Modelling of Infectious Disease Spread and Control Strategies

Dan Williams

April 2025

1 Introduction

The spread of infectious diseases has been a long-standing and complex challenge to global health which impacts millions and regularly tests the resilience of healthcare systems around the world. These diseases can severely strain health resources, which in turn increases death rates and creates long-term economic instability (*Global Report on Infection Prevention and Control* 2022). During the early stages of an infectious disease outbreak, government authorities need to be able to quickly determine how best to deal with infection and how to prevent exponential spread. This is important to ensure that the infection has the smallest possible impact on the population to prevent a significant outbreak. Recently, the COVID-19 pandemic has highlighted how quickly infections can spread in a modern world that is densely populated and highly interconnected (Teller 2021).

Mathematical modelling is a crucial method for understanding and predicting the behaviour of infectious diseases and their impact on populations without the need for testing on real populations. By examining factors like rates of infection and recovery, researchers can create accurate models that simulate disease spread and evaluate possible interventions that governments can use to reduce spread (Kucharski 2020). Traditionally, we have used compartmental models. These models have divided the population into different "compartments", representing various stages of infection, such as susceptible (individuals who could contract the disease), infected (individuals who currently have the disease), and recovered (individuals who are no longer infected). The susceptible-infected-recovered (SIR) model created by British mathematicians W. O. Kermack and A. G. McKendrick in 1927 serves as a fundamental framework for many of these models, offering a simple depiction of infection dynamics (Kermack and McKendrick 1927). Many of the models used today are derivatives of this original framework as it can be enhanced to account for additional disease factors like vaccination levels or herd immunity.

The aim of this report is to build improved models upon the original SIR framework by incorporating some of these disease factors, using this to simulate the spread of an infectious disease and assess how effective different control strategies are. This will allow for a more accurate and realistic simulation of the disease and how different factors within the population might affect the outbreak. The report will also discuss any limitations of these improvements and any factors that might affect the accuracy of the predictions made, especially the importance of using accurate and sufficient data sources.

2 Background

Compartmental models are the foundation of epidemiology models, primarily the modelling of infectious diseases and the control strategies for these diseases. These models split the population in different "compartments" each of which represents the stage of the infection that individual is in. Most commonly these are *susceptible*, the individual is susceptible to infection, *infected*, the individual is currently infected, and *recovered*, the individual is no longer infected. These form the basis of the susceptible-infected-recovered (SIR) model framework (Kermack and McKendrick 1927).

We start by defining the respective compartments of the model as functions of t:

- $S(t)$: the number of individuals that are susceptible to infection at time t.
- $I(t)$: the number of individuals that are currently infected at time t.
- $R(t)$: the number of individuals that have recovered/removed from infection at time t.

The SIR model is governed by the following system of ordinary differential equations (ODEs):

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

where β is the rate of infection and γ is the rate of recovery. The term βSI represents the number of new infections per unit time and is derived from the assumption that each susceptible individual S has a probability β of infection upon contact with an infectious individual I. Similarly, the term γI represents the number of new individuals recovered per unit time and is derived from the assumption that each infectious individual I has a probability γ of recovery (Kermack and McKendrick 1927).

An important quantity in epidemiology is the basic reproduction number R_0, sometimes referred to as the epidemic threshold. This is a measure of how many secondary infections a single infectious individual will cause in a completely susceptible population. In the SIR model, this is defined:

$$R_0 = \frac{\beta}{\gamma}$$

If $R_0 < 1$, the infection cannot grow within the population and will eventually become eradicated over time, depending on how few secondary infections occur. On the other hand, if $R_0 > 1$ then the infection will be able to grow within the population and easily spread through the individuals. The larger the number of secondary infections that occur, the more widespread the infection will be and could lead to a severe outbreak (Delamater et al. 2019).

One of the limitations with this model is that it is a *deterministic system*, which means that no randomness is used in this system. So it assumes a fixed outcome based on the initial conditions. This approach makes it difficult for the model to incorporate real-world variations that might occur. For example, the rate of infection β is a constant value in the model but in reality would change over time, it would likely increase during colder months and decrease following public health and social measures (PHSMs) (Ryu et al. 2022). In addition, β and γ are fixed for the entire population so every individual has an equal probability of becoming infected with the disease and equal probability of recovery. This does not reflect real-world situations accurately, for example an isolated individual would have a significantly lower probability as an individual in a densely populated area (Teller 2021) and high-risk individuals would have a

lower probability of recovery due to factors such as age or pre-existing conditions. The model also omits births and deaths, so it assumes that the total population N remains constant:

$$N = S(t) + I(t) + R(t)$$

Whilst this makes the equations simple it adds unrealism to the model, especially when a disease has a high mortality rate, and could deviate simulation results of long-term models where population decline affects the dynamics of the disease.

Using the SIR model as a framework, improvements can be built upon this that add additional disease factors and use stochastic variables to more accurately represent real-world dynamics. This will enable the model to better simulate the more complex variables that are present in real-world environments and produce an improved prediction of how different control strategies will work.

3 Numerical Approximation of the SIR Model

A method for solving ordinary differential equations is Euler's method. This provides a first-order approximation by iterating forward in small time steps Δt using the update formula:

$$y_{n+1} = y_n + f(y_n, t_n)\Delta t.$$

Applying Euler's method to the SIR model gives the following:

$$S_{n+1} = S_n - \beta S_n I_n \Delta t, \quad I_{n+1} = I_n + (\beta S_n I_n - \gamma I_n)\Delta t, \quad R_{n+1} = R_n + \gamma I_n \Delta t.$$

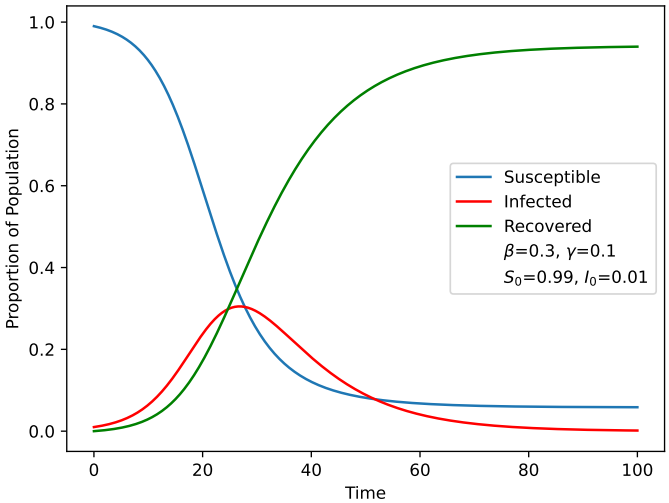


Figure 1: Numerical approximation of the SIR model using Euler's method, generated using Python and matplotlib. Source: Author.

Euler's method is widely used in epidemiological modelling and remains accurate as long as Δt is sufficiently small (Keeling 2008). A small step size improves precision by reducing truncation errors but this leads to an increased computational requirement. This is particularly relevant in real-time epidemic simulations where it is important to balance computational efficiency and numerical accuracy.

An important application of this numerical approach is determining key outbreak characteristics. $I_{\max}$ is a key outbreak characteristic, representing the maximum number of individuals infected simultaneously during an outbreak. From the SIR equations:

$$\frac{dR}{dt} = \gamma I, \quad \frac{dS}{dt} = -\beta SI \implies \frac{dR}{dS} = \frac{\gamma I}{-\beta SI} = -\frac{\gamma}{\beta S}$$

Separating variables and integrating,

$$\int dR = -\frac{\gamma}{\beta} \int \frac{1}{S} dS$$

The left hand side integrates to $R - R_0$ and the right hand side evaluates as

$$-\frac{\gamma}{\beta}[\ln S]\Big|_{S_0}^{S} = -\frac{\gamma}{\beta}\left[\ln S - \ln S_0\right]$$

So,

$$R - R_0 = \frac{\gamma}{\beta}\left[\ln S_0 - \ln S\right]$$

Thus,

$$R(t) = R_0 + \frac{\gamma}{\beta} \ln\left(\frac{S_0}{S(t)}\right)$$

The peak occurs at time $t_{I\max}$ when the number of infected individuals reaches its highest value before declining, meaning that $dI/dt = 0$. Using this, we get:

$$\beta S(t_{I\max}) I(t_{I\max}) - \gamma I(t_{I\max}) = 0 \implies S(t_{I\max}) = \frac{\gamma}{\beta}$$

Substituting this,

$$R(t_{I\max}) = R_0 + \frac{\gamma}{\beta} \ln\left(\frac{S_0}{\gamma/\beta}\right)$$

Since the total population $N = S + I + R$, solving for $I_{\max}$ gives

$$I_{\max} = N - S(t_{I\max}) - R(t_{I\max})$$

Substituting the expressions for $S(t_{I\max})$ and $R(t_{I\max})$,

$$I_{\max} = S_0 + I_0 + R_0 - \frac{\gamma}{\beta} - \left[R_0 + \frac{\gamma}{\beta} \ln\left(\frac{S_0}{\gamma/\beta}\right)\right]$$

which simplifies to

$$I_{\max} = S_0 + I_0 - \frac{\gamma}{\beta} - \frac{\gamma}{\beta} \ln\left(\frac{S_0}{\gamma/\beta}\right)$$

This is arguably the most important measure when considering control strategies during an outbreak. It can be used to determine when the maximum burden on healthcare resources will occur and can then be used to make decisions regarding PHSMs (Ryu et al. 2022). For the values used in the numerical approximation (Figure 1), with initial conditions $S_0 = 0.99$, $I_0 = 0.01$, and parameters $\beta = 0.3$, $\gamma = 0.1$, we compute $I_{\max}$:

$$I_{\max} = 0.99 + 0.01 - \frac{0.1}{0.3} - \frac{0.1}{0.3} \ln\left(\frac{0.99}{0.1/0.3}\right) = 0.303813$$

This value matches the peak infection level displayed on the graph demonstrating that the numerical approximation is correct.

4 Model Development

There are a variety of different ways that the SIR model can be extended, these can generally be put into two categories: temporal and spatial extensions. Temporal extensions are incorporating time-dependent factors into the model, such as any variables that might change over the course of the outbreak. Spatial extensions are concerned with geographical or geometric factors, such as how interactions between individuals would determine how the infection might spread. The intention of this report is to talk about a few of these extensions and the benefits these changes will have on predictions made using the model.

4.1 Demography

Demographic factors can capture how a population might naturally change during the infection and can have a significant influence on long-term disease behaviour. Allowing the model to account for natural changes in the population, such as births and deaths, enables a better representation of disease dynamics. The basic SIR model makes the assumption that the population remains a constant N, omitting natural births and deaths that occur during the infection. We can incorporate a death rate μ and birth rate Λ. We define the model as follows (Keeling 2008):

$$\frac{dS}{dt} = \Lambda - \beta SI - \mu S, \quad \frac{dI}{dt} = \beta SI - \gamma I - \mu I, \quad \frac{dR}{dt} = \gamma I - \mu R \tag{1}$$

with the basic reproductive number $R_0 = \dfrac{\beta\Lambda}{\mu(\gamma+\mu)}$, where $\gamma + \mu$ is the rate at which an infectious individual becomes recovered/removed $I \to R$. This can then be used to determine an equilibrium point, reached as $t \to \infty$, where $(S(t), I(t), R(t))$ approach constant values; in other words, when:

$$\frac{dS}{dt} = 0, \quad \frac{dI}{dt} = 0, \quad \frac{dR}{dt} = 0$$

If $R_0 \leq 1$, the infection cannot grow within the population so we approach the disease-free equilibrium (DFE). At this equilibrium, there are no infectious individuals ($I = 0$). Using $\frac{dS}{dt} = \Lambda - \beta SI - \mu S = 0$ and substituting $I = 0$:

$$\frac{dS}{dt} = \Lambda - \mu S = 0 \implies S = \frac{\Lambda}{\mu}$$

Similarly, using $\frac{dR}{dt} = \gamma I - \mu R = 0$ and substituting $I = 0$:

$$\frac{dR}{dt} = -\mu R = 0 \implies R = 0$$

Thus, the DFE is: $(S, I, R) = \left(\frac{\Lambda}{\mu}, 0, 0\right)$.

If $R_0 > 1$ and $I(0) > 0$, the infection persists and stabilises at the endemic equilibrium (EE), where all compartments remain constant. With some more calculations and algebraic manipulation we obtain the EE as:

$$(S, I, R) = \left(\frac{\gamma+\mu}{\beta}, \frac{\mu}{\beta}(R_0 - 1), \frac{\gamma}{\beta}(R_0 - 1)\right)$$

We can look at the stability of these equilibrium points to understand how demography factors into the behaviour of the outbreak. To determine the conditions for which the DFE and EE

are stable we can look at the Jacobian matrix (Keeling 2008). This is the matrix of first-order partial derivatives which represents a linear approximation for how a small difference in the variables can change the system. For our model, this is defined as:

$$J = \begin{bmatrix} \frac{\partial(dS/dt)}{\partial S} & \frac{\partial(dS/dt)}{\partial I} & \frac{\partial(dS/dt)}{\partial R} \\ \frac{\partial(dI/dt)}{\partial S} & \frac{\partial(dI/dt)}{\partial I} & \frac{\partial(dI/dt)}{\partial R} \\ \frac{\partial(dR/dt)}{\partial S} & \frac{\partial(dR/dt)}{\partial I} & \frac{\partial(dR/dt)}{\partial R} \end{bmatrix} \implies \begin{bmatrix} -\beta I - \mu & -\beta S & 0 \\ \beta I & \beta S - \gamma - \mu & 0 \\ 0 & \gamma & -\mu \end{bmatrix}$$

Substituting in the DFE $(S, I, R) = \left(\frac{\Lambda}{\mu}, 0, 0\right)$ we get:

$$J_{\text{DFE}} = \begin{bmatrix} -\mu & -\beta\frac{\Lambda}{\mu} & 0 \\ 0 & \beta\frac{\Lambda}{\mu} - \gamma - \mu & 0 \\ 0 & \gamma & -\mu \end{bmatrix}$$

We can use the Hartman-Grobman theorem which states that if the eigenvalues of the Jacobian matrix have all negative real parts then the equilibrium point is asymptotically stable. We determine that the eigenvalues of this matrix are:

$$\lambda_1 = -\mu \text{ (double root)}, \quad \lambda_2 = \beta\frac{\Lambda}{\mu} - \gamma - \mu = (\gamma + \mu)(R_0 - 1)$$

- λ_1: This eigenvalue is always negative because the death rate $\mu > 0$.
- λ_2: This eigenvalue depends on the basic reproduction number $R_0 = \frac{\beta\Lambda}{\mu(\gamma+\mu)}$.
  - If $R_0 < 1$, then $\lambda_2 < 0$, thus the DFE is stable.
  - If $R_0 > 1$, then $\lambda_2 > 0$, thus the DFE becomes unstable and so the system stabilises at the endemic equilibrium.

To investigate EE stability, we can perform similar operations as done above for DFE. We obtain the following Jacobian matrix:

$$J_{\text{EE}} = \begin{bmatrix} -\mu R_0 & -(\gamma + \mu) & 0 \\ \mu(R_0 - 1) & 0 & 0 \\ 0 & \gamma & -\mu \end{bmatrix}$$

We can easily determine that $\lambda_3 = -\mu$. To determine the conditions required for the remaining eigenvalues to be negative, we can perform some calculations to arrive at the following equation:

$$\lambda^2 + \mu R_0\,\lambda + \mu(\gamma + \mu)(R_0 - 1) = 0$$

If $R_0 > 0$, this quadratic has solutions with negative real parts. So our remaining eigenvalues will be negative. Therefore, the endemic equilibrium is stable if $R_0 > 1$ otherwise the disease-free equilibrium is stable.

Investigating the stability of these equilibrium points is an important factor when predicting disease dynamics. It enables researchers to determine whether a disease will die out completely (approaching the disease-free equilibrium point) or remain at a stable rate within the population (stabilising at the endemic equilibrium point). This can determine the measures used by public health organisations. If the DFE is stable, efforts can be focused on ensuring that R_0 remains less than 1 such as through vaccinations and improving recovery times. On the other hand, if the EE is stable efforts need to be focused on limiting interactions between individuals or detecting more individuals during the incubation period (infected but not infectious) so that the DFE stabilises (Keeling 2008).

4.2 Stochasticity

The inherent randomness that occurs during these infections cannot be modelled accurately in the deterministic SIR model as it fails to capture the fluctuations that might occur in the model's variables over time. The fundamentals of the stochasticity of the SIR model are the transitions between the states (S, I, R). These transitions are:

- Infection: $S, I \to S-1, I+1$, with rate βSI
- Recovery: $I, R \to I-1, R+1$, with rate γI

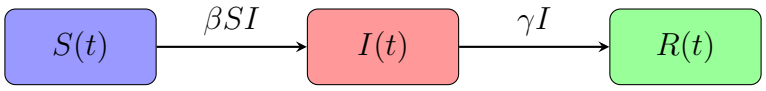


Figure 2: Flowchart of the SIR model showing transitions with βSI and γI.

The stochastic system is described by the following Chemical Master Equation (CME), which was adapted from (Pájaro et al. 2022) where $\mathbb{P}(S, I, R, t)$ denotes the probability that the system is in state (S, I, R) at time t.

$$\begin{aligned}\frac{\partial \mathbb{P}(S,I,R,t)}{\partial t} = &\underbrace{\beta(S+1)I \cdot \mathbb{P}(S+1, I-1, R, t)}_{\text{Influx due to infection}} - \underbrace{\beta SI \cdot \mathbb{P}(S,I,R,t)}_{\text{Outflux due to infection}} \\ &+ \underbrace{\gamma(I+1) \cdot \mathbb{P}(S, I+1, R-1, t)}_{\text{Influx due to recovery}} - \underbrace{\gamma I \cdot \mathbb{P}(S,I,R,t)}_{\text{Outflux due to recovery}}\end{aligned}$$

We can further improve the randomness of the system by generalising the constant rates β, γ to time-varying rates β_t, γ_t. The difficulty with explicitly stating an equation for these values is that they can be extremely unpredictable and variations can occur for a number of different reasons. Some variations might be caused by behavioural changes, for example PHSMs such as a population going into mass isolation (lockdown), or they might be caused by seasonal variations and how well the infection performs during different weather conditions. This was something that dramatically affected the COVID-19 outbreak, causing a large spike in infections in the colder winter months (Inaida, Paul, and Matsuno 2022).

$$\beta_{t1} = \begin{cases} \beta_0 & \text{if no intervention,} \\ \beta_1 & \text{if lockdown is imposed at } t > t_L \end{cases}, \qquad \beta_{t2} = \beta_0 \left(1 + \alpha \sin\left(\frac{2\pi t}{T}\right)\right)$$

Figure 3: Two example equations for β: β_{t1} a piecewise function representing the effect of a PHSM and β_{t2} a sinusoidal function representing seasonal fluctuations with period T.

We can include these time-varying rates β, γ into the stochastic system described above using the Gillespie Algorithm (Gillespie 1977). The algorithm is a method of simulating stochastic processes and will allow us to include the random fluctuations and variations that occur during an infection's outbreak. The algorithm is implemented as follows (Cole 2020):

Algorithm 1 Gillespie Algorithm

1. **Initialise**: Set initial conditions $S(0)$, $I(0)$, $R(0)$ and the time-varying rates β_t, γ_t.
2. **Calculate event rates**: Calculate rate of infection $\beta_t SI$ and rate of recovery $\gamma_t I$.
3. **Monte Carlo sampling**: Using the Poisson distributions, determine the number of infections and recoveries during time interval Δt :

$$\Delta N_I \sim \text{Poisson}(\beta_t SI \cdot \Delta t), \quad \Delta N_R \sim \text{Poisson}(\gamma_t I \cdot \Delta t).$$

4. **Update**: Update the state variables:

$$S = S - \Delta N_I, \quad I = I + \Delta N_I - \Delta N_R, \quad R = R + \Delta N_R.$$

5. **Repeat**: Repeat steps 2–4 until no infectious individuals remain ($I = 0$).

Using Python and matplotlib, we can generate a graph illustrating the impact of stochasticity on the disease dynamics, specifically examining the effect of halving β during an outbreak. At $t = 0$, the initial conditions for a closed population of $N = 1000$ are set as $S = 990$, $I = 10$, and $R = 0$.

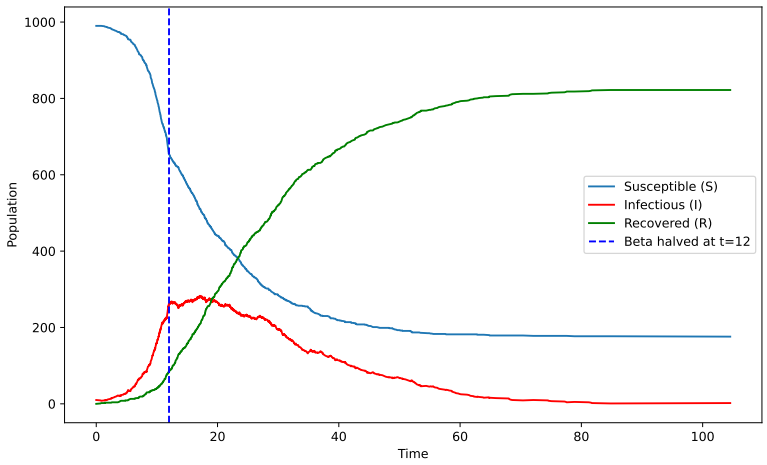


Figure 4: Stochastic SIR simulation generated using Python and Matplotlib. Source: Author.

Plotting this stochastic model on a graph, you can see the impact that a time-varying β rate has on how quickly the infection spreads. Between $t = 0$ and $t = 12$, the number of infectious individuals grows very quickly but at $t > 12$, when β is halved, the infection growth plateaus and the number of infectious individuals declines steadily. This demonstrates the effect a PHSM could have on an infection and why mass quarantines are regularly used in the control of an outbreak.

4.3 Spatial Heterogeneity

Spatial heterogeneity is the uneven distribution of various concentrations of a population within an area. We can use it to describe how disease dynamics vary across different geographical regions and how they differ within a distributed population. The SIR model assumes that a

population is well-mixed so that every individual interacts equally with others and this interaction occurs regardless of where each individual is located. However this isn't the case in a real-world scenario. In addition, spatial factors such as population density and movement of individuals influence disease transmission efficiency (Teller 2021). As a result, an accurate model needs to be able to consider these factors to determine whether an individual can actually be infected by a certain individual.

We can represent the population using a graph to consider this. Let $G = (V, E)$ be a graph, where V is the set of vertices and E is the set of edges. Each vertex $v \in V$ represents an individual in the population and has a state (S, I, R) corresponding to the compartment that the individual belongs to. The edges $e_{ij} \in E$ represent connections between individuals i and j and show a potential infection route. The structure of the graph can show different segments of the population. For example, a densely populated urban area would correspond to a highly connected subgraph.

Let $p_{ij}(t)$ represent the probability that an individual i infects j at time t. This probability can depend on factors such as the distance between nodes i and j (which relates to a real-world separation between the two individuals) or the infectiousness of the disease β (which would determine how well the disease would spread). We can incorporate this using weighted edges, where the weight w_{ij} of an edge e_{ij} represents the likelihood of an infection or strength of interaction. For example, a high weight edge indicates frequent or prolonged contact (so has a high likelihood of infecting). We can define the probability of transmission for some time step Δt as $\mathbb{P}_{ij}(t) = \beta w_{ij} \Delta t$.

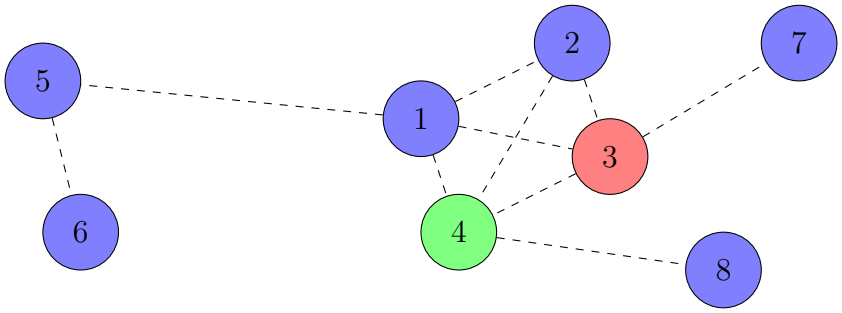
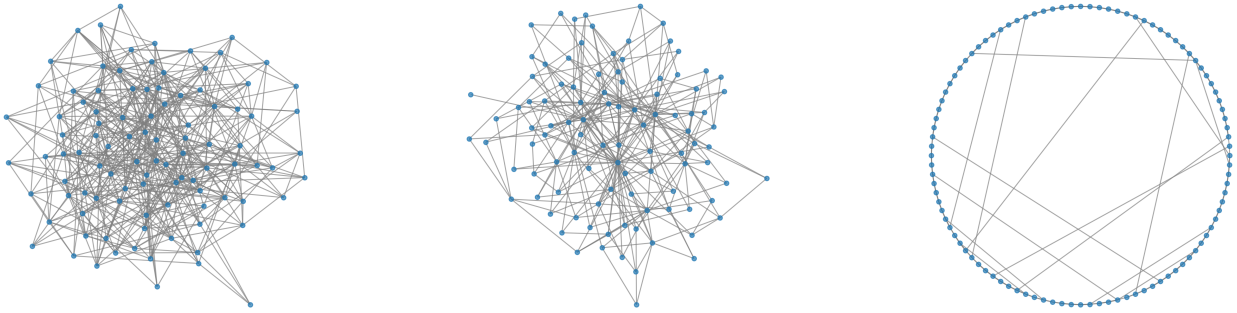


Figure 5: SIR model states visualised on a graph. Blue nodes represent susceptible individuals, red represents an infected individual, and green represents a recovered individual. The subgraph consisting of nodes 1, 2, 3, and 4 represent a densely connected population. Source: Author.

Network topologies describe the arrangement of edges between nodes in the graph and different topologies have distinct outbreak patterns. In random networks, edges are formed with equal probability so there is uniform connectivity making outbreaks generally predictable. On the other hand, in a scale-free network where there exists a few highly connected nodes (sometimes referred to as hubs) rapid and widespread transmission occurs. Once these hubs become infectious they become "super-spreaders" and will cause a significant number of secondary infections even at a low rate of infection (Moreno and Vázquez 2003).

A frequently used topology is the small-world network. This has high local clustering and short global paths which encapsulates human interaction. For example, high local clustering might represent households and workplaces where individuals frequently interact. The global paths could represent social events where individuals interact that might not be in your local cluster. These global paths enable rapid transmission across the network even if the local clusters are

well-contained. This makes the small-world network topology a very good representation when considering infections in human populations.



Random Topology | Scale-Free Topology | Small-World Topology

Figure 6: Comparison of different network topologies with 100 individuals created using Python and networkx. Source: Author.

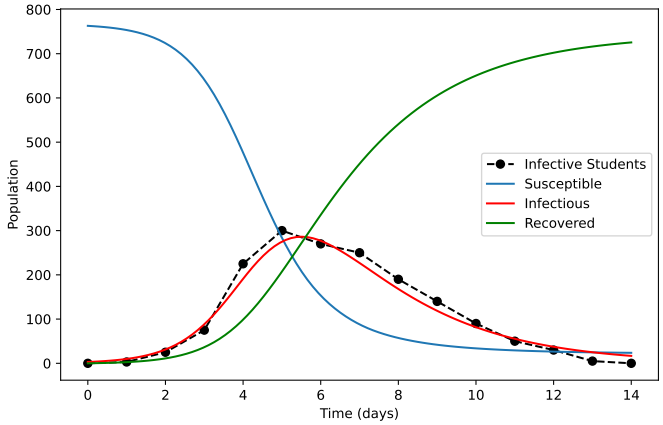
There are a wide range of network topologies that can be employed in modelling disease transmission each of which vary in heterogeneity, clustering, and average path length. As a result, these differences can capture distinct disease dynamics and patterns of interaction.

5 Applications

The models developed in this project have traditionally been applied to health settings, considering the spread of viruses that impact human health. However, the versatility of these models allow us to consider the spread of a variety of other phenomena such as behaviour, networks, informatics, and economic contagion.

5.1 Influenza in a boarding school

In early 1978, a British boarding school experienced an outbreak of influenza A virus (IAV) creating an epidemic in a closed setting (Keeling 2008). Three students went to the school infirmary displaying symptoms of influenza. Within a few days, the infection spread rapidly among the 763 students at the school. The epidemic lasted two weeks before the infection was extinguished ($I = 0$). We can plot the number of daily infective students on a graph and use this to estimate the parameters for the SIR model. By applying a simple least squares procedure, we can find the parameter values that best fit the observed infection curve.



Graph created using Python and matplotlib where the filled circles represent the number of students infected on a given day of the epidemic (data sourced from Anon 1978), along with solutions from SIR model fitted to the data using least squares procedure where $\beta = 1.66$ and $\gamma = 1/2.2$ (Keeling 2008).

Using this, we can estimate $R_0 = \frac{\beta}{\gamma} = 1.66 \times 2.2 = 3.65$. We can see on the graph that these parameters fit the observed infection curve very well. We can see that the infectious population reaches a peak around 5 days. The infection curve is a key factor in understanding how quickly the disease spreads and ensuring that the school infirmary is prepared and able to cope during the peak. The susceptible population is decreasing throughout the epidemic, as no new students joined or left the school during this period (i.e. no births or deaths) and recovered students gained immunity. This meant that the epidemic followed a classic SIR trajectory and eventually reached a point where $I = 0$ and the infection was extinguished. This suggests that the system approached a disease-free equilibrium.

However, a number of uninfected students remained susceptible at the end of the epidemic, denoted by S_∞. We can estimate this using (Miller 2012):

$$S_\infty = S_0 e^{-R_0(1-S_\infty/N)}$$

This provides an implicit relationship for the susceptible population at the end of the epidemic. Solving numerically with $S_0 = 763$ and $R_0 = 3.65$, we find that approximately $S_\infty \approx 22$ students remained uninfected at the end of the epidemic confirming that a significant proportion of the population was infected before the epidemic naturally died out. This aligns with the concept of herd immunity. This is an indirect form of protection that occurs when a large enough portion of a population has immunity to the disease, either through infection or vaccination.

5.2 Epidemic spread in social networks

The severe acute respiratory syndrome (SARS) outbreak from late 2002 to 2004 was the first epidemic of the 21st century. Originating in China, it spread to 30 countries infecting over 8000 individuals and causing 774 deaths (World Health Organization 2015). The outbreak was driven by "super-spreader" events where highly connected individuals caused a large number of secondary infections which accelerated transmission.

One of these events occurred in the Amoy Gardens residential complex, Hong Kong, where a single infected individual was linked to over 300 secondary infections within the complex. The transmission dynamics here were significantly different from those in the rest of Hong Kong, as shown in Table 1.

Parameter	**Amoy Gardens**	**Hong Kong (Feb - Mar)**	**Hong Kong (Mar - Jun)**
β	1.4850	0.2586	0.1351
γ	0.9750	0.0821	0.1923
R_0	1.5154	3.1511	0.7025

Table 1: Parameters for the SIR model during the SARS outbreak at Amoy Gardens and the Hong Kong Special Administrative Region adapted from Mkhatshwa and Mummert 2010.

We can see that the rate of infection β at Amoy Gardens was significantly higher ($\beta = 1.4850$) compared to the rest of Hong Kong in Feb - Mar ($\beta = 0.2586$) and even lower in Mar - June ($\beta = 0.1351$). This indicates that when modelling the SARS outbreak, we need to consider spatial heterogeneity. As transmission in Amoy Gardens was highly localised and amplified by dense connectivity, and given the significant role of "super spreader" events in the outbreak, a scale-free network is the most suitable modelling framework to use.

Scale-free networks are characterised by a power-law degree distribution, where a small number of highly connected nodes (hubs) dominate transmission. Namely, the fraction p_k of nodes with

k connections follows a power-law scaling for large k given by $p_k \sim k^{-\gamma}$, where $2 < \gamma < 3$ is the degree exponent which determines the skewness of the degree distribution or how common these hubs are (Pastor-Satorras and Vespignani 2001).

The epidemic threshold is a fundamental property of epidemic dynamics which determines whether a disease will die out or persist. Scale-free networks can sustain transmission even when the rate of infection falls below this threshold. This is due to the hubs that continuously propagate infections. The epidemic threshold λ_c in a scale-free network is given by (Morita 2021):

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

where $\langle k \rangle$ is the average degree and $\langle k^2 \rangle$ is the second moment of the degree distribution, given by $\langle k^2 \rangle = \sum_k k^2 p_k$, which quantifies the variance in connectivity across the network. For large k, we can approximate the summation using integrals (Integral approximation of sums):

$$\langle k^2 \rangle = \sum_k k^2 p_k \approx \sum_k k^2 k^{-\gamma} \approx \int k^2 k^{-\gamma}\, dk = \int k^{2-\gamma}\, dk.$$

Evaluating the integral:

$$\langle k^2 \rangle = \int_{k_{\min}}^{k_{\max}} k^{2-\gamma}\, dk = \frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{3-\gamma}.$$

Since $k_{\max} \gg k_{\min}$ for large networks, the term $k_{\max}^{3-\gamma}$ dominates, leading to:

$$\langle k^2 \rangle \sim \frac{k_{\max}^{3-\gamma}}{3-\gamma}.$$

In scale-free networks with $2 < \gamma < 3$, $\langle k^2 \rangle$ diverges, so $\lambda_c \to 0$. Thus, transmission persists regardless of the rate of infection. The outbreak at Amoy Gardens demonstrates the real-world dangers of these highly connected hubs which result in a large number of infections. In this case, poorly-designed sewage infrastructure acted as high-degree nodes facilitating infection spread. The localised nature of this meant that even with a lower reproductive number ($R_0 = 1.5154$) compared to the rest of Hong Kong ($R_0 = 3.1511$) during Feb - March, transmission persisted at a higher rate.

5.3 Vector-borne diseases

Vector-borne diseases are infections that require an intermediary host for transmission. Their spread follows a two-step process:

- A susceptible vector bites an infected human and becomes infected.
- An infectious vector bites a susceptible human, transmitting the disease.

As a result, both human and vector populations must be accounted for in these models. The formulation of this system is a variation of the Ross-Macdonald model (Simoy and Aparicio 2020) and extends the SIR model with demography, as defined in Equation (1). The system is governed by the equations:

$$\frac{dS_h}{dt} = \Lambda_h - \beta_h S_h I_v - \mu_h S_h, \quad \frac{dI_h}{dt} = \beta_h S_h I_v - \gamma_h I_h - \mu_h I_h, \quad \frac{dR_h}{dt} = \gamma_h I_h - \mu_h R_h$$

$$\frac{dS_v}{dt} = \Lambda_v - \beta_v S_v I_h - \mu_v S_v, \quad \frac{dI_v}{dt} = \beta_v S_v I_h - \mu_v I_v$$

We define compartments for susceptible (S_h), infected (I_h), and recovered (R_h) humans, as well as susceptible (S_v) and infected (I_v) vectors, which never recover from infection. The parameters β_h and β_v represent the human-to-vector and vector-to-human transmission rates, μ_v denotes the vector mortality rate, and Λ_h and Λ_v represent the birth rates of humans and vectors. The basic reproductive number R_0 for this model is given by:

$$R_0 = \frac{\beta_h \beta_v}{(\gamma_h + \mu_h)\mu_v} \frac{V}{H}$$

where V and H denote the total vector and human populations respectively.

Malaria is a life-threatening infection caused by *Plasmodium* parasites and is transmitted to humans through the bites of infected *Anopheles* mosquitoes. Globally, there were more than 260 million malaria cases and over 597,000 deaths in the most recent reporting year (World Health Organization 2024). As malaria requires mosquitoes to transfer parasites between humans, the dynamics of the outbreak are not only influenced by human factors but also by mosquito population dynamics.

In Figure 7, we define the rate of infection from vectors to humans as $\beta_h = 0.2$ and from humans to vectors as $\beta_v = 0.18$. The human recovery rate is $\gamma_h = 1/14$ (infectious period of 14 days). The natural death rate of humans is $\mu_h = 1/(60 \times 365)$ (average lifespan of 60 years) and mosquitoes of $\mu_v = \frac{1}{10}$ (average lifespan of 10 days). Setting initial populations of $H = 10,000$ humans and $V = 1,000$ mosquitoes of which 10 are infected. At time $t = 23$ we deploy a vector intervention simulating a realistic deployment of a short-term control strategy (e.g ULV Fogging).

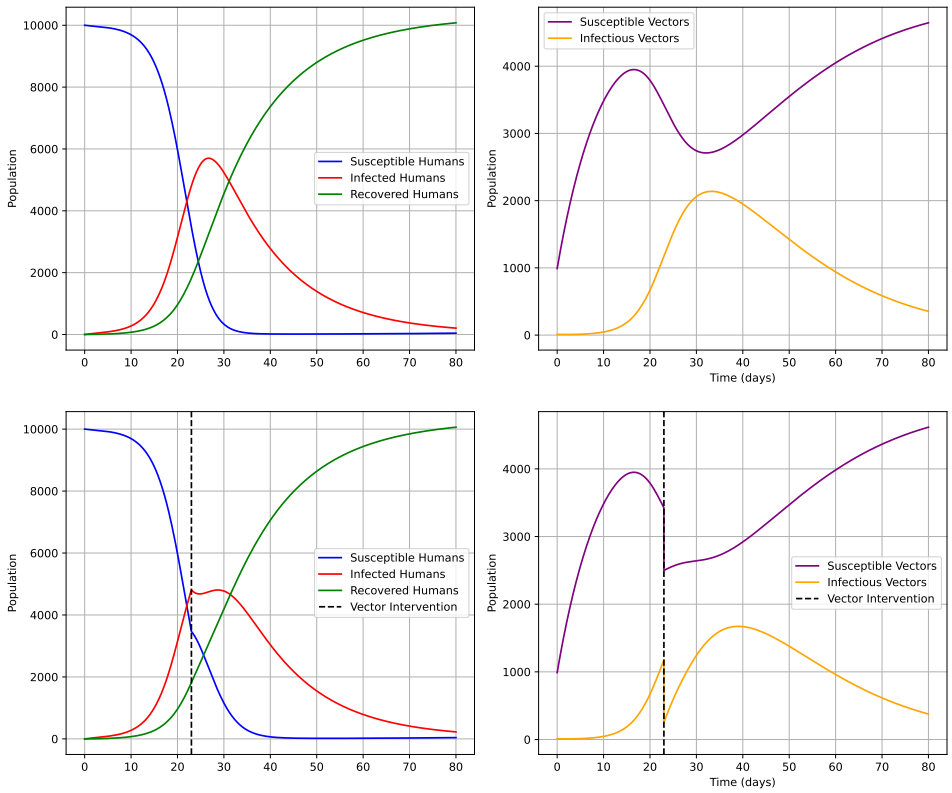


Figure 7: SIR simulations of malaria transmission with and without vector intervention generated using Python and Matplotlib. Source: Author.

Figure 7 shows the impact of vector intervention on the dynamics of human infection compared

to a scenario without intervention. One possible intervention that could produce these effects is ultra-low volume (ULV) fogging. Insecticides are aerosolised and sprayed over a large area to rapidly kill mosquitoes. This results in an immediate drop in the mosquito population but does not provide lasting protection. One example of an intervention that would give continuing protection is insecticide-treated bed nets (ITNs), which are mosquito nets that are coated with insecticides to kill or repel mosquitoes. Another example is indoor residual spraying (IRS), which is the application of long-lasting insecticides to walls that kill mosquitoes that rest indoors. These interventions, referred to as vector control strategies, differ from public health and social measures (PHSMs) as they specifically target the mosquito population (Lobo et al. 2018).

One of the limitations with the Ross-Macdonald model is that it is deterministic. This means that, given the same initial conditions, the same outcome is always produced. Malaria is inherently stochastic due to variations in mosquito populations, human movement, and environmental factors such as rainfall and temperature. Stochasticity can be introduced using Gillespie's Algorithm, as defined in Algorithm 1.

6 Conclusion

This report has demonstrated the importance and usefulness of mathematical modelling in understanding and predicting the spread of infectious diseases. Building upon the classic SIR model, we extended the model to incorporate demography, stochasticity, and spatial heterogeneity, improving its application to real-world outbreaks. We also implemented these extensions to real-world applications and demonstrated their effectiveness using simulations. Incorporating demography (births and deaths) allowed for a more realistic simulation for long-term epidemic behaviour modelling. We also demonstrated that not all populations are well-mixed so spatial heterogeneity is critical for considering how contact networks shape transmission. A key insight from this study is the the limitation of deterministic models in representing the variability of disease spread. They fail to capture the random fluctuations caused by human behaviour, environmental factors, and especially intervention strategies. The implementation of Gillespie's Algorithm allowed us to introduce stochasticity to the model and enabled it to account for this randomness. Additionally, we examined the impact of control strategies for both public health and social measures (PHSMs) in human-to-human outbreaks and vector interventions in malaria transmission. These simulations highlight the effectiveness of different intervention allowing authorities to make data-driven decisions before enforcing them.

In the future, the model could be extended further by introducing additional compartments such as exposed (E) for individuals who have been infected but are not yet infectious, quarantined (Q) for individuals who are in isolation, and vaccinated (V) for individuals who have received immunity through vaccination. This would allow for a more detailed analysis of control strategies in scenarios, such as COVID-19, where asymptomatic transmission plays a significant role (Rui et al. 2024). Further work could investigate how we can refine the estimations of infection β and recovery γ rates. Techniques such as machine learning could be employed to calibrate these parameters using data from previous studies and real-world observations (Cheng, Aruchunan, and Noor Aziz 2025).

References

Anon (Mar. 1978). "Influenza in a boarding school". In: *British Medical Journal* 1.6113, p. 587.

Cheng, Cheng, Elayaraja Aruchunan, and Muhamad Hifzhudin Noor Aziz (2025). "Leveraging dynamics informed neural networks for predictive modeling of COVID-19 spread: a hybrid SEIRV-DNNs approach". eng. In: *Scientific reports* 15.1, pp. 2043–13. ISSN: 2045-2322.

Cole, Lewis (2020). *Gillespie Algorithm*. Accessed November 22, 2024. URL: https://lewiscoleblog.com/gillespie-algorithm.

Delamater, Paul L et al. (2019). "Complexity of the Basic Reproduction Number (R 0)". eng. In: *Emerging infectious diseases* 25.1, pp. 1–4. ISSN: 1080-6040.

Gillespie, Daniel T (1977). "Exact stochastic simulation of coupled chemical reactions". eng. In: *Journal of physical chemistry (1952)* 81.25, pp. 2340–2361. ISSN: 0022-3654.

Global Report on Infection Prevention and Control (2022). eng. 1st ed. ISBN: 9789240051164.

Inaida, S., R. E. Paul, and S. Matsuno (Sept. 2022). "Viral transmissibility of SARS-CoV-2 accelerates in the winter, similarly to influenza epidemics". In: *American Journal of Infection Control* 50.9, pp. 1070–1076.

Keeling, Matthew James (2008). *Modeling infectious diseases in humans and animals / Matt J. Keeling and Pejman Rohani.* eng. Princeton ; Oxford.

Kermack, W. O. and A. G. McKendrick (Aug. 1927). "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London Series A* 115.772, pp. 700–721. DOI: 10.1098/rspa.1927.0118.

Kucharski, Adam (2020). *The rules of contagion [electronic resource] / why things spread - and why they stop / Adam Kucharski.* eng. ISBN: 9781782834304.

Lobo, Neil F et al. (2018). "Modern Vector Control". eng. In: *Cold Spring Harbor perspectives in medicine* 8.1, a025643. ISSN: 2157-1422.

Miller, Joel C. (Sept. 2012). "A note on the derivation of epidemic final sizes". In: *Bulletin of Mathematical Biology* 74.9, pp. 2125–2141. DOI: 10.1007/s11538-012-9749-6.

Mkhatshwa, Thembinkosi and Anna Mummert (2010). "Modeling Super-spreading Events for Infectious Diseases: Case Study SARS". eng. In.

Moreno, Y. and A. Vázquez (Jan. 2003). "Disease spreading in structured scale-free networks". In: *The European Physical Journal B - Condensed Matter* 31.2, pp. 265–271. ISSN: 1434-6036.

Morita, Satoru (2021). "Solvable epidemic model on degree-correlated networks". eng. In: *Physica A* 563, p. 125419. ISSN: 0378-4371.

Pájaro, M. et al. (Nov. 2022). "Stochastic SIR model predicts the evolution of COVID-19 epidemics from public health and wastewater data in small and medium-sized municipalities: A one year study". In: *Chaos, Solitons & Fractals* 164, p. 112671.

Pastor-Satorras, Romualdo and Alessandro Vespignani (2001). "Epidemic spreading in scale-free networks". eng. In: *Physical review letters* 86.14, pp. 3200–3203. ISSN: 0031-9007.

Rui, Jia et al. (2024). "MODELS: a six-step framework for developing an infectious disease model". eng. In: *Infectious diseases of poverty* 13.1, pp. 30–30. ISSN: 2049-9957.

Ryu, Sukhyun et al. (2022). "Epidemiology and Transmission Dynamics of Infectious Diseases and Control Measures". eng. In: *Viruses* 14.11, p. 2510. ISSN: 1999-4915.

Simoy, Mario Ignacio and Juan Pablo Aparicio (2020). "Ross-Macdonald models: Which one should we use?" eng. In: *Acta tropica* 207, pp. 105452–105452. ISSN: 0001-706X.

Teller, Jacques (2021). "Urban density and Covid-19: towards an adaptive approach". eng. In: *Buildings & cities* 2.1, pp. 150–165. ISSN: 2632-6655.

World Health Organization (2015). *Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003*. Accessed: 2025-01-21. URL: https://www.who.int/publications/m/item/summary-of-probable-sars-cases-with-onset-of-illness-from-1-november-2002-to-31-july-2003.

— (2024). *Malaria Fact Sheet*. Accessed: 2025-02-06. URL: https://www.who.int/news-room/fact-sheets/detail/malaria.