Virus Spread Simulation Using The SIR Model

David Clark

International Science and Engineering Fair 2021

February 17, 2021

**Introduction**

“An estimated 300 million people died from smallpox in the 20th century alone” is a terrifying statistic and even more so when taken into account that smallpox is a single virus among countless other; many of them having the possibility to cause disease, including death, among humans (). While their existence may be hard and costly to eradicate fully, understanding how they spread and what factors affect this most can save countless lives and be done with minimal resources. That’s the goal of the project described in this paper: simulating virus spread with enough accuracy to provide useful data and the ability change certain factors such as the spread radius, the infection chance, the required recovery period, the use of masks, and the effectiveness of masks and how all of these affect virus spread.

**Background Research**

To understand the effectiveness of this project one first must understood the model the simulation uses; that being the SIR model which stands for simulated, infected, and recovered (Zurich par. 2). The idea of applying mathematic models to epidemiology was brought about by Bernoulli in 1760 and later Kermack and McKendrick in 1927 furthered the framework; however, the SIR model is credited to Sir Ronald Ross and a few others in the early 20th century (Zurich par. 1). The SIR model can be simulated multiple ways, and at points can be boiled down to a few equations to find extremes in virus outbreaks, but in general a SIR model starts with an index case, which is the first introduction of an infected individual, that can spread it to other susceptible individuals and they can go on to spread the virus more and cause a chain of infections (Weiss 6).

In contrast to speaking about past research, further research regarding the study of viruses, how they spread, and how simple simulations could save millions of lives one would have to look no further than the current sars-covid-19 pandemic. A research paper by Dmitry Ivanov shows that computer simulations of viruses are used quite frequently and even goes on to prove their effectivity by demonstrating a fairly accurate one (Ivanov par. 1). It’s quite clear how data like this can be incredibly useful; projections like this could give hospitals a rough estimate of what patient load to expect, the quantity of medicine needed, and the space required.

Even more relevant, projections like this can show the necessary requirement of the general population’s reaction to a virus. For example, if 2 projections were estimated, one where people wore masks and social distanced and another where people didn’t, it could show how necessary or unnecessary practices like this are. However, these estimations are based on past research which already achieves the goal of seeing how important a population’s reaction is, it just puts it into perspective. Research goes on to say, that at least for the sars-covid-19 virus, “along with preventing someone from transmitting the coronavirus, a range of new research shows that the risk of infection to the wearer is decreased by 65%” and similar statistics can be expected when looking at air born or saliva dependent spread viruses (UC Davis Health par. 2).  
 Other factors than just the amount of susceptible, infected, and recovered individuals effect a population. For instance, certain occupations and areas will be subject to a higher risk of virus spread rather than just an entire population being at risk equally. This can be seen in a study from 2015 that stated “02% of Guinea’s population had died due to Ebola, compared with 45% of the country’s doctors, nurses, and midwives. In Liberia and Sierra Leone, the differences are more dramatic, with 11% and 06% of the general population killed by Ebola versus 07% of the health-care workers in Liberia, and 85% in Sierra Leone … health-care workers are at greater risk of contracting Ebola” (Evans et al. par. 1). The SIR model fails to take factors like occupation, age, or even general health of the infected population which offers fair reasoning to any inaccuracies from its predictions.

**How the Simulation Works**

Before going into how the procedure and goal of the project within this paper, it would be beneficial to understand how the simulation that this project utilizes works. In a computer simulation, 100 individuals represented by colored triangles wander around in an 100 by 100 pixel square. These individuals can be blue, signifying they are susceptible, red, signifying they are infected, or grey, signifying they are recovered. At the start of each experiment an index case is set, meaning a random individual becomes infected. The infected individual continues to wander among the susceptible individuals and has a chance to spread the virus if they are within the infection radius. If in a frame an infected individual and a susceptible individual are within the infect radius of each other then a calculation based off the simulations values is ran that determines if the virus spreads to that susceptible individual. When an individual becomes infected another calculation can be done to determine how many frames the individual will be infected; after that many frames pass the individual becomes recovered and can no longer spread the virus or become infected again. The calculation for how long a population member is infected is taking a gaussian distribution of the average infect period with at most 40 frames of variation. The simulation runs until there isn’t any infected individuals left.

Background pattern

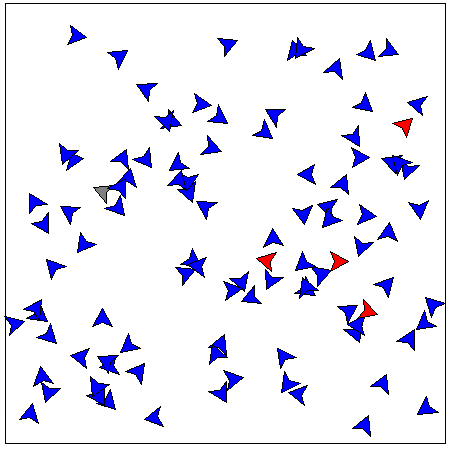
Description automatically generated

Figure 1 – Frame with many susceptible individuals

Background pattern

Description automatically generatedFigure 2 – Frame with many infected individuals Figure 3 – Frame with many recovered individuals

Looking at figures 1,2, and 3 above, these images were taken from a single control data run. Near the beginning, figure 1 was taken as the index case has barely spread to others and few have recovered. Not too far after figure 1, figure 2 was taken as the index case has spread enough to others to cause a large chain reaction and result in many infected individuals. After that, figure 3 was taken as the infected population member’s infection period runs out and they become recovered. For more explanation of how a simulation as a whole runs, look to figure 4 below for a simplified flow chart of a simulation run.

Figure 4 – How a Simulation Runs

Diagram

Description automatically generated

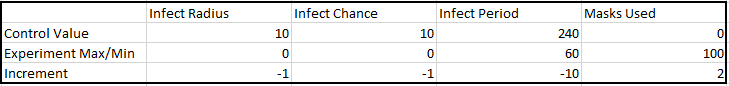
**The Goal and Hypothesis**

The goal of the described experiment is to yield realistic data that can mirror actual viruses and to see if enough change in any virus condition can lower a virus’s spread. It was hypothesized that the tested independent variables all have the ability to keep at least 50% of the population from becoming infected; however, some tested variables will meet this thresh hold before others.

**Procedures**

The experiment described in this paper began with building the used computer simulation while utilizing previous research to have realistic values. Then, a control experiment where no values were changed was ran. After that, 4 experiments were run with each only changing one variable and keeping the other variables the same as the controls. The 4 experiment’s passes consisted of them starting with the exact same values as the control experiment, but after that simulation completes a single independent variable was changed by an increment up until a final max or minimum was reached; the control values and the experiment maximums or minimums can be seen in table 1 below. Once all these experiments were completed, the average susceptible, infected, and recovered individuals for each experiment were logged. The 5 experiments completed were then ran 99 times more and their data was averaged overall to avoid sporadic data and to have more accurate results. For example, only looking at the infect radius experiment, it first ran a simulation with an infect radius of 10, an infect chance of 10, an infect period of 240, and with 0 individuals wearing a mask. After that simulation was ran and logged, another simulation was run with an infect radius of 9, an infect chance of 10, an infect period of 240, and 0 masks used. The next simulation would have the same values except the infect radius would be 8, then 7, then 6, and so on until it reached zero where the experiment would be over. This would be done 100 times and all the data for this experiment alone would be averaged and then saved. This process happened to all of the experiments except the control where it was just run 100 times and averaged, there were no changing variables in the control experiment.

Table 1



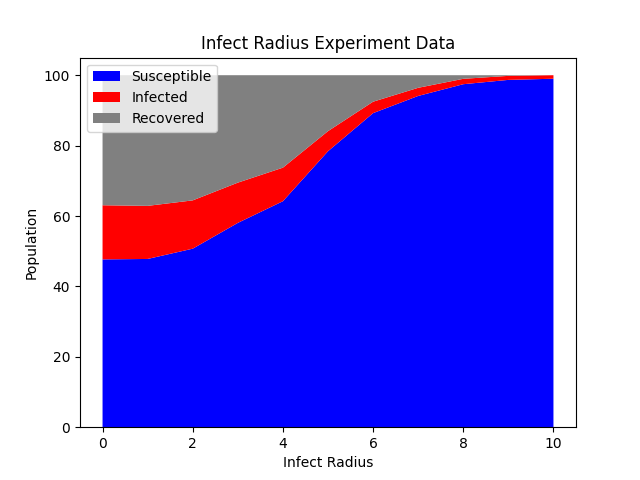
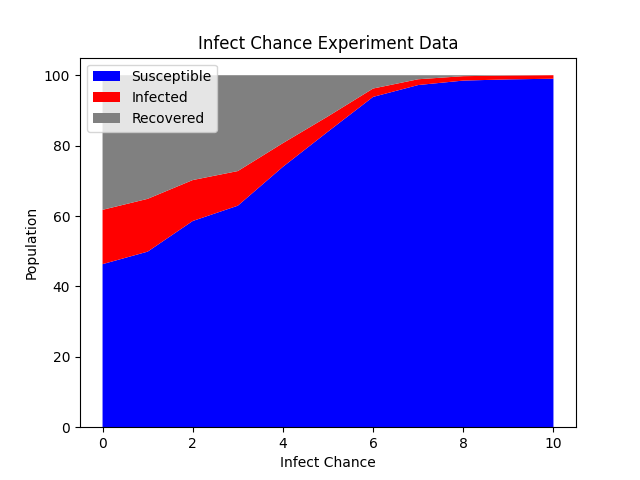
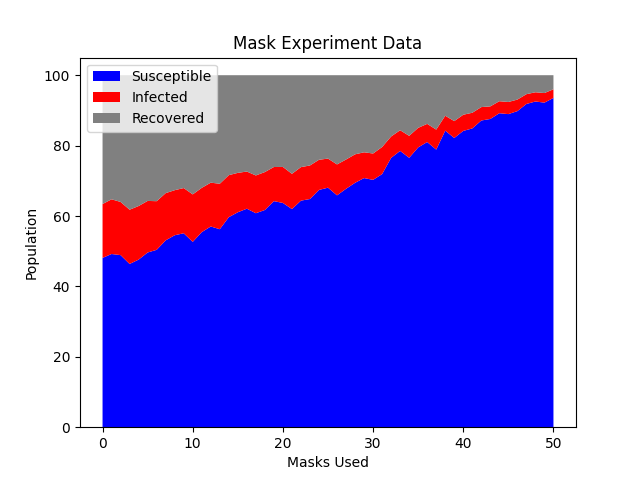
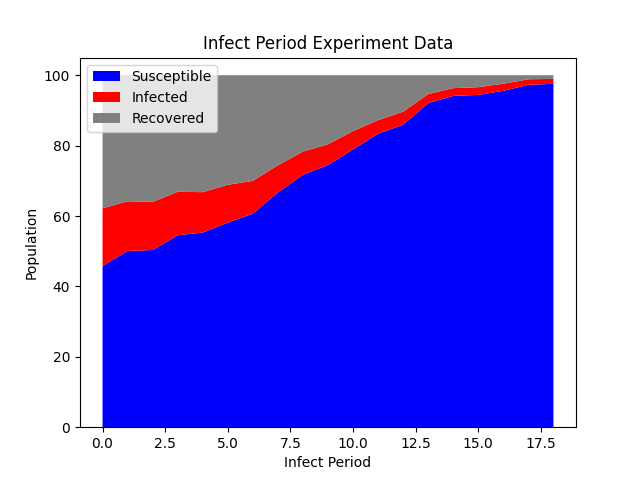
**The Collected Data**

Figure 5

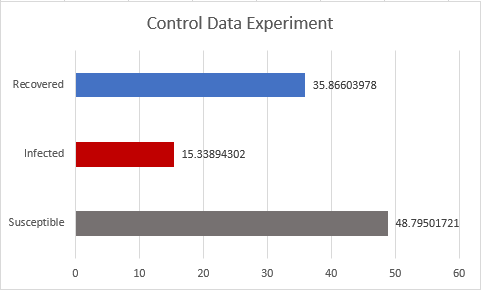
****

Note that the black bar in the graphs represent the control experiment

Figure 7

Figure 6

Figure 8

 The following paragraph references figures 5-9 above often. The collected data does meet the hypothesis and isn’t out of the ordinary by any means, but the data is still worth while and reinforces ideas already found in previous research. As any of the tested variables decrease, or in the case of masks increase, there is a direct correlation of resulting with less infected individuals. Something to point out is that all the graphs have an aggressive decline in infected people at first, but then flatten out when approaching their maximum/minimum except for the masks experiment, which has a much more static decline of infected people. It would be reasonable to assume that the mask experiment and the infect chance experiment would have nearly identical results as masks directly affected infection chance, but this isn’t the case. They do both result in less infected population members, but mask data seem to be more sporadic in their decline and even with an entirely masked population there were still noticeably more infected and recovered individuals then any other experiment, besides the control of course. A possible theory for the sporadic nature of the mask experiment data is that masks depended more on the population movement than the others; another theory for the result in more infected individuals is that masks only provided a -50% of spreading the virus to another person and a -20% chance of spreading when protecting someone, which means that even with a masked infected person exposing a masked susceptible person, the infect chance would be 4, as 10 times .5 times .8 equals 4; which is higher than the infect chance minimum which reached 1. Furthermore, one could reasonably assume that infection period would have a more drastic change as the average time it takes to recover and no longer be infectious reduces, but it’s downward decline of infected people matches the other experiments well and the increment of 10 frames was randomly selected for simplicity and time sake. In summation, even with all the minor un-anticipated results, the hypothesis was correct and furthermore the lowered results all had the ability to cause the total infected people to become less than 20%. When looking at this data it should be taken into consideration that the SIR model does have some inaccuracies when emulating real life as people don’t wander aimlessly, people have occupations that can be at a higher risk of infection, people’s health and ability to fight off viruses can vary, and when people become sick they either distance themselves from others more or feel less able to move about and spread the virus affecting them.

Note that the total of the control experiment doesn’t exactly equal 100, which is the population size, due to averaging

Figure 9

**Conclusion**

In conclusion of the entire project after the simulation was built, the 5 experiments were run repeatedly, and the data was averaged the hypothesis that at least 50% of the sample population would remain susceptible was achieved and that with enough change of any of the independent variables tested 80% of the population could remain susceptible by the end of the simulation, meaning they were never infected. Out of all the independent variables tested, such as infect radius, infect chance, infect period, and masks used; all had rather steep decline in infected individuals as the independent variable changed, except for the use of masks, which had a rather constant decline of infected people. This data proves previous research on the importance of key variables of viruses and how a population should react to a virus if they want to minimize spread. However, the SIR model used does have some inaccuracies as the population members in this experiment wandered aimlessly, all had the same health, and continued to wander when sick. This doesn’t mirror real life as people don’t wander aimlessly, some peoples occupations are more at risk to infection to certain viruses, people’s health and ability to fight off viruses are different, and generally when someone becomes sick they try to isolate themselves more or at least feel less capable of moving about and therefore can spread the virus less.

**Bibliography**

D´ebarre, Florence. “SIR Models of Epidemics.” SIR models of epidemics – Theoretical Biology | ETH Zurich, 2019. [https://tb.ethz.ch/education/learningmaterials/modelingcourse/level-1-modules/SIR.html.](https://tb.ethz.ch/education/learningmaterials/modelingcourse/level-1-modules/SIR.html.%20)

Evans, David K, Markus Goldstein, and Anna Popova. “Health-Care Worker Mortality and the Legacy of the Ebola Epidemic.” The Lancelet Global Health, July 8, 2015. [https://www.thelancet.com/journals/langlo/article/PIIS2214109X(15)00065-0/fulltext.](https://www.thelancet.com/journals/langlo/article/PIIS2214109X(15)00065-0/fulltext.%20)

Flight, Colette. “History - British History in Depth: Smallpox: Eradicating the Scourge.” BBC. BBC, February 17, 2011. <https://www.bbc.co.uk/history/british/empire_seapower/smallpox_01.shtml>.

Ivanov, Dmitry. “Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-Based Analysis on the Coronavirus Outbreak (COVID-19/SARS-CoV-2) Case.” Transportation Research Part E: Logistics and Transportation Review. Pergamon, March 24, 2020. <https://www.sciencedirect.com/science/article/abs/pii/S1366554520304300>.

UC Davis Health, Public Affairs and Marketing. “UC Davis Experts: Science Says Wearing Masks and Social Distancing Slow COVID-19 (VIDEO).” UC Davis Health, July 6, 2020. <https://health.ucdavis.edu/health-news/newsroom/uc-davis-experts-science-says-wearing-masks-and-social-distancing-slow-covid-19/2020/07>.

Weiss, Howard. “The SIR Model and the Foundations of Public Health.” Barcelona : Departament de Matemàtiques de la Universitat Autònoma de Barcelona, 2013.