2020

A Mathematical Model for the Effect of Social Distancing on the Spread of COVID-19

Anna Singley

Olympia High School, anna.singley@olympia.org

Hannah Callender Highlander

University of Portland, highland@up.edu

Follow this and additional works at: <https://ir.library.illinoisstate.edu/spora>

Part of the [Disease Modeling Commons](#), and the [Respiratory Tract Diseases Commons](#)

Recommended Citation

Singley, Anna and Callender Highlander, Hannah (2020) "A Mathematical Model for the Effect of Social Distancing on the Spread of COVID-19," *Spora: A Journal of Biomathematics*: Vol. 6, 40–51.

DOI: <https://doi.org/10.30707/SPORA6.1/ARDA7202>

Available at: <https://ir.library.illinoisstate.edu/spora/vol6/iss1/5>

This Mathematics Research is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Spora: A Journal of Biomathematics by an authorized editor of ISU ReD: Research and eData. For more information, please contact ISUReD@ilstu.edu.

A Mathematical Model for the Effect of Social Distancing on the Spread of COVID-19

Anna Singley¹, Hannah Callender Highlander^{2,*}

*Correspondence:
Prof. Hannah Callender
Highlander, Dept. of
Mathematics, University of
Portland, 5000 N. Willamette
Ave., Portland, OR 97203,
USA
highland@up.edu

Abstract

Social distancing is an effective method of impeding the spread of a novel disease such as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), but is dependent on public involvement and is susceptible to failure when sectors of the population fail to participate. A standard SIR model is largely incapable of modeling differences in a population due to the broad generalizations it makes such as uniform mixing and homogeneity of hosts, which results in lost detail and accuracy when modeling heterogeneous populations. By further compartmentalizing an SIR model, via the separation of people within susceptible and infected groups, we can more accurately model epidemic dynamics and predict the eventual outcome, highlighting the importance of societal participation in social distancing measures during novel outbreaks.

Keywords: SIR models, mathematical epidemiology, COVID-19, social distancing

1 Introduction

The stunning speed at which coronavirus disease 2019 (COVID-19) leapt from its beginning as a unique illness in Wuhan, China to a global pandemic has created the type of crisis that requires researchers to draw upon knowledge and tools from many disparate disciplines. As of yet there exists no single, simple, and clear path to containment in order to mitigate the potential destruction from this pandemic. Thus, the situation requires the implementation of a variety of strategies in order to alter the trajectory of spread, in order to buy time for the development of a comprehensive solution. COVID-19 is somewhat unique in its ability to spread through those infected individuals who are pre-symptomatic or asymptomatic. Thus far in the pandemic spread, social distancing and contact tracing have shown great promise in slowing and ultimately containing the spread of COVID-19 [4] [13] and have proven to be effective in managing past outbreaks [6]. However, a key weakness of this approach is that it is highly dependent upon understanding and willingness by the public to properly and fully adopt all of the required behaviors and actions of social distancing. Therefore, there will most certainly be a large gap between potential and actual efficacy of social distancing guidelines and directives.

The consequences of an unchecked epidemic are dire for the human population and could result in heightened mortality rates from COVID-19 and other conditions re-

quiring hospitalization [12]. Although several countries have empirically shown the results of insufficient application of measures to control the spread of COVID-19, Italy provides to date the best evidence of both the ferocity of this particular viral strain and how social distancing measures can change the trajectory of the viral impact even when measures are delayed and imperfect in implementation. As shown in other countries such as Japan and South Korea, social distancing and contact tracing done well can not only achieve the desired flattening of the curve but can also rapidly curtail the spread by robbing the virus of its human transmission vector through the application of rigorously applied social distancing [13].

The SIR model, first proposed in 1927 by Kermack and McKendrick [8], presents mathematicians with a general model for disease progression over a population. This model divides the studied population into three main compartments: *S* for *Susceptible*, or people who do not currently have symptoms of the disease and cannot spread the disease but are susceptible to infection; *Infectious* people, denoted by *I*, who have the disease or sickness being modeled, and can spread it to people within the susceptible class; infectious people have a chance of either dying from the disease, thus being removed from the population, or progressing to the last state, *Recovered*, denoted by *R*.

One drawback to the standard SIR model is the homogeneous approach it takes. Instead of operating under the assumption that different people will react in different ways and have different odds of progressing through

¹Olympia High School, ²Mathematics Department, University of Portland

the model compartments, the standard SIR model makes broad assumptions that generalize entire populations [15]. Although the SIR model can be an accurate tool in predicting the spread of a disease or the number of infections in a local outbreak, without additional modifications, the assumptions it makes are inaccurate when modeling a disease that disproportionately kills certain demographics or infects certain groups to varying degrees. COVID-19 falls into this category, both in the way it tends to leave elderly people and those with pre-existing conditions hospitalized or in the ICU[9], and in the increased rate of spread in populations who ignore social distancing orders [16]. We therefore present a model, based off of the standard SIR model, but where both the susceptible and the infectious compartments contain subgroups of people, in terms of varying levels of social distancing as well as degrees of health in the population, respectively, so that the importance of total adherence to social distancing might be illuminated. In the next section, we describe the novel aspects of our mathematical model and further reasoning behind its construction. In section 3 we discuss our methods of parameter estimation, followed by model results in Section 4 and methods and results of global and local sensitivity analysis in Section 5. Finally, we offer conclusions and future directions in Section 6.

2 Mathematical Models

The model proposed in 1927 by Kermack and McKendrick [8] has served as a template for countless biomathematical models [1]. Although the original model was comparatively simple by today's standards, the SIR model easily lends itself to modification and the introduction of complexity as necessitated by many current infectious diseases. Our model is based off of the Kermack and McKendrick equations but separates both the susceptible and the infectious compartments into three subgroups, based on the likelihood of progression to infected and recovered, respectively. This allows for a simplified model of a more heterogeneous population and also enables us to test the impact of portions of a population refraining from social distancing, as well as the dynamics of a widespread outbreak on a vulnerable populace.

Our differential equation model is further altered to divide the susceptible and infected populations into three groups based on the likelihood of progression to infected and recovered, respectively. This allows for a simplified model of a more heterogeneous population and allows us to test the impact of portions of a population refraining from social distancing, as well as the dynamics of a widespread outbreak on a vulnerable populace.

As mentioned above, in the model we establish three subgroups of S and three subgroups of I . We do this

to acknowledge that a population will not, pragmatically speaking, have the same universal traits and susceptibilities. More importantly, a population will not act in a uniform manner with respect to social distancing guidelines proposed by local governments, thus changing the dynamics of the spread of a disease. The three subgroups for S are as follows: S_1 for those who ignore social distancing or have essential, high-risk jobs; S_2 for those who largely follow social distancing but fall short in perfect adherence; and S_3 for those who follow social distancing and take additional precautions to avoid exposure and transmission, such as wearing masks in public.

An example of S_1 would be a doctor who has high exposure due to their occupation or someone who purposefully violates social distancing guidelines. At the other extreme is S_3 , which is defined as a populace that stays quarantined, and takes all necessary precautions to avoid infection as stated by local government. Between S_1 and S_3 is S_2 , which would consist of people who do not gather at work or school, but who also do not take all necessary precautions. In what follows, we will use capital letters to denote the subgroups, and lowercase to denote the fraction of the population within each group. For instance, S_1 is a subgroup of individuals, equal to $s_1 * S$, where s_1 is the fraction of the entire susceptible population contained in subgroup S_1 .

Similarly, our I compartment is divided into three subgroups, I_1 , I_2 , and I_3 , based on either a uniform distribution or a normal distribution of risk factors (both scenarios are examined in our model analysis). We define I_1 as the least healthy group at the highest risk of succumbing to COVID-19; I_2 as an average health group; and I_3 as the healthiest group. Similar to our notation for the S subgroups, i_n will denote the fraction of the I individuals who are in subgroup I_n , so that, for example $I_1 = i_1 * I$.

Currently we assume a fixed population, with no births and no non-COVID related deaths. We assume that once a susceptible individual has become infected, they immediately move into the infectious compartment. Once infectious, individuals either die from COVID-19, governed by the rate constant ϕ_n , or they recover according to the rates γ_n . We also assume that the three subgroups of S and the three subgroups from I are not strongly correlated, but in reality they may very well overlap. Our resulting model equations are as follows:

$$\frac{dS}{dt} = - \sum_{n=1}^3 \lambda_n S_n I, \quad (1)$$

$$\frac{dI}{dt} = \sum_{n=1}^3 \lambda_n S_n I - \sum_{n=1}^3 (\psi_n) I_n - \sum_{n=1}^3 \gamma_n I_n, \quad (2)$$

$$\frac{dR}{dt} = \sum_{n=1}^3 \gamma_n I_n. \quad (3)$$

Each S subgroup is assumed to have a different level of interaction with the rest of the population, as reflected in the individual infection rates λ_n . Each λ_n is determined both by the universal transmission rate, ϕ , and the variable multipliers c_n that govern the level of magnification of transmission probability corresponding to each subgroup, based off of their adherence to social distancing policies. Thus, $\lambda_n = \phi * c_n$. Similarly, each I subgroup may have a different likelihood of dying as well as recovering from COVID-19, as dictated by ψ_n and γ_n , respectively. The values of ψ_n emerge from the overall death rate, ψ , found in the current literature, with multiplier d_n . The three levels for d_n were chosen to reflect the differences in each I subgroup's ability to fight the disease. As such, $\psi_n = \psi * d_n$. Likewise, $\gamma_n = \gamma * \tau_n$, where γ is the average recovery rate derived from existing literature, and τ_n is a multiplier to take into account each I subgroup's duration of infectiousness due to their varying levels of health before contracting COVID-19. A summary of the model parameters, their default values, and references for parameter values derived from the literature are shown in Table 1.

3 Parameter Optimization

As described earlier, λ_n governs the rate at which susceptible individuals from S_n become infected. We calculate λ_n through the infection transmission (ϕ), which models how easily the disease can be spread without factoring in individual contacts, and thus remains constant for all three S subgroups, as the virus is assumed to behave the same for each group. However, the contact level (c) that these three groups maintain is grossly different and serves as a coefficient modifying the value of ϕ , and thus eventually λ , for each subgroup. For strict adherence to social distancing, $c = 0.001$; for partial adherence to social distancing, $c = 0.01$; and for little to no adherence to social distancing, $c = 1$. Thus, someone in S_2 is 10 times *more* likely to contract COVID-19 than someone in S_3 , and someone in S_2 is 100 times *less* likely than someone in S_1 to contract COVID-19.

In order to estimate the infection transmission parameter (ϕ), we originally used data from China, which reported its first case of pneumonia with an unknown origin to the WHO on December 31, 2019 [14]. Seventy-nine days later, on March 19, 2020, China recorded their first day with zero new COVID-19 deaths, which also marked the partial re-opening of Wuhan, China [5]. Therefore, we assumed that the ϕ parameter in our model, when used in combination with other properly tuned param-

eters, should result in 0 new cases after 79 days in model runs, which gave us a ϕ estimate of 0.0011. However, the validity of the Chinese data is questionable, and we did not feel comfortable using potentially fabricated data. We reached a reasonable value for ϕ via the use of infection transmission probability estimates of H1N1 spread in Italian schools during the H1N1 outbreak [3]. Although it is safe to assume that COVID-19 is considerably more contagious than H1N1, it is equally certain that transmission of a virus within a school is much more likely than within a general population [11]. Therefore, we feel justified in using 0.000952—the mean transmission probability estimate of a school in Italy—as our ϕ value.

The recovery rate, γ , was calculated based on the conservative assumption that 95% of individuals will have a case outcome within 4 weeks, giving us the parameter value of .0339 per day. The death rate from infection, ψ , was calculated in a similar fashion, based off of the assumption that an individual has a 3% chance of dying over the 4 weeks, based off of early data from Wuhan, China [17]. As mentioned earlier, τ is used to model differing health statuses among a populace. For individuals in a higher risk class (I_1) γ was cut in half by setting $\tau = 0.5$; for individuals of average health, γ was held constant ($\tau = 1$), and for individuals with above-average health, γ was doubled ($\tau = 2$). Note that decreasing γ increases the time an individual's duration of infectiousness, thus increasing the probability of death.

4 Model Analysis and Results

Our initial population of susceptible individuals is set to 1,000 people, and our initial infected population is set to 10 people, or 0.99% of the total population. We estimate that the most probable distribution of risk factors for affecting the chance of recovery, dictated by the parameters i_n , follows a normal distribution, based on data of age distribution in the U.S. [2] In addition, we tested the model in the case of all individuals having the same, average health, as a means for measuring the importance of separating the infectious class into the three subgroups. The S distribution could realistically be shifted by government intervention, which would hold long-term ramifications for the dynamics and outcome of the outbreak. It is with this in mind we undergo an analysis to estimate the epidemic outcome from various S distributions, after first optimizing our model parameters via real COVID-19 data, as described in the previous section. Table 2 provides the parameter values associated with the distributions we tested in our analysis for the S and I subgroups.

For all tested scenarios, we kept our unperturbed parameters constant according to Table 1. Thus, the only parameters being altered in this initial analysis are s_n

Parameters	Definition	Nominal Value
λ_n	infection rate	$\lambda_n = \phi c_n \left(\frac{1}{\text{day} \cdot \text{person}} \right)$
γ	rate of recovery	$0.0339 \left(\frac{1}{\text{day}} \right)$ [17]
ψ	death rate from infection	$0.00107 \left(\frac{1}{\text{day}} \right)$ [17]
ϕ	transmission rate	$0.000952 \left(\frac{1}{\text{day} \cdot \text{person}} \right)$ [3]
c_1	multiplier for S_1 interactions	1
c_2	multiplier for S_2 interactions	0.01
c_3	multiplier for S_3 interactions	0.001
τ_1	health multiplier of I_1	0.5
τ_2	health multiplier of I_2	1
τ_3	health multiplier of I_3	2

Table 1: Descriptions and nominal values for the model parameters.

Scenario	s_1 high exposure	s_2 medium exposure	s_3 low exposure	i_1 low health	i_2 average health	i_3 high health
1a	0.05	0	0.95	0	1	0
1b	0.05	0	0.95	0.16	0.68	0.16
2a	0.10	0.45	0.45	0	1	0
2b	0.10	0.45	0.45	0.16	0.68	0.16
3a	0.50	0.25	0.25	0	1	0
3b	0.50	0.25	0.25	0.16	0.68	0.16

Table 2: Parameter settings for S and I subgroups that set the proportion of individuals in each subgroup for each of the six scenarios tested.

and i_n . Sensitivity analysis on all model parameters will be the focus of the next section. All simulations were run using the ode45 solver in MATLAB.

4.1 Scenario 1: Severe social distancing measures

For this hypothetical, and likely impossible scenario, we assume that only 5% of the population falls into our S_1 category, and the rest practice harsh social distancing, thus becoming part of our S_3 category. It took roughly 629 days to get from approximately 1% infection to 0 cases when simulating with a uniform distribution of average risk factors (Scenario 1a, Figure 1a), and it took roughly 680 days for the number of infections to drop to 0 in a population with health distributed following normal distribution (Scenario 1b, Figure 1b). At the conclusion of the universal average health simulation, roughly 1.54% of the initial susceptible population is dead, compared to roughly 2.36% in the normal distribution health simulation. The peak number of infections for the average health simulation occurred at 195 days, with 4.72% of the initial population infected. For the normal distribution of health factors, a peak occurred at 203 days, wherein 2.72% of the population was infected.

4.2 Scenario 2: Partial social distancing

For this scenario, we assume 10% of the population falls into our S_1 category, 45% fall into S_2 , and 45% fall into S_3 . This is an approximation of what would happen if 10% of the population continued to attend social gatherings and other high-risk events, 45% of the population would take partial precaution but fall short from complete adherence, and 45% of the population would take all necessary precautions. It took roughly 305 days to get from 1% infection to 0 cases when simulating with a population consisting of only average-health people (Scenario 2a, Figure 2a), and 310 days in a population with health distributed following normal distribution (Scenario 2b, Figure 2b). At the conclusion of the universal average health simulation (defined as when infections reach 0) 2.54% of the initial susceptible population is dead. At the conclusion of the simulation in which health factors follow a normal distribution, 5.18% of the initial susceptible population is dead. The peak number of infections for the normally distributed risk factors infected 23% of the initial population and occurred at 81 days. The peak for the universal average health simulation varied slightly, with a peak number of infections reaching 26.5% of the initial population at 79 days.

Figure 1: Time series plot of severe social distancing with the assumption of a uniform, average health population (panel a, Scenario 1a) and normally distributed health population (panel b, Scenario 1b).

Figure 2: Time series plot of partial social distancing with the assumption of a uniform, average health population (panel a, Scenario 2a) and normally distributed health population (panel b, Scenario 2b).

4.3 Scenario 3: Minimal Social Distancing

For this scenario, we assume 50% of the population falls into S_1 , 25% of the population is S_2 , and the remaining 25% is S_3 . With a population of universal average health (Scenario 3a, Figure 3a), it takes 209 days for the infections to reach 0. With a normal distribution of health (Scenario 3b, Figure 3b), it takes 190 days for the infections to reach zero infections from 1% infected populace initially. Once the infections reach 0 days in the average health population, 2.9% of the initial population is dead. In the normal distribution simulation, 5.81% of the initial population is dead. A major difference between these minimal social distancing scenarios and other scenarios is the peak infection size. In the universal average health simulation, a peak occurs within 16 days, where 67% of the initial population is simultaneously infected. In the normal distribution scenario, a peak occurs at 18 days, where 64% of the population is simultaneously infected. This scenario would most certainly prove to be catastrophic for the healthcare system, likely leading to even higher death rates than predicted by this simplified model.

For ease of comparison among the six scenarios, a summary table is provided in Table 3.

5 Sensitivity Analysis

As many characteristics of the spread of COVID-19 remain unknown, many of our model parameters were necessarily estimated. It is therefore crucial to understand which parameters, when varied, lead to the greatest model variation by conducting a global sensitivity analysis. Such analysis can both help determine which parameters we should focus on obtaining accurate estimates for, as well as which interventions might cause the most drastic outcome changes in terms of, for instance, disease duration, and public health outcomes.

5.1 Global sensitivity analysis using eFAST

The global sensitivity analysis procedure called Extended Fourier Amplitude Sensitivity Testing (eFAST) was used to partition model output to the input parameters. eFAST is capable of deriving a mathematical relationship between the outcome of a model run, and the model parameters even when nonlinear interactions exist between parameters. Each of the model's output variables were most sensitive to different parameters, at three different points in time, as reported through first- and total-order sensitivity indices, shown in Tables 4 through 7. The three time points—5 days, 95 days, and 300 days—were

chosen based off of the dynamics of the outbreak for the same set of parameters as defined in Section 3. For these parameter settings, 5 days gives an idea of sensitivity at the beginning of the outbreak; 95 days was the peak of the outbreak, and the number of infections had approached 0 by day 300. For further review of eFAST, see [10].

As demonstrated in Table 4, S is greatly sensitive to the values of both s_1 (percentage of individuals in S_1 , the group with highest exposure) and λ_1 (infection rate for S_1) at all tested time points, while I is also most sensitive to the values of these same parameters at the 5-day and 300-day marks (see Table 5). This suggests that managing the number of people who ignore social distancing is crucial to mitigating the damage that an outbreak will cause. The percentage of individuals who have average health, i_2 , and the rate at which those individuals recover, γ_2 , caused most output uncertainty for I at the peak of the outbreak (see Table 5) and for R at the beginning of the outbreak (see Table 6), which indicates that the preexisting health of the average person in the population plays a crucial role in managing the size of the peak and the initial number of recovered individuals. As the disease progresses, s_1 and λ_1 reemerge as the most important parameters for the recovered compartment. These two parameters remain at the forefront of dictating the number of deaths at the peak of infection, while i_1 and γ_2 report high sensitivity indices at the beginning and end of the outbreak, respectively, as shown in Table 7. Finally, D was also sensitive to ψ_1 , the death rate for individuals in the lowest health category, in the earlier stages and, to a lesser extent, in the later stages of the outbreak.

5.2 Local sensitivity analysis

The results from this global sensitivity analysis enable us to identify the model parameters that are most crucial to the overall outcome. However, the prior sensitivity analysis is not sufficient to understand how drastically outcomes can vary due to only one parameter. Thus, in addition to the global sensitivity analysis, we also undertook a one at a time (OAT) sensitivity analysis of several key parameters, chosen on the criteria that the parameters had large sensitivity indices from the global analysis. The parameters studied in our OAT include those with the highest two sensitivity indices for each of the four model output variables (S, I, R, D), at each of the three time points tested from eFAST (5 days, 95 days, and 300 days), and are as follows: s_1 , λ_1 , i_2 , γ_2 , i_1 , and ψ_1 . In addition to having one of the highest sensitivity indices for several outputs, λ_1 is also important to study because of its association with ϕ , the infection transmission rate. As we could not find an accurate estimate for ϕ in existing COVID-19 literature, we used an estimate from H1N1

Figure 3: Time series plot of severely lacking social distancing with the assumption of a uniform, average health population (panel a, Scenario 3a) and normally distributed health population (panel b, Scenario 3b). Note that in this simulation, the epidemic is not controlled and thus becomes drastically worse.

Scenario	days until $I = 0$	% of population who have died	day of peak infection	% of population infected at peak
1a	629	1.54%	195	4.72%
1b	680	2.36%	203	2.72%
2a	305	2.54%	81	23.00%
2b	310	5.18%	79	26.50%
3a	209	2.90%	16	67.00%
3b	190	5.81%	18	64.00%

Table 3: Summary of results from each of the scenarios listed in Table 2.

Sensitivity to S						
Parameter	$t = 5$ days		$t = 95$ days		$t = 300$ days	
	first-order	total-order	first-order	total-order	first-order	total-order
s_1	0.4962	0.4982	0.4455	0.4477	0.3355	0.3396
s_2	0.0002	0.0002	0.0002	0.0004	0.0001	0.0007
s_3	0.0000	0.0000	0.0000	0.0002	0.0000	0.0005
λ_1	0.4962	0.4988	0.4472	0.4489	0.3390	0.3427
λ_2	0.0002	0.0002	0.0002	0.0004	0.0001	0.0007
λ_3	0.0000	0.0001	0.0002	0.0002	0.0005	0.0005
ψ_1	0.0000	0.0000	0.0000	0.0003	0.0001	0.0011
ψ_2	0.0000	0.0000	0.0000	0.0002	0.0000	0.0006
ψ_3	0.0004	0.0000	0.0115	0.0002	0.0351	0.0006
i_1	0.0017	0.0005	0.0421	0.0118	0.1295	0.0361
i_2	0.0000	0.0017	0.0006	0.0425	0.0017	0.1317
i_3	0.0003	0.0000	0.0089	0.0007	0.0276	0.0022
γ_1	0.0016	0.0004	0.0402	0.0091	0.1239	0.0285
γ_2	0.0000	0.0016	0.0006	0.0405	0.0017	0.1260
γ_3	0.0000	0.0000	0.0000	0.0007	0.0000	0.0023

Table 4: First-order and total-order sensitivity indices obtained by running eFAST for the S compartment's sensitivity to model parameters at 5 days, 95 days, and 300 days. In the eFAST analysis, parameters were each varied by $\pm 5\%$ of their nominal values.

Sensitivity to I						
Parameter	$t = 5$ days		$t = 95$ days		$t = 300$ days	
	first-order	total-order	first-order	total-order	first-order	total-order
s_1	0.4534	0.4554	0.1979	0.2290	0.4812	0.5061
s_2	0.0002	0.0003	0.0001	0.0010	0.0002	0.0003
s_3	0.0000	0.0002*	0.0000	0.0008	0.0000	0.0001
λ_1	0.4536	0.4560	0.1997	0.2296	0.4840	0.5062
λ_2	0.0002	0.0003	0.0001	0.00010	0.0002	0.0003
λ_3	0.0000	0.0002*	0.0000	0.0008	0.0000	0.0001
ψ_1	0.0001	0.0003	0.0009	0.0018	0.0000	0.0001
ψ_2	0.0000	0.0002	0.0002*	0.00010	0.0000*	0.0001
ψ_3	0.0000	0.0002	0.0000	0.0009*	0.0000	0.0001
i_1	0.0099	0.0100	0.0640	0.0653	0.0011	0.00013
i_2	0.0367	0.0370	0.2271	0.2296	0.0037	0.0040
i_3	0.0005	0.0006	0.0029	0.0038	0.0001	0.0001
γ_1	0.0076	0.0078	0.0497	0.0509	0.0008	0.0010
γ_2	0.0345	0.0347	0.2219	0.2243	0.0036	0.0039
γ_3	0.0005	0.0006	0.0031	0.0040	0.0000	0.0001

Table 5: First-order and total-order sensitivity indices obtained by running eFAST for the I compartment's sensitivity to model parameters at 5 days, 95 days, and 300 days. In the eFAST analysis, parameters were each varied by $\pm 5\%$ of their nominal values. Parameters with an * indicate those whose p-values failed to fall below the threshold of significance, 0.01.

Sensitivity to R						
Parameter	$t = 5$ days		$t = 95$ days		$t = 300$ days	
	first-order	total-order	first-order	total-order	first-order	total-order
s_1	0.0684	0.0699	0.4792	0.4805	0.3511	0.3554
s_2	0.0000	0.0015	0.0002	0.0003	0.0001	0.0006
s_3	0.0000	0.0012	0.0000	0.0001	0.0000	0.0004
λ_1	0.0683	0.0699	0.4803	0.4817	0.3552	0.3590
λ_2	0.0000	0.0014	0.0002	0.0003	0.0001	0.0006
λ_3	0.0000	0.0013	0.0000	0.0001	0.0000	0.0005
ψ_1	0.0000	0.0014	0.0003	0.0004	0.0047	0.0052
ψ_2	0.0000	0.0013	0.0001	0.0001	0.0009	0.00013
ψ_3	0.0000	0.0013	0.0000	0.0001	0.0000	0.0005
i_1	0.0749	0.0764	0.0054	0.0055	0.0460	0.0469
i_2	0.3476	0.3497	0.0155	0.0157	0.1149	0.1171
i_3	0.0048	0.0061	0.0002	0.0003	0.0013	0.0018
γ_1	0.0773	0.0785	0.0030	0.0031	0.0216	0.0224
γ_2	0.3502	0.3517	0.0136	0.0138	0.0970	0.0991
γ_3	0.0049	0.0062	0.0002	0.0003	0.0013	0.0019

Table 6: First-order and total-order sensitivity indices obtained by running eFAST for the R compartment's sensitivity to model parameters at 5 days, 95 days, and 300 days. In the eFAST analysis, parameters were each varied by $\pm 5\%$ of their nominal values.

Sensitivity to D						
	$t = 5$ days		$t = 95$ days		$t = 300$ days	
Parameter	first-order	total-order	first-order	total-order	first-order	total-order
s_1	0.0518	0.0525	0.4155	0.4171	0.1013	0.1036
s_2	0.0000	0.0004	0.0002	0.0004	0.0000	0.0011
s_3	0.0000	0.0004	0.0000	0.0003	0.0000	0.0012
λ_1	0.0518	0.0525	0.4170	0.4187	0.1015	0.1035
λ_2	0.0000	0.0004	0.0002	0.0004	0.0000	0.0012
λ_3	0.0000	0.0004	0.0000	0.0003	0.0000	0.0012
ψ_1	0.4041	0.4058	0.0253	0.0260	0.1712	0.1731
ψ_2	0.0729	0.0735	0.0046	0.0049	0.0309	0.0322
ψ_3	0.0000	0.0004	0.0000	0.0003	0.0000	0.0011
i_1	0.3683	0.3696	0.0011	0.0016	0.0209	0.0224
i_2	0.0429	0.0434	0.0399	0.0403	0.1572	0.1590
i_3	0.0000	0.0004	0.0009	0.0012	0.0040	0.0051
γ_1	0.0009	0.0013	0.0160	0.0163	0.0735	0.0749
γ_2	0.0040	0.0044	0.0719	0.0723	0.3311	0.3326
γ_3	0.0001	0.0005	0.0010	0.0012	0.0045	0.0056

Table 7: First-order and total-order sensitivity indices obtained by running eFAST for the D compartment's sensitivity to model parameters at 5 days, 95 days, and 300 days. In the eFAST analysis, parameters were each varied by $\pm 5\%$ of their nominal values.

in an Italian school. We cannot, at this time, ascertain the accuracy of our ϕ estimate, so the OAT sensitivity analysis is essential for providing a picture of different trajectories present with differing levels of transmission. For this OAT analysis, all of our parameters were varied by $\pm 25\%$ of their nominal values to create the differing trajectories shown in Figures 4 through 6. In each OAT analysis, the default distribution for S was the same as in Scenario 2, and the default distribution for I was the normal distribution.

As you can see from Figure 4, all trajectories are largely impacted by varying s_1 . Note that since s_1 and λ_1 are multiplied by one another at each instance in which they appear in the model, it makes sense that varying each of these parameters individually would result in a nearly identical output, which is exactly what we observed and why we elected to omit the figure for varying λ_1 . A 25% increase in either s_1 or in λ_1 resulted in an earlier and larger peak of infection as well as a quicker rate of recovery, while a 25% decrease in either of these parameters had exactly the opposite effect. When s_1 was decreased by 25%, the peak number of infections decreased almost 64% compared to the baseline simulations, to a peak simultaneously infecting 16% of the population. However, the decrease in the s_1 parameter also resulted in the outbreak lasting 382 days. When s_1 was increased by 25%, the peak number increased by 132% compared to baseline simulations, but the outbreak reached zero infection within 265 days. The number of deaths was virtually the same in these simulations. These outcomes indicate the importance of social distancing in regards both to its affect on the timing and size of peak infection as well as the overall duration of the outbreak. It is important

here to note that these two parameters caused the largest quantitative variations in model outputs of all the model parameters.

Figure 5 reveals differences that are apparent when the recovery rate γ_2 is varied. Similar to the relationship between s_1 and λ_1 , since i_2 (the percentage of the infectious population with average levels of health) and γ_2 are multiplied by one another in equations for I and R , the OAT analysis results for varying γ_2 were nearly identical to those from varying i_2 . Here, an increase in γ_2 (or in i_2) did not greatly affect the timing of peak infection but did cause a reduction in the magnitude of the peak. Likewise, a decrease in γ_2 (or in i_2) resulted in an equal-sized increase in the magnitude of the peak infection. Reducing i_1 resulted in negligible effects to the outbreak length, but had a large impact on the death rate - decreasing i_1 by 25% resulted in only 4.9% of the population dying, compared to 6.4% when i_1 is increased by 25%. Similar results were true for the final size of the recovered compartment due to the altered number of deaths that can be partitioned to i_1 . Though the effects were not as large for the deaths compartment, γ_2 played the largest role in determining the total number of deaths by the end of the outbreak. These results suggest that an effective way to minimize severity of outbreak, without affecting timing, is to find ways to either improve the existing health of the population (likely impossible in the short term) or to find mechanisms to increase the rate of recovery, such as finding novel therapeutics to reduce duration of infectiousness.

Varying i_1 produced only minor changes in comparison to the effects of the other parameters studied in this OAT analysis (see Figure 6). These outcomes were nearly iden-

Figure 4: Effects of varying s_1 (parameter controlling the proportion of the population in S_1 , those who have highest social contact) $\pm 25\%$ on epidemic trajectories. Solid curves represent the default values; dotted curves represent a 25% reduction in s_1 ; and the dashed curves represent a 25% increase in s_1 . Note that the results for varying s_1 are nearly identical to those varying λ_1 (infection rate for S_1 ; figure not shown).

Figure 5: Effects of varying γ_2 (the recovery rate of S_2) by $\pm 25\%$ on epidemic trajectories. Solid curves represent the default values; dotted curves represent a 25% reduction in γ_2 ; and the dashed curves represent a 25% increase in γ_2 . Note that the results for varying γ_2 are nearly identical to those varying i_2 (percentage of I population with average health levels; figure not shown).

Figure 6: Effects of varying i_1 (parameter controlling the percentage of the population who are at highest risk of infection due to poor health) by $\pm 25\%$ on epidemic trajectories. Solid curves represent the default values; dotted curves represent a 25% reduction in i_1 ; and the dashed curves represent a 25% increase in i_1 . Note that the results for varying i_1 are nearly identical to those varying ψ_1 (death rate for I_1 , the subgroup of the I population with lowest health levels; figure not shown).

tical to those observed from changing ψ_1 , the death rate for I_1 (figure not included).

6 Conclusions and Future Work

As shown by the preceding analysis, the establishment of three different subgroups of S and I can have critical impacts on the dynamics of an outbreak. Further, compartmentalizing an SIR model as we have demonstrated here allows for a more accurate and realistic representation of the more complex behaviors of populations when modeling this type of infectious disease, where human behavior and health levels play such a vital role in the outbreak dynamics. This analysis makes clear that behavioral actions at the population level provide an effective non-medical intervention to alter the course of a viral-driven catastrophe.

Severe social distancing was shown to be, on average 2.23 times more effective in our model at reducing deaths than minimal social distancing but also took the longest for the infection to die out, at an average of 654.5 days. Minimal social distancing, on the other hand, only required 198.5 days on average to reach 0 infections. A major difference, complimenting death rate as a metric proving the importance of social distancing, is the infection curve. Minimal social distancing produced a peak where an average of 66% of the initial population was si-

multaneously infected, occurring at 17 days. However, severe social distancing produced a much more manageable peak, where 3.7% percent of the population on average was simultaneously infected at 199 days.

Proper application of a holistic program to contain the spread of highly infectious outbreak offers a powerful tool in blunting the total damage from a rapidly unfolding viral pandemic. Models which clearly delineate the necessary measures needed to control viral spread while remaining easily understandable provide a powerful tool to both the general public and to public policy makers. This class of model also is quite applicable to decisions balancing the trade-off between economic devastation and excessive viral spread.

The results of our sensitivity analysis indicate the most productive interventions include those that might reduce the proportion of the population with highest social contact or, equivalently, reducing the infection rate of that population, and also maximizing the health of the general population. Minimizing the percentage of society that falls into the highest social contact categories is essential for reducing the spread of COVID-19, minimizing the fatalities, and blunting the infection peak. Scenarios in which s_1 is allowed to become a majority may result in shorter outbreak times, but they also result in considerably higher number of fatalities and large percentages of the population being simultaneously infected, increasing the likelihood of overwhelming hospitals. [12].

Further investigation into the sensitivity of the model to level of adherence with guidelines will allow selecting a more practical balance between guidelines effective in controlling viral spread and guidelines likely to be adopted by a large sector of the population. Additional analysis into the sensitivity of the model to the initial population of infected persons could allow insight into the likelihood of a second wave of COVID-19 infections arising after social distancing regulations are relaxed. The development of a network-based compartmental model, such as in the Netlogo model Infection On the NeTWorks [7] could also offer a different means of analysis and insight into the problem at hand.

We are additionally interested in creating a SIRS model with classes within each compartment for different portions of the population. It is currently unknown whether or not COVID-19 patients become immune for long periods of time after recovering. This result could have far-reaching consequences for the epidemic dynamics both in our model and in real-life applications. The more heterogeneous approach we took to modeling disease spread in a population could potentially also be applied to modeling herd immunity, via vastly decreasing the transmission probabilities for varying percentages of the studied population.

Lastly, we are interested in using our more heteroge-

neous model to study the effects of relaxing social distancing orders at varying points in time. This work will be more meaningful, however, once this type of data becomes available as various regions around the world begin to reopen their economies. This could ultimately allow for more insight into the optimal time to re-open communities, allowing both for the personal safety of citizens, and the economic survival of various industries.

Author Contributions

A.S. conceived the conceptual idea and conducted research to find model parameters. H.H. helped in model refinement and wrote MATLAB code necessary for model simulations. Both A.S. and H.H. performed analytical simulations. A.S. and H.H. contributed equally to writing the manuscript.

References

- [1] Akman, O., Corby, M. R., & Schaefer, E. (2016). Examination of models for cholera: insights into model comparison methods. *Letters in Biomathematics*, 3(1), 93–118.
- [2] CIA. (2013). The CIA World Factbook 2013.
- [3] Clamer, V., et al. (2016). Estimating transmission probability in schools for the 2009 H1N1 influenza pandemic in Italy. *Theor. Biol. Med. Model.*, 13(1), 19.
- [4] Fisher, D., & Wilder-Smith, A. (2020). The global community needs to swiftly ramp up the response to contain COVID-19, *The Lancet*, 395(10230), 1109–1110.
- [5] General Office of Hubei Provincial People's Government. (2020). China demands unremitting containment efforts as Wuhan lockdown lifted, http://www.china.org.cn/china/2020-04/08/content_75906281.htm Accessed: April 10, 2020.
- [6] Horney, J.A., et al. (2010). Intent to receive pandemic influenza a (H1N1) vaccine, compliance with social distancing and sources of information in NC, 2009. *PLoS ONE*, 5(6), e11226.
- [7] Just, W., Callender, H., & Drew Lamar, M. (2015). Exploring distances with IONTW. <https://qubeshub.org/resources/742/download/ModuleDMQ.pdf> Accessed: April 10, 2020.

- [8] Kermack, W.O., & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772), 700–721.
- [9] Liu, K., et al. (2020). Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection*, 80(6), E14–E18. [https://www.journalofinfection.com/article/S0163-4453\(20\)30116-X/fulltext](https://www.journalofinfection.com/article/S0163-4453(20)30116-X/fulltext) Accessed: June 2, 2020.
- [10] Marino, S., et al. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1), 178–196.
- [11] Montesclaros J.M., & Luis, P. (2020). Beyond COVID-19: Global Priorities Against Future Contagion. *RSIS Commentaries*, 030(20).
- [12] Murray, C. J. (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *medRxiv*.
- [13] Muto, K. et al. (2020). Japanese citizens' behavioral changes and preparedness against COVID-19: How effective is Japan's approach of self-restraint? *medRxiv*.
- [14] Patel, A. et al. (2020). Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak - United States, December 31, 2019-February 4, 2020. *Morbidity and Mortality Weekly Report*, 69(5), 140.
- [15] Roberts, M. et al. (2015). Nine challenges for deterministic epidemic models. *Epidemics*, 10, 49–53.
- [16] Small, M., Tse, C.K., & Walker, D.M. Super-spreaders and the rate of transmission of the SARS virus. *Physica D: Nonlinear Phenomena*, 215(2), 146–158.
- [17] Zhou, F., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395, 1054–1062.