Utilizing the SIR Model to Simulate the Spread of an Epidemic: An Agent-Based Modeling Approach

Thi Hoa Hien La, Joanna Rashid, Arnav Chandra Sarma
Georgia Institute of Technology, USA

Abstract: In light of recent global pandemics such as COVID-19, a thorough understanding of how quickly an epidemic can spread under various conditions remains critical**.** While both mathematical (differential equation-based) and computational (agent-based) models are commonly utilized in the simulation of large-scale epidemics, the agent-based modeling approach has proven to be more flexible, especially when there are varying input parameters involved. This paper demonstrates how to simulate the spread of an epidemic using the agent-based approach combined with the SIR model framework. Additionally, the effects of various parameters including the infection probability, length of the infectious period, and sample size on the epidemic outcomes were also measured. It was found that in all three cases, an increase in each parameter led to a higher number of daily and total infections, as well as a longer epidemic overall.

Index terms: agent-based, SIR, simulation, epidemic

I. Introduction

Despite numerous medical and technological breakthroughs made in recent years, global pandemics continue to bring long-lasting impacts to our society. Notable major epidemics that have occurred in history include the plague, cholera, influenza (flu), severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), etc. (Piret and Boivin [1]). Most recently, in 2019, a new type of coronavirus known as COVID-19 quickly spread across the globe and was estimated to have caused 640,395,651 infections and 6,618,579 deaths to date (WHO Covid-19 Dashboard [2]). The vast scale at which such diseases may impact our lives has propelled scientists and policymakers to further study their transmission capabilities in various settings, as well as consider a number of intervention strategies to mitigate their effects (Wang et al. [3], Aleman, Wibisono, and Schwartz [4]).

Given that these diseases have the potential to rapidly evolve and infect a large number of the human population, a thorough understanding of how quickly they can spread remains critical, especially when little information is known. In this regard, mathematical or computational models can be very effective (Ozmen et al. [5]) at simulating important measures such as the length of the epidemic, the number of people in the population that may get infected daily and over time, etc.

The first major modeling approach that has been popularized by existing literature employs a deterministic approach of differential equation-based models, or EBMs [5]. EBMs focus on modeling different segments of the population over the course of the disease, using the

transition rates from one segment (i.e., state) to another as a way of demonstrating how the disease evolves and eventually becomes extinct. EBMs are suited to large-scale modeling scenarios where a sizable population (e.g., a country) is at risk of infection and tend to require very little information as inputs [5]. While computationally economical, the results from EBMs can be overly simplistic. The deterministic nature of an EMB does not account for any of the randomnesses that take place in real-world disease spread. EMBs also lack information at the agent level. While well suited for large population sizes, these two limitations make EMBs a poor choice for small population sizes. Because an EMB is deterministic, the model produces a single value for epidemic length, prevalence, and incidence, rather than a more realistic interval of expected values.

Agent-based (or discrete-event) models are a methodology that addresses the limitations of EMBs. . Agent-based models, also known as ABMs [5], use a stochastic approach by simulating interactions between agents at the individual level. Such information can range from the person's characteristics, behaviors, and activities (e.g., work, school, exercise) to interactions with other individuals within the population (Auchincloss and Diez Roux [6], Van Wave, Scutchfield, and Honore [7]). ABMs account for the randomness of real-world disease spread by utilizing Bernoulli trials with a pre-determined probability of transmission for interaction between agents at every discrete time increment. ABM models, however, are not without criticisms: they are more complicated to develop than EBMs, produce "random" outputs that may necessitate using a more careful approach in analysis, and in many cases, can get computationally expensive quickly as the number of required trials increases (Wilson, Alabdulkarim, and Goldsman [8]).

Regardless of approaches, the SIR model (Kermack and McKendrick [9]) model is commonly used as a conceptual framework to simulate the spread of disease infection [5], especially in cases where the population size remains constant (Allen and Burgin [10]). As depicted in Figure 1 below, the SIR model divides the population into three main groups: Susceptible (S), Infected (I), and Recovered (R, sometimes also referred to as Removed). Over the course of the epidemic, an individual will move from S (not yet infected) to I (when they get exposed and infected) and finally to R (either recovered or dead and can no longer be infected again). In the simplest case, the transition is one-directional and linear; in other words, an individual cannot move from the infected back to the susceptible state. However, there are other variations of the SIR model where the individual does not gain immunity and therefore becomes susceptible again after recovery, such as the SIS model [10].

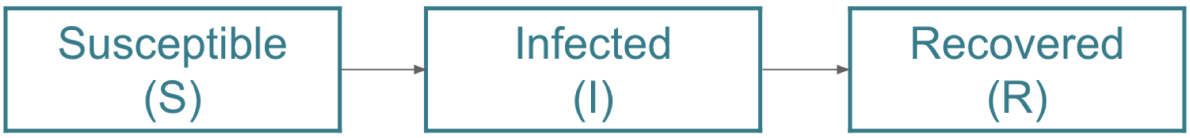


Fig 1. The compartmental states in the basic SIR model

Depending on the approach (EBM vs ABM), the SIR model can be adapted to consider the analysis from its respective angle [5]. In the EBM approach, the population's rate of change between each state is calculated based on the following deterministic equations:

Rate of change of the S (susceptible) group: $\frac{dS(t)}{dt} = -\beta S(t)I(t)$

Rate of change of the I (infected) group: $\frac{dI(t)}{dt} = \beta I(t)S(t) - \gamma I(t)$

Rate of change of the R (recovered) group: $\frac{dR(t)}{dt} = \gamma I(t)$

where S(t), I(t), and R(t) signify the number of individuals in each state at time t, and β and γ, respectively, represent the rates of transmission and recovery.

While the differential equation-based (EBM) SIR model is simple to implement and can capture large-scale disease spread with relatively little effort, the EBM approach raises a number of concerns among many critics. For example, the EBM's "top-down" approach means it is unable to capture individual-level information [5]. Such information can be especially important in impacting the outcomes of the epidemic. Furthermore, the EBM approach fails to account for randomness, limiting our capabilities to predict when there are more complicated scenarios [8]. Lastly, the EBM approach rightfully requires the number of individuals in each state to be a whole integer, but this is not always possible when the population is small and thus leads to poorer performance overall [8].

This paper primarily explores an example epidemic case for a small sample. Given the known limitations of the EBM approach for small populations, this paper employs a stochastic, agent-based simulation model. Since the SIR model assumes that there is no reinfection, the susceptible population will likely decrease over time, allowing the epidemic and the simulation to end after a reasonable length of time. This makes it computationally viable to conduct a large number of simulation trials, which yields more accurate results.

In the following Methods section, the problem in the simulation, along with any assumption that was made, will be described in detail. Each of the steps that were carried out in the experiment will also be explained. Next, the Results section covers the different analyses performed to understand the effect of different model parameters, including the probability of infection, length of infection period, and the size of the test sample on the epidemic outcomes. Finally, the Discussion section will consider the limitations of the current experiment, as well as recommend further possible research or revision ideas for future studies.

II. Methods

In the base version of the experiment, the test sample is a classroom of 21 elementary school kids (or individuals). On the first day, out of these 21 individuals, there is one infected individual who can infect others while the remaining 20 are susceptible. Infected individuals (including the first person) stay in the experiment and can infect others for t = 3 days. The

probability that an infected individual will infect any susceptible person is p = 0.02. All individuals and days are considered independent.

An initial comparison of a deterministic EBM and stochastic ABM using the R package, EpiModel (Jenness, et al.[13]) was conducted in order to confirm the selection of a stochastic ABM as the best approach for this experiment with a small population.

A number of assumptions were made in the process of running this version of the experiment. First, each susceptible individual will have one interaction a day with the infected group. Second, each interaction between one susceptible individual and the infected group is considered a Bernoulli (p) trial, with p equaling the probability that the susceptible person will become infected as a result of that interaction. Under the current experiment's setup, this probability is p = 0.02. In cases where there is more than one infected person on any given day, each susceptible person will still only have one probability indicating whether or not they will get infected associated with them on that day. As a result, within each day, if there are n susceptible persons, there will be n possible interactions in the sample. These n interactions are then modeled as n iid Bernoulli (p) trials.

Another assumption made during this experiment is that the overall population remains constant and no death occurs due to infection. Furthermore, there is no latent period in which exposed individuals remain noninfectious i.e., the SEIR model (Hethcote [11]); once an individual has been exposed, they will immediately be moved to the infected state starting the next day. Finally, each trial in the simulation ends when there are no more infected people who can spread the virus. In other words, once the last infected person has recovered from the virus, the epidemic has "ended" and the length of the epidemic will be measured accordingly.

Once the above assumptions had been defined, Python was utilized to build customized functions for the simulation trials. The first function, `flu_bern`, simulates the series of iid Bernoulli (p) trials. On the first day of the experiment, the first infected person was randomly picked and moved to the "Infected" state, leaving 20 remaining susceptible individuals in the `susceptible` set.

At the beginning of each day, a uniform random number between 0 and 1 was generated and assigned to each person in the `susceptible` set. Although such numbers were, in fact, pseudo-random numbers (PRNs), they appeared "random" and had a big enough cycle length such that we could carry out multiple "random" trials (Law [12]). The number of PRNs generated each day would be equal to the number of people in the "Susceptible" state on that day.

Since each susceptible person had their unique PRN associated with them each day, they also each had a different probability of becoming infected on that day. Whether each susceptible person would become infected on day X was treated as a single Bernoulli trial, with a probability of success p equalling the pre-determined probability of transmission (i.e p = 0.02). Therefore, if the generated PRN (U) associated with the person is less than p or 0.02 (U < p), the susceptible individual is considered to have become infected on that day and moved from

the “Susceptible” state to the “Infected” state starting the next day. In total, on any given day, we ran n iid Bernoulli (p) trials, with n equalling the number of people in the “Susceptible” state at the beginning of the day.

After the Bernoulli trials had run, the number of new infections for a given day, `num_daily_exposed`, was determined. Each person in the “Infected” state was also assigned a variable that records how many days they remain in this state; this information was stored in a dictionary named `infectious`. When a person was first infected, the value of this variable would equal the length of the infectious period chosen in the experiment i.e., d = 3 days. At the end of each day, we subtracted 1 day from this variable; if the new value equals 0, the “infected” individual would then be moved to the “Recovered” set. This logic effectively updated the number of individuals in the “Recovered” state. The number of daily total infected people was tallied, equalling the number of new infections plus the number of previously infected individuals with a non-zero value for the infectious period length. Conversely, the number of susceptible individuals would also be updated by removing the new infections from the “Susceptible” set, and 1 was added to the total number of days in the experiment. This entire process is detailed in the pseudo-code below:

```
susceptible = set([1,pop_size]) # set of integers ranging inclusive from 1 to pop_size
recovery = set([]) # empty set initially

# initialize infectious dictionary with the first_infected
# p is the probability of transmission (i.e 0.02)
# d is the infection period (i.e 3 days)
first_infected = randint([1,pop_size])
infectious = {first_infected: {‘p’: p, ‘d’:d}}
susceptible.remove(first_infected) # remove first_infected from susceptible

# Start of the epidemic! Initialize epidemic_timeseries to empty list
# epidemic_timeseries stores daily infection
day=0
epidemic_timseries=[]

while 0 < length(infectious) < pop_size:
       # Beginning of the day
       day = day + 1
       # for every person in susceptible, generate a PRN and store in sample array
       sample = rand.unif(size = length(susceptible))

       # Count number of PRNs such that U_i < p to be considered exposed and infected
       # Add that count to epidemic_timeseries list
       num_daily_exposed = sum(U_i<p for i in sample)
       epidemic_timeseries.append(num_daily_exposed)

       # Update infectious based on the following rules for every infected person

       # Decrement infected person’s infection time ‘d’ by one day
       # If infected person’s infection time reaches zero, that person recovered
       # Add infected person to recovery and remove recovered person from infectious
```

```
for k in infectious.keys(): # k denotes an infected person
       infectious[k]['d'] = infectious[k]['d'] - 1
       if infectious[k]['d'] == 0:
              recovery.add(k)
              delete(infectious[k])

# Randomly select a person from susceptible set now infected
# Remove them from susceptible and add their entries into infectious.
# Repeat this a number of times equal to num_daily_exposed
for i in [0,num_daily_exposed):
       inf_person = 0
       while inf_person not in susceptible:
              inf_person = randint(1,pop_size)
       infectious[inf_person] = {'p': p, 'd':d}
       susceptible.remove(inf_person)
# End of the day
```

The steps outlined above were repeated every day until there were no more individuals left in the "Infectious" state (i.e., they have recovered and can no longer infect other individuals). This effectively concluded one replication of the simulation. The trial outcomes - including the time series of the number of daily infections, the total number of infections, and how long the epidemic lasted - would be saved for further analysis. Another independent replication would then be carried out following all the same steps. Overall, we ran a total of one million independent replications for each set of experiment parameters using the function `run_trials`. With a custom function `run_experiment`, we generated a data frame containing all the trial data and outputs, such as the number of total infections and the epidemic length in each trial. From these metrics, we used another custom function, `basic_stats`, to analyze the distribution, the expected value, and important measures of the epidemic length including the minimum, maximum, standard deviation, 25th - 50th - 75th percentile, as well as the same metrics for the total number of infections during the epidemic. We also built a function called `plot_histogram` to generate histograms of these distributions for visualization purposes.

In addition to the base case described above, we also simulated a number of additional scenarios to understand their effects on epidemic outcomes. First, we modeled the impact of different transmission probabilities on the distribution of the number of daily infections over time, as well as the length of the epidemic. We tested 5 different values of infection probabilities, ranging from .01 to .75. The remaining model parameters were held constant, with the infectious period lasting 3 days and the total class size equalling 21 individuals. We hypothesize that the probability of infection will be positively correlated with the number of infections and the length of the epidemic.

Next, various lengths of infection were simulated to understand how a longer infectious period may contribute to the overall length of the epidemic, as well as the number of daily infections. Since the value tested in the base experiment for this metric was 3 days, values ranging from 1 to 5 days were modeled in these additional analyses. The probability of infection

was kept at p = 0.02 and the class size was 21. Given these parameters, we hypothesize that a longer infectious period will lead to a longer epidemic. However, it may not have a similar effect on the number of daily infections during the earlier days of the epidemic, but only during the later stage when the number of daily infections starts trending down. In other words, a longer infectious period may only result in a higher number of daily infections towards the second half of the epidemic.

Last, we simulated different class sizes to observe how having a larger sample would affect the same two key outcomes from other experiments: the number of daily infections and the length of the epidemic. Values were tested in increments of 10, starting from 10 and ending at 100. Other model parameters were kept consistent: the probability of infection p = 0.02 and length of the infectious period I = 3 days. We hypothesize that having a larger class size will result in a higher number of daily infections, which also contributes to a longer epidemic overall.

III. Results

1. EBM vs. ABM Comparison

A comparison of a deterministic EBM and stochastic ABM yielded significantly different results due to the small population size of this experiment (Fig 2). This is in alignment with the literature and confirmed the choice to use a stochastic ABM for this paper.

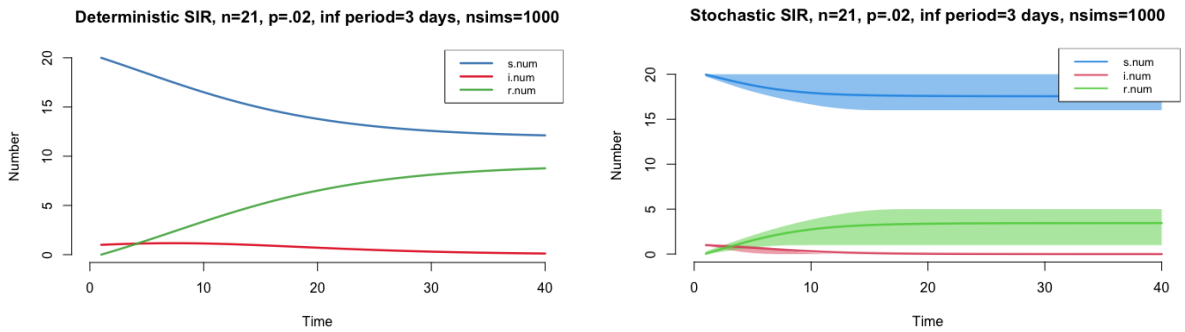


Fig 2. Initial Deterministic Model and Stochastic Model Comparison (EpiModel R package)

2. The base experiment

First, we conducted a total of 10,000 trials. The total infections over these 10,000 replications of the epidemic simulation are skewed heavily left (fig 3) with 75% of all simulations having a total of 3 infections or less. The values for epidemic length are also skewed left with 75% of all simulations being 8 days or less (fig 4).

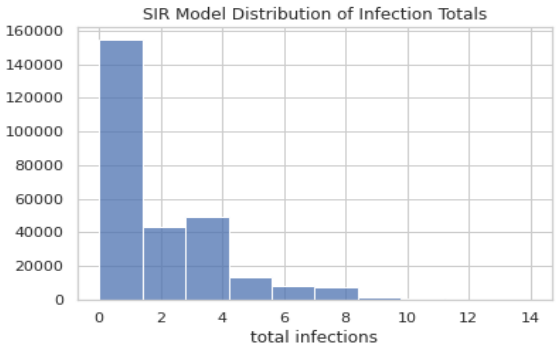


Fig 3. The Distribution of Total Infections

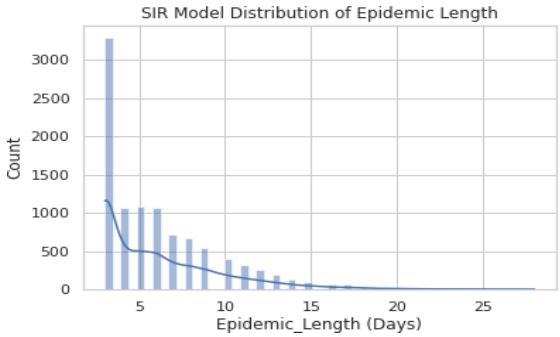


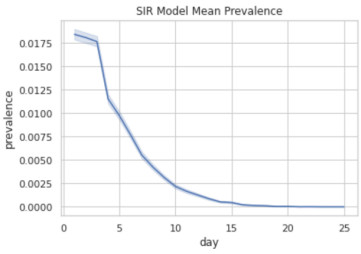
Fig 4. The Distribution of Epidemic Length

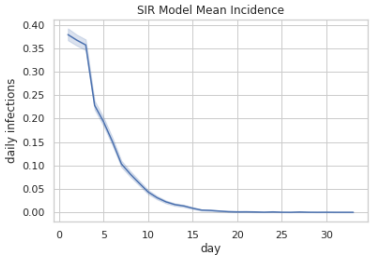
Over 10,000 replications of the initial simulation parameters, the mean infection total is 1.871 and the mean epidemic length is 6.198 days with only 2 or fewer students infected at a level of 95% confidence. All confidence intervals are calculated using the t-distribution. The confidence intervals of these expected values are very narrow given the high number of replications conducted (Table 1).

Table 1: Descriptive Statistics of Simulation Results (10,000 runs, n=21, p(infection) = .02, infection length = 3 days, number of initial infections = 1)

	susceptible	incidence (daily infections)	total infections	prevalence (proportion infected)	recovered	epidemic length (days)
mean (expected value)	19.129	0.063	1.871	0.003	1.682	6.198
std	2.066	0.281	2.066	0.013	2.021	3.613
min	6.000	0.000	0.000	0.000	0.000	3.000
25%	18.000	0.000	0.000	0.000	0.000	3.000
50%	20.000	0.000	1.000	0.000	1.000	5.000
75%	21.000	0.000	3.000	0.000	3.000	8.000
max	21.000	4.000	15.000	0.190	15.000	35.000
95% Confidence Interval	(19.126, 19.132)	(0.062, 0.064)	(1.868, 1.874)	(0.003, 0.003)	(1.679, 1.685)	(6.194, 6.202)

Further analysis of the epidemic timeseries shows that the expected disease incidence and prevalence both peak within the first two days of the epidemic and then drop sharply (figs. 5a and 5b).





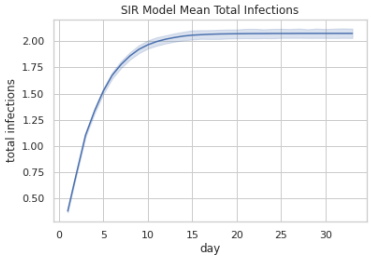


Fig 5a. Expected Prevalence Fig 5b. Expected Incidence Fig 5c. Expected Total Infections

The expected number of total infections on day 1 is .379, on day 2 .746, on day 3 1.103, and so on as shown in Fig 5c (refer to Table 3 in Appendix for a more detailed view). The distribution of total infections for each day is skewed heavily left with the distribution of values growing wider with each passing day in the epidemic.

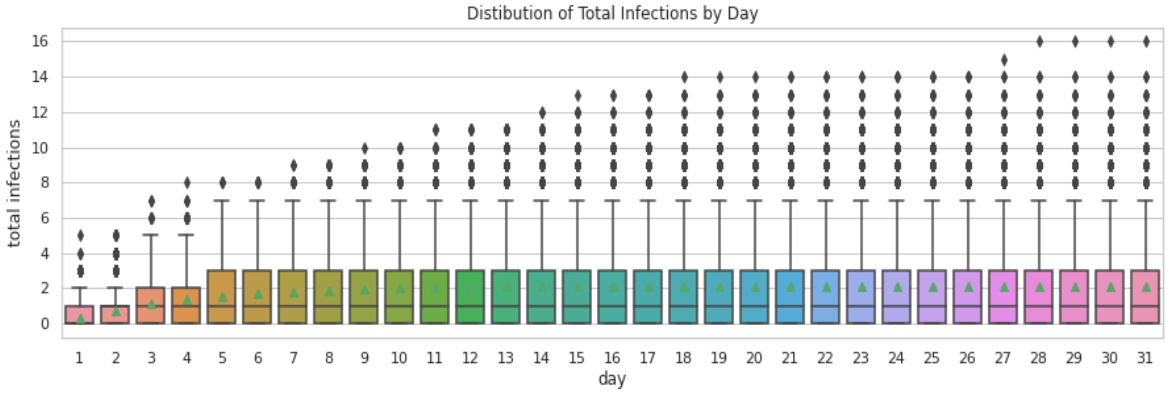


Fig 6. The Distribution of Total Infections Per Day
(Expected values are represented by a green triangle on each bar)

Next, we expanded the scope of the experiment to 1 million trials. For all 1 million trials, the number of daily infections was recorded for each day as each trial's epidemic varied in length. This amounted to a total of nearly 6.2 million days worth of daily infection data. The distribution of daily infections is shown below in Fig 7 and Table 2. We recorded zero daily infections 70.85% of the time, 1 daily infection 24.59%, 2 daily infections 4.09%, and 3+ infections taking up the remaining 0.43%. The maximum number of daily infections recorded was 6.

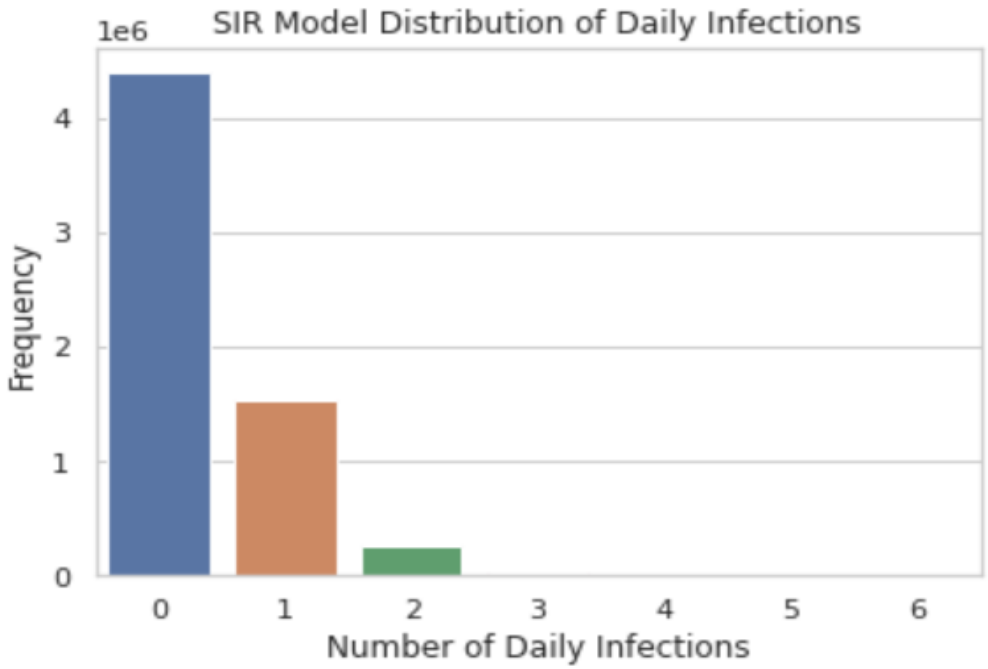


Fig 7. SIR Model Distribution of Daily Infections and their Frequencies (in millions)

Table 2: Daily Infections Simulation Results (1,000,000 runs, ~6.2 million days, n=21, p(infection) = .02, infection length = 3 days, number of initial infections = 1)		
Number of Daily Infections	Frequency	Percentage
0	4393165	70.85
1	1525021	24.59
2	253910	4.09
3	266380	0.43
4	1945	0.03
5	115	negligible
6	2	negligible

Some may think that the distribution of daily infections follows the binomial(n,p) distribution as each day consists of i.i.d Bernoulli(p) trials and that the convolution of all the Bernoulli trials' results is being performed every day. However, since the number of susceptible people can vary daily instead of being constant, that assertion is proven false.

3. Additional Analysis

Iterations on an array of infection probabilities, infectious period lengths, and sample sizes using the same simulation model illustrate the effect of each parameter. Epidemic length as a function of infection probability follows a parabola form that peaks at 15 days when the

infection probability is .1 per interaction between an infected and susceptible agent. Infectious period length and classroom sample size were both positively correlated with epidemic length over the range of values considered in this paper (across one to five-day infectious periods and 10 to 100 students respectively) (Fig 8).

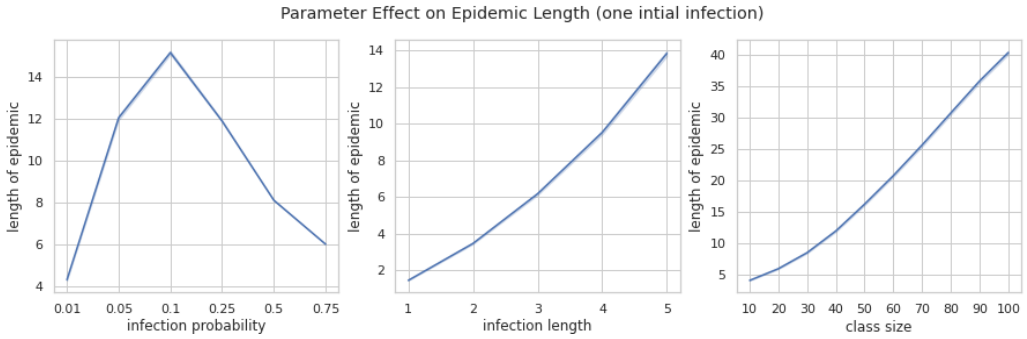


Fig 8. The Effects of Infection Probability, Infectious Length & Class Size on epidemic Length

As the infection probability values increase, the disease incidence spikes and falls sharply early in the epidemic. Epidemic lengths are widely distributed for low infection probabilities, with a more narrow distribution of values for low infection probabilities (figs 9 & 10). This suggests that this simulation model will produce more accurate expected values for extremely high or low infection probabilities. However, for mid-range infection probabilities, expected values will likely have larger confidence intervals and therefore be less accurate. In other words, the outcomes of epidemics with infection probabilities between .05 and .25 will be more difficult to accurately predict.

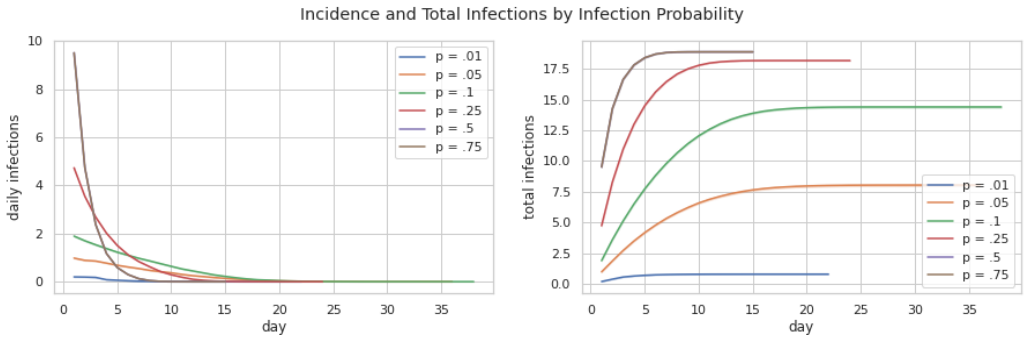


Fig 9. Incidence and Total Infections - By Infection Probability

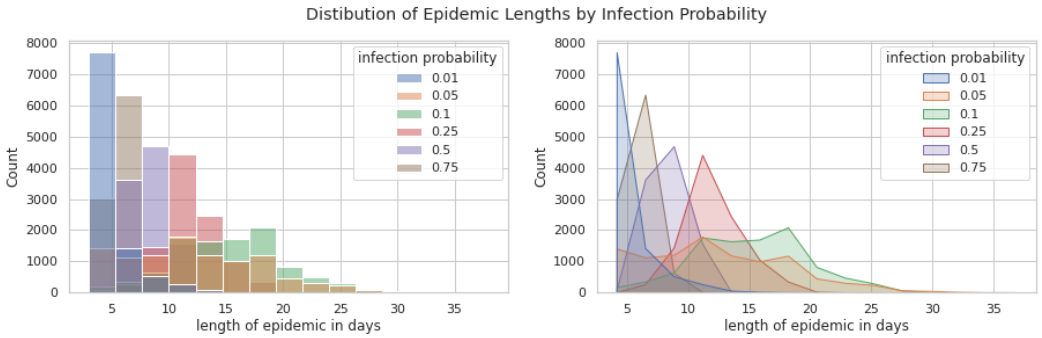


Fig 10. The Distribution of Epidemic Lengths - By Infection Probability

This paper considers a discrete range of infection lengths from one to five days. These values show that there is a clear and constant positive correlation between the number of infections and epidemic length (figs 11 & 12). The distribution of values for epidemic length is more narrow for shorter infectious periods. This implies that simulations of epidemics of diseases with shorter infectious period lengths will produce more accurate expected values with more narrow confidence intervals. Conversely, it is more difficult to obtain accurate expected values in the simulation of epidemics of disease with longer infectious periods.

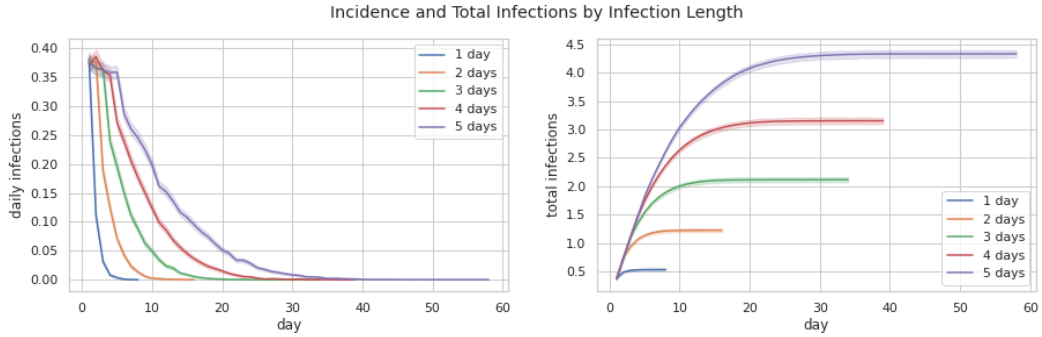


Fig 11. Incidence and Total Infections - By Length of Infection Period

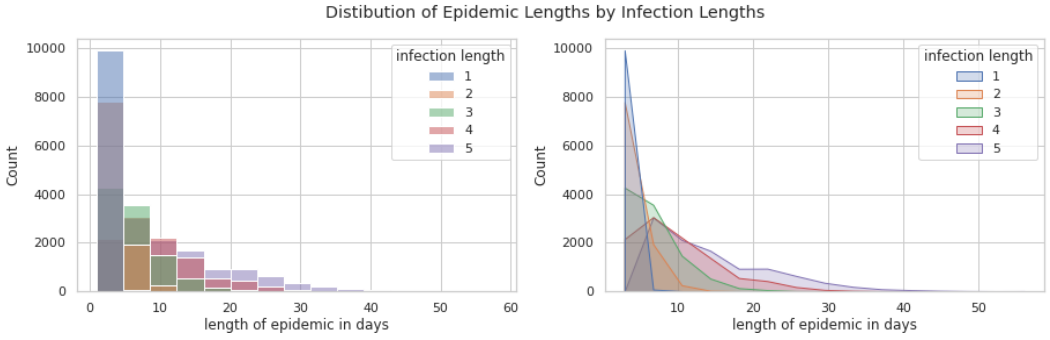


Fig 12. The Distribution of Epidemic Length - By Length of Infection Period

Classroom sample size also has a clear positive correlation with both disease incidence and epidemic length. The distribution of epidemic lengths follows a similar pattern as for infectious period lengths (figs 13 & 14). Epidemic lengths are more narrowly distributed for small

classroom samples, suggesting this simulation model will produce more accurate expected values and have narrow confidence intervals. Generally, epidemics in larger classroom populations will be not only longer in duration, but more difficult to predict.

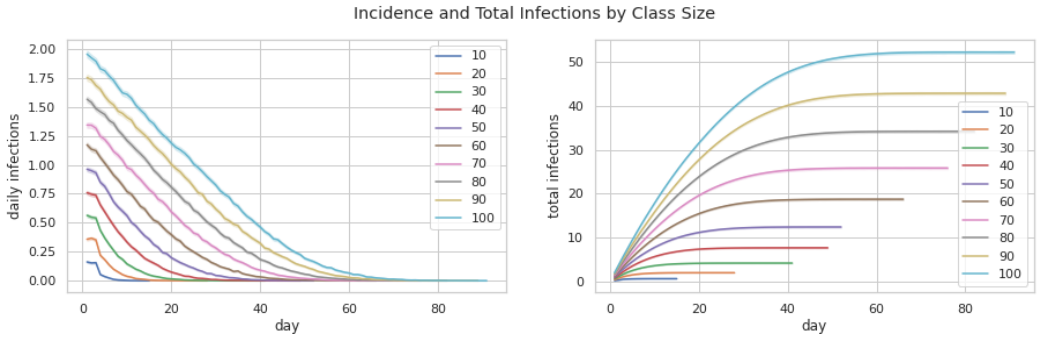


Fig 13. Incidence and Total Infections - By Class Size

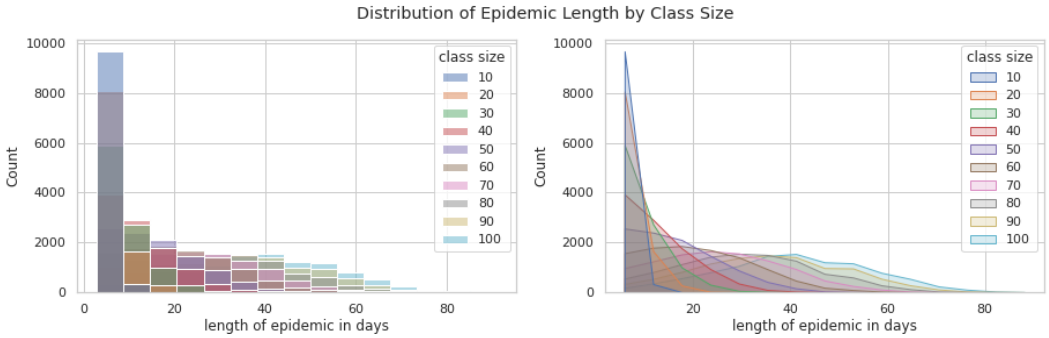


Fig 14. The Distribution of epidemic Length - By Class Size

IV. Discussion and Future Research

In this paper, we have shown how to conduct a simple epidemic spread simulation using the SIR model as the framework, as well as discussing the effects of different epidemic parameters such as the probability of infection, length of the infectious period, and sample size on the model outcomes. While not included in the current experiment, another analysis that could be relevant would be to measure the impact of having multiple infected persons (versus one) during each day of the epidemic. Our study assumes the probability that each susceptible person will get infected each day is represented by only a single Bernoulli (p) trial. However, if there are multiple (n) infected persons in the sample on a given day, there could be more than one interaction between each susceptible person and the infected group. As a result, there will be more opportunities for infection; if we assume that there are m possible interactions that could result in the susceptible person getting infected, there could be up to m Bernoulli (p) trials per susceptible person per day instead of only one in our experiment. It could be very feasible that having multiple infected persons in the sample may increase the chance of someone getting infected, thus increasing the number of infections daily and overall, as well as the length of the epidemic.

In addition to measuring the impacts of having multiple vs. one infected person in the sample, another analysis that may prove worthwhile is quantifying how the experiment outcomes may change if the epidemic progresses under the assumptions of the SIS model, instead of the SIR model. A key difference between the SIS and the SIR models is the addition of reinfection among the recovered group [10]. Introducing the possibility of reinfection to the experiment would increase the size of the susceptible sample, allowing the epidemic to progress for a longer period. While this would make the experiment computationally expensive to run, there are existing solutions to make the simulation more efficient. For example, instead of generating a new uniform random number for every single Bernoulli (p) trial which results in n Bernoulli (p) trials per day, a better method would be to utilize a single Bin (n, p) random variable, especially if np is “small” [8]. We could also consider reusing random numbers between trial runs, especially when the input parameters vary such as in our additional analyses above.

Furthermore, all of the experiments so far involved a constant probability of infection and infection period length. In a more complex simulation, the probability of transmitting or contracting a virus can also vary based on demographic factors, such as age groups. Additionally, people can remain infected for a range of days instead of a constant number of days. The results of an early attempt to conduct an experiment involving dynamic infection probabilities and infection periods can be found in Figs 15 through 25 (see Appendix). To determine whether a susceptible person would get infected, we selected the most contagious among those infected i.e., the person with the highest probability of transmitting the virus as the threshold. However, more research and experimentation would be required to understand how varying the probability of infection and infectious period length could impact the epidemic outcomes.

Last but not least, all experiments thus far assumed that all the interactions occur only in one place, namely a classroom. People interact with one another in many settings and venues, as well as indoors and outdoors. Therefore, the likelihood of contracting any virus can vary based on the place and the number of people in the vicinity.

V. References

1. J. Piret and G. Boivin, “Pandemics Throughout History,” Frontiers in Microbiology, vol. 11, no. 1, Jan. 2021, doi: 10.3389/fmicb.2020.631736.
2. WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: https://covid19.who.int/ (last cited: 12-03-2022).
3. Y. Wang, H. Xiong, S. Liu, A. Jung, T. Stone, and L. Chukoskie, “Simulation Based on an Agent-Based Model to Demonstrate the Transmission of COVID-19 and Effectiveness of Different Public Health Strategies,” Frontiers in Computer Science, vol. 3, Sep. 2021, doi: 10.3389/fcomp.2021.642321.
4. D. M. Aleman, T. G. Wibisono, and B. Schwartz, “A Nonhomogeneous Agent-Based Simulation Approach to Modeling the Spread of Disease in a Pandemic Outbreak,” Interfaces, vol. 41, no. 3, pp. 301–315, Jun. 2011, doi: 10.1287/inte.1100.0550.

5. Ö. Özmen, J. J. Nutaro, L. L. Pullum, and A. Ramanathan, “Analyzing the impact of modeling choices and assumptions in compartmental epidemiological models,” SIMULATION, vol. 92, no. 5, pp. 459–472, Apr. 2016, doi: 10.1177/0037549716640877.
6. A. H. Auchincloss and A. V. Diez Roux, “A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health,” American Journal of Epidemiology, vol. 168, no. 1, pp. 1–8, May 2008, doi: 10.1093/aje/kwn118.
7. T. W. Van Wave, F. D. Scutchfield, and P. A. Honoré, “Recent Advances in Public Health Systems Research in the United States,” Annual Review of Public Health, vol. 31, no. 1, pp. 283–295, Mar. 2010, doi: 10.1146/annurev.publhealth.012809.103550.
8. S. Wilson, A. Alabdulkarim, and D. Goldsman, “Green Simulation of Pandemic Disease Propagation,” Symmetry, vol. 11, no. 4, p. 580, Apr. 2019, doi: 10.3390/sym11040580.
9. W. O. Kermack and A. G. McKendrick, “A Contribution to the Mathematical Theory of Epidemics,” Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 115, no. 772, pp. 700–721, Aug. 1927, doi: 10.1098/rspa.1927.0118.
10. L. J. S. Allen and A. M. Burgin, “Comparison of deterministic and stochastic SIS and SIR models in discrete time,” Mathematical Biosciences, vol. 163, no. 1, pp. 1–33, Jan. 2000, doi: 10.1016/s0025-5564(99)00047-4.
11. H. W. Hethcote, “The Mathematics of Infectious Diseases,” SIAM Rev., vol. 42, no. 4, pp. 599–653, 2000, doi: 10.1137/s0036144500371907.
12. Law, A.M. Simulation Modeling and Analysis, 5th ed.; McGraw-Hill Education: New York, NY, USA, 2015.
13. Jenness, S. M., Goodreau, S. M., & Morris, M. (2018). EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks. Journal of statistical software, 84, 8

VI. Appendix

Table 3: Mean Values for Epidemic (10,000 runs, n=21, p(infection) = .02, infection length = 3 days, number of initial infections = 1)

day	susceptible	incidence (daily infections)	total infections	prevalence (proportion infected)	recovered
1	20.621	0.379	0.379	0.018	0.000
2	20.254	0.367	0.746	0.018	0.000
3	19.897	0.357	1.103	0.017	0.000
4	19.669	0.228	1.331	0.011	0.379
5	19.476	0.193	1.524	0.009	0.746
6	19.325	0.151	1.675	0.007	1.103
7	19.222	0.103	1.778	0.005	1.331
8	19.140	0.082	1.860	0.004	1.524
9	19.078	0.062	1.922	0.003	1.675

10	19.035	0.043	1.965	0.002	1.778
11	19.004	0.031	1.997	0.001	1.860
12	18.981	0.022	2.019	0.001	1.922
13	18.965	0.016	2.035	0.001	1.965
14	18.951	0.014	2.049	0.001	1.997
15	18.943	0.009	2.057	0.000	2.019
16	18.938	0.005	2.062	0.000	2.035
17	18.934	0.004	2.066	0.000	2.049
18	18.931	0.003	2.069	0.000	2.057
19	18.930	0.001	2.070	0.000	2.062
20	18.929	0.001	2.071	0.000	2.066
21	18.928	0.001	2.072	0.000	2.069
22	18.927	0.001	2.073	0.000	2.070
23	18.927	0.000	2.073	0.000	2.071
24	18.926	0.001	2.074	0.000	2.072
25	18.926	0.000	2.074	0.000	2.073
26	18.926	0.000	2.074	0.000	2.073
27	18.926	0.001	2.074	0.000	2.074
28	18.926	0.000	2.074	0.000	2.074
29	18.926	0.000	2.074	0.000	2.074
30	18.926	0.000	2.075	0.000	2.074
31	18.926	0.000	2.075	0.000	2.074
32	18.926	0.000	2.075	0.000	2.074
33	18.926	0.000	2.075	0.000	2.075

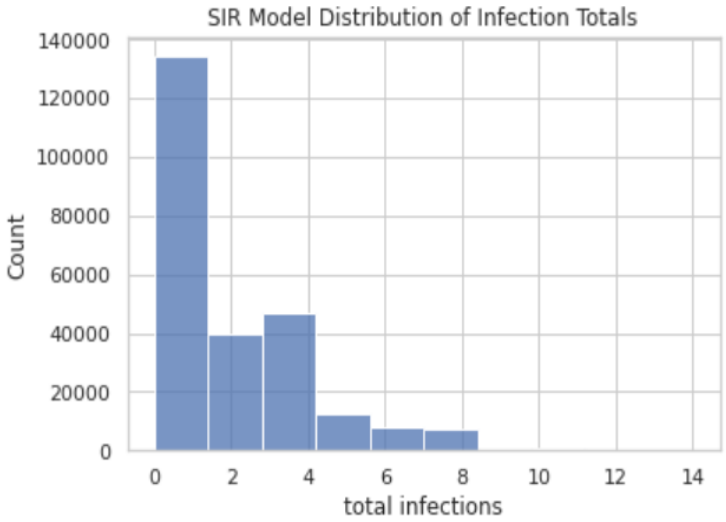


Fig 15. The Distribution of Total Infections with p=(0.05,0.15) and Infection Period [3,7)

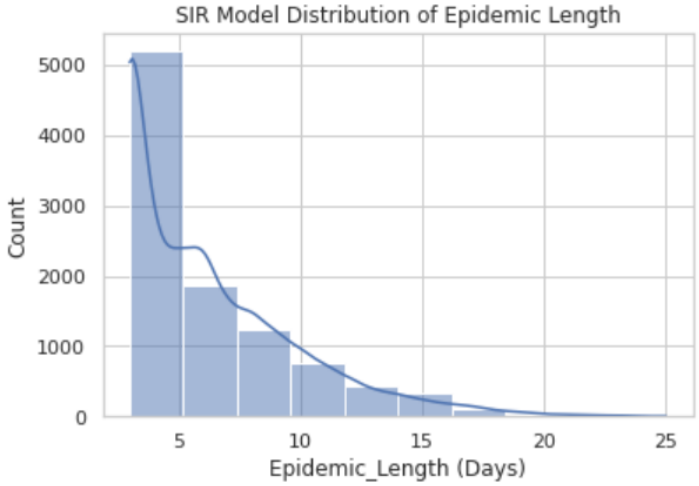


Fig 16. The Distribution of Epidemic Length with p=(0.05,0.15) and Infection Period [3,7)

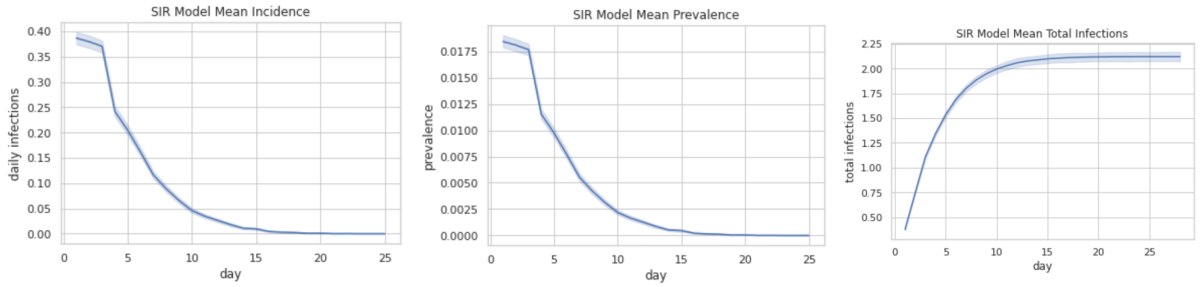


Fig 17. The Expected Prevalence, Expected Incidence, and Expected Infections with p=(0.05,0.15) and Infection Period [3,7)

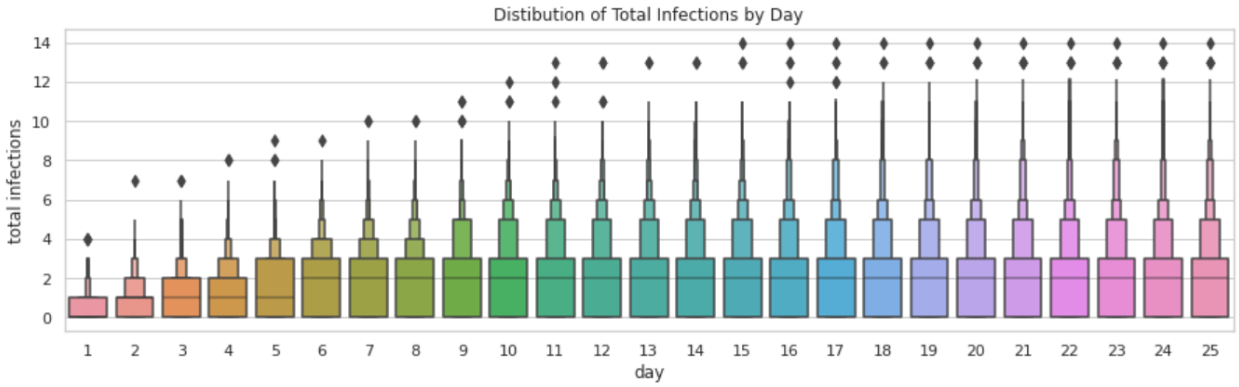


Fig 18. The Distribution of Total Infections Per Day with p=(0.05,0.15) and Infection Period [3,7) (Expected values are represented by a green triangle on each bar)

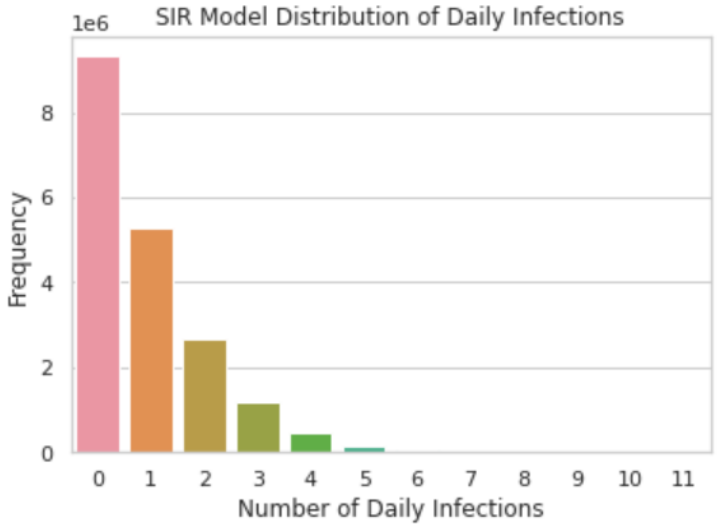


Fig 19. The Number of Daily Infections and their Frequencies (in millions) with p=(0.05,0.15) and Infection Period [3,7)

Table 4: Daily Infections Simulation Results (1,000,000 runs, ~19.1 million days, n=21, p(infection) = (0.05,0.15), infection length = [3,7) days, number of initial infections = 1)

Number of Daily Infections	Frequency	Percentage
0	9323982	48.72
1	5298043	27.69
2	2667117	13.94
3	1197131	6.26
4	454763	2.38
5	145743	0.76

6	38961	0.20
7	8789	0.05
8	1733	negligible
9	275	negligible
10	36	negligible
11	5	negligible

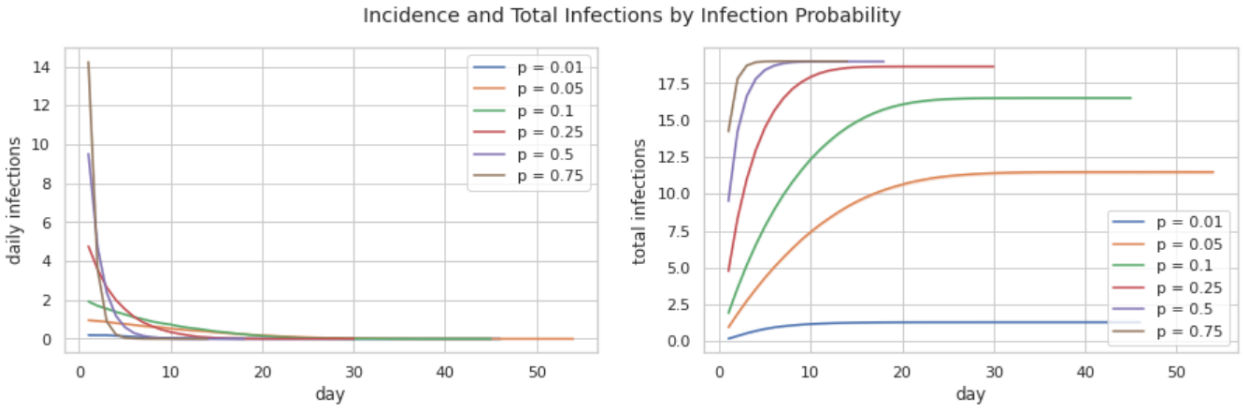


Fig 20. Incidence and Total Infections - By Infection Probability with Infection Period [3,7)

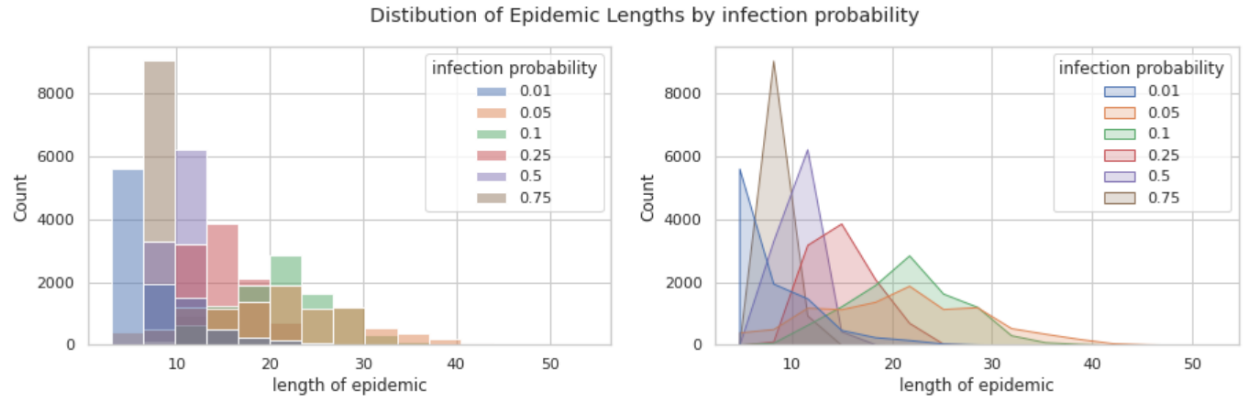


Fig 21. The Distribution of Epidemic Lengths - By Infection Probability with Infection Period [3,7)

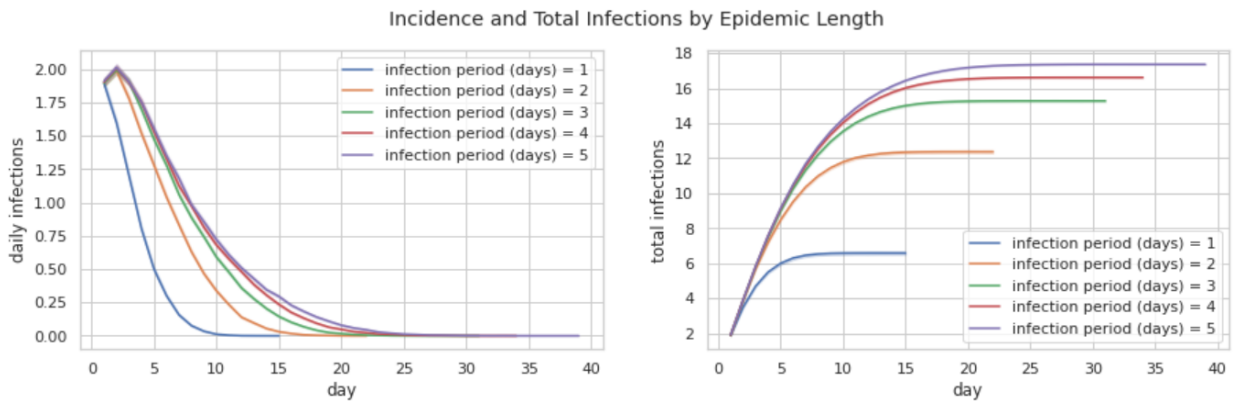


Fig 22. Incidence and Total Infections - By Length of Infection Period with p=(0.05,0.15)

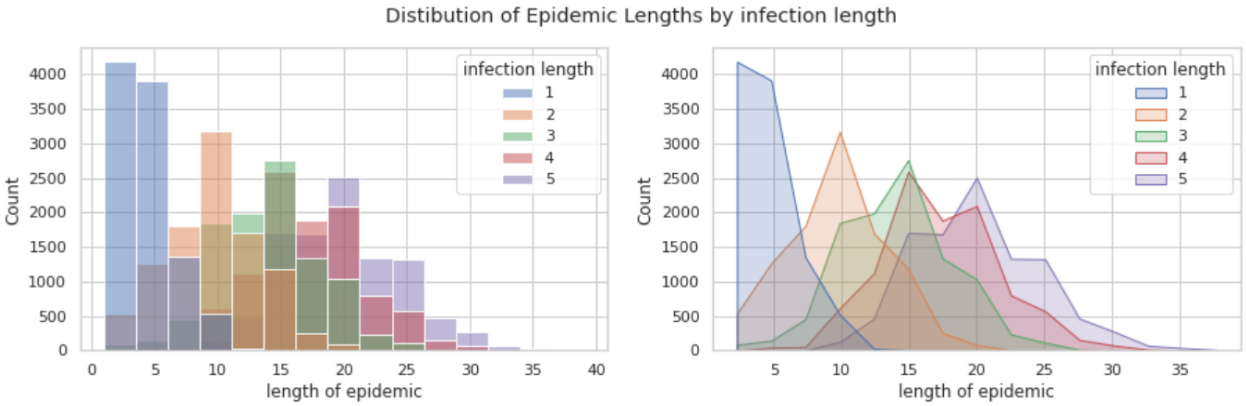


Fig 23. The Distribution of Epidemic Length - By Length of Infection Period with p=(0.05,0.15)

Fig 24. Incidence and Total Infections - By Class Size with Infection Period [3,7) and p=(0.05,0.15)

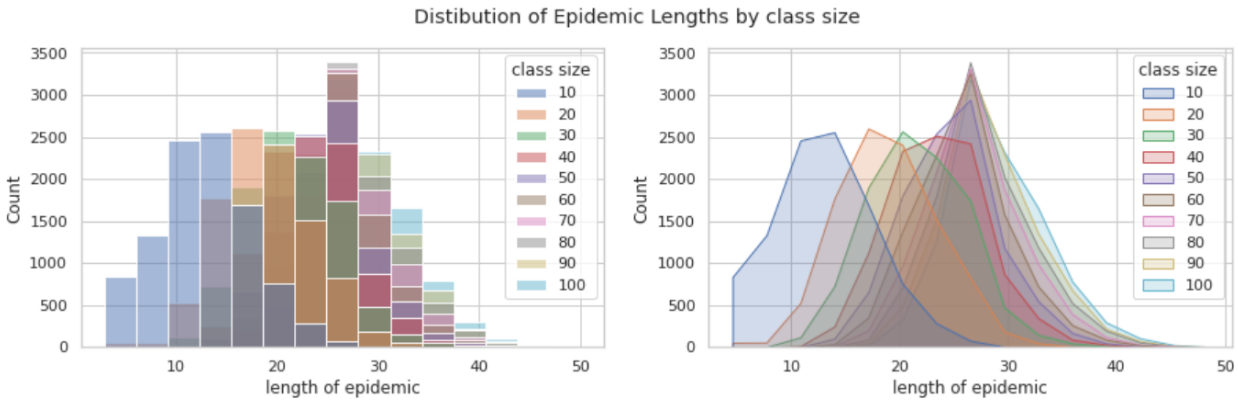


Fig 25. The Distribution of epidemic Length - By Class Size with Infection Period [3,7) and p=(0.05,0.15)