

Replicating and Improving a Random Covid-19 Walk

Isabelle Munsell, Thomas Harris, Sophie Giacobbe

2022-12-01

Contents

1 Motivation:	1
2 About the Study	1
2.1 Methodology:	2
3 Colorado	3
3.1 Results:	3
3.2 Conclusion:	7
3.3 References:	8

1 Motivation:

A few short years ago the Covid-19 pandemic broke out—infesting and killing millions of people. At the onset of any new virus, the effects are catastrophic because not much is known about the virus and there is unlikely to be a suitable vaccine to prevent the spread of the virus. Thus, it is imperative for the scientific community to continually improve upon early data models. As such, we believe it is important to look at a paper that was released at the beginning of the pandemic to see how it can be improved upon, as well as use the model to see if it can keep up with a changing environment pandemics are known for. This devastating virus has continued to persist among our societies in the form of other variants such as Alpha, Beta, Delta, Gamma, and Omicron. Since the onset, there have been several variants that differ in the level of contagiousness. This level of contagiousness is known as R_0 , the number of secondary infections that result from one infected person. For instance, in the early stages of Covid-19, scientists estimated the R_0 value to be between 2.24 and 3.58, which means that for every one person infected, 2.24 to 3.58 people were infected from them. This is a way for scientists to measure the spread of the virus in a universal value. Not only can this value tell us about how contagious a disease is, but it can also tell us if a disease is likely to cause an epidemic or not. In fact, for a disease to become an epidemic, the R_0 value has to be greater than 1. Given that previous studies have explored the spread and level of contagiousness at the onset of Covid-19, it is important to understand both the beginning of this pandemic, as well as to estimate how it changes over time. As such our goal is twofold, we are interested in recreating the paper’s early prediction of Covid-19 spread in New York, as well as use the same model, with updated parameters, to see if it could accurately predict Covid-19 spread after two years and a new infection rate.

2 About the Study

The study we are interested in is called “A random walk Monte Carlo simulation study of Covid-19-like infection spread. Although this study was published in April, 2021, it was finalized and set to publishers in July, 202. Because relatively little was known about Covid at the time, there was not much information

about Covid. The purpose of the study is to use parameters they did know about Covid, like R_0 values, and simulate the spread of the disease. Once they built the simulation, they ran many different types of experiments to better understand how the disease might spread. To run their simulation, they used what is known as a random walk. A random walk is a way to simulate interacting with agents, or people. A random walk is a process in which an agent takes random steps in a direction, determined by a random variable. Lastly, there is a boundary, meaning the agents cannot walk past a certain limit. In this study, this was square by nature, but could realistically be any shape. Through the random walk, the scientists can simulate and estimate the spread of Covid, because Covid is transmissible through infected people coming into contact with non infected people. When the simulation starts, there is one infected person and 9,999 uninfected people. As a trial run, each person is 100% susceptible, meaning that if they come close enough (0.02) meters, then they have a 100% chance of getting infected. The infected person is placed at the middle of the square, and as time goes on the infection slowly spreads outward. This paper can be split into two distinct parts. The first part of the paper shows how Covid could spread when they artificially change parameters like population size, how fast the agents can move and reinfection rate. The second part of the paper takes real life populations, and predicts how Covid can spread in real world cities. The key takeaways from the first part is that as agents walk faster (or move more freely without Covid restrictions), the rate of infection spreads faster. However, as the agents walk slower (or have more Covid restrictions imposed on them), infection rates flatten out, and it takes longer for the infection to spread. Lastly, they were interested to see how reinfection plays a part in the continuing of Covid, with a 5% chance of being susceptible to reinfection. From visual inspection alone, it seems that reinfection does increase the number of Covid cases, although not by much. Another reason for this is that they only ran 300 time steps, and so if they ran the simulation longer the results may have been different.

2.1 Methodology:

Unfortunately for us, the researchers not release the source code for the paper, and so we had to make many decisions on how to run the study that may not have been inline with how the researchers ran it. First, we were computationally limited and so we had to decrease the population, as well as the area. Fortunately one of the most important parts of the paper for us was that they ran a simulation in which they did this exact thing. The researchers changed the population size, but not the density⁷. They showed that infection rates were almost identical. This was crucial for us because it shows that even though he had 1/10 the population, it is likely the case that our results should be similar when looking at rate statistics. Secondly, they did not describe what specific random variable they used for the random walk part of the simulation. They mention that each agent takes a step size 'l' which is a multiple of the average separation between individuals. To us, it was unclear why a step size would be in relation to the average distance between each agent. Furthermore, it is unclear why they chose the specific step size 'l', as there appears to be no real world justification, rather, they used this 'l' parameter to view different step sizes. In our simulation, we spread out the agents on a uniform distribution, and they would walk based on a normal distribution. For each agent, the normal distribution has a mean of their current location, and the standard deviation was 0.005. This number was chosen so the average distance between each agent was proportionally equidistant to the 'l' value they used of 0.02 meters. Lastly, we also had to change what a time step meant. We decided that one time step would equal one day, which made our simulation run faster than the paper simulation. We did this for two reasons. The first, and more obvious answer, is that a day is a good way to measure the spread of Covid, because it allows us to closely monitor the spread, while also not updating our plots with needless information. The second reason we chose the time step to be a day is so we had less computations. The problem is, as more people get infected, the more times we have to check to see if an infected person is close to a non-infected person. The real problem is that computers cannot 'see' if a person is close to another one. Thus for each infected person, the infected person needs to check around all people who are not infected to see if they are close. If we have a population of one thousand people, and twenty people are infected, there will be twenty thousand checks per day. While these simplifications are unfortunate because it means we are less accurate to the original paper, we believe that each simplification was necessary for the sake of the recreation.

Next, we decided to improve upon their infection spread. In the study, once an agent got infected, they stayed infected for the rest of the simulation. However, we know this is not realistic, and so we decided to have a

rate of recovery. Importantly this simulation does not take into effect the effects of Covid-19 on an agent, for instance in the real world if a person is sick they are more likely to stay at home and not move. This means we need to estimate how long a person can no longer spread the virus, not how long it takes for them to recover. We found that it takes on average 7-14 days to no longer be contagious once infected ⁸. Thus, we decided that after 11 days of the infection, a person would automatically recover from Covid-19. Importantly, however, we do think a reasonable improvement is to change this static ‘11’ number into a random variable, because different people recover at different rates. Lastly, we also have an immunity period for each person. Unlike the study, where there is a five percent chance of reinfection, we assumed that everyone could be infected, although there is an immunity period of 60 days. 60 days seems to be the minimum amount of time it takes before immunity wears off from a Covid infection ².

3 Colorado

When it came to implementing our model with Colorado’s data, we did so by researching approximate parameters. For instance, based on a study conducted in 2021, researchers found that 1 individual recovered on average every 8 days (Buchwald). The probability of becoming infected was kept at 100% because there is little to no data on actual infection probabilities. Along with this, the model simulates the onset of Covid-19 before people began changing their behaviors—which is more realistic.

One thing that we had to take into account was Denver, Colorado’s density. Denver has a density of approximately $1,784 \text{ people}/\text{km}^2$. Given that there are computational limitations, the size of the grid was 0.8 km by 0.8 km, so the starting population reflected this grid size. Taking $[(1,784) * (0.8)]$, we determined a starting population of 1,427 people. Lastly, the corresponding $\langle r \rangle$ value, which is the average distance of separation between any given walkers, was calculated using $\sqrt{A/N}$. If the distance between any given walkers was less than 0.024 km than the person would become infected.

To compare the simulation results with real Covid-19 reported data we used the Covidcast package created by the Carnegie Mellon University Research group. This package tracks and provides various measurements relative to Covid-19 data. Using the John Hopkins data source and cumulative Covid-19 cases signal, we collected the data for Denver county from March 2, 2020 to April 21, 2020 (50 day difference).

3.1 Results:

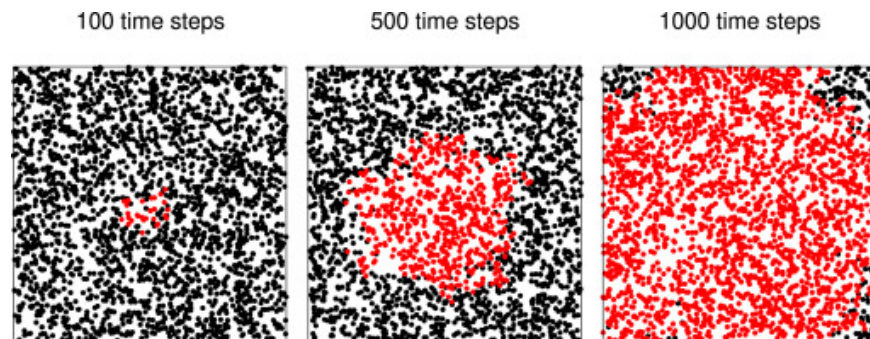
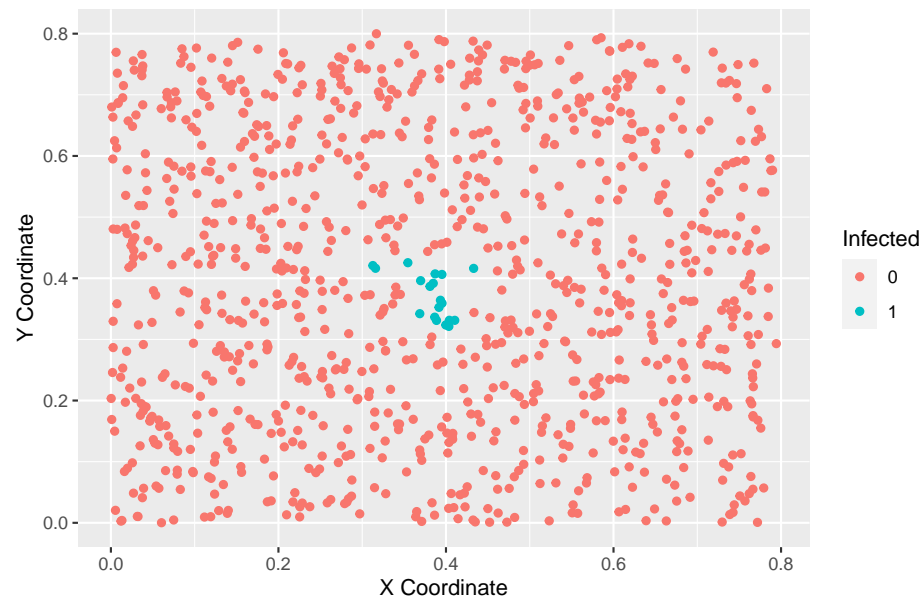
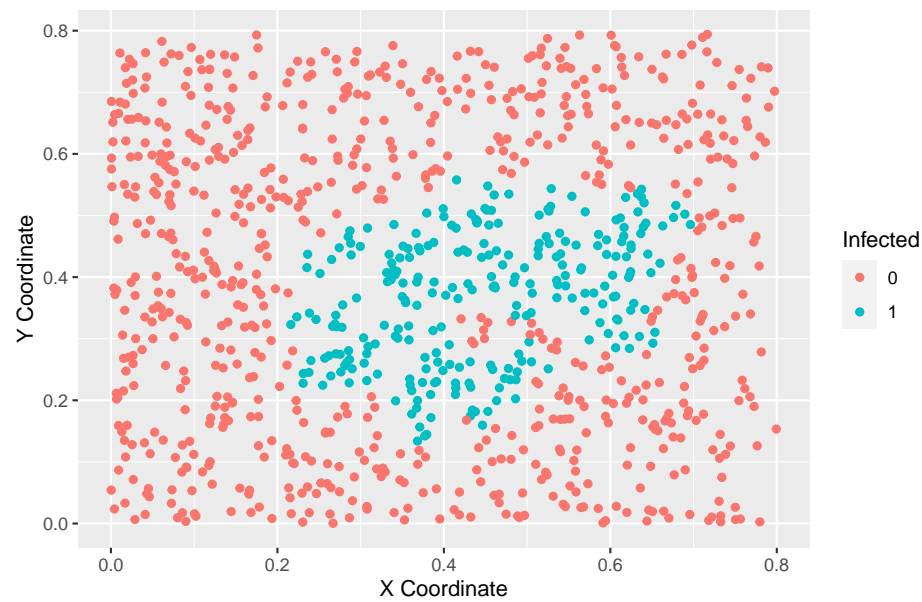


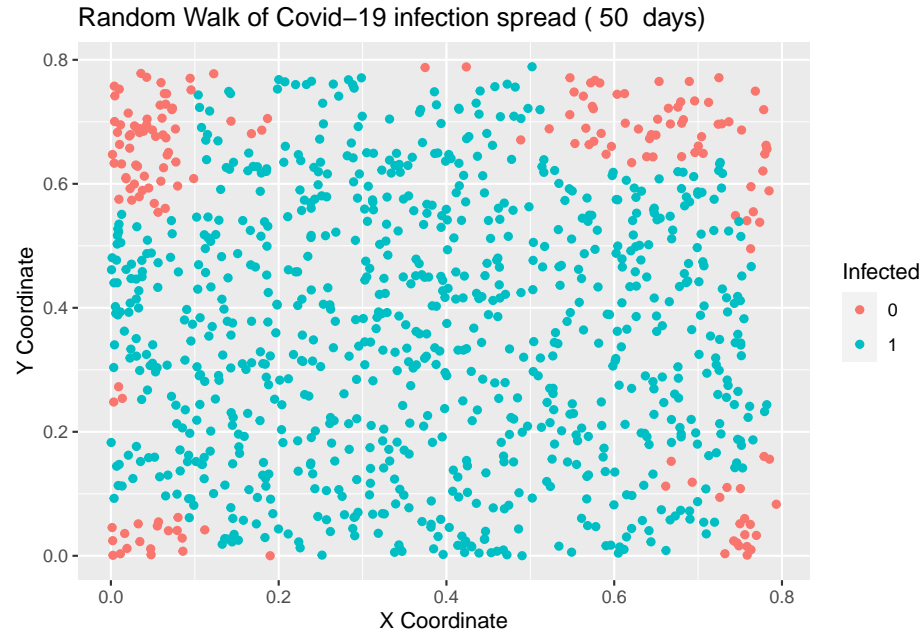
Figure 1: Original Results

Random Walk of Covid-19 infection spread (10 days)

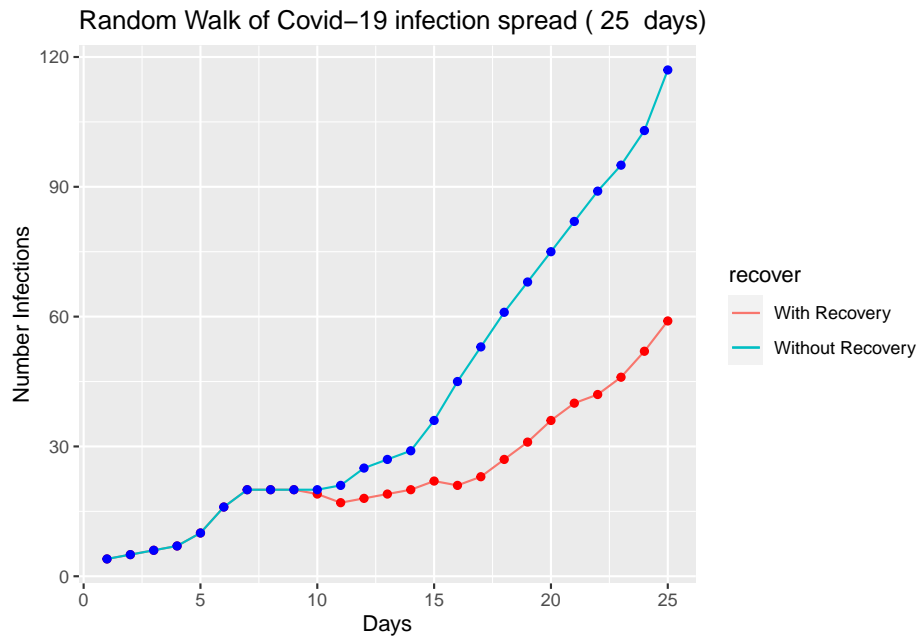


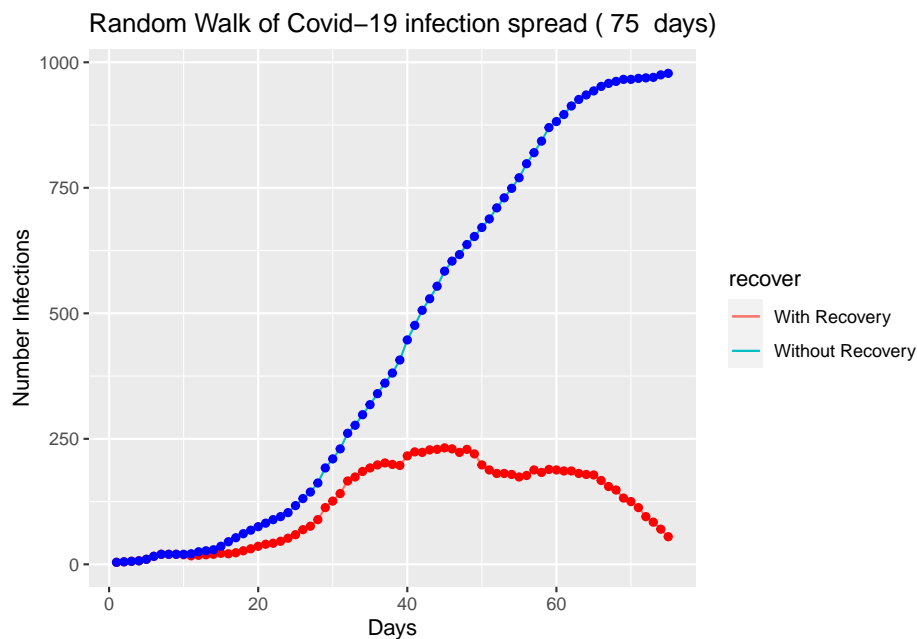
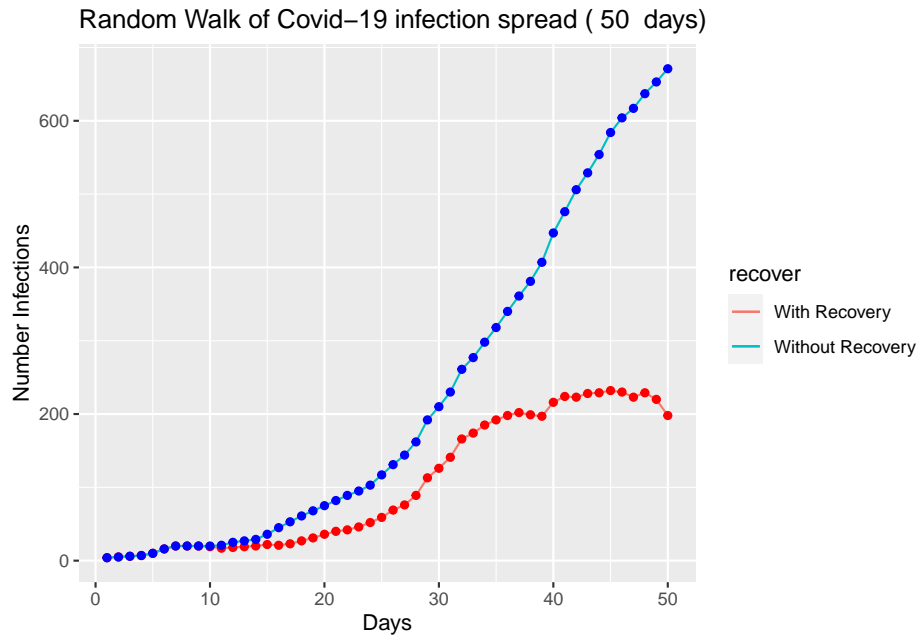
Random Walk of Covid-19 infection spread (25 days)



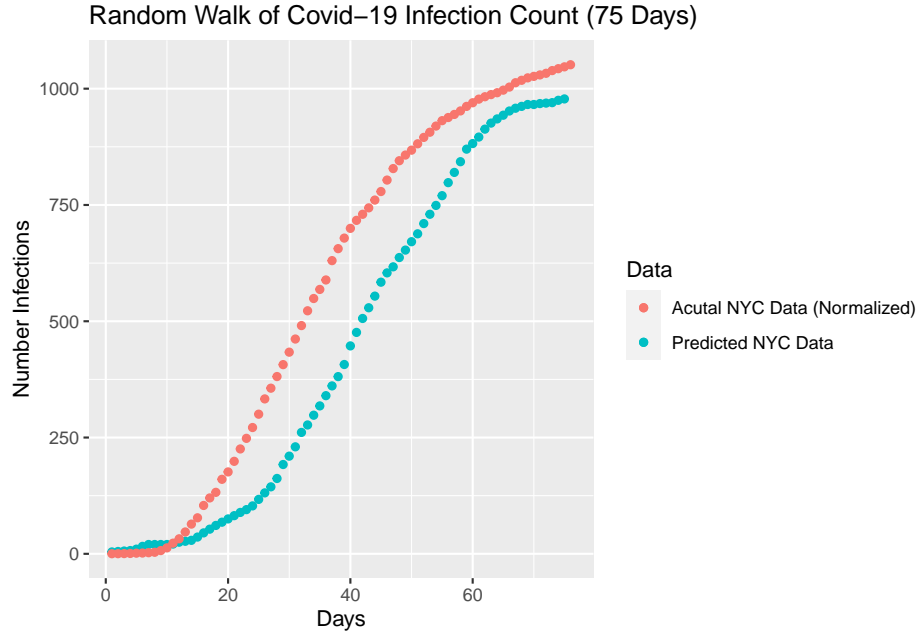


We believe our results are in alignment with the paper, as well as the fact that we improved the paper in some areas. We were able to recreate their spread of Covid, as seen in fig. 1. Figure 1 is the original paper's experiment, and below is our recreation. While we could not recreate their findings one to one due to simplifications and assumptions, we were able to show that their results are reasonable.





Next, we can look at the difference between the number of infections compared to the number of people currently infected. For the first 11 days, there is no difference which makes sense: people do not have time to recover from the virus. However, after day twelve days, there is a huge difference. We can see that the overall spread of the disease is much lower, and that not everyone in our simulation gets infected. After 75 days, the infection almost disappears. However, this does not seem reasonable with what we know now. Covid did not disappear after 75 days, and so we must investigate why. It may be the case that using a grid in which people cannot leave is a large assumption. For a disease to spread, it needs to find new communities and new people to infect. Really, what our graph shows is what a successful lock down measure would do. We can think of this simulation as an entirely locked down community, in which no one in or out is allowed to leave. Thus, because there are no new people to infect, and the immunity is 60 days, this graph shows that if dramatic lockdown laws took place, that covid would almost disappear.



Finally, we are able to compare our results to the real world New York Data. We took this data from the R package “covdata”, in which the New York Times updates daily⁵. We can see from the plot above that our model and the actual infection spread are close, although with some caveats. First, our X axis starts at 0, but this day corresponds to March 1st, and then goes to 75 days (May 15th). In general, our model of Covid spread in New York tended to under predict the amount of infections. This means that if this model was to be actually used, this may be dangerous because our model would underplay how infectious Covid truly is. The other idea to keep in mind is that we do not have an accurate representation of the population of New York in our model, due to computational limitations. Thus, the true infection count is not actually 1,000 after 75 days. Instead, that number is around 19,000. The reason why we scaled the truly values of New York City down, was so that both plots would fit on the graph. In reality, it is not about the numbers on the graph, but rather the shape of the infection rate. We think there are two main ways our predictions and the actual data are different. First, our model takes longer for the disease to start spreading rapidly (from 0 - 20 days). In reality, the real life data shows that Covid spread more quickly than our model predicts from 0 - 20 days. However, we can start to see the curve flattening out earlier in the real life data, around day 45. In our model, the curve only flattens around day 55. The data we found is the summation of the total number of cases, and so we used our model in which people do not recover from the infection. As such, it may be the case in the real world, the flattening of the curve means more people stayed at home, social distancing occurred, or any of the other measures taken early in the Pandemic. The reason why the curve flattens out in our data is because there are less people to infect, because the number of people who are not infected is close to 0.

3.2 Conclusion:

We believe our results are in alignment with the paper, as well as the fact that we improved the paper in some areas. We were able to recreate their predicted spread of Covid, as seen in fig. 1 and figure 2. Figure 1 is the original paper’s experiment, and Figure 2 is our recreation. While we could not recreate their findings one to one due to simplifications and assumptions, we were able to show that their results are reasonable. Furthermore, by adding a time of immunity as well as a recovery rate, we believe that we improved this paper to be more accurate with Covid research. We were also able to use our model to estimate the spread of Covid in New York in 2020. While our results are not perfect, we believe that our estimations are not too far off, and our simple model was able to predict how fast Covid spread in New York.

Overall, we believe that using a random walk to simulate the spread of Covid, and more generally infectious

diseases is a great way to estimate the spread of infection early on into the life of a new disease. However, there are a lot of assumptions that have to be made. One of the problems with a random walk is that people do not move randomly through space. In general, it is quite the opposite: people have set weekly routines. For instance, if you are a college student, it is likely that your weekdays have a consistent schedule where, for the most part, some portion of your day will be spent on campus. While there is some “randomness” like going to a grocery store, or running miscellaneous errands, most people have a consistent schedule. Because of this, a random walk is limited because it is random. Furthermore, only running a Monte Carlo simulation once is limiting, because there is more variance in running one simulation, compared to running ten thousand. Unfortunately, due to computing limitations, we could not reasonably run this simulation more than once.

3.3 References:

1. 9—September 2021 - Emerging Infectious Diseases Journal - CDC. [wwwnc.cdc.gov](https://wwwnc.cdc.gov/eid/article/27/9/21-0909), <https://doi.org/10.3201/eid2709.204167>. Accessed 14 Dec. 2022. Buchwald, Andrea G., et al. Estimating the Impact of Statewide Policies to Reduce Spread of Severe Acute Respiratory Syndrome Coronavirus 2 in Real Time, Colorado, USA - Volume 27, Number
2. Buchwald, Andrea G., et al. Estimating the Impact of Statewide Policies to Reduce Spread of Severe Acute Respiratory Syndrome Coronavirus 2 in Real Time, Colorado, USA - Volume 27, Number
3. Cassata, Cathy. “How Long Does Immunity Last after COVID-19? What We Know.” Healthline, Healthline Media, 22 July 2022, <https://www.healthline.com/health-news/how-long-does-immunity-last-after-covid-19-what-we-know>.
4. CFpaper/Params Curve Fitting Scenarios.Csv at Main · Agb85/CFpaper. <https://github.com/agb85/CFpaper/blob/main/params%20curve%20fitting%20scenarios.csv>. Accessed 14 Dec. 2022.
5. Nytimes. “Covid-19-Data/US-States.csv.” GitHub, New York Times, 15 Dec. 2022, <https://github.com/nytimes/Covid-19-data/blob/master/us-states.csv>.
6. Sheikhi, Farnaz, et al. “Estimation of the Basic Reproduction Number of Alpha and Delta Variants of Covid-19 Pandemic in Iran.” PloS One, vol. 17, no. 5, 2022, p. e0265489. PubMed, <https://doi.org/10.1371/journal.pone.0265489>.
7. Triambak, S., and D. P. Mahapatra. “A Random Walk Monte Carlo Simulation Study of Covid-19-like Infection Spread.” Physica A: Statistical Mechanics and Its Applications, vol. 574, July 2021, p. 126014. ScienceDirect, <https://doi.org/10.1016/j.physa.2021.126014>.
8. “The 14-Day Quarantine: Understanding the Coronavirus Incubation Period.” Pfizer, <https://www.pfizer.com/news/articles/14-Day-Quarantine-Incubation-Period>.