infectious}}$

Finally, we also track two other populations, which are the people currently in the hospital, and the people who will die.

$$\frac{dHospital}{dt} = \frac{Severe}{D_{HospitalLag}} - \frac{Hospital}{D_{recovery_{severe}}} + CFR * \gamma * I - \frac{Fatal}{D_{death}}$$

$$\frac{dFatal}{dt} = \frac{Fatal}{D_{death}}$$

The level of sickness is also included in the model,

$$\frac{dMild}{dt} = P_{mild} * \gamma * I - \frac{Mild}{D_{recovery_{mild}}}$$

$$\frac{dSevere}{dt} = P_{severe} * \gamma * I - \frac{Severe}{D_{HospitalLag}}$$

where $Severe$ is a measure of severe cases that are not currently in the hospital, and $p_{mild} = 1 - p_{severe}$. $D_i = Duration(days)$, or in the case of death, the length of time until death.

The model was implemented in python, and all runs were done in a bash environment. This allowed for easy use of multiple runs to collect data. The model runs from Jan. 15, 2020 until Jan. 14, 2024, or 1462 days. This was the date chosen to run the model to completion in order to combat endpoint behavior that was exhibited by running the model shorter. The model was run using the estimated Whatcom County population of 226387.

Table 1: Input Parameter Ranges

Input	Range of Input	Step Size	Default
I_0	1-100	1	1
R_0	0-7	0.1	2.85
Incubation period	0.5-20	0.1	5
Infectious period	0.5-50	0.5	3
P-severe	0-0.8	0.01	0.04
Hospital lag	1-15.0	0.1	8
CFR	0-0.5	0.001	0.01
Recovery period (mild)	0.5-50	0.5	11
Recovery period (severe)	0.5-50	0.5	21
TTD	4-50.0	1	25

Table 1 shows the different ranges used for inputs into the model. These input ranges were generated by looking at values for papers to get estimates, but also just values that were chosen to test how the model behaves, such as using large psevere or cfr values. The examined outputs were the number of infectious people at the peak of the curve, as well as the date that the curve occurs at. Next, there was also the peak for exposed and the date, the peak for the number of people in the hospital and the date, and finally the total number of dead and total number of recovered people at the end of the model.

To analyze the data, Tukey's five number summary was used to summarize the one-variable data. This was done as it captures linearity/non-linearity, as well as providing ranges of values. For two-variable data, visualizations by heatmaps and other graphs were used. The maximum of the two-variable data was then compared against the one-variable runs to see how the interactions modify the model behavior.

5 RESULTS

5.1 Single Variable

To decide which factors the model was most sensitive to, we counted the number of output parameters that were modified by the input variable. If the output was modified slightly, then we decided that that constituted a change by that input parameter. The parameters that the model was found to be the most sensitive to were I_0, R_0, the incubation period, and the infectious period (Table 2). This makes sense because these parameters are used in the SEIR equations. The parameter that had no effect on the model was the recovery period for mild cases, but that is due to the model running to completion, as all mild cases survived. Other variables such as P-severe and CFR only affected one parameter, namely the parameter that they defined, hospitalization and death respectively. An analysis looking at the interaction of these two variables is performed later.

Table 2: Number of parameters affected

Input Variable	# of parameters affected
I_0	3
R_0	3
Incubation period	3
Infectious period	3
CFR	2
TTD	1
P-severe	1
Hospital lag	1
Recovery period (severe)	1
Recovery period (mild)	0

Looking at how much certain variables change the outputs is also important, and especially those parameters which can be changed by humans, versus those that are more intrinsic to the virus. Looking at how the maximum number of infected people changes (Table 3), only the parameters infectious period, I_0, R_0, and incubation period affected it. The incubation periods and infectious periods affect the peak of infectious people in opposite directions, which would

Table 3: Number of infected

Input Variable		Min	Q1	Median	Q3	Max
R_0	Output	1	8419.18	28577.03	39569.22	45771.25
	Date	Jan-15-20	Jun-12-20	Mar-15-20	Feb-25-20	Feb-16-20
	Input	0	1.7	3.5	5.2	7
I_0	Output	23428.17	23443.625	23452.77	23459.2625	23466.9
	Date	Mar-01-20	Mar-7-20	Mar-14-20	Mar-21-20	Mar-27-20
	Input	1	25	50	75	100
Incubation	Output	61642.31	23455.05	14535.08	10546.28	8274.09
	Date	Feb-07-20	Mar-27-20	May-04-20	Jun-08-20	Jul-12-20
	Input	0.1	5	10	15	20
Infectious	Output	5780	45859	52916	56017	57803
	Date	Feb-25-20	Jun-28-20	Sep-28-20	Dec-25-20	Mar-21-20
	Input	0.5	13	25.5	37.5	50

be expected. Both also have non-linear effects, where the longer the period, the smaller the change for every increase in step size. Both of them also push the date of the peak back. I_0, on the other hand, has a minor linear effect on the output. I_0 is defined as the number of people at the start of the model that are infected with the disease, so a small effect is somewhat surprising. This means, however, that isolating the first case is very important, but isolating all the other cases may not have a large effect. It also suggests that one case can spread the disease very far. R_0 affected the peak both by increasing the value, and also causing the peak to occur sooner. This appeared to be a non-linear effect, as larger values of R_0 caused smaller changes in peak size. Overall, a smaller R_0 value causes the peak to be smaller, but causes the epidemic to occur into longer periods of the year. This effect shows the popular mantra "Flatten the Curve", where keeping the number of infections down should decrease the death rate, but the public has to live through the pandemic.

Next, looking at the effect of parameters on the peak number of people in the hospital (Table 4). This was affected the most, showing that the model is most sensitive to this area. This may also be because more parameters are a part of the hospital equation. The variables that affected the hospitalization peak are I_0, R_0, Hospital Lag (or the time it takes for a patient to become sick before they enter the hospital), the Infectious period, the Incubation period, the probability of a severe case, and the length of time it takes to recover from a severe case. I_0 also had almost no effect on the peak number in the hospital, and there appeared to be some randomness involved. It did push forward the peak, however, causing it to occur earlier. So, an increasing I_0 causes the hospital to become full earlier but does not change the value very much. As expected, hospital lag changes the hospitalization rate, but not in a way that would be expected. It both decreases the curve and pushes it back later. So, when people wait to go to the hospital, less people end up in the hospital. Hospital lag, however, has no effect on the death rate, which is a product of how the model is implemented. The incubation period decreases the hospitalization rate linearly, as well as pushing the peak towards the end of the year.

P-severe and the recovery period of severe cases both only affect the hospitalization rate. Surprisingly, P-severe affects it linearly, and only has a minor effect on the date. The higher the probability is, the more people end up in the hospital. This variable had the largest effect on the number of any of the other variables. The recovery period mostly affected the date, pushing the peak later into the month, but changed the value in a non-linear fashion. R_0 had a non-linear effect on the hospitalization rate as well and increasing R_0 caused the peak to occur sooner. Finally, the infectious period affected the hospitalization rate as well. However, it only behaved well when the infectious period was less than the time to death. Therefore, for infectious period lengths greater than 25, the values are very large. For the well-behaved bit (infectious period < time to death = 25), it did seem to be linear, while also causing the peak to occur later in the year.

Finally, the number of deaths was quite robust in the model, only having two parameters that affected it. However, only one parameter was well-behaved. That parameter was case fatality rate, which modified the total death in an increasing linear fashion. The other parameter that affected the number of dead was the length of the infectious period, but when the infectious period was less than time to death, it did not affect the total dead. When the infectious period was greater than the time to death, the model behaved poorly with very large negative numbers. This suggests some sort of limit on the length of the infectious period, which is discussed further in the discussion.

5.2 Multi-Variable

To see the model behavior on combinations of variables, a subset of the input variables was chosen. This subset was chosen as variables that could be affected or changed by human activity. For example, better care in the form of more respirators could lower the case fatality rate leading to less deaths overall. Hospital lag was chosen because the medical industry or governments could develop methods to identify cases sooner and decrease the wait time of admittance to a hospital. The length of recovery time of a severe case could be lowered with hospital treatment, as well as possible drugs such as antivirals that are currently in development. And

Table 4: Parameter change on hospitalization

Input Variable		Min	Q1	Median	Q3	Max
R_0	Output	0.03	2214.86	4789.67	5294.16	5464.14
	Date	Jan-29-20	Jul-03-20	Mar-30-20	Mar-12-20	Mar-03-20
	Input	0	1.7	3.5	5.2	7
I_0	Output	4317.02	4317.55	4317.99	4317.12	4318.79
	Date	Apr-13-20	Mar-26-20	Mar-22-20	Mar-19-20	Mar-18-20
	Input	1	25	50	75	100
P-severe	Output	973.7	17968.5	35056.3	52144.2	69232
	Date	Apr-07-20	Apr-15-20	Apr-15-20	Apr-15-20	Apr-15-20
	Input	0	0.2	0.4	0.6	0.8
Recovery (severe)	Output	1095	3474	4643	5340	5810
	Date	Apr-06-20	Apr-10-20	Apr-14-20	Apr-17-20	Apr-19-20
	Input	0.5	12.5	25	37.5	50
Incubation period	Output	5424.47	4317.02	3414.17	2800.54	2364.71
	Date	Feb-21-20	Apr-13-20	May-24-20	Jun-30-20	Aug-05-20
	Input	0.1	5	10	15	20
Hospital lag	Output	4769.5	4594.94	4317.02	4034.86	3775.98
	Date	Apr-08-20	Apr-11-20	Apr-13-20	Apr-15-20	Apr-17-20
	Input	1	4.5	8	11.5	15

finally, R_0 can be lowered by public health measures such as hand washing and mask wearing, as well as by social distancing.

Most variables combinations are independent of each other however there are a few exceptions: R_0 and case fatality rate, P-severe and severe recovery period, and P-severe and case fatality rate. P-severe and case fatality rate (Figure 1) when combined have a devastating impact the number of people hospitalized, increasing the maximum by 77% and 137% over its single variable counterparts P-severe and CFR respectfully. This is the two variable combination that have the most drastic impact out of any of the two variables combination tested.

When R_0 and case fatality rate (Figure 2) are combined the death total reached a maximum of 95,674 which is a 7% increase over CFR alone, and a 45% increase on R_0 alone. Recall that by itself R_0 does not affect the total death count. This combination is expected because as both variables increase, the number of people infected and the number of infected that die from the disease both increase as well. This combination also increases the peak number of people hospitalized by 14% over CFR and 980% over R_0 when compared to when they were ran independently, since the more people are in critical condition there are, the more people will be seeking out medical attention.

Lastly, we have P-severe and severe recovery period (Figure 3). As more people are getting worse symptoms and it takes longer for them to get better it comes as no surprise that the maximum number of people in the hospital increases, in this case 46% over an individual run of P-severe. Recovery period, while not having an impact on the maximum number of people hospitalized, does have an impact on when the peak occurs. When comparing the two dates the two-variable combination happens two days later, April 21st, 2020, and just recovery period which occurs on April 19th, 2020. While it is not shown within the model, having a large percent of the populous in the hospital for an extended period of time would surly overwhelm the hospital resulting in insufficient space to give everyone the treatment that they need, directly increasing the death count.

6 DISCUSSION

Time to death and the infectious period were linked in our model by the variable D_{death}. This caused some strange behavior when certain values for either time to death or the infectious period were used. When time to death = infectious period, the program crashes out. When time to death < infectious period, the death output would become very large (100+ digits) and negative or NaN. We would also see the hospitalization output become very large, or else it would report infinity. This is a possible artifact of the model. If it is an artifact of the model, then the infectious period must always be less than the time it takes for a patient to die. In terms of real-world analysis, this makes sense as very few diseases, especially not respiratory viruses, kill their host before the host is able to infect a new one.

Another interesting aspect of the model was that the hospitalization rate and death rate were independent. This is a limitation of the model, because in the real world, these values would be linked, such that if the hospitalization rate increases, especially past a maximum threshold value, then the number of deaths would also increase. Revising the model to reflect this would be advised, as the hospitals can only take a maximum number of patients before having to turn away new patients or rationing supplies. A revised model would allow for hospital managers to make accurate predictions as to when this would occur.

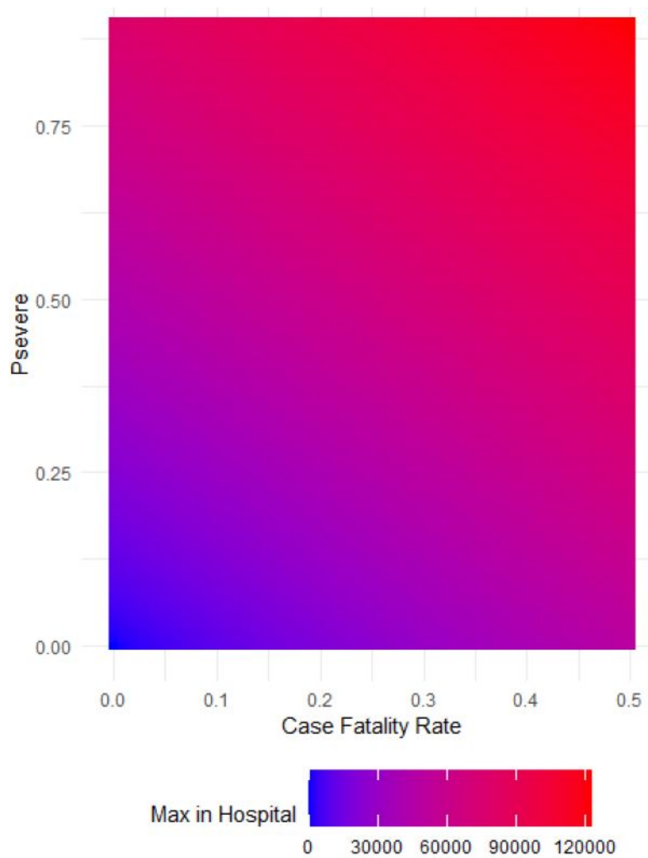

Figure 1: P-severe and CFR vs. Maximum number of people in hospital. This shows an increase in the hospitalization rate as P-severe and CFR increase.

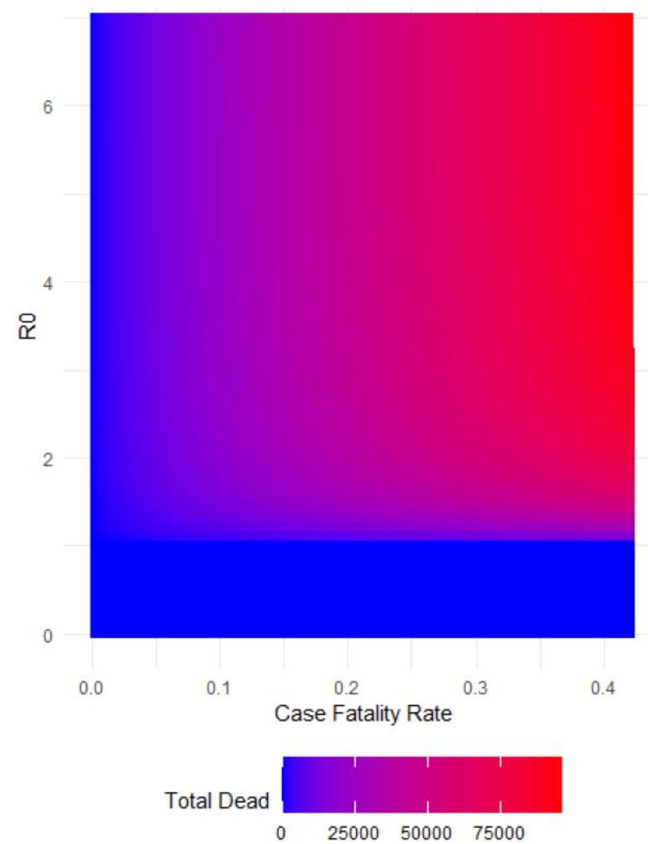

Figure 2: R_0 and CFR vs. total dead. This shows that when R_0 is small (1 or less), then the number of deaths is low. However, for R_0 greater than 1, the total deaths are controlled by CFR.

The data presented here shows that some interventions may be more effective at controlling the spread and minimizing the death rate than others. A multi-pronged approach aimed at reducing the controllable parameters such as R_0 or hospital lag can help reduce the number of people that die. We see that interventions targeting R_0 are the most effective. However, possible future interventions allowed by new pharmaceuticals show that they would be very effective, but it would be best to minimize the damage until those interventions are created.

7 NEXT STEPS

There were some issues during the experimentation. Not all the results for two variables combinations are complete and we could not continue to three variable combinations due to the time constraint of the project. Therefore, a full analysis of the model with all combinations would need to be completed. The model has limitations that need to be investigated and fixed to produce more accurate results. The biggest limitation seems to be that the death rate is only dependent on the case fatality rate, as well as the inconsistencies associated with time to death and the infectious period. It is possible that the equations need to be revised so that the infectious period and time to death do not cause the model to behave unpredictably.

Through our speculation on the results, there are some external factors that also influence the result from running the model. First is investigating the lack of responsiveness in total deaths for most variables. The result on total number of dead remains the same for most variables, which does not correlate to our expectation. This is because of the independence of hospitalization and death. Creating a new model that links these two variables would be of interest. Second, introducing randomness to the model. This could be done using stochastic methods, or a probability that somebody may be infected if they come into contact with an infected person. There are some factors which could not be defined that also influence the result. For example, adding birth and natural death rates would affect the total number of dead to the model because there are also people who die but tested negative. Collecting numbers of infected, hospitalized, and dead in Whatcom country to determine the accuracy of the results would help us understand how efficient the model is. The model would also need to be checked for accuracy by using data from past coronavirus outbreaks such as SARS and MERS to see how well the model predicts the course of events.

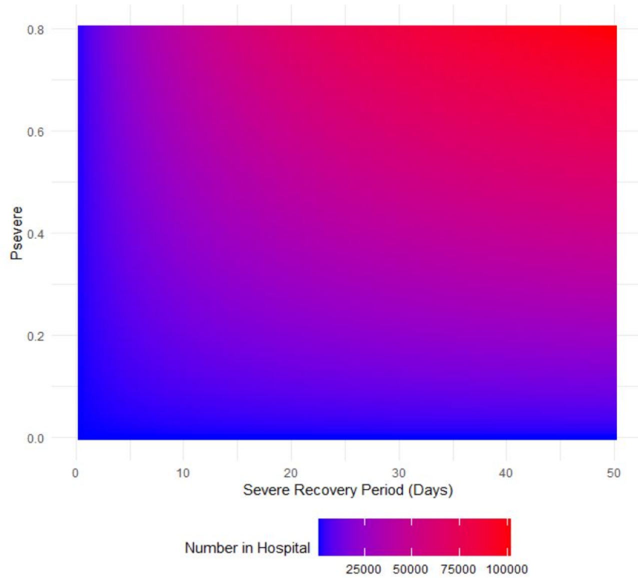

Figure 3: P-severe and recovery length vs. Maximum number of people in hospital. This shows that at the very low values for either parameter, the number of people in the hospital is low. It increases with larger values.

8 REFERENCES

[1] Biswas, M. H. A., Paiva, L. T., De Pinho, M. (2014). A SEIR model for control of infectious diseases with constraints. Mathematical Biosciences and Engineering, 11(4), 761-784. Doi:10.3934/mbe.2014.11.761

[2] Fang, Y., Nie, Y., & Penny, M. (2020). Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. Journal of Medical Virology, 92(6). Doi: 10.1002/jmv.25750

[3] Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., … Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine, 172(9), 577–582. Doi: 10.7326/M20-0504

[4] Kermack, W. O. & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society A, 115(772). 700-721. Doi: 10.1098/rpsa.1927.0118

[5] Okaïs, C., Roche, S., Kürzinger, M.-L., Riche, B., Bricout, H., Derrough, T., … Ecochard, R. (2010). Methodology of the sensitivity analysis used for modeling an infectious disease. Vaccine, 28(51), 8132-8140. Doi: 10.1016/j.vaccine.2010.09.099

[6] Park, M., Cook, A., Lim, J. T., Sun, Y., Dickens, B. (2020). A systematic review of COVID-19 epidemiology based on current evidence. Journal of Clinical Medicine, 9(4), 967. https://doi.org/10.3390/jcm9040967

[7] Whatcom County. (2020, May 22). Novel Coronavirus. Whatcom County. https://www.whatcomcounty.us/3329/Novel-Coronavirus-COVID-19