Modelling the spread of infectious diseases with a compartmental model

H. J. C. Oliver[1, a)]
Department of Physics, The University of Texas, Austin, TX 78712, USA

(Dated: 10 December 2018)

A compartmental model for the spread of infectious diseases has been developed. In our model, a disease with probability of transmission p can spread through a population of N individuals that interact with 6 other individuals each time step. Infectious individuals remain contagious for 5 days, during which they can infect other people. We investigate the effect of population size, transmission probability, and vaccination rate on the spread of infectious disease. A relationship between transmission probability and the herd immunity threshold for large populations is found. The implications of these results and the limitations of our simulation are then discussed.

I. INTRODUCTION

In 2010, 15 million people died from infections caused by pathogens and parasites[1]. The World Health Organization (WHO) estimates 13 million people will die because of infectious diseases in 2050. Programmes designed to control infectious diseases can be optimised by improving our understanding of how these diseases spread through the population.

We can model the transmission of infectious disease using compartmental models[2]. Compartmental models divide the population into groups of people with equal characteristics. In this article, we use the SIR compartmental model to calculate the spread of an infectious disease. The SIR model splits the population into three compartments: (i) **s**usceptible people, who are healthy and can be infected with the disease; (ii) **i**nfectious people, who carry the disease and can infect susceptible people; and (iii) **r**ecovered people, who had the disease, but are now healthy and immune to further infection. The SIR model performs well for models of diseases that are transmitted between individuals and where recovered patients are immune to further infection, for example measles, mumps, and rubella.

In this report, we detail the code development in Section II. In Section III, we examine how the spread of infectious disease depends on properties of the disease and population. We discuss the consequences of these findings in Section IV before concluding in Section V.

II. CODE DEVELOPMENT

The compartmental code was developed in four stages: (i) a model for a single patient becoming sick and recovering; (ii) a model for a population of people, still with a single patient; (ii) a model where the infected can be contagious to their neighbours, and individuals can be vaccinated; and (iv) a model where the infectious spread the disease to random people, and individuals can be vaccinated. The development and testing of these four models will be discussed in detail.

A. A single patient

We begin by modelling a single person that begins the simulation as healthy, later catches the disease, and finally recovers after 5 time steps. This is implemented using a `Person` class with methods to return the person's status (susceptible, infected, or recovered), update the integer that describes the individual's status for the current time step, infect the person with a disease that causes them to be sick for n time steps, and to report whether the person has been sick and is now recovered. A typical output of this model is:

```
On day 1, Joe is susceptible
On day 2, Joe is susceptible
On day 3, Joe is susceptible
On day 4, Joe is susceptible
On day 5, Joe is sick (5 to go)
On day 6, Joe is sick (4 to go)
On day 7, Joe is sick (3 to go)
On day 8, Joe is sick (2 to go)
On day 9, Joe is sick (1 to go)
On day 10, Joe is recovered
```

This model is limited because we can only model a single person, so therefore it is not possible to study the dynamics of a disease within a population of many individuals. Therefore, we now introduce a population of multiple people.

B. Population introduced

We now add a `Population` class that is constructed with N individuals. The `Population` class contains methods that infect a random individual, count the number of infected people, update the statuses of all individuals within the population, and print the status of all individuals in the population for a given time step. A typical output of this model is:

```
# people in the population?  7
In step  0, 1 sick: ? ? ? ? + ? ?
```

a)Electronic mail: hjcoliver@utexas.edu

```
In step  1, 1 sick: ? ? ? ? + ? ?
In step  2, 1 sick: ? ? ? ? + ? ?
In step  3, 1 sick: ? ? ? ? + ? ?
In step  4, 1 sick: ? ? ? ? + ? ?
In step  5, 0 sick: ? ? ? ? - ? ?
```

However, currently patient zero never spreads the disease to other individuals in the population. Therefore, we introduce contagion.

C. Contagion introduced

For contagion, we implement a simple model of infection based on the disease spreading to neighbours only. The probability of this transmission per time step per interaction is given by p, where $0 \leq p \leq 1$. A typical output of this model is:

```
How many people are in your population?
7
Enter a probability of disease transfer
(between 0 and 1):  0.7
In step  0, 1 sick: ? ? ? ? ? + ?
In step  1, 1 sick: ? ? ? ? ? + ?
In step  2, 2 sick: ? ? ? ? + + ?
In step  3, 3 sick: ? ? ? ? + + +
In step  4, 4 sick: ? ? ? + + + +
In step  5, 3 sick: ? ? ? + + - +
In step  6, 3 sick: ? ? ? + + - +
In step  7, 5 sick: + ? + + + - +
In step  8, 4 sick: + ? + + - - +
In step  9, 4 sick: + + + + - - -
In step 10, 3 sick: + + + - - - -
In step 11, 3 sick: + + + - - - -
In step 12, 3 sick: + + + - - - -
In step 13, 1 sick: - + - - - - -
In step 14, 1 sick: - + - - - - -
In step 15, 0 sick: - - - - - - -
```

With individuals now able to infect neighbours, we must account for the edge conditions. The 1st individual can infect the Nth individual and vice versa, where N is the size of the population.

D. Vaccination introduced

Next, vaccination is introduced. Individuals can now be immune to disease from the beginning of the simulation. A typical output is shown below:

```
How many people are in your population?
7
Enter a probability of disease transfer
(between 0 and 1):  0.7
What proportion of the population are
vaccinated (between 0 and 1)?  0.7
In step   0, 1 sick: - ? - ? - + -
In step   1, 1 sick: - ? - ? - + -
In step   2, 1 sick: - ? - ? - + -
In step   3, 1 sick: - ? - ? - + -
In step   4, 1 sick: - ? - ? - + -
In step   5, 0 sick: - ? - ? - - -
```

In this example, the phenomenon of herd immunity protects all susceptible individuals from infection. This will be discussed in more detail in Section III C. Next, we move to a more realistic model of disease transmission.

E. Random spreading of the disease

The final step of development replaces the neighbour-only interactions with a fixed number of random interactions. Throughout this report this number remains fixed at 6. Now we move to a statistics-based study on disease transmission, so the display of the previous three sections is replaced with a summary at the end of the simulation for brevity. When the code is executed with `./spread`, a typical output is:

```
Enter population size:  1000000
Enter a probability of disease transfer
(between 0 and 1):  0.4
What proportion of the population are
vaccinated (between 0 and 1)?  0.4
In step 35:  0 % are sick, 0.0275 % are
susceptible, 59.9972 % have recovered, and
39.9753 % are inoculated.
```

Alternatively, the code can be executed with the population size, probability of disease transfer and vaccination rate as arguments on the command line, for example, `./spread 1000000 0.4 0.4`. This is particularly useful for scanning over a range of parameter space.

The statistics for each time step are saved to a file `disease.out` that can further analysed.

As the simulation progresses, the proportion of the population that are susceptible to the disease drops as the disease spreads. Conversely, the number of infected individuals increases up to a maximum (I_{max}), before decreasing. 5 time steps after the infection begins, the number of recovered people begins to rise. This delay is due to the duration of the disease for a given individual ($d = 5$ time steps). An example of this process is shown in Figure 1.

III. INVESTIGATION OF DISEASE TRANSMISSION

We now use the most realistic model of disease transmission, `spread`, to draw conclusions about the effect of disease and population properties on the spread of contagious disease.

A. Variation of probability of disease transmission

Unsurprisingly, varying the probability of disease transmission dramatically affects the spread of the disease.

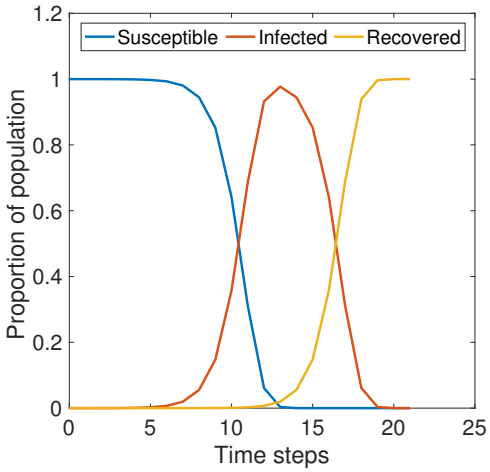

FIG. 1: A figure showing the typical evolution of the proportion of the population that are susceptible (blue), infected (red), and recovered (yellow) for a population of $N = 10^5$ people, with probability of disease transmission $p = 0.5$ and no vaccination.

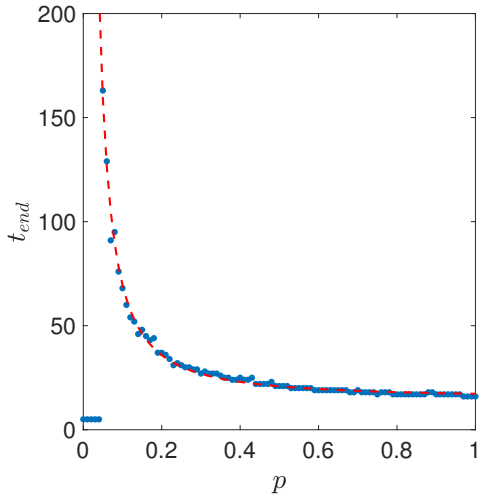

FIG. 2: The time steps required to reach a disease-free population (t_{end}) as a function of probability of disease transfer (p) for a population of $N = 10^5$ people with no vaccination, for a disease that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step. The decrease in t_{end} with p can be modelled by $t_{end} = 2.204p^{-1.394} + 15.08$ (shown in dashed red).

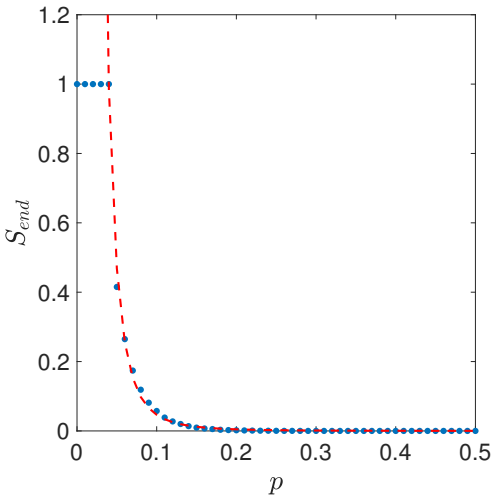

FIG. 3: The proportion of the population that is still susceptible at the end of the simulation (S_{end}) as a function of probability of disease transfer (p) for a population of $N = 10^5$ people with no vaccination, for a disease that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step. The relationship can be described by $S_{end} = 2.159 \times 10^{-5} p^{3.333}$ (shown in dashed red).

There is a threshold probability before the disease can gain any traction, rather than terminating after patient zero heals. This threshold probability typically lies in the range $0.05 \leq p_c \leq 0.15$ for a population of $N = 10^5$ people without vaccination. This threshold is visible in Figures 2 to 4.

Increasing the probability of disease transmission past this probability threshold causes the number of time steps required to reach recovery (t_{end}) to dramatically increase, as shown in Figure 2. This is because the disease spreads slowly, and more time steps are required to reach a disease-free population. Increasing the probability past this threshold causes the number of time steps required for a disease-free population to decrease. The increased probability means the disease spreads more easily, covering more of the population in a given number of time steps. The decay in number of time steps required for a health population is described well by $t_{end} = 2.204p^{-1.394} + 15.08$, where p is the probability of disease transfer. The coefficient of determination (R^2), which provides a measure for how well the dependent variable can be replicated by the model for a given value of the independent variable, for this fit is $R^2 = 0.9922$.

Above the threshold probability, the proportion of susceptible people who escape infection by the end of the simulation (S_{end}) drops off rapidly with increasing probability of transmission p, as shown in Figure 3. By $p = 0.2$, only 1 in 1000 people escape infection. By $p = 0.4$, no member of the population escapes infection. The drop in proportion of susceptible people left at the end of the simulation can be modelled by $S_{end} = 2.159 \times 10^{-5} p^{3.333}$ with $R^2 = 0.9964$.

As the probability is increased above the threshold probability, the maximum in the proportion of infected people (I_{max}) increases rapidly with the probability of disease transfer, as shown in Figure 4. As $p \to 1$, $I_{max} \to 1$.

In populations where everyone is eventually infected (for $p > 0.4$), the maximum occurs at $(0.59 \pm 0.02)t_{end}$, i.e. the maximum occurs at approximately 60% of the time required to reach a fully-recovered population.

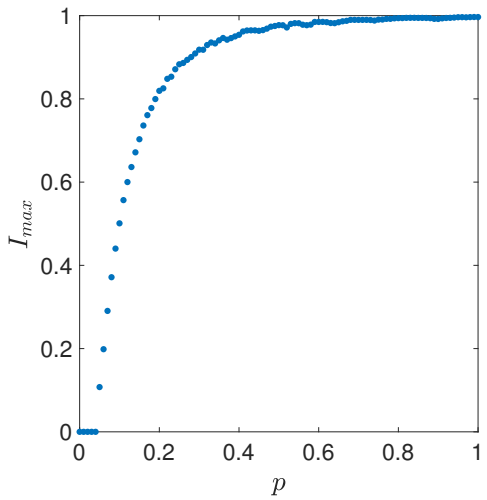

FIG. 4: The maximum in the proportion of infected people (I_{max}) increases as a function of probability of disease transfer (p) for a population of $N = 10^5$ people with no vaccination, for a disease that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step.

B. Dependence of transmission on population size

As expected, increasing the total number of people in our population (N) results in the time steps required to reach a disease-free population to increase, as shown in Figure 5. This is because there are more people that need to be infected before the population can reach the end state. There is a logarithmic relation between the population size and time steps required for a disease-free population. This relationship is well described by: $t_{end} = 6.363 \log_{10}(N) + 5.158$ with $R^2 = 0.9866$. For $N = 1$, the 5 time steps required for the single patient to recover from disease is roughly reproduced.

Scanning over the population size reveals that in small populations ($N < \mathcal{O}(10^3)$), nobody escapes the disease, as Figure 6 shows. Conversely, for large populations ($N > \mathcal{O}(10^5)$), approximately 1 in 1750 people will escape the disease, for a disease with transmission probability $p = 0.2$ and no vaccination. This trend only becomes apparent when population size is statistically large enough to show this trend.

For small population sizes, the statistics are too poor to consider the dynamics of disease spread, as shown in Figure 7. Below $N \sim \mathcal{O}(10^2)$, the maximum in the proportion of infected people can not be accurately determined. Hence, for statistically meaningful results, we should use populations comprised of more than 100 individuals.

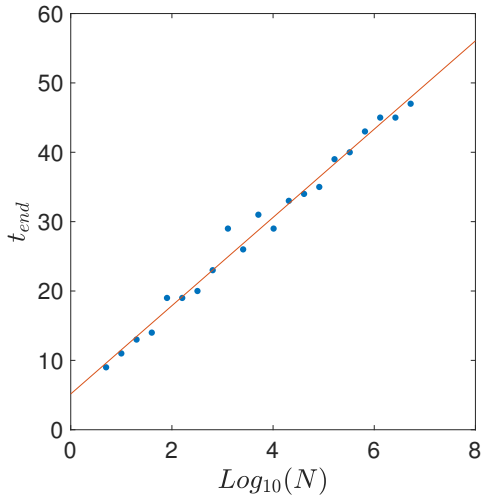

FIG. 5: The time steps required to reach a disease-free population (t_{end}) as a function of population size (N) with no vaccination, for a disease with transmission probability $p = 0.2$ that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step. The time steps required increases as $t_{end} = 6.363 \log_{10}(N) + 5.158$.

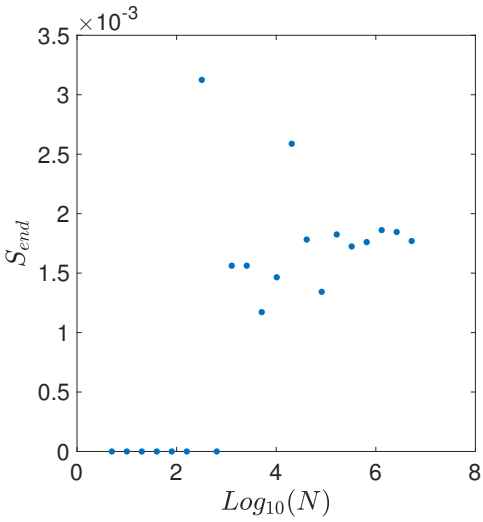

FIG. 6: The proportion of the population that is still susceptible at the end of the simulation (S_{end}) as a function of population size (N) for an unvaccinated population and a disease with transmission probability $p = 0.2$ that causes the patient to be sick for 5 timesteps and infectious to 6 random people in each time step.

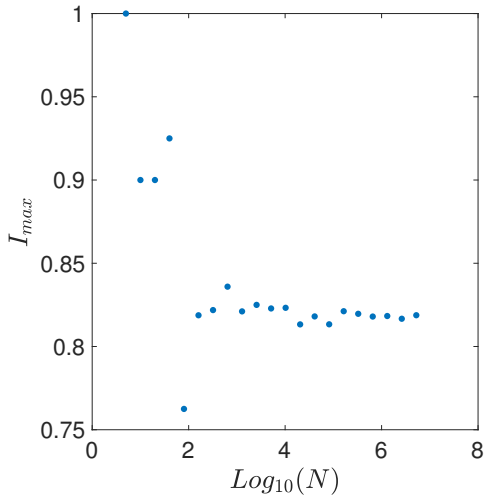

FIG. 7: The maximum in the proportion of infected people (I_{max}) is not accurately determined for small population sizes ($N < \mathcal{O}(10^2)$). For good convergence, larger populations are required.

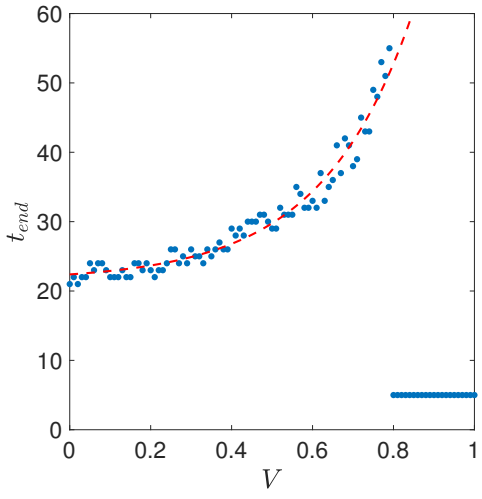

FIG. 8: The time steps required to reach a disease-free population (t_{end}) as a vaccination rate (V), for a population of 10^5 people and a disease with transmission probability $p = 0.5$ that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step. Above a critical vaccination rate (V_c), herd immunity takes effect and the disease does not spread past patient zero.

C. Herd immunity

Increasing the vaccination rate increases the number of time steps required to reach a disease-free population, as shown in Figure 8. This occurs because it is more difficult for the disease to spread to new patients. As the proportion of the population that is vaccinated (V) becomes significant, some of the limited number of random interactions between people will occur with vaccinated (i.e. immune) individuals. There is no chance of disease transmission in these encounters, which become more likely with increasing vaccination rate. This increase in t_{end} is approximately exponential with vaccination rate, and can be well described by: $t_{end} = 21.49 + 0.9013\exp(4.434V)$, with $R^2 = 0.9599$. Above threshold vaccination rate (V_c), the time required for a disease-free population drops immediately to 5, the time taken for patient zero to recover. For $V \geq V_c$, $t_{end} = 5$. Hence, for vaccination rates above a critical value (V_c), the disease is unable to spread. For a population of size $N = 10^5$ people and for a disease with transmission probability $p = 0.5$, the critical vaccination rate $V_c = 0.8$. Above this value, the phenomenon of "herd immunity" protects even members of population that are unvaccinated.

For low vaccination rates, no members of the population escape illness, as shown in Figure 9. As the vaccination rate approaches the critical value, $V \to V_c$, a small proportion of the population make it to the end of the simulation without catching the disease. Above the critical vaccination rate, the number of people reaches the end jumps as herd immunity takes effect. Increasing the vaccination rate further causes the number of people susceptible at the end of the simulation to decrease linearly. This decrease is due to the effective replacement of these susceptible people with vaccinated individuals.

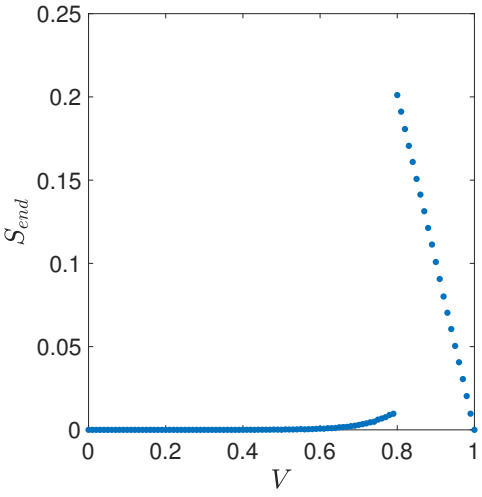

FIG. 9: The proportion of the population that is still susceptible at the end of the simulation (S_{end}) as a function of vaccination rate (V), for a population of 10^5 people and a disease with transmission probability $p = 0.5$ that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step.

By varying the probability of transmission, we can de-

duce the minimum vaccination rate for a population to exhibit herd immunity for a given disease. As Figure 10 shows, the critical vaccination rate required for herd immunity increases with the disease transmission probability. As $p \to 1$, $V_c \to 1$. The numerical values were obtained by running the simulation 10 times for each transmission probability, and taking the mean. The error bars are given by the standard deviation.

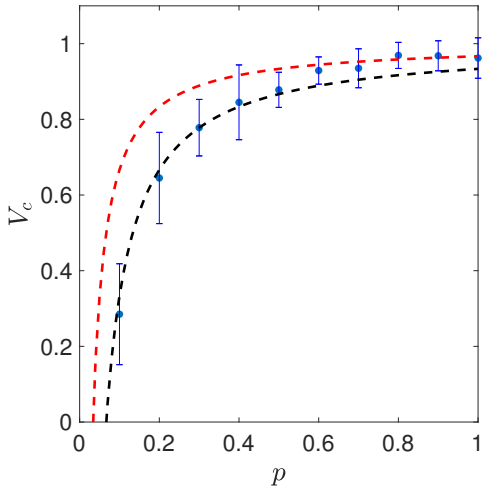

FIG. 10: The vaccination rate required for herd immunity (V_c) as a function of the probability of disease transmission (p), for a population of 10^5 people and a disease that causes the patient to be sick for 5 time steps and infectious to 6 random people in each time step. The analytical estimate of the critical vaccination rate (dashed) is calculated using Equation 1 with duration of infectiousness $d = 5$ and average rate of contact between susceptible and infected people $c = 6$ (red) or $c = 2.5$ (black).

The calculated relationship can be compared to the analytical estimate for the critical vaccination rate[3]:

$$V_c = 1 - \frac{1}{R_0}, \tag{1}$$

where the basic reproduction number (R_0) is given by[4]:

$$R_0 = pcd, \tag{2}$$

where p is the probability of disease transmission, c is the average rate of contact between susceptible and infected people and d is the duration of infectiousness. The basic reproduction number gives the number of people that one infected patient can generate over the course of the infectious period (d). In our simulation, the infectious period $d = 5$.

The average rate of contact between susceptible and infected people c is harder to determine. In our simulation, each individual interacts with 6 individuals per time step. However, this is not equivalent to c because some of the random interactions are between susceptible people and vaccinated, recovered, or other susceptible individuals. Hence $c \leq 6$. In Figure 10, we use Equation 1 with $R_0 = 5 \times 6 \times p$ to give an upper limit. This explains the discrepancy between the analytical estimate (shown in dashed red) and the numerical result (shown as scatter points). Varying c, we find that $c = 2.5$ gives a good fit for the analytical estimate to the numerical results, suggesting this may be a more accurate value for low values of p.

IV. DISCUSSION

In Section III A, we found a threshold in the probability of disease transmission. Below this threshold, diseases are not sufficiently contagious to gain traction in a population. This suggests disease that are only very weakly contagious ($p < 0.05$) will never be able to spread to epidemic proportions among the population. Therefore, these diseases are not a priority in the study of containment of infectious diseases.

Above this threshold, the time taken for a disease to spread through the population decreases with increasing probability of transmission. Therefore, diseases with high probability of transmission should be priorities when considering the management of diseases. This should be weighed against other factors, including severity of disease. Unsurprisingly, diseases with high probability of transmission will affect more people, and infect a greater proportion of the population.

We found the maxima in the proportion of the population that are in the infected stage of the disease peaks at approximately 50–60% of the lifetime of the disease. This suggests that early intervention is crucial in preventing the spread of the disease.

In Section III B, it was seen that in small populations with no vaccination, nobody escapes the disease, even with even modest probabilities of transmission. In large populations with no vaccination, only a very small proportion will escape the disease due to random luck. This underlines the importance of vaccinating effectively.

In Section III C, we studied the effect of vaccination rates on the disease dynamics. Increasing the vaccination rate caused an exponential increase in the time taken for a disease to spread through the population. However, below the herd immunity threshold, it had only a small effect on the number of people that escaped the disease by the end of the simulation. Above the herd immunity threshold, the effect was dramatic, with the entire population escaping the disease.

Varying the probability of disease transmission, the critical vaccination rate for herd immunity rose with the probability of transmission. This highlights the importance of high vaccination rates for diseases with high

probabilities of transmission, for example, Measles ($R_0 \approx 12 - 18$), Diphtheria or Rubella ($R_0 \approx 6 - 7$)[5].

We consider only random vaccination in our simulation. This could be improved by vaccinating high-risk groups. For example, targeting HIV vaccination at drug injectors, or gay and bisexual men will be much more cost effective compared to random vaccination. Targeting MMR vaccines at children will be more effective than randomly vaccinating the entire population. As well as targeting groups that are more likely to spread the disease, it is useful to target groups that are more likely to suffer adverse consequences as a result of catching the disease. For example, vaccinating the elderly and young against flu is crucial, as these groups are more likely to suffer severe effects of the flu, compared to the general population.

The implementation of non-random vaccination requires the addition of further detail to the individuals in the population. This could be achieved using the `Person` class. For example, age could be introduced. Using age, imperfect immunity could be introduced. For example, introducing immunity at age X with immunity lasting for Y years. This would be useful for modelling diseases for which children are a high risk group. For example, the implementation of the measles, mumps and rubella (MMR) vaccine.

We could also improve our simulation by introducing a more sophisticated model for the interaction of individuals in the `Population` class. Currently individuals can spread the disease to random individuals. In reality, individuals are often friends, co-workers or family. A combination of the random and neighbour systems (used in the model development) could address this.

V. CONCLUSIONS

A compartmental model for the spread of infectious diseases was developed. In the model, a disease with probability of transmission p spreads through a population of N individuals that interact with 6 other individuals each day. Infectious individuals remained infectious for 5 days, during which they could pass on the disease to other people. A threshold on the minimum transmission probability was found, below which a disease is too ineffective to propagate through a population. A minimum vaccination rate required to safeguard the rest of the population was found. This herd immunity threshold was found to agree well with existing analytical estimates for an average rate of contact of 2.5 per time step. The simulation could be improved by considering nonrandom vaccination, age-based vaccination, or more sophisticated interactions.

ACKNOWLEDGEMENTS

The author of this article would like to thank the instructors and teaching assistants of the class SDS392 for their guidance and advice in constructing this simulation.

[1] C. Dye. After 2015: infectious diseases in a new era of health and development. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369:20130426, Jun 2014.

[2] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemicsi. *Bulletin of Mathematical Biology*, 53(1-2):33–55, 1927.

[3] P. Fine, K. Eames, and D. L. Heymann. "Herd Immunity": A Rough Guide. *Clinical Infectious Diseases*, 52(7):911–916, 2011.

[4] J. H. Jones. Notes on R_0. 2007.

[5] Center for Disease Control and the World Health Organization. History and epidemiology of global smallpox eradication. from the training course titled "Smallpox: Disease, Prevention, and Intervention", Slide 17, 2015.