Extension of the SEIR Epidemic Model through an Age-Stratified Stochastic Individual Agent Model for COVID-19

Hannah Phan, Jessica Tian, Nina Uzoigwe, Lan Zhang

Abstract

This project aims to extend the SEIR model for COVID-19 through a stochastic individual agent model that incorporates undetected (U) and fatal (F) populations to better capture the nuances of the virus' effect on different communities. This model also stratifies by age to account for the role age plays in population movement, immunity, symptom development, and prognosis, leading to differing exposure, infection, and mortality rates across age groups. Our results suggest that elderly populations are significantly more susceptible to developing symptoms after exposure to COVID-19 compared to each younger age group and are less likely to recover after infection (fatality instead of recovery).

1 Introduction

The spread of infectious disease is a complex process influenced by societal, economic, environmental, biological, and behavioural factors. Given the detrimental impact of these illnesses, it comes at no surprise that mathematical models are used to understand the dynamics of their evolution in a population over a given time period to better prevent, minimize, and study its impact. However, it is important to note that the use of complex models will involve a large number of estimated parameters, affecting model accuracy and conclusiveness due to the lack of comprehensive data and difficulty of determining true population parameters to supply to a stochastic model.

Fortunately, compartmentalization is an essential technique that can be used to divide a population into simplified and workable compartments, where every individual in the same compartment is assumed to share a certain characteristic. These compartments are then investigated using deterministic and stochastic ordinary differential equations. One mathematical model that uses this technique is the susceptible-exposed-infected-removed (SEIR) model. Here a population is divided into four basic homogeneous compartments: (1) An individual is susceptible (S), (2) An individual is exposed to the disease but not yet infectious (E), (3) An individual is infectious (I), or an individual has been removed through recovery or death (R). This model has been used for a variety of human infections and many variations have been implemented to improve its representation of the evolution of the illness in question.

The goal of this report is two-fold. First, we plan to extend the SEIR stochastic individual agent model to include an undetected population (U), which are the individuals within the population who are asymptomatic but can spread the virus and can only recover, and a fatality population (F), which are individuals that have died. This is to better capture the complexity of COVID-19, specifically the ability for asymptomatic carriers infect and recover at different rates and intensities compared to symptomatic carriers, who sometimes do not recover. Second, we plan to stratify our stochastic model by age, as previous studies have identified age to be a critical distinguishing factor in varying exposure, immunity, and infectiousness parameter values due to biological and behavioral conditions corresponding to each age group[1]. Age stratification will help to construct more insightful models to study the impact COVID-19 has had and

[1]Anca Rădulescu, et. al, 2020, "Management strategies in a SEIR-type model of COVID 19 community spread", https://www.nature.com/articles/s41598-020-77628-4.pdf

will have on the community types that comprise the global population. Understanding these differences can improve and guide our tailored responses to future pandemics in terms of policy-making and vaccine distribution.

This paper is organized as follows: in Section 2, we will propose a SEIURF compartmental model with stratification of parameters according to age. Dynamic analysis of the current model will be examined in Section 3, our findings from this model along with sensitivity analysis of our parameters will be discussed in Section 4, with a conclusion following in Section 5.

2 Description of Model

2.1 Extended SEIR (SEIURF) Individual Agent Model

We chose to pursue a stochastic individual agent model (IAM) to observe the natural evolution of a population impacted by COVID-19 under certain parameters over time. Furthermore, we looked to identify any potential population patterns that may arise from the randomness of individual interactions that would have been excluded in a deterministic model. The stochastic individual agent model is also more flexible in that it is easier to add and manipulate the modeling environment to test our hypotheses as well as scale the complexity of each of the individual agents' behavioural traits.

In the model, the entire population begins in the susceptible (S) population. Initially, a random agent in the population is infected. Then for each timestep until the exposed (E), infected/symptomatic (I), and undetected/asymptomatic (U) populations are 0, we iterate through each individual in the population. If a given individual is a member of the exposed, infected, or undetected groups, we randomly generate a number of contacts for the individual according to the Poisson distribution with different rates λ depending on whether the given individual is symptomatic or asymptomatic. Each of the randomly generated contacts enters the exposed population with probability `p_e`. We also keep track of and update the amount of time each individual who is exposed, infected, or undetected has left in that state at each timestep. With probability `p_i`, an exposed individual becomes infected otherwise undetected, with probability `p_f`, an infected person dies or otherwise recovers, and all undetected individuals recover. A flowchart outlining the movement through the different states of the model can be found in Figures 6 and 7 of the Appendix.

In the case of our SEIURF model, the stochastic IAM allowed us to partition the rate of contacts per day into rates λ_1 and λ_2 for exposed as well as asymptomatic people (`nc1_lam`), and symptomatic infected people (`nc2_lam`), respectively. We assume $\lambda_1 > \lambda_2$, where $\lambda_1 = 1.665$ and $\lambda_2 = 1$, based on our hypothesis that depending on an individual's personal knowledge whether they had been infected with COVID-19, they would self-impose restrictions on contact with other people, the community would limit their contact rates, and the degree of illness would physically prevent an individual from contacting people. The Poisson distribution was selected to model the number of contacts of an individual agent due to its discrete output and versatility in capturing independent, randomly occurring events.

We moved the exposed population into the symptomatic (I) and asymptomatic (U) compartments after a set incubation period (`incubate_time`) of 2 days. Individuals would become exposed after a contagious contact with an exposed, asymptomatic, or symptomatic individual with probability of `p_e = 0.201`. Exposed individuals would either become symptomatic/infected with a probability of `p_i = 0.484`, and asymptomatic/undetected with a probability of `p_u = 1 - p_i = 0.516`. We chose a higher `p_u` value than a `p_i` because we assumed, given previous studies[2], that more carriers of COVID-19 experience mild to

[2]Balabdaoui, Mohr, 2020, "Age-stratifed discrete compartment model of the COVID-19 epidemic with application to Switzerland", https://www.nature.com/articles/s41598-020-77420-4.pdf

undetectable symptoms as opposed to severe ones and that the data on undetected carriers are incomplete due to the difficulty in identifying the asymptomatic population. Then, symptomatic individuals would die with a probability of `p_f = 2.66%`, otherwise they recover after a given countdown period. Here, we assume that only symptomatic individuals can experience severe enough symptoms to become fatal. Finally, we set the number of days an infected or undetected person is contagious for to be uniformally distributed between a minimum value of `min_infect` = 5 days and maximum value of `max_infect` = 15 days. We chose these parameter values based on the data published by the CDC[3]. Values for the parameters in our baseline model were found by taking the weighted average of the corresponding parameters for each age group in Table 1, and they were then applied onto the entire population. The corresponding deterministic version of the non-age-stratified model is available in section 7.1 of the Appendix.

2.2 Stratification by Age Group

We then chose to stratify this model by age due to the overwhelming evidence that people of different ages respond to COVID-19 in distinct ways physiologically (e.g. how severe symptoms are) and socially (e.g. social-distancing policies). We segment the population according to the following age brackets: Children (`C`, 0–18 years of age), Young adults (`Y`, 19–34 years of age), Adults (`A`, 35–64 years) and Elderly (`E`, 65+ years), assigning each individual in our population vector an age based on the United States population distribution by age. Children (`C`) compose 25% of the population, young adults (`Y`) compose 22% of the population, adults (`A`) compose 39% of the population, and the elderly (`E`) compose the remaining 14% of the population[4]. While there is some overlap and similarities in responses across age groups, in general, each age group has unique epidemiological profiles, as the severity of symptoms and mortality rates vary based on age. Additionally, members of varying age groups interact with different numbers of individuals per day[5]. Thus, we assign separate values for the parameters `p_e` (probability of being exposed), `p_i` (probability of being symptomatic), `p_f` (mortality rate), and `nc1_lam` (expected number of contacts when exposed/asymptomatic) by age group as follows:

Table 1: Parameters by Age Group

Age Group	`p_e`	`p_i`	`p_f`	`nc1_lam`
C	0.1 (30%), 0.3 (70%)	0.2	Unif$(0, 0.002)$	2
Y	0.1 (40%), 0.3 (60%)	0.4	Unif$(0.002, 0.003)$	2
A	0.1 (60%), 0.3 (40%)	0.6	Unif$(0.003, 0.036)$	1.5
E	0.1 (70%), 0.3 (30%)	0.8	Unif$(0.06, 0.2)$	1

We chose the above age-stratified parameters based largely on the parameter values in Rădulescu et al. We chose two different values for `p_e`, 0.1 for probability of exposure when socially distancing and 0.3 for probability of exposure otherwise, varying the distribution of `p_e` based on age group due to different social behaviors and levels of maturity across the brackets. Different `p_i` values were chosen based on varying rates of asymptomatic incidence across age groups. Similarly, different `p_f` values were chosen based on varying mortality rates from COVID-19 between different age groups. Finally, different `nc1_lam` values were chosen since different age groups come into contact with varying expected number of individuals per time unit (1 day) due to diverse educational, social/familial, and occupational responsibilities. Note that `nc2_lam` is set to 1 across all age groups.

[3]Centers for Disease Control and Prevention, Feb. 23, 2021, "Interim Guidance on Duration of Isolation and Precautions for Adults with COVID-19"

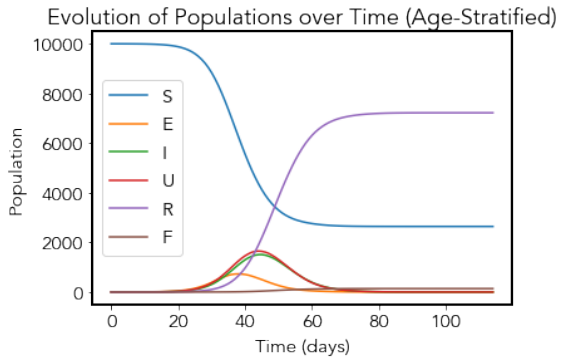
[4]https://www.kff.org/other/state-indicator/distribution-by-age

[5]Anca Rădulescu, et. al, 2020, "Management strategies in a SEIR-type model of COVID 19 community spread", https://www.nature.com/articles/s41598-020-77628-4.pdf

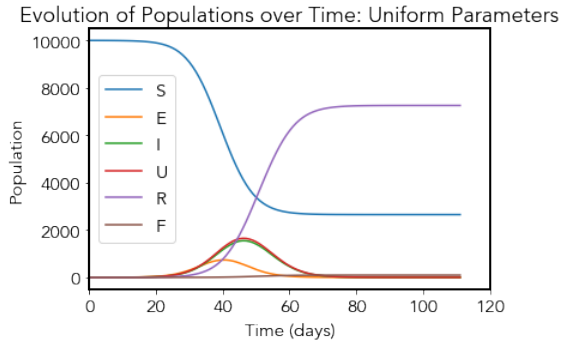
3 Results and Analysis

3.1 Age-Stratified Results

Figure 1: System Dynamics (averaged over 100 realizations)



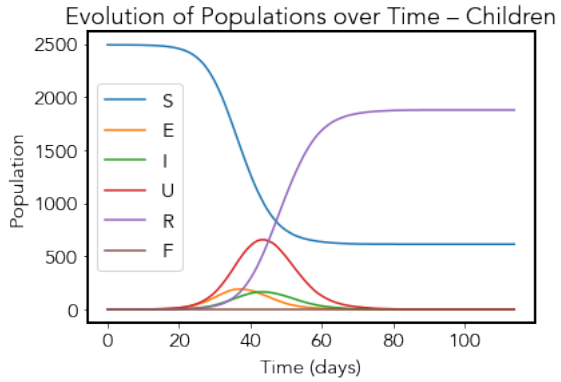
(a) Using Age Specific Parameters

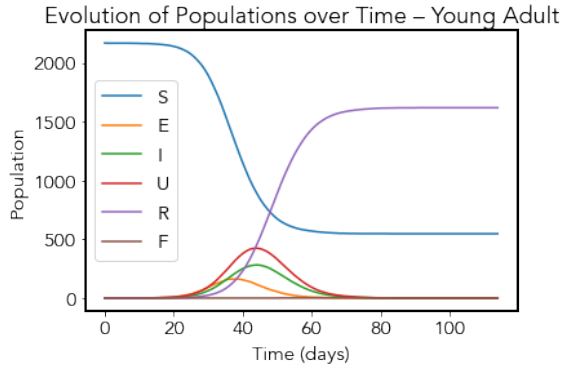


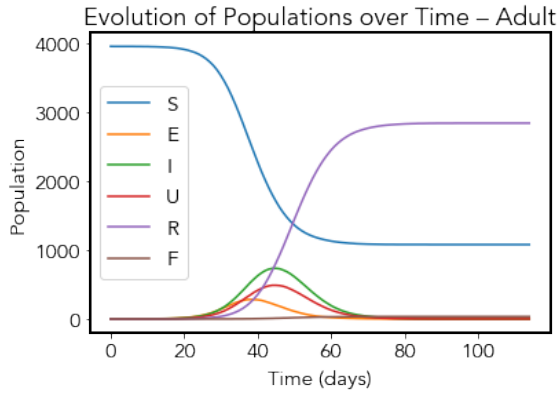
(b) Using Uniform Parameters

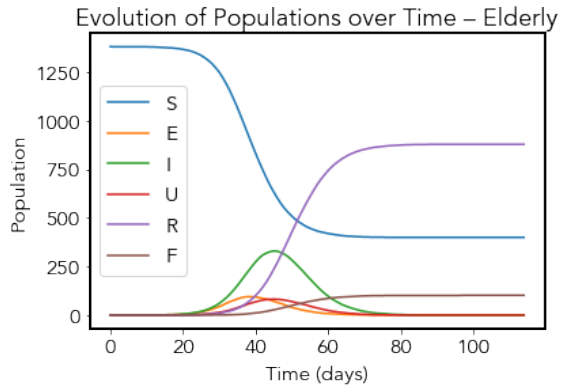
We can see from the above two plots that when we assign each individual to an age group and use age-specific parameters when moving the agent through our SEIURF model, we observe very similar system dynamics to our baseline SEIURF model where we use uniform parameters across all individuals. However, upon plotting system dynamics by age group, shown in Figure 2 below, we are able to better isolate the effect of age with respect to the impact of COVID-19 epidemic on a population.

Figure 2: System Dynamics by Age Group (averaged over 100 realizations)









From Figure 2, we observe that across age categories, the two curves that are most distinct amongst the groups are the U (undetected/asymptomatic) and I (infected/symptomatic) curves. The children age

group (0-18 years) experiences a significantly larger number of asymptomatic cases over symptomatic ones, as evidenced by the steeper U curve in red and the flatter I curve in green. The young adult population (19-34 years) behaves similarly with a higher number of asymptomatic cases rather than symptomatic ones, however the difference in cases between the two compartments is less than that of the children age group. On the other hand, the adult population (35-64 years) experiences a larger number of symptomatic rather than asymptomatic cases, and this difference is even more stark in the elderly population (65+) whose infected/symptomatic curve I in green is significantly steeper than the undetected/asymptomatic curve U in red. We also observe that the curve corresponding to fatalities F in brown plateaus at a number corresponding to a larger fraction of the age group at the end of the epidemic, aligning with the higher mortality rate of COVID-19 in elderly patients.

Figure 3: Evolution & Composition of S,E,I,U,R,F Curves by Age Group (averaged over 100 realizations)

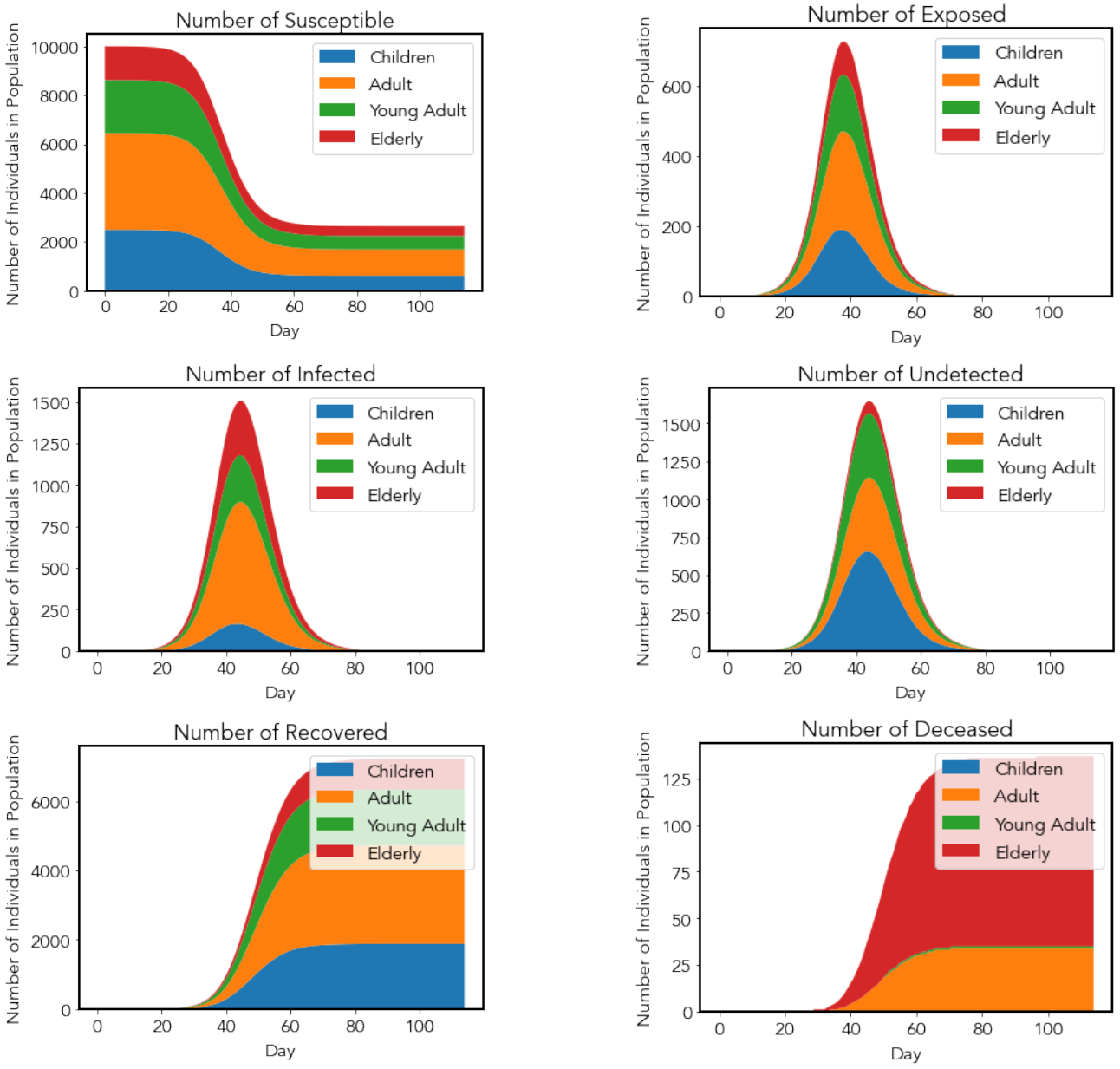


Figure 3 further highlights the composition of each curve in our SEIURF model by age group. We observe that the infected curve I is largely composed of the adult and elderly age groups, and that the

undetected/asymptomatic curve U is largely composed of the children and young adult age groups. The deceased curve F is nearly entirely composed of the adult and elderly age groups, with a few cases from the young adult group.

3.2 Sensitivity Analysis Using Uniform Parameters

Figure 4: Effect of Social Distancing in Varying λ_1

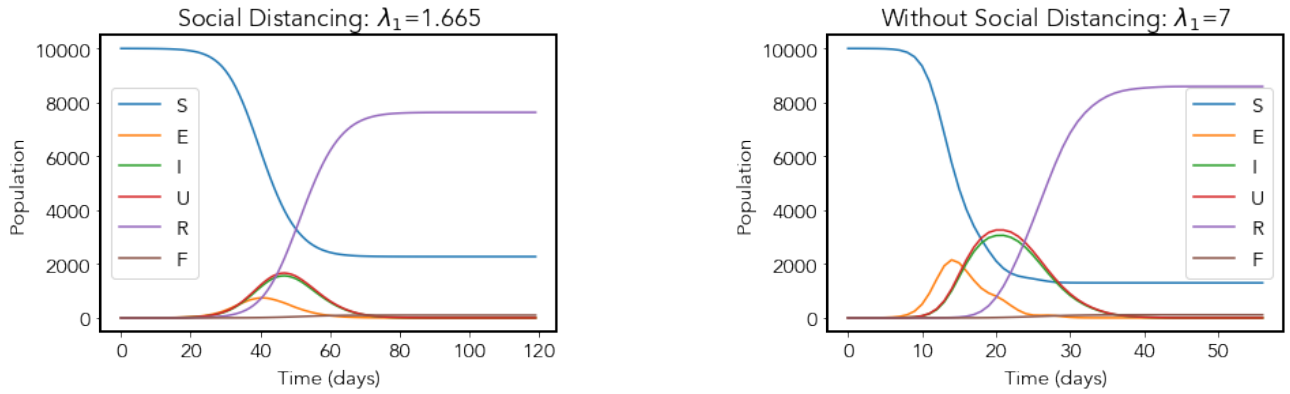


Table 2 highlights the changes in the infection I and fatality F curves, averaged over 100 realizations, given changes in `p_e` and `p_i` values, specifically the average time in which a given population peaks and the corresponding population value over 100 realizations, with all other parameters held constant.[6] `max_time(I)` and `max_time(F)` correspond to the time at which the infected and fatality populations, respectively, reached their peak in our model, and `max_pop(I)` and `max_pop(F)` corresponds to the population of the infected and fatality populations, respectively, at that time. Note that we especially focus on symptomatic individuals (I population) since much of the data, particularly early data, on COVID-19 infection rates captured only this demographic.[7]

As we observe in Table 2, when `p_e` (the probability of becoming exposed after a contagious contact) increases from 0.1 to 0.5, the infected and fatality populations peak at increasingly greater values and peak faster. In other words, the populations experience a higher infection rate and higher fatality rate. When `p_i` (the probability an exposed individual becomes symptomatic) increases from 0.1 to 0.5, the I and F populations peak at greater values, but I continues to peak at around 45 days each time. In Figure 4, averaging across 100 realizations, we observe a similar trend when `nc1_lam` increases from 1.665, as set in our baseline model, to 7. With a higher average number of contacts among individual agents, the exposed population increases quickly earlier on, hitting its peak at around day 13, as opposed to around day 40. As the result, we observe a higher infection rate.

4 Discussion

The sensitivity analysis of parameters `nc1_lam` (i.e., λ_1), `p_e`, `p_i`, and `p_f` is useful in estimating the effect of specific measures on slowing the spread of disease. For example, understanding the relationship of each parameter with the evolution of the S, E, I, U, R, and F populations can enable policymakers to develop

[6] Animations are available here

[7] We expect the evolution of the asymptomatic population (U) to be similar to that of the symptomatic population (I), assuming the ratio of asymptomatic to symptomatic individuals stays relatively constant. Within our model, this ratio would be $\frac{\texttt{p\_u}}{\texttt{p\_i}}$.

Figure 5: Sensitivity Analysis Varying p_e, p_i, and p_f

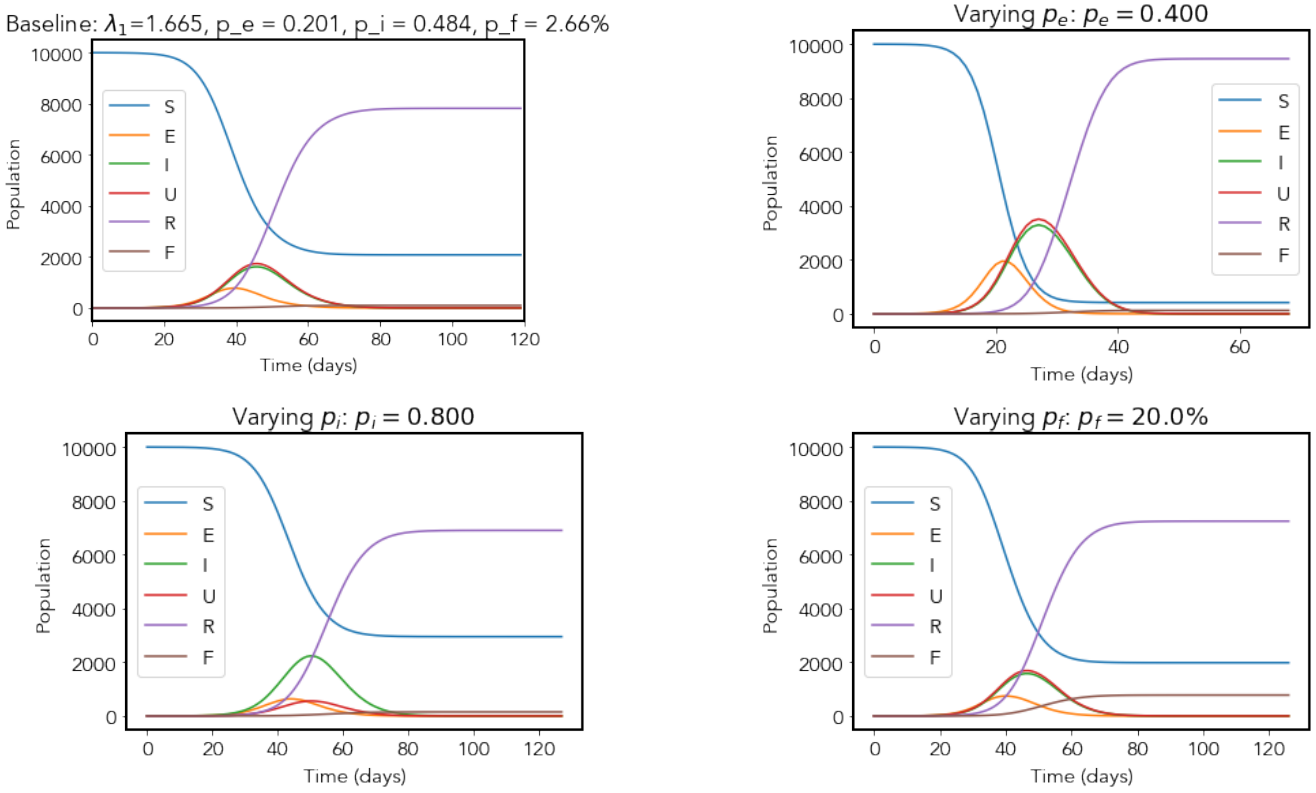


Table 2: Sensitivity Analysis

p_e = 0.1	p_e = 0.2	p_e = 0.3	p_e = 0.4	p_e = 0.5
max_pop(I) = 178.18; max_time(I) = 103; max_pop(F) = 32.02; max_time(F) = 221	max_pop(I) = 1479.11; max_time(I) = 47; max_pop(F) = 96.45; max_time(F) = 95	max_pop(I) = 2455.58; max_time(I) = 33; max_pop(F) = 114.18; max_time(F) = 66	max_pop(I) = 3294.79; max_time(I) = 27; max_pop(F) = 124.55; max_time(F) = 54	max_pop(I) = 3511.85; max_time(I) = 24; max_pop(F) = 123.52; max_time(F) = 48
p_i = 0.1	**p_i = 0.2**	**p_i = 0.3**	**p_i = 0.4**	**p_i = 0.5**
max_pop(I) = 353.89; max_time(I) = 43; max_pop(F) = 20.78; max_time(F) = 84	max_pop(I) = 654.95; max_time(I) = 45; max_pop(F) = 40.18; max_time(F) = 90	max_pop(I) = 964.39 ; max_time(I) = 44; max_pop(F) = 58.31; max_time(F) = 100	max_pop(I) = 1338.14; max_time(I) = 45; max_pop(F) = 81.89; max_time(F) = 102	max_pop(I) = 1668.24; max_time(I) = 45; max_pop(F) = 106.87; max_time(F) = 106

a more targeted plan (e.g. social distancing policies, mask mandates) in combating the epidemic. For instance, as seen in Figure 4, a higher `nc1_lam` results in a noticeably higher infection rate, suggesting that social distancing policies that encourage less potentially contagious contacts between individuals may be effective in containing an epidemic. Similarly, a higher `p_e` corresponds to a higher infection rate. By decreasing `p_e` values while keeping other parameters constant, we observe lower infection and fatality numbers in our model, which could inform policy makers of the efficacy of stringent mask mandate adher-

ence. Decreasing `p_i` values while keeping other parameters constant also appears to lower max infection numbers and speeds, which helps us more easily see how diseases of different degrees of infectiousness spread throughout a population. Finally, testing `p_f` sensitivity while keeping other parameters constant is important for understanding the impact of deadly diseases on population dynamics and composition, such as whether infection rates slow or whether surviving strains become less dangerous if infected people do not recover and are removed from the population. Additionally, `p_f` may vary depending on the efficacy of the health care system in a nation, region, or town, which has a variety of policy implications. A shortcoming of the sensitivity analysis is that it does not account for the possibility of reinfection after recovery, which may in turn increase `p_e` and `p_i`.

Analysis of our age-stratified stochastic individual agent SEIURF model allows us to understand how the epidemic affects each age group individually rather than the aggregated population as a whole, upon taking into account varying physiological responses to COVID-19 and social behaviors. These results can help inform current policy decisions necessary to control the spread of the COVID-19 epidemic; for instance, the larger number of undetected/asymptomatic cases for children and young adults suggests that policies like creating shopping hours for senior citizens at household essentials stores are necessary to limit the number of infections and fatalities in the elderly population. Additionally, these results can help inform vaccination strategies. Our age-stratified SEIURF individual agent model can be used to project how system dynamics will be impacted if the vaccine is distributed to certain age groups first, as we can simply change parameters associated with the vaccinated group to reflect their newly immune status.

Strengths & Limitations of Model:

Overall, by using an IAM, we are able to capture the characteristics of different groups in a population, providing a sense of realism in capturing the evolution of each population, which a deterministic model would lack. Furthermore, in the case of a deterministic model, one must assume that the population is large enough to be approximated as a real number, as opposed to an integer. However, in the case of our model, we are able to also capture what a small town may be able expect. Moreover, in running many realizations of our IAM, we were able to gain a better sense of the different outcomes that could play out, as the result of randomness. Our model was also able to evaluate the sensitivity of each parameter on the evolution of populations, and with this information, provide suggestions regarding what policies should be targeting (e.g. reduction of contacts between individuals with social distancing policies and reduction of the probability of becoming exposed with mask mandates). Finally, the IAM model may be generalized to capture the spread of other infectious diseases. That said, the primary limitation of the stochastic SEIURF model is that it is computationally expensive, with a time complexity of $O(n^2)$.

5 Conclusion

The spread of infectious disease is a complex process influenced by many factors within today's society, namely those that are economic, environmental, biological, and behavioural in nature. Through compartmentalization, as seen in our SEIURF model, these factors can be represented to observe the evolution of COVID-19. This model suggests that children that are aged 0-18 years and young adults that are aged 19-34 years experience higher numbers of asymptomatic cases than symptomatic ones, while the adults aged 35-64 years and elderly aged 65+ experience the vice versa. This model further suggests that the infected and deceased are largely composed of the adult and elderly populations. These results can ultimately help inform current policy decisions necessary to control the spread of the COVID-19 epidemic and help inform vaccination strategies across varying age groups to mitigate the spread of this illness.

When reflecting on how to expand this current model, there are several paths for future work that can

be taken. First and foremost, it would be of great interest to capture the different behaviors of the agents involved, more specifically their location and mobility. This can also be expanded to account for the movement of the agents from place to place, such as from home to work. This would allow for more accurate mapping of how the disease spreads within a population. Additionally, the deterministic model can be expanded. Currently, the run time for this model, explained in Appendix 7.2, poses a significant limitation and needs to be faster as well as less computationally expensive. By adhering to these characteristics, this would allow for further translation and application to larger populations with more stratified age groups.

6 Attribution of Effort

- Lan contributed to this project by conducting background research to help formulate the SEIURF model and the stratification by age group extension. She implemented code associated with age-group stratification to the baseline SEIURF model implemented by Jessica (`age_stratified_model.ipynb`) and generated relevant visualizations. She also helped write the paper and create the slide deck, largely focusing on sections related to age-group stratification, but contributed to the writeup of other portions of this project as well.
- Hannah contributed by conducting background research to find reasonable parameter values for the stochastic individual agent model. She attended office hours to discuss group questions for the project. She helped create the sensitivity analysis tables and graphs with the baseline SEIURF model implemented by Jessica (`IAM.ipynb`) and wrote the abstract, and parts of the introduction, description of model (section 2.1), and discussion sections. She also helped with creating the pitch deck.
- Nina contributed by conducting background research and finding publications on how to manipulate the standard SEIR model to reflect the complexity of disease evolution over a population. She also performed background research on the sensitivity analysis, exploring several options that would best reflect the sturdiness of the proposed model. She contributed to writing parts of the introduction as well as the conclusion and played a role in creating and presenting the slide deck associated with this paper, expanding on the background and motivation and the future works slides.
- Jessica contributed to this project by extending the SIR individual agent model code, as provided by David Wei. In particular, she developed the baseline SEIURF model (`IAM.ipynb`) of this paper, running the model on uniform parameters applied to the entire population. Within the paper, she contributed parts of to section 2.1 (Model) and section 4 (Discussion). Additionally, she developed the Appendix (flowchart and deterministic model). Finally, she helped with creating the slide deck, particularly the sections discussing the baseline SEIURF model's mechanics and limitations of the model.

7 APPENDIX

7.1 Flowcharts for the SEIURF Model

Figure 6: Different states of the SEIURF individual agent model in simulating the progress of epidemics in a human population

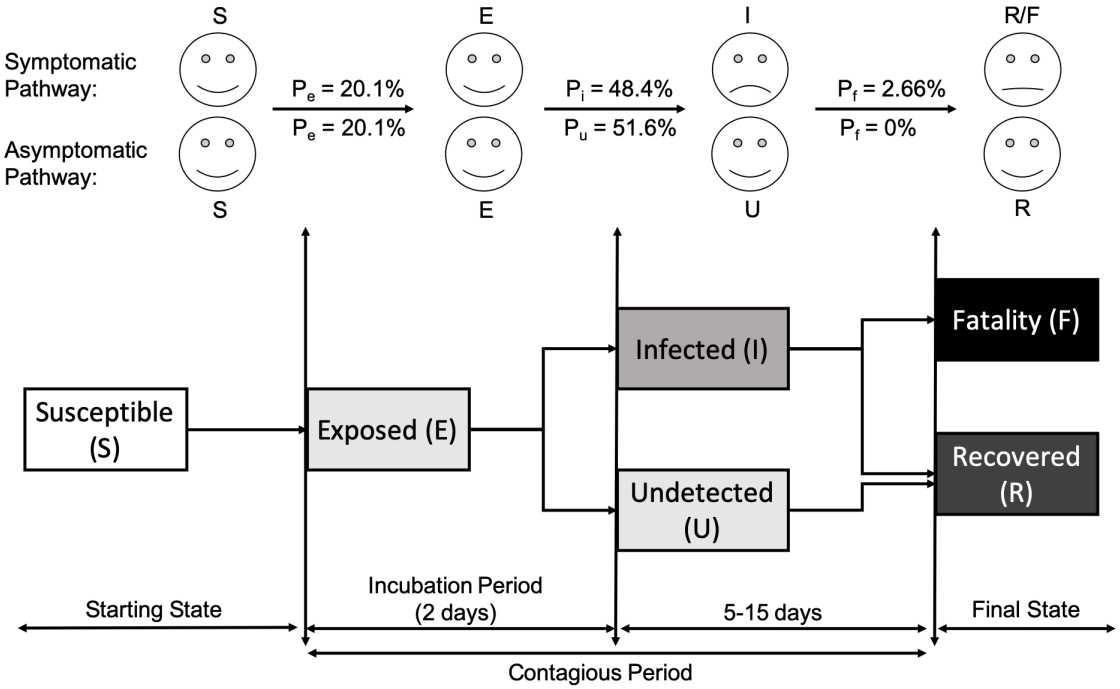
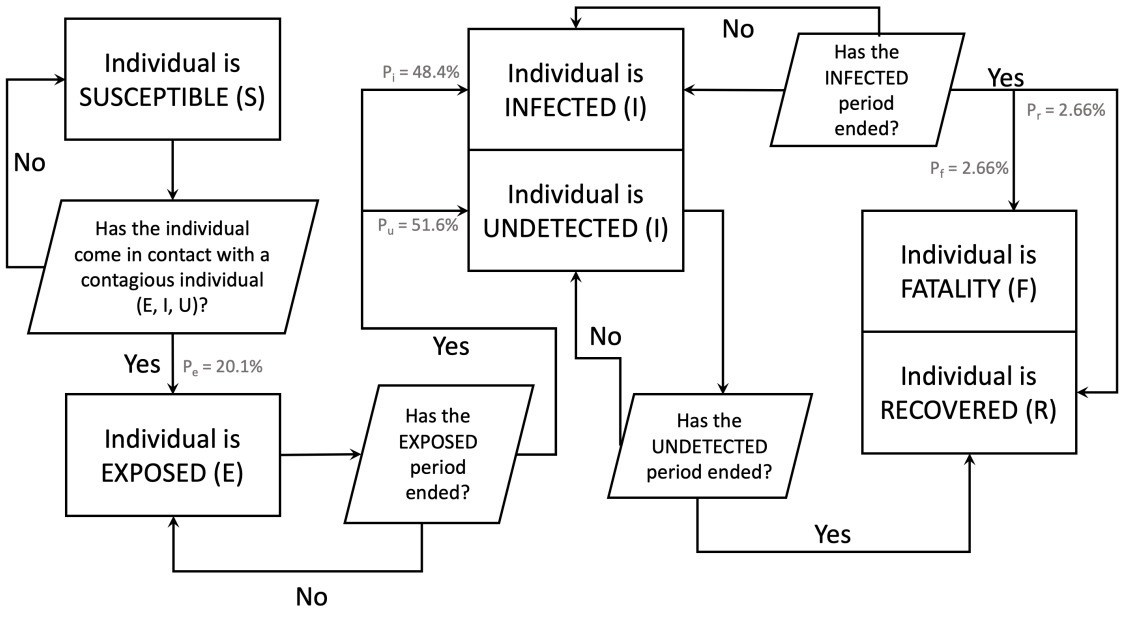


Figure 7: Flow diagram of states in the SEIURF individual agent model



7.2 Deterministic Model

The deterministic model, like our individual agent model, is an extension of the SEIR model. The population size N is divided into six classes: the susceptible $S(t)$, the exposed (in the latent period) $E(t)$, the symptomatic/infected $I(t)$, the asymptomatic/undetected $U(t)$, the recovered $R(t)$, and the deceased $F(t)$. The corresponding deterministic formulation of the individual agent model is as follows:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta_1 \frac{SI}{N} - \beta_2 \frac{S(U+E)}{N} \\
\frac{dE}{dt} &= \beta_1 \frac{SI}{N} + \beta_2 \frac{S(U+E)}{N} - E\delta_1 - E\delta_2 \\
\frac{dI}{dt} &= E\delta_1 - I\gamma_1 - If \\
\frac{dU}{dt} &= E\delta_2 - U\gamma_2 \\
\frac{dR}{dt} &= I\gamma_1 + U\gamma_2 \\
\frac{dF}{dt} &= If
\end{aligned} \tag{1}$$

The parameters β_1 and β_2 represent the the average number of contacts per person per time unit (1 day), multiplied by the probability that the virus is transmitted via a contact between a susceptible individual and a contagious individual. β_1 corresponds to the rate given the symptomatic population I, whereas β_2 corresponds to the rate given the population in the incubation period E and the asymptomatic, undetected population U. This model operates under the assumption that $\beta_2 < \beta_1$, as symptomatic individuals are more likely to change their behavior and reduce the number of contacts they make with other individuals in the population.

The parameters δ_1 and δ_2 refer to the rate at which exposed individuals, who are undergoing the incubation period, become symptomatic (I) or asymptomatic (U). The parameters γ_1 and γ_2 represent the rate of recovery for symptomatic, infected individuals (I) and asymptomatic, undetected individuals (U). Finally, the parameter f corresponds to the fatality rate. Here, our model assumes that only symptomatic ave a chance of dying.