FIT3139 ASSIGNMENT 3: STOCHASTIC SIR MODEL

Sharon Zhou

Sharon Zhou

Student ID: 3316040

SECTION 1: Specification Table

Base Model	The original continuous SIR model which divides the population into susceptible, infected and recovered individuals to describe the spread of a contagious disease, utilising a set of ordinary differential equations.
Extension Assumptions	The extension in which this report aims to investigate targets the population structure of the SIR model. The original SIR model assumes a homogenous population. That is, everyone contains the same characteristics. Within this case, this means that their infection rates and recovery rates remain the same. The extension modifies the base SIR model by incorporating age-specific contact rates and recovery rates, dividing the population into three distinct age groups: children, adults, and elderly. Each group has its own transmission and recovery dynamics, represented by a system of stochastic differential equations. This allows for a more realistic representation of disease spread by accounting for differences in interaction patterns and immune response across age groups. The extension introduces randomness in the transmission and recovery processes to capture the inherent stochasticity of infectious disease dynamics.
Techniques Utilised	**Gillespie Algorithm:** Analyse how disease dynamics vary between age groups and how interactions between groups affect overall disease spread. **Monte Carlo:** Used in conjunction with the Gillespie's Algorithm in order to obtain a probabilistic understanding of the disease spread across age groups. **Heuristics:** Utilised in order to find the maximum coverage of vaccination campaigns, given a certain budget threshold.
Modelling Question 1	How does the presence of age-structure within the population affect the spread of disease?
Modelling Question 2	In the case in which age structure affects the spread of disease, what is the most optimal vaccination campaign, when considering budget, and differentiating vaccination costs for each age group? i.e. Are mass vaccination campaigns better, or targeted vaccination campaigns?

Section 2:
Proposed by William O, Kermack and A.G. McKendrick 1927, the S.I.R. model is a classic framework in epidemiology used to understand the spread of infectious diseases within the population. The model assumes that everyone falls within one of the three following compartments:

- Susceptible (S) – individuals who are vulnerable to the infection
- Infected (I) – Individuals who have become infected
- Recovered (R) – Individuals who are now immune and recovered from the infection

Within the SIR model, an assumption is made that the population simulated is homogenous. This indicates that all individuals regardless of race, age-structure, gender are identical in terms of susceptibility, infection rates and recovery rates. However, in a real world context, children, adults and elderly have distinct interaction patterns and responses. For example, children often act as super-spreaders due to high contact rates in school, whilst elderly might be more vulnerable to severe infections due to a weaker physique.

Hence, this paper aims to investigate the impact of these following questions:

1. How does the presence of age-structure within the population affect the spread of disease?
2. In the case in which age structure affects the spread of disease, what is the most optimal vaccination campaign, when considering budget, and differentiating vaccination costs for each age group? i.e. Are mass vaccination campaigns better, or targeted vaccination campaigns?

In order to answer these questions, the following modelling techniques will be utilised:

-

Gillespie's Algorithm- Analyse how disease dynamics vary between age groups and how interactions between groups affect overall disease spread.
The Gillespie algorithm's capacity to capture stochasticity and manage detailed interactions is especially well-suited for examining how disease dynamics differ among age groups and how their interactions impact overall disease transmission and population recovery within the SIR model. It effectively models random events and heterogeneous contact patterns, crucial for age-structured populations where different groups have varied interaction rates and sizes. Additionally, it is able to efficiently track granular epidemic trajectories. Hence, this method takes into consideration the intricacies of interactions that occur in the real world and provides comprehensive insights on the spread of diseases across various age groups.

-

Monte Carlo - Used in conjunction with the Gillespie's Algorithm in order to obtain a probabilistic understanding of the disease spread across age groups.
The Gillespie algorithm utilises Monte Carlo to generate random numbers at each stage of the simulation to predict the next occurrence. Within this case, the events are infection or recovery. Furthermore, the monte carlo approach further determines the time until the next event.

•

Heuristics - Utilised in order to find the maximum coverage of vaccination campaigns given a certain budget.
This is done by utilising a heuristic grid search over vaccination proportions to explore a range of combinations of vaccine proportions. This is in order to complete the objective of maximising vaccination coverage and minimise infections.

Section 3:

First, let's take a look at the original S.I.R. continuous model.

$$\frac{dS}{dt} = -\beta SI$$

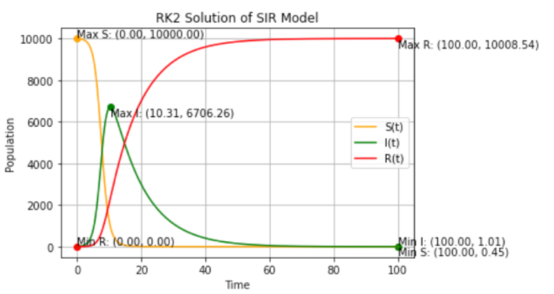
$$\frac{dI}{dt} = \beta SI - \lambda I$$

$$\frac{dR}{dt} = \lambda I$$

Within this case, S, I and R refer to the number of individuals who are in the susceptible, infected and recovered respectively.

β and λ refer to the rate of infection and rate of recovery respectively.

$\frac{dS}{dt}, \frac{dI}{dt}$ and $\frac{dR}{dt}$ refers to the rate of change of the susceptible, infected and recovered population over time respectively.



These equations can be integrated utilising an ODE solver to produce a plot that defines the population over a continuous time interval. It can be seen that the SIR model follows the trend, that as the number of susceptible individuals decrease, the number of infectious individuals start to increase. This is until a certain point until the maximum number of infectious individuals is reached, and members of the population begin to move to a recovered state.

The original S.I.R. model takes on the following key assumptions:

1. The number of individuals within this model will remain constant, not accounting for deaths and births.
2. The population is assumed to be homogenous, meaning that individuals interact with each other randomly and with equal probability.
3. There is no latent or exposed period in which individuals are exposed but not yet infected.
4. Individuals who are recovered are fully immune and may not become infected or susceptible again.
5. Transmission rate and recovery rate remain constant.

This report aims to aims to target the second assumption, that the population is assumed to be homogenous. Hence, the following extension is proposed:

Children

$$\frac{dS_C}{dt} = -\beta_1 S_C I_C - \beta_2 S_C I_A - \beta_3 S_C I_E$$
$$\frac{dI_C}{dt} = \beta_1 S_C I_C + \beta_2 S_C I_A + \beta_3 S_C I_E - \lambda_1 I_C$$
$$\frac{dR_C}{dt} = \lambda_1 I_C$$

Adult

$$\frac{dS_A}{dt} = -\beta_4 S_A I_C - \beta_5 S_A I_A - \beta_6 S_A I_E$$

$$\frac{dI_A}{dt} = \beta_4 S_A I_C + \beta_5 S_A I_A + \beta_6 S_A I_E - \lambda_2 I_A$$

$$\frac{dR_A}{dt} = \lambda_2 I_A$$

Elderly

$$\frac{dS_E}{dt} = -\beta_7 S_E I_C - \beta_8 S_E I_A - \beta_9 S_E I_E$$

$$\frac{dI_E}{dt} = \beta_7 S_E I_C + \beta_8 S_E I_A + \beta_9 S_E I_E - \lambda_3 I_E$$

$$\frac{dR_E}{dt} = \lambda_3 I_E$$

Where:

- c = children
- a = adults
- e = elderly
- β_{1-9} = rates of infection
- λ_{1-3} = rates of recovery

$\frac{dS}{dt}$, $\frac{dI}{dt}$ and $\frac{dR}{dt}$ refers to the rate of change of the susceptible, infected and recovered population over time respectively. The extended model takes a similar idea to the original SIR model. However, it separates the original the population into 3 ages groups, children, adults and elderly, taking into account interactions, between both individuals within the same age group, as well as individuals within different age groups. This is depicted by the terms such as $S_C I_C$ and $S_C I_A$, representing the interaction between susceptible children and infected children, and susceptible children and infected adults respectively.

Additionally, the β_{1-9} and λ_{1-3} remain constant. That is, transmission rates between varying interactions remain constant, and recovery rates per age group remain constant. This differs from the original S.I.R. model assumption, as a homogenous population assumed that all individuals had the same infection rate and recovery rate.

Within this model, the following assumptions will be made.

The original S.I.R. model takes on the following key assumptions:

1. The number of individuals within this model will remain constant, not accounting for deaths and births.
2. The population is assumed to be heterogenous, meaning that individuals will not interact with equal probability.
3. There is no latent or exposed period in which individuals are exposed but not yet infected.
4. Individuals who are recovered are fully immune and may not become infected or susceptible again.
5. Transmission rate and recovery rate remain constant.

Furthermore, the rates of infection and recovery in order to simulate real world interactions. For example, for recovery rates, the older the individual, the lower their recovery rate. This is due to the idea that older individuals tend to take a weaker physique.

PROPENSITIES OF THE GILLESPIE MODEL

The propensities that were generated are as follows:

$$infection\ rate_{children} = S_c * (\beta_0 * \frac{(CM_{00}*I_c)}{C_{population}} + \beta_1 * \frac{(CM_{01}*I_A)}{A_{population}} + \beta_2 * \frac{(CM_{02}*I_E)}{E_{population}})$$

$$infection\ rate_{adult} = S_A * \left(\beta_3 * \frac{(CM_{10} * I_c)}{C_{population}} + \beta_4 * \frac{(CM_{11} * I_A)}{A_{population}} + \beta_5 * \frac{(CM_{12} * I_E)}{E_{population}}\right)$$

$$infection\ rate_{elderly} = S_E * \left(\beta_6 * \frac{(CM_{20} * I_c)}{C_{population}} + \beta_7 * \frac{(CM_{21} * I_A)}{A_{population}} + \beta_8 * \frac{(CM_{22} * I_E)}{E_{population}}\right)$$

$$recovery\_rate_{children} = \lambda_1 I_C$$

$$recovery\_rate_{adult} = \lambda_2 I_A$$

$$recovery\_rate_{elderly} = \lambda_3 I_E$$

Where CM refers to the contact matrix presented within the results section. That is, it will contain the number of individuals who come in contact. The numbers refer to the position on the matrix.
For example, CM_{00} = 30 and CM_{10}= 2

Within this case, the infection rates have been normalised. This is in order to ensure the model is scaled with population size in order to ensure accurate infection rates.

MODEL ANALYSIS

Much like that of the original SIR model, the extended SIR model presented within this report is a continuous model. This is due to the idea that the differential equations describe how the number of people within each compartment change over time.

Additionally, both models are numerical in nature. The ordinary differential equations within the original SIR continuous model, are usually solved with methods such as the Euler Method or Runge-Kutta method. These methods approximate the solutions of the differential equation. Similarly, the Gillespie's algorithm also provides an approximation in order to solve the differential equations of the stochastic model.

Both the original and extended SIR models are considered to be non-linear. This is due to the idea that the amount of people within each compartment differ between different periods. Hence, interactions between people, as depicted by the SI terms, will also be non-linear.

However, in comparison, to the deterministic nature of the original SIR model, the extended SIR model is considered to be stochastic. The original SIR model utilises a set of ordinary differential equations to describe the rate of change of the population of different compartments over time. The SIR model assumes a homogenous population where rates of infection and recovery are consistent across the whole population and randomness within events and recovery are not considered.

Conversely, within the case of the age-structured S.I.R. model, random events which influence the spread of disease are simulated. This includes the random nature of contact between individuals, or the probabilistic nature of transmission and recovery.

Section 4:
QUESTION 1: What is the effects of age-structure on the spread of disease within the population?

In a real world context, age structure plays a significant role in understanding the spread of disease. This is due to the idea that different age groups may exhibit varying susceptibilities towards the disease due to factors such as immunity levels, as well as varying contact rates due to different social behaviours and social interactions.

This question aims to explore the effects of age structure on the spread of disease within the entire population utilising a population of 243 children, 615 adults and 140 elderly. Note that this is the proportion of children, adults and elderly within Melbourne.

The contact rates per person/ contract matrix is presented as follows.

	Children	Adults	Elderly
1 child	30	10	2
1 adult	2	20	2
1 elderly	2	2	1

For example, one child may come in contact with 30 children during school, 10 adults and 2 elderly at home per day.

The infection rates are as follows:
B1: 0.8, B2: 0.6, B3: 0.4, B4: 0.6 ,B5: 0.7, B6: 0.5, B7: 0.4, B8: 0.5, B9: 0.6

And recovery rates are as follows assuming that the older an individual is, the weaker their physique:
C = 0.7, A = 0.6, E = 0.3

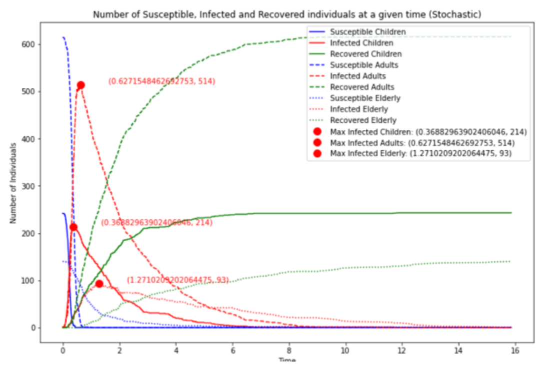


Table 1: Splits population into children, adults and elderly

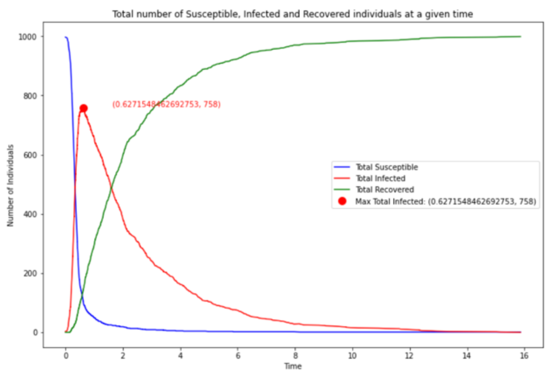


Table 2: Total population of susceptible, infectious and recovered compartments

Now let us compare this graph to three homogenous populations which considers all individuals to contain the same statistics adults, children and elderly respectively.

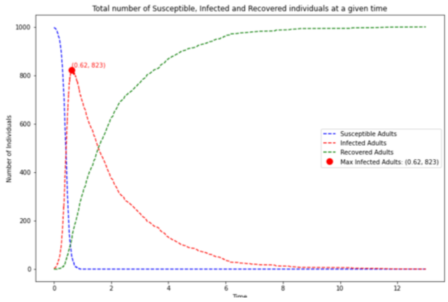


Table 3: Assuming all 1000 individuals are adults

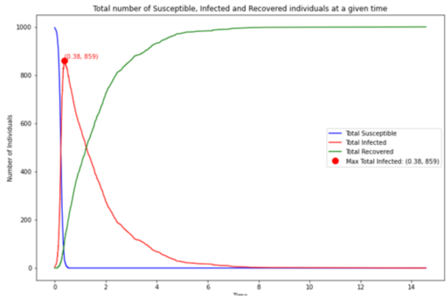


Table 4: Assuming all 1000 individuals are children

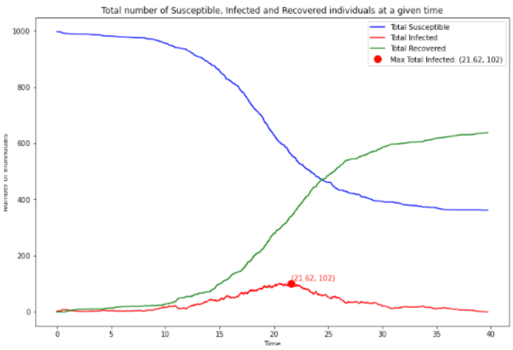


Table 5: Assuming all 1000 individuals are elderly

When comparing Table 2 against Table 3, 4 and 5, it can be seen that the number Table 2 is most similar to Table 3. In stating this, there are still slight differences within the model. Within table 2, it can be seen that the maximum number of infected individuals at any given time when considering a mixture of age groups, 758 people, is lower than that of the maximum number of infected individuals, assuming all individuals are adults. However, the point in which the peak infections are reached remain extremely similar.

Note that Table 2 is most similar to Table 3 due to the idea that a large proportion of individuals tested within the age-structured SIR model are adults. Hence, it can be inferred that the largest group within a heterogenous population, within this case, an age-structured population, will have the largest effect on the overall spread of disease.

Now let us take a look at a homogenous population of 243 people if they all contained characteristics of children.

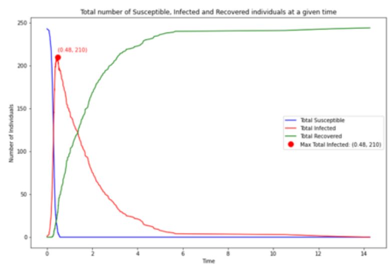


Table 6: Homogenous population of 243 children

When looking at the lines depicting the patterns of infected children Table 1, compared to table 4 and 6, it can be observed that the lines within Table 1 seem to have more fluctuations and inconsistencies within the shape. This may be due to the more intricate and complex nature of interactions within an age-structured population. That is, in a heterogenous population where interactions are more dynamic, contact patterns vary significantly, which in turn, results in a higher variability of infection numbers Conversely, if the entire population had the same statistics, the contact rates will remain uniformly high. This may lead to the spread of infection to become more evenly distributed, leading to less fluctuation within the curves.

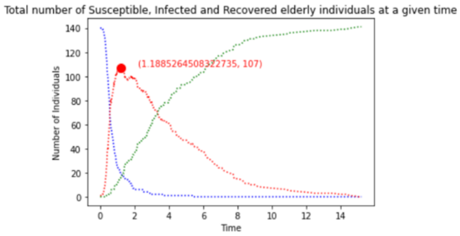


Table 7: Homogenous population of 140 elderly, who come in contact with one other elderly person, as well as adults and children

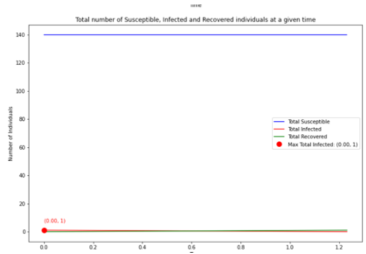


Table 8: Homogenous population of 140 elderly, who come in contact with one other elderly person

Now, let us compare the S.I.R. distribution of elderly who come in contact with adults and children, to the S.I.R. distribution of elderly individuals who only come in contact with other elderly individuals (Table 7 and Table 8). When comparing these two tables, a stark difference between the patterns can be observed. Whereas Table 7 takes on a pattern in which individuals move from susceptible, towards infected and eventually recovered, in table 8, all individuals seem to stay within the same compartment. This indicates that interactions amongst differing age groups has induced a quicker spread of disease amongst the elderly population. Hence, it can be concluded that in incorporating age-structure and varying interaction rates, an increase or decrease within the rates of infection may be produced within a particular age group. Furthermore, through these comparisons, it can be seen that even if a homogenous SIR model aims to capture the dynamics of interactions between one certain age group, these results are unrepresentative of the overall interactions in which they may have within a real-world context, further emphasising the importance of capturing the interactions between varying groups. Hence, an age-structured SIR model will more accurately model the spread of disease for each age group to simulate the dynamics of the real world.

Taking all of this into account, that effect in which age-structure may have on the number of peak infections and the herd immunity may differ depending on the initial parameters assumed. However, in splitting the population into heterogenous groups, it can be observed that within a real world context, the overall spread of disease is largely effected by the largest group. Furthermore, in incorporating interactions between various groups, contact patterns become more complex allowing us to observe that interactions between various age groups, may lead to more variability within infections, as compared to a groups which contain homogenous interactions. This may further indicate that an age-structured model may lead to more random infection outbreaks. Additionally, interactions between different age groups may lead to increases and decreases within the spread of disease amongst a particular group.

As in incorporating interactions between various groups, we are able to gain a better understanding of the conditions and dynamics of each specific age group , the importance of age-structure to model the spread of disease cannot be overstated.

Question 2: In the case in which age structure affects the spread of disease, what is the most optimal vaccination campaign, given a certain budget, as well as accounting for varying costs of vaccinations? i.e. Are mass vaccination campaigns better, or targeted vaccination campaigns?

In order to conduct this analysis, the following parameters were defined.

Number of adults: 10000
Number of children: 5000
Number of elderly: 3000

Cost of adult vaccination: $20
Cost of child vaccination: $10
Cost of elderly vaccination:$30

The cost of vaccinations varies between age group due to assumption that older individuals have a weaker physique. Hence, they will need a higher dosage of the vaccination.

Heuristics and Targetted Vaccination Heat map Analysis

In order to answer the given question, the concept of heuristics is utilised in order to obtain the maximum vaccination coverage for both vaccination strategies, given a certain budget. Whereas mass vaccinations do not discriminate against the proportion of individuals vaccinated per age group, the targeted vaccination campaign does. Hence, it is important to showcase how the targeted vaccinations work, and which groups are prioritised in order to achieve maximal coverage.

The below heat maps showcases the relationship between the proportion of vaccinated individuals per age group which would allow for the maximum vaccination coverage within a given budget. Note that within certain cells, a value of 0 is given. This may be due to the idea that the budget is too low to account for the given combination of vaccinations.

For example, within Table 1, the bottom right cell containing information about the combination of 100 percent of children being vaccinated and 100 percent of adults being vaccinated, contains the value of 0. This is due to the idea that vaccinating 100 percent of these two age groups would cost $25000. This exceeds the budget of $10,000 that was utilised within the example to produce these heatmaps.

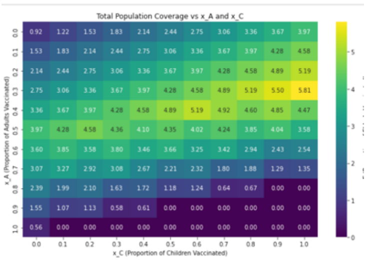


Table 1: Children vs Adults

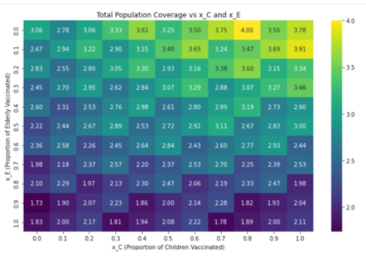


Table 2: Children vs Elderly

Table 3: Adults vs Elderly

Table 1 models the relationship between the proportion of children who should be vaccinated and the proportion of adults who should be vaccinated in order to gain maximal vaccine coverage, when taking into account various proportions of elderly vaccinations.

Within the case of a $10,000 budget, it can be seen that children vaccinations should be prioritised over adult vaccinations in order to gain the optimal vaccination coverage.

Table 2 models the relationship between the proportion of children who should be vaccinated and the proportion of elderly who should be vaccinated in order to gain maximal vaccine coverage, when taking into account various proportions of adult vaccinations.

Within the case of a $10,000 budget, it can be seen that children vaccinations should be prioritised over elderly vaccinations in order to gain the optimal vaccination coverage.

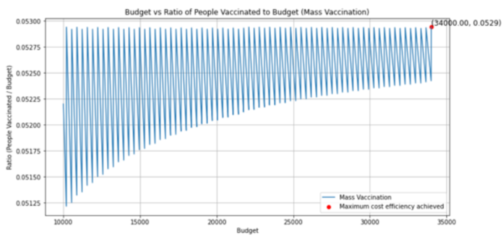
Table 3 models the relationship between the proportion of adults who should be vaccinated and the proportion of elderly who should be vaccinated in order to gain maximal vaccine coverage, when taking into account various proportions of children vaccinations.

Within this case of a $1000 budget, it can be seen that adult vaccinations should be prioritised over elderly vaccinations in order to gain the optimal vaccination coverage.

Hence, when conducting targeted vaccination campaigns, the order of prioritisation should be to vaccinate all children first, then adults and finally elderly individuals.

<u>Cost efficiency analysis</u>

Given a range of spending from 10,000- 34,000 dollars, let us now conduct a cost efficiency analysis on mass vaccinations. That is, let us find the optimal amount of money spent for mass vaccinations, that would allow us to achieve the most number of people vaccinated per dollar spent.



Here, it can be observed that the spending of $34000 will result in the most number of people vaccinated per dollar spent.

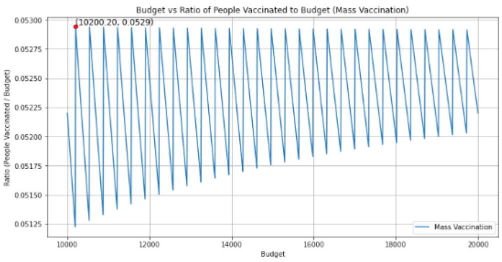


Figure 1: Cost efficiency analysis of mass vaccination strategy

Given a budget from the range 10,000-20,000, it can be observed that the spending of $10200.20 dollars will result in the most number of people vaccinated per dollar spent. When observing the trend of this graph, a fluctuating zig-zag pattern can be seen. This is due to the idea that at certain points, increasing the budget will not result in a higher number of individuals who become vaccinated. Hence, these points are potentially less cost efficient than previous points. Whereas the local maximums within this graph tends to stay around the same point along the y-axis, the y-values of the local minimums increase in a logarithmic pattern. This is due to the idea that when the budget increases, between 2 local minimums, the proportion of increase in people vaccinated is larger than the increase of budget. Additionally, as the budget increases , the difference in y-values between the local maximum points, as well as their adjacent local minimum points decrease. This indicates that at higher budgets, the amount of money allocated towards vaccinations matters less in terms of cost efficiency.

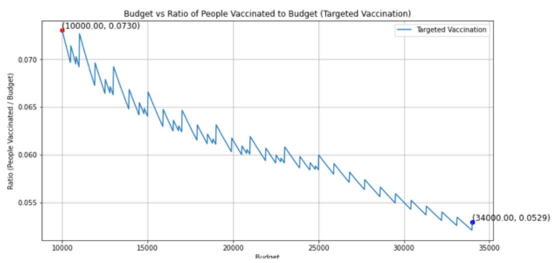


Figure 2: Cost efficiency analysis of targeted vaccination strategy

Now let us conduct the same cost efficiency analysis on targeted vaccination campaigns. Within this case, it can be seen that a spending of $10000 is most cost efficient, resulting in the most number of people vaccinated per dollar spent when targeted vaccinations are used. In comparison to the mass vaccination campaign, Figure 2 takes a general decreasing trend with fluctuations, indicating that the higher the budget set, the less cost effect the vaccination campaign is. However, much like Figure 1, Figure 2 also presents the idea that the higher the budget, the amount of money allocated towards vaccinations matters less in terms of cost efficiency. This is emphasised by the decreasing size of the fluctuations.

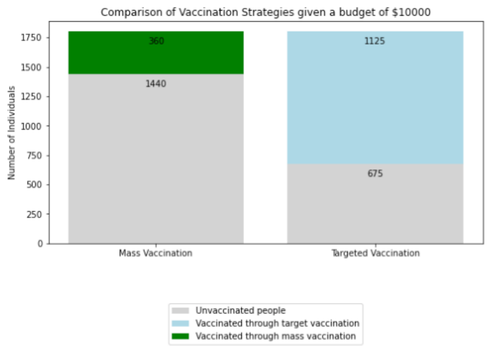
When comparing the maximum points, it can be seen that in spite of the decreasing trend shown by Figure 2, targeted vaccinations are more cost efficient. This is due to the idea that when conducting targeted vaccinations, spending $10,000 will result in 0.072 people becoming vaccinated per dollar spent. However, spending $10200 on the mass vaccination campaign only results in 0.0529 people becoming vaccinated per dollar spent. Furthermore, spending $34000, the amount of money which achieves the maximum amount of cost efficiency when utilising mass vaccinations, will also only result in 0.0529 people becoming vaccinated per dollar spent.

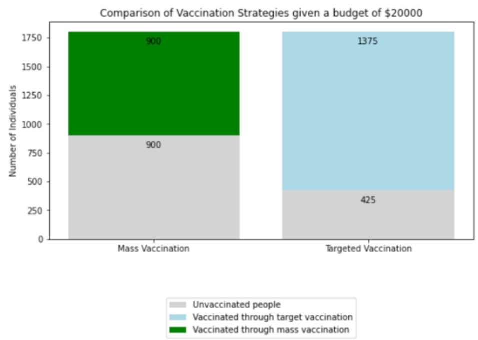
In fact, when comparing these two graphs, it can be seen that a higher number of people will be vaccinated per dollar spent when utilising a targeted vaccination campaign, even if the same budget is spent on the two campaigns.

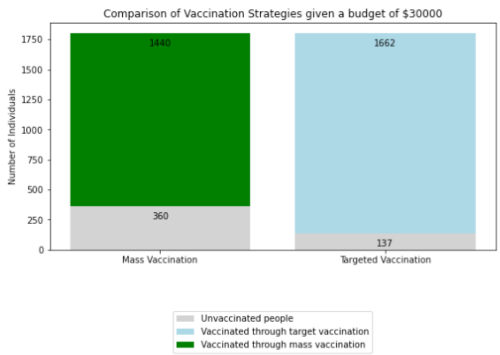
It is only slightly before the budget is set to 34000 dollars , that the cost efficiency of the targeted vaccination campaign falls below a rate of 0.0529, which is the maximum cost efficiency within mass vaccination campaigns.

As the overall cost efficiency of the targeted vaccination campaign is higher than that of the mass vaccination campaign, this indicates that targeted vaccinations may be more effective.

<u>Comparison of vaccination strategies in terms of vaccination coverage</u>







Another way in which vaccination strategies may be compared, is by comparing the total vaccination coverage. Within all three graphs, regardless of budget, the vaccination coverage achieved by utilising the targeted vaccination campaign is higher than that of the mass vaccination campaign.

In completing an analysis of the cost efficiency, as well as total vaccination coverage, it can be concluded that when accounting for age structure, a targeted vaccination campaign is most efficient when considering costs of vaccination and total budget are considered as well.

Section 5: List of algorithms used

Gillespie's algorithm- Used to model the spread of distribution when age-structure is accounted for.

Monte-Carlo algorithm- Used in conjunction with the Gillespie's Algorithm in order to obtain a probabilistic understanding of the disease spread across age groups

Heuristics-Used to determine the proportion of populations vaccinated in order to gain maximal coverage within a given budget.

Normalisation- Normalisation was used in order to ensure the model is scaled so that infection calculations are correctly made

Heat maps- containing information about how the targeted vaccination campaign works

Bar charts- Used to model the proportion of individuals vaccinated

Line graphs-Used to model the cost-efficiency analysis, as well as the spread of the SIR model.

Seaborn package- Utilised to produce heat maps

Matplotlib – Plots variety of different graphs

Numpy