EXPLORING THE IMPACT OF BUBBLE STRATEGIES ON THE SPREAD OF INFECTIOUS DISEASES

by

Jingyu Li

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2023

Table of Contents

List of Tables

List of Figures

Abstract

The emergence of the COVID-19 pandemic has led to a great deal of scientific interest in strategies to effectively deal with epidemics. Compartmental models, such as the Susceptible-Exposed-Infectious-Removed (SEIR) model, are popular tools that allow us to study the behaviour of various epidemics. The SEIR model classifies people into groups according to their health status. However, this model assumes that the population is well-mixed which does not take into account public health policies such as social distance and bubble strategies. In this thesis we will construct the Bubble SEIR model based on the continuous-time Markov chain SEIR model. The Bubble SEIR model divides the population into subpopulations called bubbles. The model assumes that each subpopulation is well-mixed and also allows the possibility of movements between bubbles. The purpose of the thesis is studying the impact of the bubble strategies on the epidemic by simulations using the proposed model. Our simulations demonstrate that bubble strategies are effective in controlling the spread of infectious diseases. In the simulation study, we will discuss the effects of parameters of the model including rate of movement between bubbles, basic reproduction number, latency period, population size and initial number of infecteds, on the final size and the duration of the epidemic. Specifically, although changes in the infection rate and the removed rate could influence the final size of the epidemic, the variations in the basic reproduction number have a significantly greater impact. The increase of the latency period and the rate of movement between bubbles increases the final size and the duration respectively. When the population size increases, having more bubbles prove more effective in controlling epidemics. Furthermore, the final size of the epidemic is smaller when the initial number of infectious individuals in the population is small.

Acknowledgements

I would like to extend my heartfelt gratitude to the following individuals who played instrumental roles in making this research project a reality. Thanks to my two supervisors, Dr. Edward Susko and Dr. Lam Ho, for their unwavering support, invaluable guidance and encouragement throughout the research process. They have provided me with precious guidance at every stage of this research as well as the thesis, from formulating the initial research questions to analysing the results and drawing meaningful conclusions. Their patience and willingness to invest time in detailed discussions greatly enriched the quality of my work. I was honoured to be able to study under their supervision! I would like to thank my family for their understanding and support of my studies, and my friends for taking care of me and helping me. I would also like to thank my cats, Darla and Gill, for being there for me during my working days and nights.

Chapter 1

Introduction

The spread of infectious diseases is a significant public health concern that affects millions of people worldwide each year. There is a great deal of concern about the frequent international outbreaks of infectious diseases such as Ebola, SARS, coronavirus and many others. With public health and life safety at stake, understanding and predicting epidemic behaviour as well as informing decisions related to disease control and prevention is invaluable. Epidemic modelling, in turn, is a tool for knowing the dynamics of disease transmission and assessing the effectiveness of various control measures. One class of models, referred to as compartmental models, classifies people into several compartments, such as susceptible compartment S, exposed compartment E, infected compartment I and removed compartment R, and then different models are created based on feasible compartment transitions. Classical models of infectious disease include the Susceptible-Infectious-Susceptible(SIS) model, which allows a person to return from an infected compartment to a susceptible compartment and then be re-infected. The Susceptible-Infectious-Removed(SIR) model allows for an individual who has removed from an infection and is not re-infected. The Susceptible-Exposed-Infectious-Removed(SEIR) model allows that the infectious disease under study has an incubation period and that once in contact with a susceptible person one can become a potential carrier of the virus with a certain incubation period. Variations of SEIR models will be a focus of this thesis.

Compartmental models like SEIR are frequently expressed deterministically through a set of differential equations. Such models implicitly assume relatively large population sizes. In deterministic models, if the rate of secondary infections, referred to as the basic reproduction number, is high enough, everybody in the population is eventually expected to get infected. Stochastic versions of SEIR models apply regardless of population sizes. Due to their stochastic nature, even with large reproduction numbers, there is some chance that not everyone in the population will be infected. By

simulating the dynamics of disease transmission under different scenarios, stochastic epidemic models can help us to better understand the potential impact of interventions, and to design strategies for mitigating the spread of infectious diseases. Scholars at Dalhousie University, for instance, used a modified age-stratified SEIR model and genetic algorithms to investigate optimal budgetary solution for Nova Scotia's closure policies, travel restrictions and other policies (Gillis et al., 2021). This provides a pragmatic direction for the application of epidemiological models.

In this thesis, we will develop stochastic epidemic models to analyse the impact of bubble strategies. Specifically, we focus on the Bubble SEIR model where the population is divided into subpopulation, individuals can move between compartments in subpopulations or individuals can move between subpopulations. Our objectives are twofold: first, to develop a stochastic SEIR model that can segment the overall population and accurately capture the dynamics of disease transmission within and between population groups; second, to use the model to study the effects of various parameters on the number of infecteds, such as the rate of movement between bubbles, basic reproduction number, the latency rate, the total population size and the initial number of infecteds. We then explore whether bubble strategies help to control the spread of the epidemic.

To achieve these objectives, we begin by reviewing the existing literature on stochastic epidemic models and the SIS model in particular. We then develop a continuous-time Markov chain Bubble SEIR model to simulate the spread of disease which we have implemented in R. We use measures like the final number of people infected by the disease and the expected duration. Overall, this thesis contributes to our understanding of the dynamics of infectious disease transmission and provides valuable insights into the design of effective control measures.

Chapter 2

Background

2.1 Markov Chain

The Russian mathematician Andrey Markov studied and proposed a model to try to explain mathematically the general laws of natural variation, which was named Markov Chain (Gagniuc, 2017). Markov chains are stochastic processes that assume that, given the entire past history of the process, the probability distribution of the next state is determined by the current state, and that the events preceding it in the time series are irrelevant. This is a special property of Markov chains, which we call the Markov property.

We may explain Markov chains in a mathematical way. If we have a set of random variables and define each random variable as X_t, then the set $\{X_t, t \in T\}$ represents stochastic process (Ross, 2014). We consider the index t as time and X_t as the state of the process for each time t. When T is a countable set, we have a *discrete-time* stochastic process. When T is an uncountable set, we have a *continuous-time* stochastic process. We define the state space as the set of all possible values. Suppose we have non-negative, integer-valued sequential states X_0, X_1, X_2,..., X_n and X_n is the present state, then the process is a Markov chain if the conditional probability of X_{n+1} only depends on the X_n. Formally, if we have

$$P\{X_{n+1} = j \big| X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0\} = P\{X_{n+1} = j \big| X_n = i\} \tag{2.1}$$

whenever $P\{X_n = i, X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0\} > 0$, we call this discrete stochastic process a discrete-time Markov chain (DTMC). Furthermore when a DTMC has probabilities $p_{ij} = P\{X_{n+1} = j \big| X_n = i\}$ that are independent of time n, we say that this Markov chain is time homogeneous. A transition probability matrix

is one way to represent a time-homogeneous Markov chain. If is defined as

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \\ p_{i0} & p_{i1} & p_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

where the term p_{ij} in the matrix is the probability that the process moves to state j at time $n+1$ given that the process is in state i at time n.

By analogy with the previous definition, we can refer to a continuous stochastic process $\{X(t), t \geq 0\}$ having the Markov property as a continuous-time Markov chain (CTMC). A CTMC is a system that evolves continuously, allowing transitions between compartments to occur at any time rather than just at discrete intervals. The probability of moving from one state to another depends solely on the present state of the chain and Δt, which is referred to as the Markov property. In this case, the Markov property will be

$$P\{X(t+\Delta t) = j \big| X(t) = i, X(u) = x(u), 0 \leq u < t\} = P\{X(t+\Delta t) = j \big| X(t) = i\} \tag{2.2}$$

for all $t, \Delta t \geq 0$ and nonnegative integers i and j. Here we let $X(u)$ be the past state, $X(t)$ be the present state and $X(t+\Delta t)$ be the state at time $t+\Delta t$. The CTMC is time homogeneous if $P\{X(t+\Delta t) = j \big| X(t) = i\} =: p_{ij}(\Delta t)$ which does not depend on t.

Let $P(t)$ be the matrix with ijth entry $p_{ij}(t)$. The Markov property can be used to show that $P(t)$ is determined by a time-independent rate matrix Q. We define the rate at which the chain moves from state i to j as $q_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t}$ for $j \neq i$ and we define $q_{ii} = -\sum_{j \neq i} q_{ij}$. Then, $Q = (q_{ij})$ is the instantaneous rate matrix of the CTMC. The rate matrix is used to calculate the probability of the chain being in a particular state at a given time, and how that probability changes over time.

By the Kolmogorov's equations (Pinsky and Karlin, 2010), we know that the relationship between rate matrix Q and transition probability matrix $P(t)$ is

$$P(t) = e^{Qt} = \sum_{i=0}^{\infty} \frac{t^i}{i!} Q^i \tag{2.3}$$

where Q is

$$Q = \begin{pmatrix} q_{00} & q_{01} & q_{02} & \cdots & q_{0n} \\ q_{10} & q_{11} & q_{12} & \cdots & q_{0n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{n0} & q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix}.$$

2.1.1 The Embedded Markov Chain

In order to simulate from the CTMC model, there is an alternative characterization of a CTMC that will be useful. We can consider the CTMC as a combination of two parts. One is the existence of a DTMC, known as a jump chain or an embedded Markov chain. The other is the time that the individual stays in the current state until moves to next state, we call the time as holding time. We let Y_n denote the state of the system at the nth transition and let T_n denote the holding time between the $n-1$st and nth transition. Then $\{Y_n\}$ and $\{T_n\}$ provide an alternative, equivalent description of the process to $\{X(t)\}$. We have $Y_0 = X(0) = i, Y_1 = X(T_1) = j, Y_2 = X(T_1 + T_2) = k, ...$ The $\{Y_n\}$ are a DTMC with transition probabilities p_{ij} that are determined by the q_{ij} through

$$\begin{aligned} P[y_n = j \big| y_{n-1} = i] = p_{ij} = \tfrac{q_{ij}}{\sum_{j \neq i} q_{ij}} \\ p_{ii} = 0. \end{aligned} \tag{2.4}$$

The random variables $T_1, T_2, ...$ satisfy that, conditional on the past history of the process, T_n has a distribution that depends only on y_{n-1}, given that $y_{n-1} = i$ has an exponential distribution with rate $\sum_{j \neq i} q_{ij}$.

2.2 Stochastic Epidemic Models

2.2.1 CTMC SIS Epidemic Model

In this study we mainly use the CTMC SEIR epidemic model. We start with a review of the simple CTMC SIS model and then based on a similar theoretical basis we describe the SEIR model. Much of the development here follows that of Allen (2008).

We denote observation time as $0 \leq t_0 < t_1 < ... < t_{n+1}$, and denote the number of susceptible and infected individuals at time t as $S(t), I(t) \in \{0, 1, 2, ..., N\}$. The SIS

model assumes that the total population N remains constant. To construct the SIS stochastic model, we only need to consider the number of individuals in the infected compartment I as the only independent random variable $I(t)$, because the number of individuals in the other compartment S can be represented as $S(t) = N - I(t)$. In this case, the number of infected individuals $I(t)$ at continuous time is a stochastic process with state probabilities $p_i(t) = P\{I(t) = i\}$ and it has Markov property

$$P\{I(t+\Delta t) = j \big| I(t) = i, I(u) = i(u), 0 < u < t\} = P\{I(t+\Delta t) = j \big| I(t) = i\}. \quad (2.5)$$

To simplify the formula, we make the assumption that the time period Δt is set small enough that the probability of two or more events occurring in the time period Δt is approximately 0. Thus under this assumption the SIS model has only three possible transitions in a time period Δt. The number of infecteds, $I(t)$, can increase by one, decrease by one or remain the same. We can denote the current infected compartment *I(t)* by *i*. Then the next state *I(t+Δt)* = *j* can be *j*=*i*+1, *j*=*i*-1 and *j*=*i*, which indicate a new infection, a remove or no change respectively. In summary we have that

$$p_{ij}(\Delta t) = \begin{cases} \frac{\beta i(N-i)}{N}\Delta t + o(\Delta t), & j = i+1 \\ \gamma i \Delta t + o(\Delta t), & j = i-1 \\ 1 - \frac{\beta i(N-i)}{N}\Delta t - \gamma i \Delta t + o(\Delta t), & j = i \\ o(\Delta t), & j \neq i+1, i-1, i. \end{cases} \quad (2.6)$$

In (2.6), the infection rate β is a combination of the rate at which each individual comes into contact with the other individual and the rate at which it causes infection. Allen (2008) does not mention the derivation of β in the chapter, which may be expressed differently than in (2.6). Specifics on the derivation of the transition probabilities in the model are in Appendix A. The most intuitive way to understand the dynamic development of the model is from its compartmental diagram in Figure 2.1.

According to the transition probabilities in (2.6), the transition rate matrix of the stochastic SIS model can be formed. We use the entry q_{ij} to denote the transmission rate for infected individual from state i to state j,

$$q_{ij} = \begin{cases} \frac{\beta i(N-i)}{N}, & j = i+1 \\ \gamma i, & j = i-1. \end{cases} \quad (2.7)$$

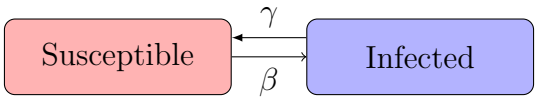


Figure 2.1: Compartmental diagram of SIS model. β is the infection rate, which is the rate of change from a susceptible individual to an infected individual. γ is the removed rate, which is the rate of change from an infected individual back to the susceptible compartment.

When the states are ordered from 0 to N, the state $\{0\}$ is an absorbing state since there will be no more transition between compartments when the number of infecteds is 0.

There is a crucial coefficient in epidemic modelling known as the basic reproduction number R_0. The basic reproduction number is the number of secondary infections caused by an already infected individual in a susceptible population before they are removed from the infected compartment (Anderson and May, 1991; Hethcote, 2000). The R_0 under the SIS model is formed as

$$R_0 = \frac{\beta}{\gamma}. \tag{2.8}$$

This equation depends on the infection rate β and the average time for an infected individual to be cured, which is $\frac{1}{\gamma}$. When $R_0 \leq 1$, the epidemic will die out quickly. When $R_0 > 1$, the epidemic will persist and majority of people or everyone will be infected.

2.2.2 CTMC SEIR Epidemic Model

We use the same principles to construct the SEIR model. Since we do not take into account the natural birth and death rates, the transition probabilities we need are succinct and there are only three transitions: from a susceptible compartment to an exposed compartment, from an exposed compartment to an infected compartment, and from an infected compartment to a removed compartment. Since the total population is constant, we can focus on the multivariate process $\{S(t), E(t), I(t)\}_{t\geq 0}$ and its joint probability function

$$p_{(s,e,i)}(t) = P\{(S(t), E(t), I(t)) = (s, e, i)\} \tag{2.9}$$

where $S(t), E(t), I(t) \in \{0, 1, 2, ..., N\}$ and $0 \leq S(t) + E(t) + I(t) \leq N$.

Because at the beginning of an epidemic there are only susceptible and infected people in the population, when the susceptible and infected people have been in contact, the susceptible people move to the exposed compartment. Thus the probability of transition of an individual from the susceptible compartment to the exposed compartment is the same as the probability of infection in the SIS model. For an infected individual removed, the probability is the same as the one calculated for the SIS model. Relative to the SIS model we add the probability of the individual's health status from an exposed to infected. We use κ as the latency rate, the derivation is in the Appendix A.

We let x represents the current state (s, e, i, r) and x' represents the next state (s', e', i', r') that the individual could move to. The transition probabilities are formulated as,

$$p_{xx'}(\Delta t) = \begin{cases} \frac{\beta si}{N}\Delta t + o(\Delta t), & s' = s-1, e' = e+1 \\ \kappa e \Delta t + o(\Delta t), & e' = e-1, i' = i+1 \\ \gamma i \Delta t + o(\Delta t), & i' = i-1, r' = r+1 \\ 1 - \frac{\beta si}{N}\Delta t - \kappa e \Delta t - \gamma i \Delta t + o(\Delta t), & (s', e', i', r') = (s, e, i, r) \\ o(\Delta t), & \text{otherwise.} \end{cases} \tag{2.10}$$

We can give the compartmental diagram for the SEIR model,

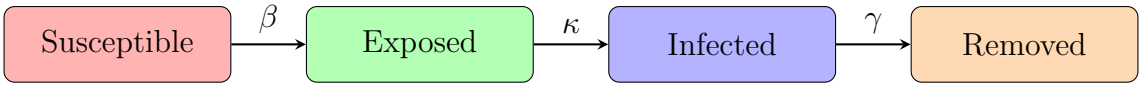


Figure 2.2: SEIR Compartmental diagram. β is infection rate, κ is latency rate and γ is removed rate. κ^{-1} means the incubation period.

We can use (2.10) to get the transition rate matrix Q by denoting the entry $q_{xx'}$,

$$q_{xx'} = \begin{cases} \frac{\beta si}{N}, & s' = s-1, e' = e+1 \\ \kappa e, & e' = e-1, i' = i+1 \\ \gamma i, & i' = i-1, r' = r+1. \end{cases} \tag{2.11}$$

DTMC SEIR Model

Once we have derived the CTMC epidemic model, it is easy to understand the DTMC model. In contrast to CTMC models, DTMC models define time as a countable set

of time periods where $t \in \{0, \Delta t, 2\Delta t, ...\}$. The DTMC model simply assumes that each time period is small enough that only one inter-individual transfer can occur in each time period. This means that the transition probabilities of the DTMC model is the transition probabilities of the CTMC model without the $o(\Delta t)$ term. For DTMC SEIR model, the transition probabilities are

$$p_{xx'}(\Delta t) = \begin{cases} \frac{\beta si}{N}\Delta t, & s' = s-1, e' = e+1 \\ \kappa e \Delta t, & e' = e-1, i' = i+1 \\ \gamma i \Delta t, & i' = i-1, r' = r+1 \\ 1 - \frac{\beta si}{N}\Delta t - \kappa e \Delta t - \gamma i \Delta t, & (s', e', i', r') = (s, e, i, r) \\ 0, & \text{otherwise.} \end{cases} \tag{2.12}$$

Deterministic Epidemic Model

When discussing stochastic models, it is essential to relate to deterministic models because the dynamics of the deterministic model are widely known (Hethcote, 1976). We already know the difference between the two models, but the connection between the two models is also worth mentioning.

The dynamics of the compartmental model can be explained by differential equations, as the SEIR model is formulated as

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N}SI \\ \frac{dE}{dt} &= \frac{\beta}{N}SI - \kappa E \\ \frac{dI}{dt} &= \kappa E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{2.13}$$

where $\beta > 0, \gamma > 0, \kappa > 0, N = S(t) + E(t) + I(t) + R(t)$. The initial conditions are $S(0) > 0, E(t) > 0, I(0) > 0, R(0) > 0, N = S(0) + E(0) + I(0) + R(0)$.

We can use the SIR model as an example to explore the connection between deterministic models and stochastic models. Note that the SIR model is a special case of SEIR model when $\kappa \to \infty$. The SIR model expressed as a system of differential equations is,

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N}SI \\ \frac{dI}{dt} &= \frac{\beta}{N}SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{2.14}$$

where $\beta > 0, \gamma > 0, N = S(t)+I(t)+R(t)$. The initial conditions are $S(0) > 0, I(0) > 0, R(0) > 0, N = S(0) + I(0) + R(0)$.

Britton et al. (2019) shows the SIR model will be characterized by the Poisson processes. Given the history of the process up to the time t, we have

$$\begin{aligned} S(t) &= S(0) - P_1\big(\tfrac{\beta}{N}\int_0^t S(u)I(u)du\big) \\ I(t) &= I(0) + P_1\big(\tfrac{\beta}{N}\int_0^t S(u)I(u)du\big) - P_2\big(\gamma\int_0^t I(u)du\big) \\ R(t) &= R(0) + P_2\big(\gamma\int_0^t I(u)du\big) \end{aligned} \tag{2.15}$$

where $\{P_1(t_1')\}$ and $\{P_2(t_2')\}$ are two independent Poisson processes with rate 1.

We can check if (2.15) gives the same process with SIR model. To illustrate we will calculate $P[S(t + \Delta t) - S(t) = 1]$ conditioned upon the history up to time t. This probability is the probability that the Poisson process $\{P_1(t_1')\}$ has a single event in the interval of time $[\frac{\beta}{N}\int_0^t S(u)I(u)du, \frac{\beta}{N}\int_0^{t+\Delta t} S(u)I(u)du]$. Properties of the standard Poisson process give that this probability is the width of the time interval with the term $o(\Delta t)$,

$$\begin{aligned} P[S(t+\Delta t) - S(t) = 1] &= \frac{\beta}{N}\left\{\int_0^{t+\Delta t} S(u)I(u)du - \int_0^t S(u)I(u)du\right\} + o(\Delta t) \\ &= \frac{\beta}{N}\int_t^{t+\Delta t} S(u)I(u)du + o(\Delta t) \\ &= \frac{\beta}{N}S(t)I(t)\Delta t + o(\Delta t) \end{aligned} \tag{2.16}$$

which matches the probabilities from the SIR model.

We let $s(t), i(t), r(t)$ denote the proportions $\frac{S(t)}{N}, \frac{I(t)}{N}, \frac{R(t)}{N}$, then

$$\begin{aligned} s(t) &= s(0) - \tfrac{P_1\left(\beta N\int_0^t s(u)i(u)du\right)}{N} \\ i(t) &= i(0) + \tfrac{P_1\left(\beta N\int_0^t s(u)i(u)du\right)}{N} - \tfrac{P_2\left(\gamma N\int_0^t i(u)du\right)}{N} \\ r(t) &= r(0) + \tfrac{P_2\left(\gamma N\int_0^t i(u)du\right)}{N}. \end{aligned} \tag{2.17}$$

The Poisson process $\{P_j(\tau)\}$ at time Nt can be expressed as

$$P_j(Nt) = \sum_{k=1}^{N} X_k \tag{2.18}$$

where X_k is the number of events in $\big((k-1)t, kt]$. By the properties of a Poisson process, the events are independent and they have Poisson distributions with rate

parameter t. The law of large numbers (LLN) gives

$$\frac{P_j(Nt)}{N} = \frac{\sum_{k=1}^{N} X_k}{N} \to t \qquad as \quad N \to \infty. \tag{2.19}$$

Applying the LLN in (2.17) gives

$$\begin{aligned} s(t) &= s(0) - \beta \int_0^t s(u)i(u)du \\ i(t) &= i(0) + \beta \int_0^t s(u)i(u)du - \gamma \int_0^t i(u)du \\ r(t) &= r(0) + \gamma \int_0^t i(u)du. \end{aligned} \tag{2.20}$$

Taking derivatives of both sides of (2.20) gives the differential equations

$$\begin{aligned} s'(t) &= -\beta s(t)i(t) \\ i'(t) &= \beta s(t)i(t) - \gamma i(t) \\ r'(t) &= \gamma i(t) \end{aligned} \tag{2.21}$$

which are same as (2.14).

2.3 Properties of Stochastic Epidemic Models

There are many ways in which we can analyse the information brought to light by epidemiological models. Two of the main summary quantities that we consider are the final size of the epidemic and the duration of the epidemic. The final size is the number of people initially infected plus the number of additional infected people during the epidemic. The duration of the epidemic corresponds to the time until the number of infecteds is zero, that is, the time from the start of the epidemic to the end of the epidemic. In this thesis, instead of the usual definition, we will scale the final size to be a proportion. In subsequent simulation sections, we will design experiments to compare how different initial values of infection will affect a bubble strategies, and for consistency we will always use the following notation of the final size:

$$\frac{R(\infty) - I_n}{N - I_n} \tag{2.22}$$

where $R(\infty) = N - S(\infty)$ is the final size of the epidemic as defined by Allen (2008) and I_n is the initial number of infected individual. We note that an epidemic is more severe if the final size is larger.

When the final size of two epidemics appears to be similar, we can track the epidemic duration to compare the trends of the two epidemics. A short duration is usually associated with a large increase in the number of infecteds in a short period of time. Such an increase can overwhelm health care resources. Allen (2008) suggests that the duration depends on the initial value of the number of infecteds, the total population and the basic reproduction number. We can verify this through simulations. However, when comparing two epidemics, if the two final sizes are different, the duration is sometimes not as meaningful. Example will be explained later in the simulation section.

Allen (2008) shows how to manually calculate the probability distribution of the final size for the SIR model. The method was originally developed by Foster (1955) and is based on the embedded Markov chain. Because the SIR model has independent variable pair (s,i), Allen (2008) lists the possible pairs as a set, which is $\{(s,i) : s = 0,1,2,...,N; i = 0,1,2,...,N-s\}$. The author gave an example with a total population of 3 and listed 10 pairs of (s,i), then listed a 10×10 transition matrix.

However, we will not be using this method next because it would be very computationally intensive. If applied to the SIR model with a larger population size, the growth of the pairs will result in a voluminous matrix and the power of the matrix will be calculated, which is a limitation of this method. Instead, we use simulation to approximate distribution of the final size of an epidemic. Let us test this method under the SIR model and compare it with Allen's method to see if the two final size distributions derived from the two methods are the same.

Figure 2.3 shows the final size distribution of the simulation. It is similar to Figure 3.13 from Allen (2008), which provides evidence that our simulation method is working correctly.

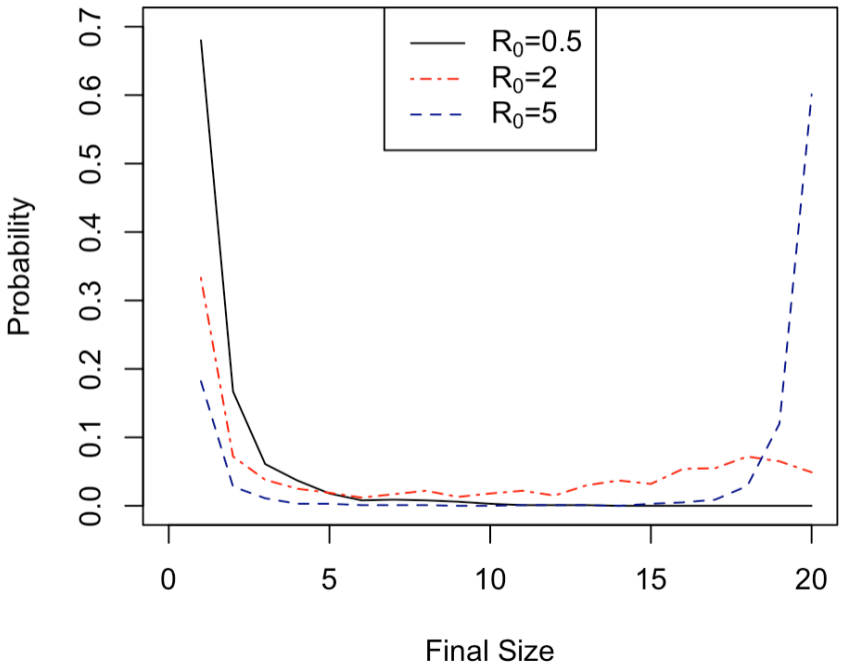


Figure 2.3: Final size distribution simulated by stochastic SIR model. We use γ=1, N=20 and I_0=1 with three R_0 values 0.5, 2, 5.

2.4 Bubble Strategies

The bubble strategies, as the way of helping to control epidemics that we consider in our research, derives its concept from the social distance that people are asked to maintain during an epidemic outbreak. A bubble usually consists of two or three households. Each individual is required to have contact only with those in the same bubble and to avoid contact with individuals in other bubbles. During the outbreak, many governments tried to control the spread of the virus by reducing person-to-person contact through bubble strategies. In Nova Scotia, the government introduced the household bubble on May 15, 2020 and allowed two families to meet without social distance (B. Roth, 2021). However, after a month the bubble restrictions were loosened.

2.4.1 Related Works

In 2020, the COVID-19 pandemic broke out worldwide, and many scholars conducted a lot of studies on the public health policies including bubble strategies.

New Zealand has had success in terms of its bubble strategies. New Zealand reported no new cases more than a month after the policy was implemented (Cousins, 2020). Kearns et al. (2021) did an online survey to study the size characteristics of bubbles. The survey was implemented quickly, with 14,876 surveys used to calculate the average size of a household unit or bubble, the average number of households in each bubble, the proportion of bubbles containing essential staffs and/or vulnerable groups, and the average number of times per week that a person left the house. However the experiment was a cross-sectional study mainly based on data obtained from online platforms such as Facebook, email lists etc. and has the limitation of selection bias. They performed simple calculations on the data and did not use the SEIR model. Our study will also analyse bubble size. Unlike this survey, we use simulated data instead of actual data. The impact of bubble strategies on the spread of pandemics is illustrated by comparing large bubble sizes to small bubble sizes.

The UK also released a Bubble Strategy in 2020, but unlike New Zealand's policy, the UK is divided into a Support Bubble, a Childcare Bubble, a Christmas Bubble and so on, depending on the period and status (Danon et al., 2021). Instead of an SEIR

model, Danon et al. (2021) used percolation theory and networks of households to explore the effect of bubble strategies. They used the network without family bubble as a baseline network, then merged multiple households into larger units, thereby creating additional external connections. Once the merged network is established, they examined its resilience to random removal of these external links. They set up eight scenarios and through percolation analysis concluded that a single-person household bubble had the least impact on the spread of the epidemic. It is also believed that the greater number of families forming bubbles will make it difficult to control the epidemic, but that the likelihood of transmission can be reduced by controlling outside contact. Hill (2023) has conducted a simulation study of the Christmas bubble. He used a stochastic individual-based SEIR model, while classifying infection status as asymptomatic and symptomatic. He simulated five scenarios to estimate the epidemiological impact of the Christmas bubble and the continued expansion of the bubble. In the simulations, he also divided the population into three age groups and also studied the effects of the bubble strategies on different age groups. In contrast, the study in this thesis has done very little scenario modelling of the bubble strategies itself; we have focused more on whether the presence or absence of a bubble strategy has an effect on the spread of the epidemic and the impact of the parameters of the model.

Willem et al. (2021) have also conducted research on bubble strategies. They used the open source individual-based model (STRIDE) to simulate interactions between Belgium's 11 million inhabitants. The contact tracing strategy and the impact on hospitalisation rates are explored by increasing the connection between household bubbles and the outside world. They believe that the tracking strategy has great potential to reduce the spread of the epidemic and believe that the bubbling is effective if people are open to contact. Individual-based model is used to track one person if he is susceptible, exposed, infected, or removed. This increases computational complexity. Unlike their study, we did not track the dynamics of each individual, but recorded the overall number of infecteds. Our main focus was on the impact of the bubble strategies. However, a similar limitation to the above study is that our model is a simple one, but the actual situation is more complex due to regional and climatic influences.

2.5 Mann-Whitney U Test

In this thesis we will simulate different scenarios for populations with and without bubbles. When we want to compare the difference between the two populations in the final size distribution and the duration distribution, the Mann-Whitney U test is an appropriate choice. Suppose we define P_1 and P_2 are random variables from each of the two populations. The U-test statistic approximates the probability that a random observation from the first population is larger than the second population. We define the U-statistic as

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(P_{1i}, P_{2j})$$

with

$$S(P_1, P_2) = \begin{cases} 1, & P_1 > P_2 \\ \frac{1}{2}, & P_1 = P_2 \\ 0, & P_1 < P_2 \end{cases}$$

where $P_{1i}, i = 1, \ldots n$ and $P_{2j}, j = 1, \ldots, m$ are random samples from the two populations. In this thesis we normalized the statistical value by using $\frac{U}{nm}$. That normalized value gives an estimate of the effect size $P[P_1 > P_2] + \frac{1}{2}P[P_1 = P_2]$ where P_1 is a randomly chosen final size without bubbling and P_2 is a randomly chosen final size with bubbling.

To test whether the distributions are the same for the two populations, we can use the Mann-Whitney hypothesis test which is based on the value $|U/(mn) - 1/2|$. The hypothesis of Mann-Whitney U test is that $H_o : P[P_1 > x] = P[P_2 > x]$ for all x and $H_A : P[P_1 > x] > P[P_2 > x]$. The test that we use computes a two-sided p-value calculated as

$$P\left[|Z| > \frac{|\frac{U}{mn} - \frac{1}{2}|}{SD(\frac{U}{mn})}\right]$$

where $Z \sim N(0, 1)$ and $SD(\frac{U}{mn})$ is the standard deviation of the test statistic under H_o.

We will use the function *wilcox.test*() in R since it gives that the Mann-Whitney test statistic and p-value. The Mann-Whitney test is equivalent to the Wilcoxon rank-sum test. We have used a simple example to check if this function works, it gives the same statistical values as Mann-Whitney test and found that the returned

value `statistic` is the same as $\frac{U}{mn}$. In the simulation we will use symbol to indicate the significance of the test.

Symbol	P-value
$***$	$p \leq 0.001$
$**$	$p \leq 0.01$
$*$	$p \leq 0.05$
ns	$p > 0.05$

Table 2.1: The meaning of the symbol used to indicate statistical significance.

Chapter 3

Methodology

3.1 The Bubble SEIR

This thesis focuses on whether bubble strategies have an impact on the spread of epidemics. In this section we develop an extension of the SEIR model that has bubbling and we derive an algorithm to simulate from it. We would like to simulate the process of epidemic change using a CTMC stochastic SEIR model to investigate the effects of a bubble strategy by varying different parameter settings. The proposed model keeps track of the number of individuals in each bubble. We name this model the Bubble SEIR. As a stochastic model, the probability of having two transitions within a period of Δt is $o(\Delta t)$ and movements can be divided into two different types in our model. One category is transition within a bubble, where a person can move from a susceptible compartment to an exposed compartment, or from an exposed compartment to an infected compartment, or from an infected compartment to a removed compartment. The other type of movement is between bubbles. For example, an infected person can move from group i to group j. Because two or more events in a small interval of time are unlikely to occur, when an individual moves from one group to another, that individual's health status remains the same.

Since the Bubble SEIR model divides the population into K bubbles, then we denote the current state $x = (s_1, e_1, i_1, r_1, s_2, e_2, i_2, r_2, ..., s_K, e_K, i_K, r_K)$, where $(s_k, e_k, i_k, r_k) \in \{0, 1, ..., N\}$ for $k = 1, 2, \ldots, K$. There are on the order of N^{4K-1} states in the model with $N = \|x\| - 1$. We denote $x' = (s'_1, e'_1, i'_1, r'_1, s'_2, e'_2, i'_2, r'_2, ...,$ $s'_K, e'_K, i'_K, r'_K)$. When individuals move within the bubble, the probabilities of an individual in the current state moving between compartments in the model are the same as the regular SEIR model in Section 2.2.2. Note that the infection rate is dependent on the number of individuals in each bubble N_k rather than the overall population size N. Restricting attention to transitions within a bubble k for which

$x' \neq x$, the transition probabilities are

$$p_{xx'}(\Delta t) = \begin{cases} \frac{\beta s_k i_k}{N_k}\Delta t + o(\Delta t), & s'_k = s_k - 1, e'_k = e_k + 1 \\ \kappa e_k \Delta t + o(\Delta t), & e'_k = e_k - 1, i'_k = i_k + 1 \\ \gamma i_k \Delta t + o(\Delta t), & i'_k = i_k - 1, r'_k = r_k + 1 \\ o(\Delta t), & \text{otherwise.} \end{cases} \tag{3.1}$$

When individuals move between bubbles, the transition parameters in the regular SEIR model no longer apply. We let $M(\Delta t)$ be the event that some individuals move from one bubble to another in a period of Δt unit of time. Let $\mathcal{I}_{Mkl}$ be the event that an individual M in bubble k moves to the bubble l. We assume that the probability that any particular individual moves from bubble k to the bubble l is $\frac{r}{K-1}\Delta t + o(\Delta t)$. Then

$$\begin{aligned} P\{M(\Delta t)\} &= \sum_{k=1}^{K} \sum_{M \in B_k} \sum_{l=1, l\neq k}^{K} P\{\mathcal{I}_{Mkl}\} + o(\Delta t) \\ &= \sum_{k=1}^{K} N_k (K-1) \frac{r}{K-1}\Delta t + o(\Delta t) \\ &= N(K-1)\frac{r}{K-1}\Delta t + o(\Delta t) \\ &= Nr\Delta t + o(\Delta t) \end{aligned} \tag{3.2}$$

where B_k is the set of individuals from bubble k. Thus the total rate of movement is Nr and the rate of movement per individual is r. There will be more movements in large population N than in small population. Restricting attention to movements from bubble k to bubble l, or equivalently to x' that differ from x because $(s'_k, e'_k, i'_k, r'_k) \neq (s_k, e_k, i_k, r_k)$ and $(s'_l, e'_l, i'_l, r'_l) \neq (s_l, e_l, i_l, r_l)$, the transition probabilities are

$$p_{xx'}(\Delta t) = \begin{cases} s_k \frac{r}{K-1}\Delta t + o(\Delta t), & s'_k = s_k - 1, s'_l = s_l + 1 \\ e_k \frac{r}{K-1}\Delta t + o(\Delta t), & e'_k = e_k - 1, e'_l = e_l + 1 \\ i_k \frac{r}{K-1}\Delta t + o(\Delta t), & i'_k = i_k - 1, i'_l = i_l + 1 \\ r_k \frac{r}{K-1}\Delta t + o(\Delta t), & r'_k = r_k - 1, r'_l = r_l + 1 \\ o(\Delta t), & \text{otherwise.} \end{cases} \tag{3.3}$$

Every other choice of $x' \neq x$ either has probability 0 of occurring or requires two or more events and thus has $p_{xx'}(\Delta t) = o(\Delta t)$. The first set of non-zero entries of the

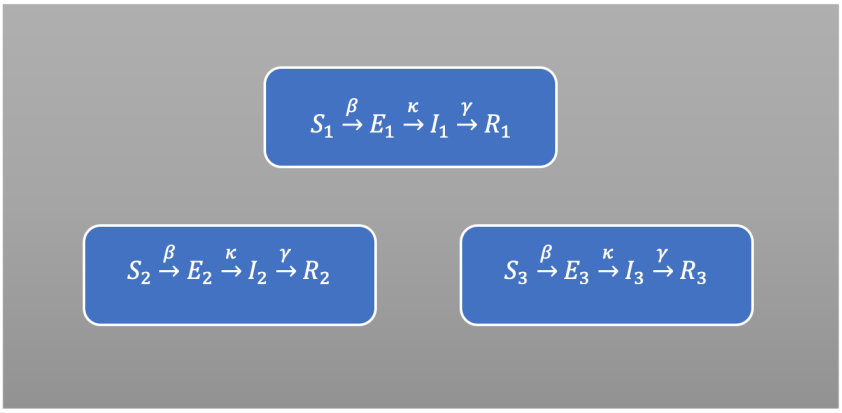
rate matrix Q correspond to transitions within a bubble. These rates are obtained from (3.1) as

$$q_{xx'} = \begin{cases} \frac{\beta s_k i_k}{N_k}, & s'_k = s_k - 1, e'_k = e_k + 1 \\ \kappa e_k, & e'_k = e_k - 1, i'_k = i_k + 1 \\ \gamma i_k, & i'_k = i_k - 1, r'_k = r_k + 1 \end{cases} \tag{3.4}$$

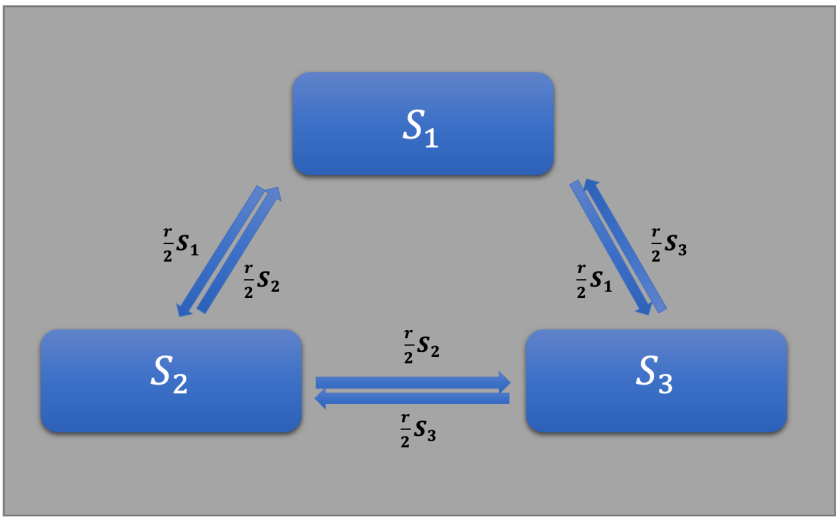
for all $k = 1, 2, \ldots K$. The other non-zero entries corresponding to movements between bubbles are obtained from (3.3) as

$$q_{xx'} = \begin{cases} s_k \frac{r}{K-1}, & s'_k = s_k - 1, s'_l = s_l + 1 \\ e_k \frac{r}{K-1}, & e'_k = e_k - 1, e'_l = e_l + 1 \\ i_k \frac{r}{K-1}, & i'_k = i_k - 1, i'_l = i_l + 1 \\ r_k \frac{r}{K-1}, & r'_k = r_k - 1, r'_l = r_l + 1 \end{cases} \tag{3.5}$$

for all $k \neq l$. The Figure 3.1 gives an example of how the Bubble SEIR works.



(a) Three bubbles in the population. β, κ and γ could be different, but here we assume the parameters are constant for every bubble.



(b) Movements between susceptible compartments within the three bubbles.

Figure 3.1: We divide the population into three bubbles as shown in (a), and in each bubble there exists a SEIR model to represent the epidemic compartment that occurs, which is what we call intra-group transition. We also allow for contact and movement between bubbles as shown in (b), where the rate of movement per individual between the three susceptible compartments is $\frac{r}{K-1}$, $K = 3$.

3.1.1 Embedded Markov Chain Characterization

The order of N^{4K-1} states poses a difficulty in constructing the transition probability matrix $P(t)$ if we use the equation $P(t) = e^{Qt}$. The rate matrix is of such a

high dimension, a matrix exponentiation is infeasible except for very small population sizes. We will utilize the embedded Markov chain theory in Section 2.1.1 for simulation. Given the current state of the system, x, we can generate the next state x' according to all the possible non-zero probabilities $q_{xx'}/\sum_{x'\neq x} q_{xx'}$. The waiting time before the next transition can be generated from an exponential distribution with rate $\sum_{x'\neq x} q_{xx'}$.

A complication for the method described above is that the summation is across all states which can be massive. Fortunately, given the current state of the system x, the number of states, $x' \neq x$, with a non-zero rate $q_{xx'} > 0$ is small. This is because the probability of two or more events occuring in a small interval of time is assumed small.

Recall that for the embedded Markov chain, $P[y_n = x' \big| y_{n-1} = x] = p_{xx'} = \frac{q_{xx'}}{\sum_{x'\neq x} q_{xx'}}$. Although we have described both the rate of movement within a bubble and the rate of movement between bubbles separately above, in the simulation both rates are the chances that an individual can move. To use the embedded Markov chain characterization we need to compute the normalizing constant $\mu = \sum_{x'\neq x} q_{xx'}$ where x is the current state of the system. In the current state, the possible moves are at most $3K$ ways within the bubble, and at most $4K(K-1)$ ways for individuals to move to another bubbles, respectively. Let W_k denote the set of $x' \neq x$ that give non-zero $q_{xx'}$ and that correspond to transitions within bubbles. Let B_{kl} denote the set of $x' \neq x$ that give non-zero $q_{xx'}$ and that correspond to movements between bubbles.

$$
\begin{aligned}
\mu &= \sum_{k=1}^{K} \sum_{x'\in W_k} q_{xx'} + \sum_{k=1}^{K} \sum_{l=1,l\neq k}^{K} \sum_{x'\in B_{kl}} q_{xx'} \\
&= \sum_{k=1}^{K} \big(\frac{\beta s_k i_k}{N_k} + \kappa e_k + \gamma i_k\big) + \sum_{k=1}^{K} \sum_{l=1,l\neq k}^{K} \big(s_k \frac{r}{K-1} + e_k \frac{r}{K-1} + i_k \frac{r}{K-1} + r_k \frac{r}{K-1}\big) \\
&= \sum_{k=1}^{K} \big(\frac{\beta s_k i_k}{N_k} + \kappa e_k + \gamma i_k\big) + \sum_{k=1}^{K} \big(s_k \frac{r}{K-1} + e_k \frac{r}{K-1} + i_k \frac{r}{K-1} + r_k \frac{r}{K-1}\big)\big(K-1\big) \\
&= \sum_{k=1}^{K} \big(\frac{\beta s_k i_k}{N_k} + \kappa e_k + \gamma i_k\big) + \sum_{k=1}^{K} \big(s_k r + e_k r + i_k r + r_k r\big) \\
&= \sum_{k=1}^{K} \big(\frac{\beta s_k i_k}{N_k} + \kappa e_k + \gamma i_k\big) + \sum_{k=1}^{K} N_k r
\end{aligned}
$$

$$= \beta \sum_{k=1}^{K} \frac{s_k i_k}{N_k} + \kappa e + \gamma i + Nr. \tag{3.6}$$

Then the transition probabilities of the jump process under the Bubble SEIR model can be expressed as

$$p_{xx'} = \begin{cases} \frac{\beta i_k s_k}{N_k}/\mu, & s'_k = s_k - 1, e'_k = e_k + 1 \\ \kappa e_k/\mu, & e'_k = e_k - 1, i'_k = i_k + 1 \\ \gamma i_k/\mu, & i'_k = i_k - 1, r'_k = r_k + 1 \end{cases} \tag{3.7}$$

for all $k = 1, 2, \ldots, K$ and

$$p_{xx'} = \begin{cases} s_k \frac{r}{K-1}/\mu, & s'_k = s_k - 1, s'_l = s_l + 1 \\ e_k \frac{r}{K-1}/\mu, & e'_k = e_k - 1, e'_l = e_l + 1 \\ i_k \frac{r}{K-1}/\mu, & i'_k = i_k - 1, i'_l = i_l + 1 \\ r_k \frac{r}{K-1}/\mu, & r'_k = r_k - 1, r'_l = r_l + 1 \end{cases} \tag{3.8}$$

for all $k \neq l$. Once we have defined all the transition probabilities, we can then randomly sample the next state of the system based on the probabilities. Recall that we can sample the waiting time t via an exponential distribution with rate μ.

3.2 Implementation

In Section 3.1.1 we addressed the way to simulate the Bubble SEIR through embedded Markov chain. Next we deal with how to indicate the path of an individual moving through the model in the implementation. We can implement the Bubble SEIR in programming language by the Algorithm 1.

Algorithm 1 Bubble SEIR Pseudo-code

Input: K: The number of groups (bubbles).

s_k, e_k, i_k, r_k: The initial number of susceptible, exposed, infected, and removed individuals in each group.

β: The number representing infection rate of the disease.

γ: The number representing recovery rate of the disease.

κ: The number representing latency rate of the disease.

r: The number representing the movement rate per individual between groups.

Output: Lists of the number of s_k, e_k, i_k, r_k for each group and at each time at which an event occurs as well as the corresponding cumulative random time period.

1: **while** $\sum E + \sum I \neq 0$ **do**
2: Create the non-zero transition rate vector q and fill in the rate values.
3: **for** $k \in 1 : K$ **do**
4: The population N_k in each group: $N_k \leftarrow s_k + e_k + i_k + r_k$.
5: The first $3K$ elements are the rates that an individual moves within the bubble k, the rates are $\frac{\beta}{N_k} s_k i_k, \kappa e_k, \gamma i_k$.
6: The remaining $4K(K-1)$ elements are the rates that an individual moves from bubble k to others, the rates are $s_k \frac{r}{K-1}, e_k \frac{r}{K-1}, i_k \frac{r}{K-1}, r_k \frac{r}{K-1}$.
7: **end for**
8: Generate the next event index y from $1 : 3K + 4K(K-1)$ according to the probabilities $q/\sum q$.
9: Generate a random time t for the event occurrence based on an exponential distribution with rate $\sum q$.
10: Update the s_k, e_k, i_k, r_k and time according to the selected event.
11: **end while**

This is a complete simulation. To approximate key quantities via simulation, like the distribution of the final size of the epidemic, in the simulation, we will simulate a thousand epidemics to obtain the data. With the SEIR model we get changes in the compartment of the system at irregular time intervals. In order to normalise the data patterns in the way that they would be seen in practice we sometimes converted this raw data to daily counts. With the framework above, we can analyse the effect of bubble strategies in various different scenarios.

Chapter 4

Simulations Study

In the simulations, we will use the estimated parameter values from (Read et al., 2021), as a baseline to investigate the impact of the bubble strategies. Our baseline is with a total population N of 100, divided into 5 groups of 20 people each. In each group the initial values are $S_0 = 19, I_0 = 1, E_0 = R_0 = 0$. The transition rates between compartments are $\beta = 1.94$, $\kappa = 0.25$, and $\gamma = \frac{1}{1.61}$, and at baseline our basic reproduction number R_0 is 3.12. In each simulation scenario, we explore the impact of a parameter on the spread of an epidemic by changing its value while keep other parameters fixed. For each scenario we will run 1000 simulations.

Scenario A - changing movement rate r: We keep the other parameter values in the baseline. In each bubble we have $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, $N = 100$, $K = 5$, $S_0 = 19$, $I_0 = 1$. We set the value of the movement rate r to 0, 0.01, 0.05, 0.1. According to the result of this scenario, we will use $r = 0.01$ in the later scenarios. One way to explain the movement rate is that for a population of size 100, this means 1 movement per day is expected. For a population of size 1000, we expect 10 movements per day.

Scenario B - Changing infection rate β and recovery rate γ:

B1 - Changing infection rate β: We set R_0 to four different values, 0.8, 1, 2.5 and 5. Only the value of β is changed to match the R_0 value. In this case, our β values are 0.5, 0.62, 1.55 and 3.11. The other parameter values remain the same in each bubble, $\kappa = 0.25, \gamma = \frac{1}{1.61}, N = 100, K = 5, S_0 = 19, I_0 = 1, r = 0.01$.

B2 - Changing recovery rate γ: We set R_0 to four different values, 0.8, 1, 2.5 and 5. Only the value of γ is changed to match the R_0 value. In this case, our γ values are 2.425, 1.94, 0.776 and 0.388. Other parameter values remain the same in each bubble, $\beta = 1.94, \kappa = 0.25, N = 100, K = 5, S_0 = 19, I_0 = 1, r = 0.01$.

B3 - Scaling β and γ: We keep the R_0 to remain at 3.12, scale the $R_0 = \frac{1.94}{0.62}$ by changing β and γ. We use $\frac{0.485}{0.155}$, $\frac{0.97}{0.31}$, $\frac{3.88}{1.24}$, $\frac{7.76}{2.48}$. Other parameter values remain the

same in each bubble, $\kappa = 0.25, N = 1000, K = 5, S_0 = 19, I_0 = 1, r = 0.01$.

Scenario C - Changing latency rate κ: In this section we only change the values of κ to 0.05, 0.1 and 0.5. We still use the baseline parameter values in each bubble, $\beta = 1.94$, $\gamma = \frac{1}{1.61}$, $N = 100$, $K = 5$, $S_0 = 19$, $I_0 = 1$, $r = 0.01$.

Scenario D - Changing population N: We will increase the population to 500, 1,000, 5,000, and 10,000, but we consider two situations based on the bubble strategies.

D1 - Equal number of bubbles: Each population is divided into 5 bubbles, where $K = 5$ with different the number of individuals n in each bubble. In addition to these we use baseline parameter values, $\beta = 1.94, \kappa = 0.25, \gamma = \frac{1}{1.61}, I_0 = 1, S_0 = \frac{N}{K} - 1, r = 0.01$.

D2 - Equal size of bubbles: Each population is divided into bubbles with 20 people, where $n = 20$ with different K. In addition to these we use baseline parameter values in each bubble, $\beta = 1.94, \kappa = 0.25, \gamma = \frac{1}{1.61}, I_0 = 1, S_0 = 19, r = 0.01$.

Scenario E - Changing infected initial values I_0: We set different initial values for the infection: only 1 infected individual in the population, population with 5% infected individuals and population with 10% infected individuals. We also consider large population sizes in this scenario and the transition rates are same as baseline.

Histograms and density curves will be plotted for detailed analysis, primarily for the final size and duration of the number of infecteds. Here our final size is the number of infecteds growing in the epidemic as a percentage of the number of people not initially infected, that is $\frac{R(\infty)-I_n}{N-I_n}$. It is generally accepted that the smaller the final size the less damaging the epidemic. When the final size distributions of two epidemics are similar, we can look at the duration distributions of the epidemics. The duration distribution allows us to determine whether the epidemic spreads faster in one scenario or another. If there is a significant increase in the number of people infected in a short period of time, it will pose a challenge to the health care system. The rapid increase in cases can place a heavy burden on health care facilities and services. We are concerned mostly with the final size distribution. When there is a significant difference between the final size distributions of two populations, comparing duration distributions may be meaningless. When everyone is eventually infected, a setting

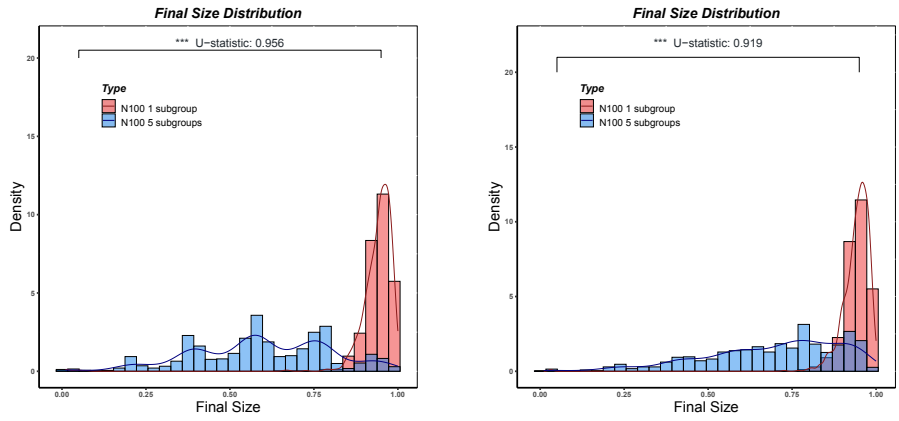
with a short duration is less desirable than a setting with a long duration because hospital resources are more likely stretched in the former case. However, when only a fraction of the population is infected, one might have a short duration simply because the epidemic died out quickly, without there ever being a period with large number of infected.

4.1 The Movement Rate r

In this section we will simulate the data according to different settings of the movement rate (**Scenario A**). We begin by investigating the impact of bubbling, i.e., dividing the population into groups, on the spread of epidemics. Using the baseline parameters, we will compare the scenarios with and without bubbles. In the absence of bubbles, the population is treated as a single unit, whereas with bubbles, the population is divided into subpopulations where the movements between these groups are limited. To examine the effect of bubbles, we will vary the movement rate r, with values of 0, 0.01, 0.05, and 0.1. The movement rate determines the probability of infection transmission from one individual to another. By incorporating group-to-group transmission, bubbles enable more targeted control measures. Our simulations indicate that bubbles are effective in controlling epidemic transmission, as not everyone in the population will eventually be infected, and the epidemic will die out early.

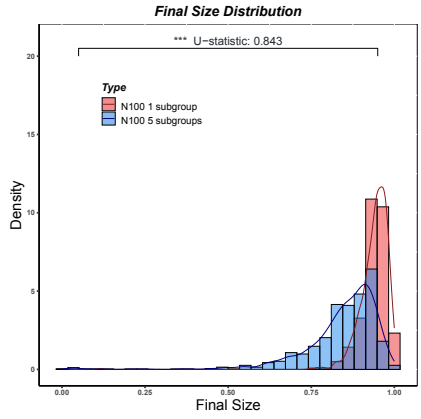
Let us look at the final size distribution for the number of infecteds given in Figure 4.1. Here the value of R_0 is 3.12, a relatively large value. This means that eventually the vast majority will be infected and accordingly final size distribution is left-skewed. When the movement rate r is 0, no transmission can take place between the 5 bubbles and the infected people can only move within the bubble. Each bubble has a single initial infection, the stochastic nature of the epidemic can cause it to end within some bubbles with few individuals in the bubble getting infected. Because there is not movement between bubbles, this results in 5 modes in the final size distribution. In this case, not everyone ends up infected. In contrast, the population without bubbles has only one peak and most everyone ends up infected. While as the movement rate increases, the final size distribution of the population with 5 bubbles gets closer to the distribution of the population without bubbles. However, we can see that the final number of infecteds in the population with bubbles is consistently

smaller than in the population without bubbles. The U-statistics approximate the probability that the final size for the population without bubbles is greater than the final size for the population with bubbles. The U-statistic becomes smaller as r increases, while p-value is consistently smaller than 0.05, indicating that the two populations are still significantly different.

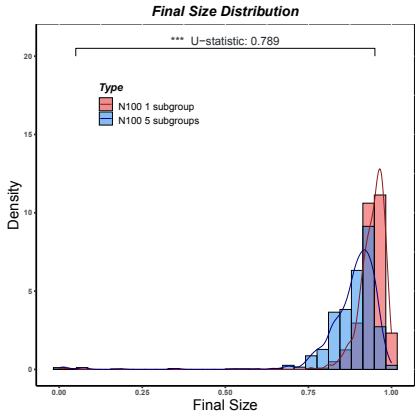


(a) r=0

(b) r=0.01



(c) r=0.05



(d) r=0.1

Figure 4.1: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, N=100. We compare two populations, one that treats the population as a whole and the other that divides the population into 5 bubbles. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

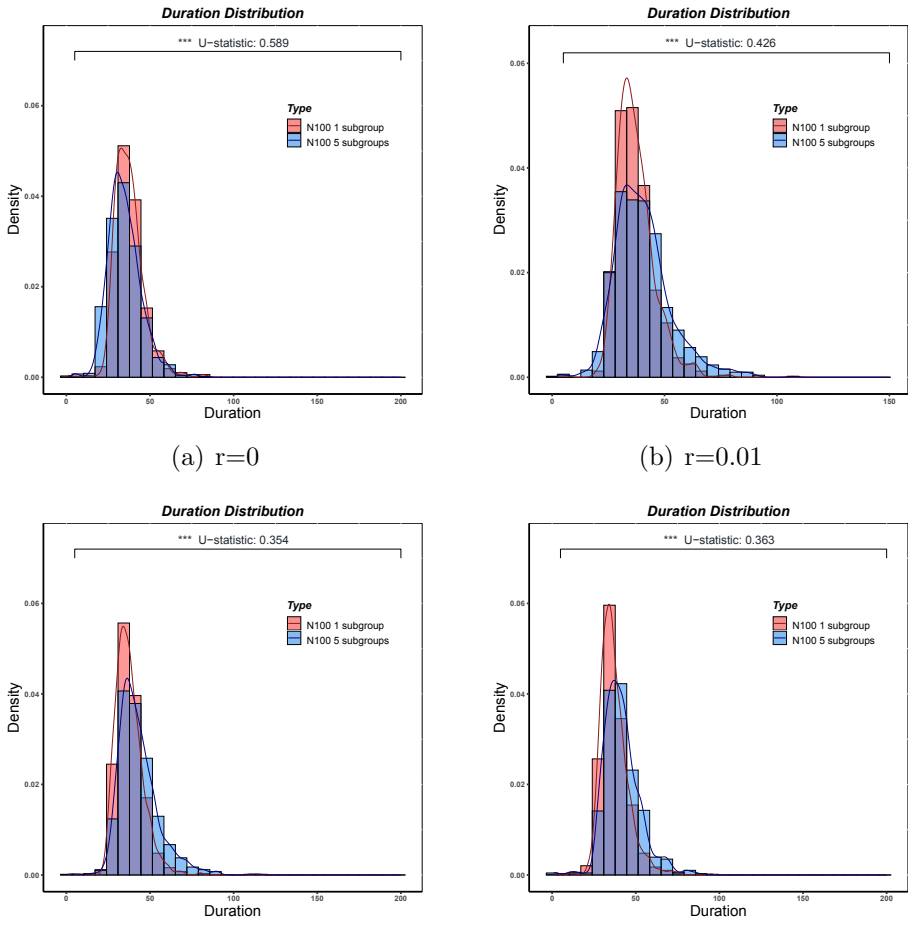
$I(\infty)$ / r	1	2-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40
0	25.82	9.42	2.38	4.5	57.88	0	0	0	0
0.01	13.52	11.04	3.64	9.94	42.08	19.26	0.52	0	0
0.05	1.1	4.74	8.68	20.1	34.88	24.12	5.66	0.66	0.06
0.1	0.62	1.34	7.32	23.12	35.8	23.8	6.94	1.02	0.04

Table 4.1: The distribution of infected individuals within a bubble at the end of the epidemic. The numbers in the table represent the percentage of the final number of infected individuals in each bubble of 1000 simulations for different movement rates.

To better understand the reasons for the differences between bubbling and not in Figure 4.1, we looked at the distribution of the final number of infected over all bubbles and all simulations (Table 4.1). The distribution of the number of infecteds has a bimodal distribution if there is no movement between the bubbles. There is a 25.82% chance that there will end up being only one infected individual within the bubble, which means that no other individual is infected in these bubbles and this one infected individual is the initial infection. With 20 individuals in each bubble, in more than half of the cases the vast majority even all of the individuals in the bubble were infected. Thus upon partitioning the population into five bubbles, the appearance of five modes in Figure 4.1. (a) signifies the distribution as: one bubble with the entirety of its individuals infected, two bubbles with every individual infected, three bubbles, four bubbles, and finally, five bubbles. As the rate of movement increased, so did the mobility between bubbles. At the end of epidemic, the number of bubbles with only one infected individual decreased rapidly, while the number of infected individuals in some bubbles exceeded 20. This change shifted the distribution of infecteds from bimodal to unimodal. Due to the presence of individuals moving between bubbles, there is a situation for infected individuals to move into another bubble, resulting in continued infection within the bubble. The difference in the distribution of the number of infecteds for $r = 0.01$ and $r = 0.05$ in the table can show that allowing more movements per day leads to more people being infected.

The histograms in Figure 4.2 illustrate the duration of the baseline under different movement rates. When there is no movement of infected individuals between groups, the epidemic duration is shorter for the population with five bubbles compared to the population without bubbles. However, when infected individuals move between

groups, the opposite is observed. The shorter epidemic duration of the bubbled population does not necessarily indicate poor performance, as evidenced by the bumps in the final size distribution at 0.2, 0.4, 0.6, and 0.8, respectively. This suggests that the small size of each bubble and the lower number of infected individuals at the end of the epidemic still make bubbles more effective for epidemic control. Based on the final size of the population with and without bubbles, we can conclude that larger final size tails correspond to longer epidemic durations when almost everyone is infected.



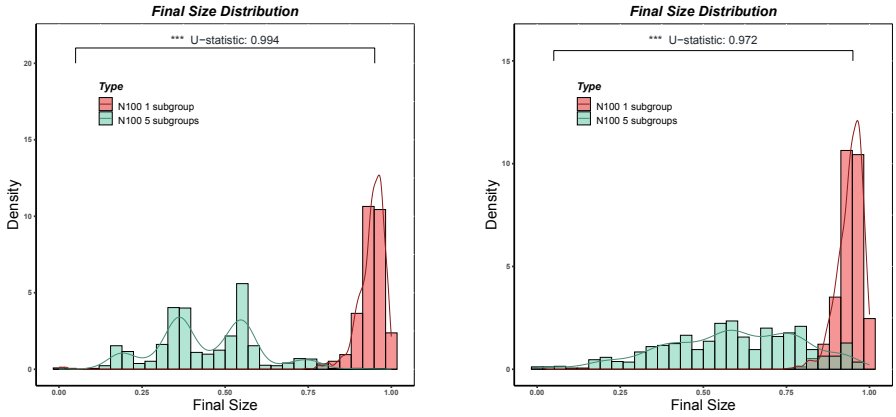
(a) r=0 (b) r=0.01

(c) r=0.05 (d) r=0.1

Figure 4.2: Duration distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

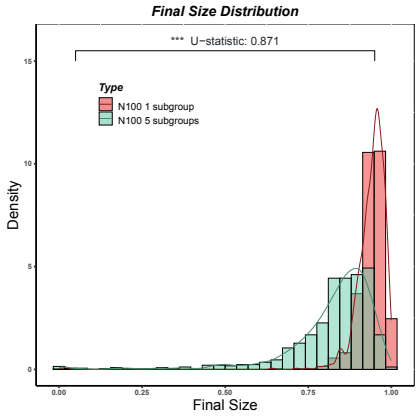
However, the assumption of assigning only one initially infected individual to each bubble might lean toward idealism. In a more realistic scenario, it is plausible that five

initial infected individuals could be allocated randomly among the bubbles. Figure 4.3 shows the final size distributions for different values of movement rate (**Scenario A**) with randomly arranged initial values. When the rate of movement is 0, the final size distributions for the two arrangements of initial values are somewhat different. The randomization of initial values introduces instances where certain bubbles commence with zero initially infected individuals. Consequently, bubbles devoid of initially infected individuals remain disease-free throughout the simulation. However, when movement between bubbles is allowed, the two final size distributions are similar. Even if some bubbles are initially free of infected individuals, there will be instances where infected individuals move into these bubbles, which may result in individuals within the bubbles getting infected. Thus, the final size distribution of a population with fixed initial values exhibits similar behaviour as a population with randomly arranged initial values. Since the two ways of assigning initial values did not have a substantial effect, in subsequent simulations we use constant initial sizes as described in the introduction to this chapter.

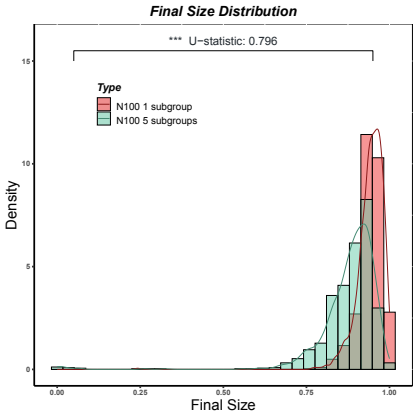


(a) r=0

(b) r=0.01



(c) r=0.05



(d) r=0.1

Figure 4.3: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, N=100. We compare two populations, one that treats the population as a whole and the other that divides the population into 5 bubbles. We set 5 initial infected individuals in total and randomly assign to each bubbles.

4.2 The Basic Reproduction Number R_0

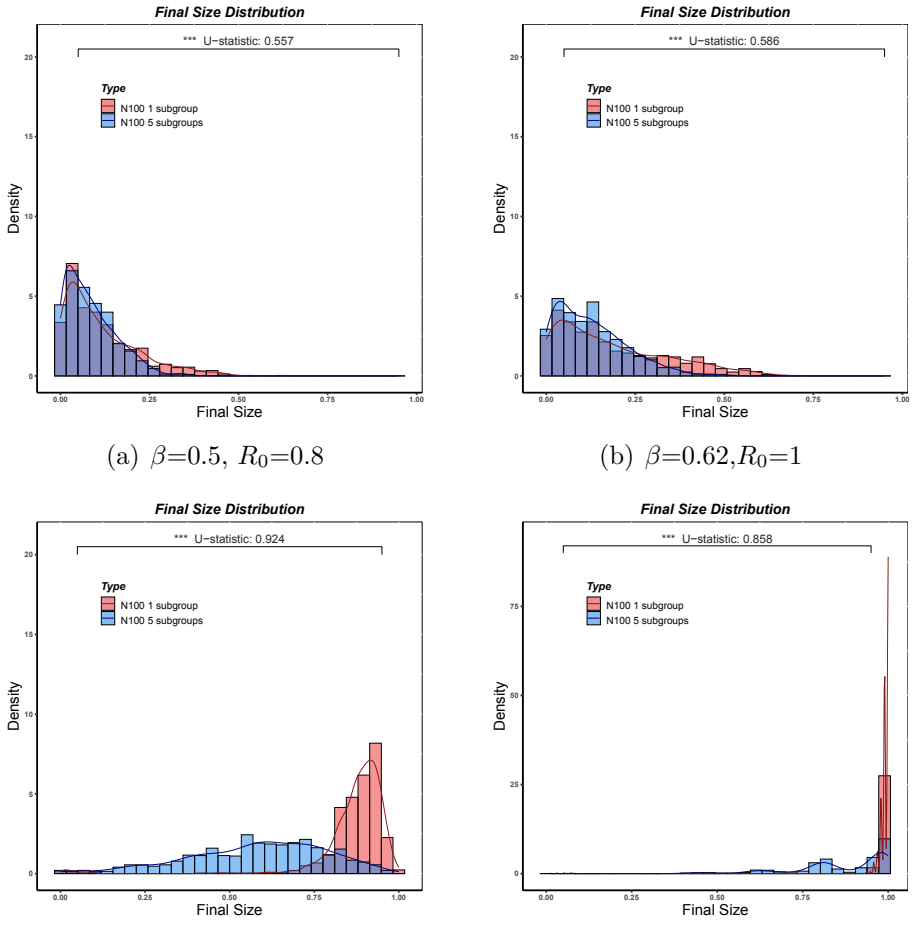
In this section, we look at the impact of basic reproduction number R_0 in the spread of epidemics (**Scenario B**). Recall that $R_0 = \frac{\beta}{\gamma}$, is the average of how many other people an infected person can infect. So we consider R_0 to be an important parameter in determining the size of the epidemic, playing a dominant role in the distribution of the final size. In the coming subsections we look at epidemic trends by fixing several values of R_0, such as the four values 0.8, 1, 2.5 and 5, varying the parameter β (**Scenario B1**) and varying the parameter γ (**ScenarioB2**). And finally we keep

the value of R_0 as 3.12 by scaling β and γ (**Scenario B3**). We explore how changing the R_0, as well as keeping it, would affect the spread of the epidemic.

4.2.1 Varying β

The parameter β in the epidemic model, is the rate of infection for an individual whose health status changes from susceptible to exposed. In this part, we keep the other parameter values constant and change the value of β according to R_0. In Section 4.1, with Figure 4.1 we can observe that the the final size distribution of bubbled population is similar to the final size distribution of population without bubbles when r is large. The big value of r means that an individual in each compartment within each bubble can have a lot of movements per day. While the simulation is unrealistic when r is 0, since it is difficult to control the absence of contact between individuals across bubbles in the bubble strategies. When r is 0.01, individuals do not move between bubbles a lot. It can give us a clear view of difference between the population with bubbles and population without bubbles, we will use $r = 0.01$ in subsequent scenarios.

At the same rate of recovery γ, the distributions of final sizes giving in Figure 4.4 (a) and (b) are similar with bubbling and without when the $R_0 \leq 1$. In both cases, there is a more concentrated distribution of final sizes with 5 groups than with 1 group. The distributions of the final size are right-skewed because of the smaller R_0 values, meaning that the epidemic often results in a small number of people being infected. Figure 4.4 (c) gives results when $R_0 = 2.5$ and exhibits a behaviour similar to the corresponding baseline case in Figure 4.1 (b) where $R_0 = 3.12$. Whereas Figure 4.4 (d) shows a very extreme performance, with the population without bubbles having almost no tail and almost everyone eventually being infected. The population with 5 groups has a longer tail but all to the extent of 50% or more. The disparity between the (b), (c) and (d) plots in Figure 4.4 suggests that bubbling is particularly effective with $R_0 > 1$.



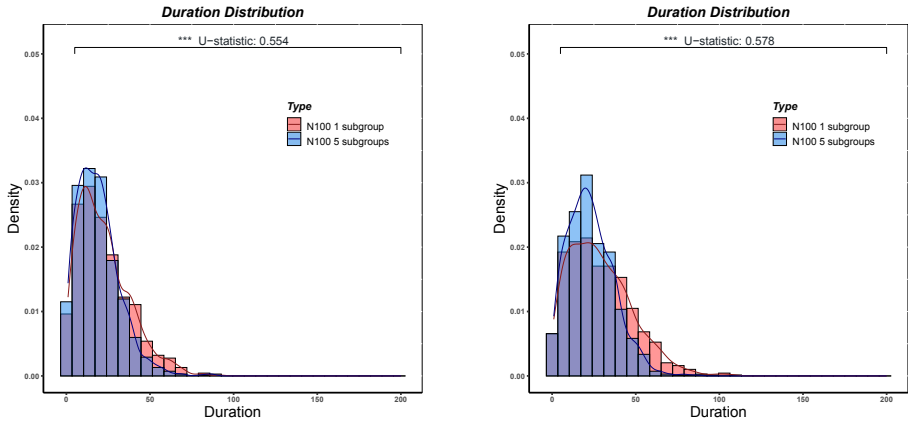
(a) β=0.5, R_0=0.8

(b) β=0.62,R_0=1

(c) β=1.55,R_0=2.5

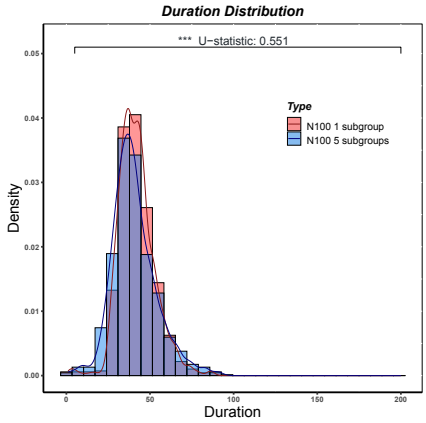
(d) β=3.11,R_0=5

Figure 4.4: Final size distribution plots for simulations with $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

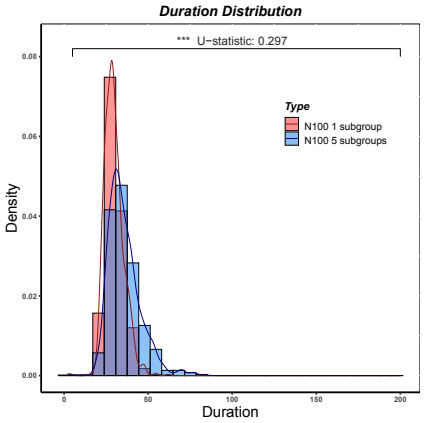


(a) β=0.5, R_0=0.8

(b) β=0.62,R_0=1



(c) β=1.55,R_0=2.5



(d) β=3.11,R_0=5

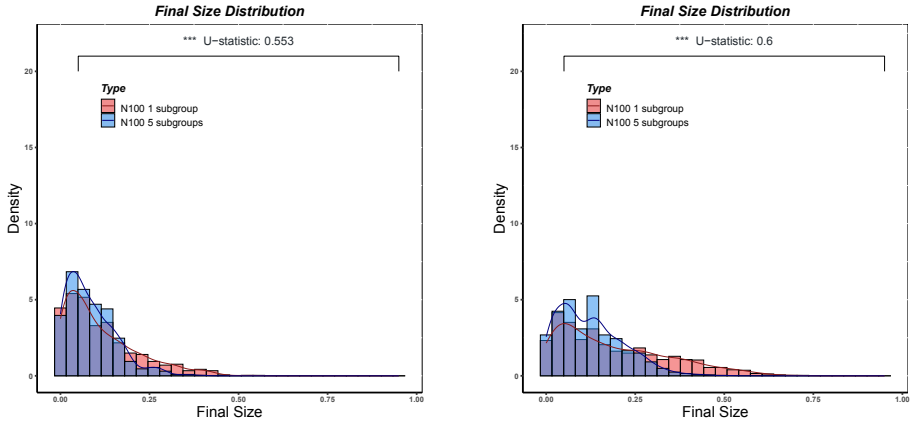
Figure 4.5: Duration distribution plots for simulations with $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

For smaller R_0 values, most simulated epidemics end up with only a small number of people infected. Smaller final size distributions indicate that epidemics die out quickly and will be of short duration. Figures 4.5. (a) and (b) show the mode of duration is about 20 days. For large R_0 values, more people end up being infected, the larger final size distribution cause the duration shifts to the right. The mode of duration is around 40 days in (c) and (d).

4.2.2 Varying γ

In the previous subsection for varying β, the R_0 value increases and so does the β. In this subsection, the opposite is true, as R_0 increases and γ decreases.

Figure 4.6 gives the final size distribution for this scenario. Comparing it to the final size distribution when β was varying but gamma was fixed, given in Figure 4.4, we see that the distributions are very similar in both cases across different values of R_0. It is because although the values of β and γ change, the value of R_0 does not, and from these two sets of plots we can confirm that the value of R_0 is the more important determinant of differences in the final size distribution.

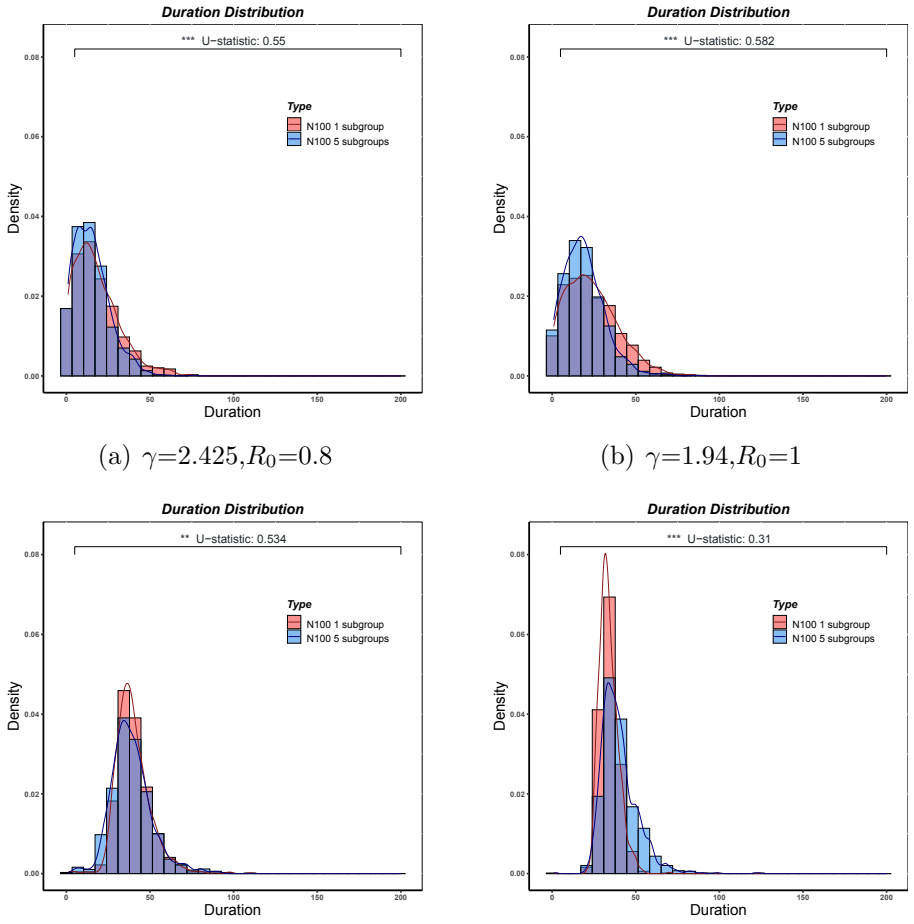


(a) γ=2.425,R_0=0.8 (b) γ=1.94,R_0=1

(c) γ=0.776,R_0=2.5 (d) γ=0.388,R_0=5

Figure 4.6: Final size distribution plots for simulations with $\beta = 1.94$, $\kappa = 0.25$, r=0.01, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

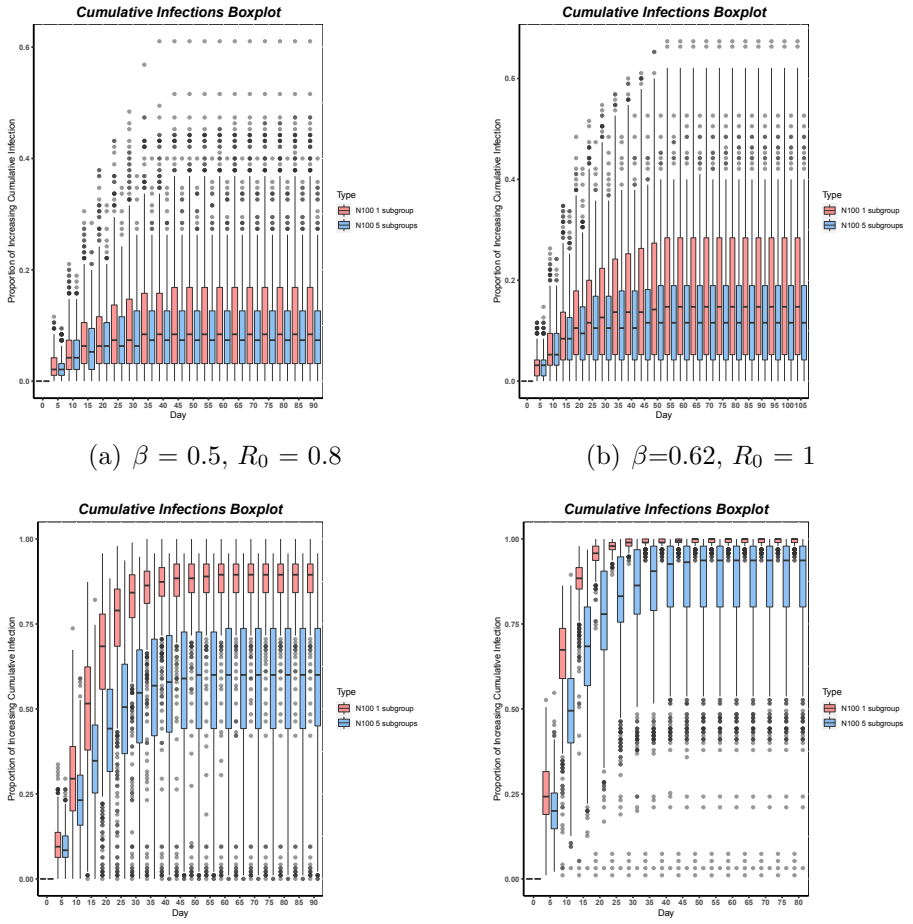
Next we look at the distributions of duration in Figure 4.7, and since the histograms and density curves are also extremely close to the plots in Figure 4.5 for varying β, we will use the boxplots to compare the duration of epidemic for these two groups.



(a) γ=2.425,R_0=0.8 (b) γ=1.94,R_0=1

(c) γ=0.776,R_0=2.5 (d) γ=0.388,R_0=5

Figure 4.7: Duration distribution plots for simulations with $\beta = 1.94$, $\kappa = 0.25$, r=0.01, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.



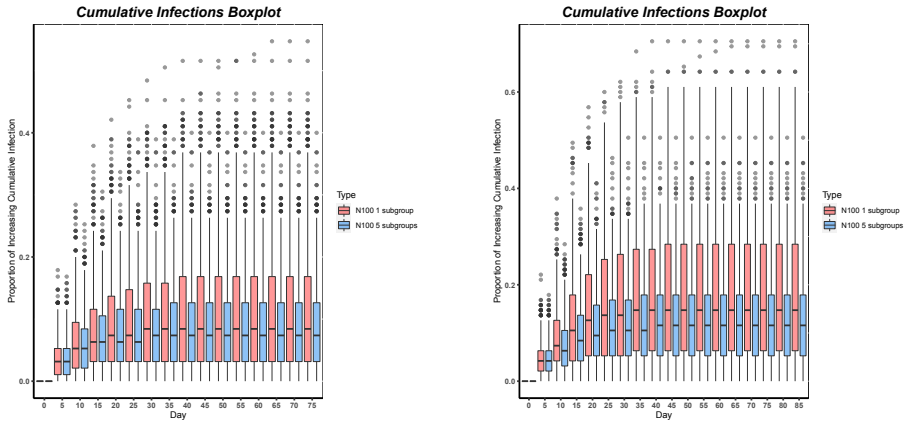
(a) $\beta = 0.5$, $R_0 = 0.8$

(b) β=0.62, $R_0 = 1$

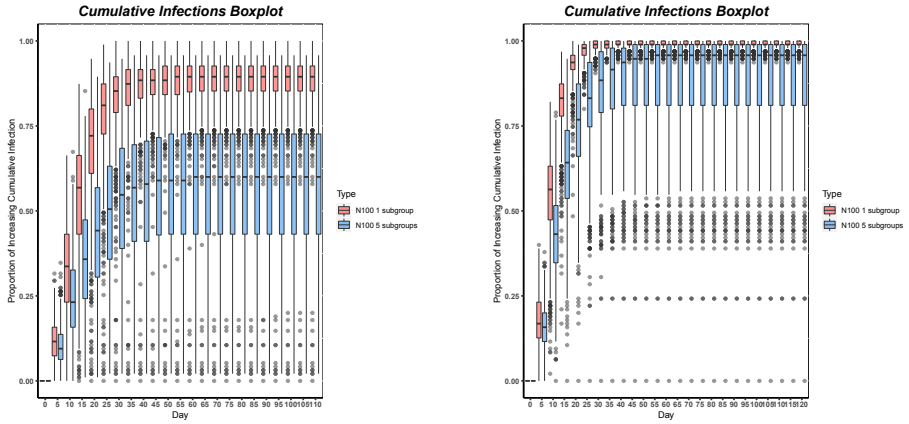
(c) β=1.55, $R_0 = 2.5$

(d) β=3.11, $R_0 = 5$

Figure 4.8: Boxplots for simulations with varying the value of β, $\gamma = \frac{1}{1.61}$, $\kappa = 0.25$, r=0.01, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.



(a) γ=2.425, $R_0 = 0.8$ (b) γ=1.94, $R_0 = 1$



(c) $\gamma = 0.776$, $R_0 = 2.5$ (d) $\gamma = 0.388$, $R_0 = 5$

Figure 4.9: Boxplots for simulations with varying value of γ , $\beta = 1.94$, $\kappa = 0.25$, r=0.01, N=100. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

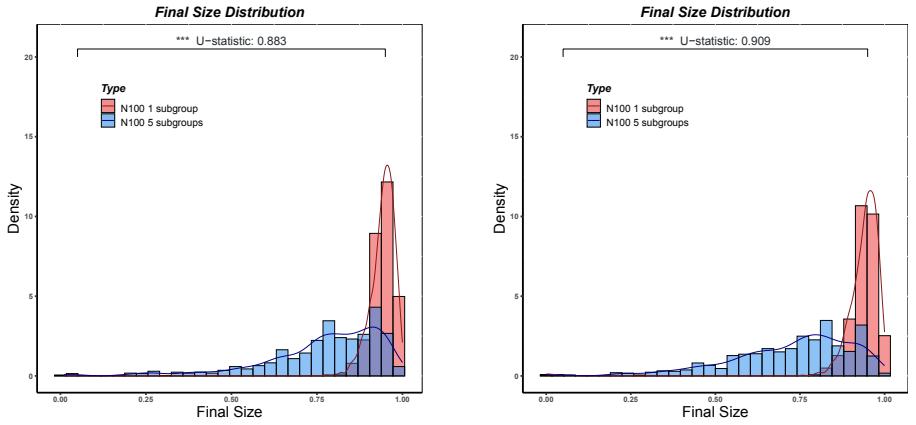
Comparing Figure 4.8 and Figure 4.9, the epidemic lasts longer when the $R_0 \leq 1$ for smaller β values, and longer when the $R_0 > 1$ for smaller γ values. This phenomenon occurs because when $R_0 \leq 1$, the values of β are also less than 1. Smaller infection rate means that it takes more time for an individual to be infected, which leads to longer durations. The values of γ are inversely small when $R_0 > 1$, which are small recovery rates. Individuals take longer to remove from infection status, again making epidemics last longer. The boxplot was constructed with the x-axis sufficiently large that the epidemic had ended at some time on the x-axis for every simulated epidemic, which means there were no infections in the population. In order to provide a clear view of the boxes, the boxplots in Figures 4.8 and 4.9 show

the trajectory of epidemic in every 5 days. The boxplot not only looks at the number of days the epidemic lasts, but also shows the overall trend of the epidemic and the variability of the number of infecteds per day. But we do not focus on that in this thesis.

4.2.3 Scaling β and γ

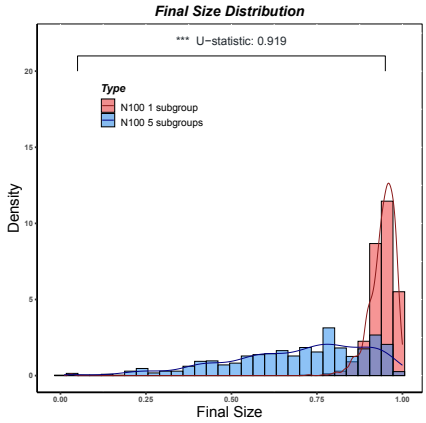
In the previous subsections, we only considered changing the value of β or γ depending on the different R_0 values. In this subsection, we will fix the R_0 value by scaling up and down the values of β and γ. We consider doubling and quadrupling both beta and gamma, as well as reducing them to half and a quarter of their original values. The plots in Figure 4.10 will be arranged from smallest to largest values of β and γ.

From the Figure 4.10 above we can see that the final size distributions for the population without bubbles and the population with five bubbles are very different for each of the five different β and γ values. In each scenario, the population without bubbles behaves similarly, with three-quarters and more of the population ending up infected in almost all of the thousand simulations. In contrast, the distribution of the bubbled population gradually shifts to the left and has a longer tail. Consistent with this, the U-statistic exhibited a continuous increase. As β and γ values increase, the rate of movement from susceptible compartment to exposed compartment and from infected compartment to removed compartment increases, and we can see through Figure 4.11 that the duration of the epidemic will be shorter. Increasing β and γ speeds up the occurrences of events leaving the number of secondary infections fixed. This provides a greater chance for the epidemic to dissipate due to stochastic effects, even if the R_0 is relatively high. Such stochastic effects have a larger end result on bubbled populations because the bubble sizes are smaller.

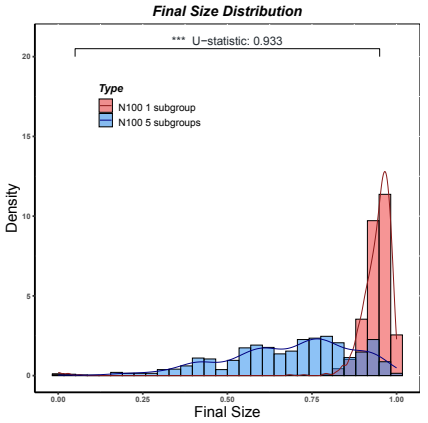


(a) β=0.485, γ=0.155

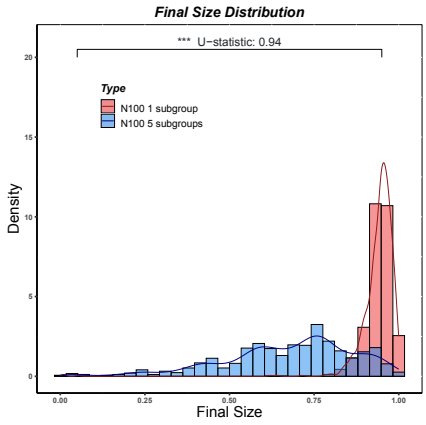
(b) β=0.97, γ=0.31



(c) β=1.94, γ=0.62

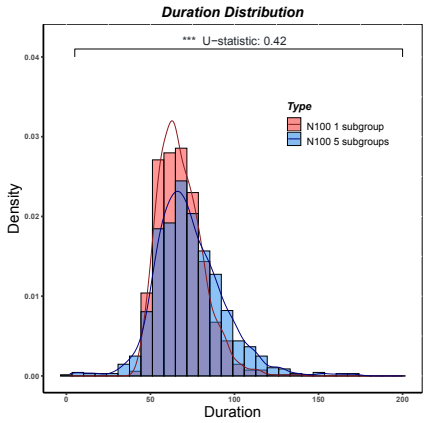


(d) β=3.88, γ=1.24

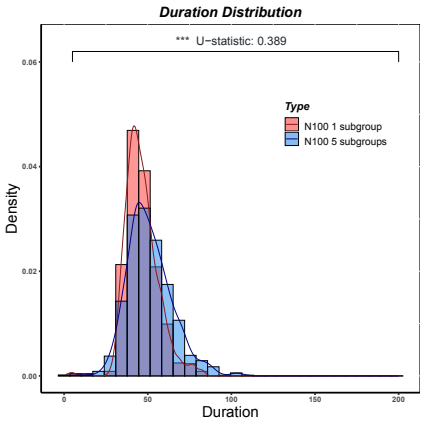


(e) β=7.76, γ=2.48

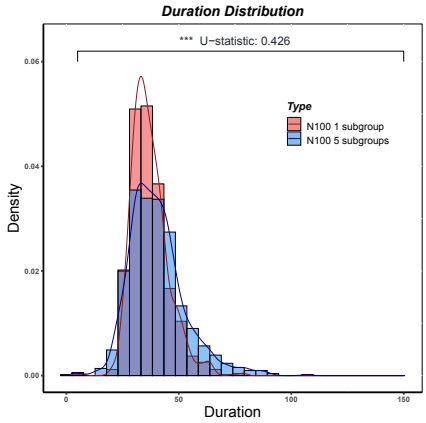
Figure 4.10: Final size distribution plots for simulations with $\kappa = 0.25$, r=0.01, R_0=3.12. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.



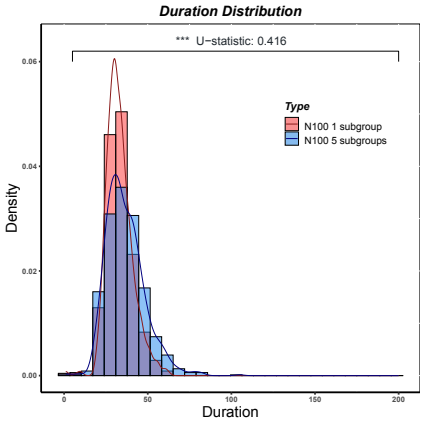
(a) β=0.485, γ=0.155



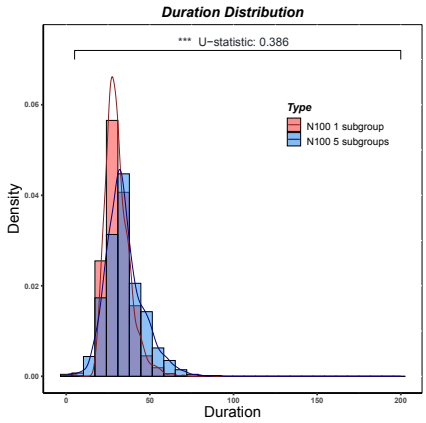
(b) β=0.97, γ=0.31



(c) β=1.94, γ=0.62



(d) β=3.88, γ=1.24

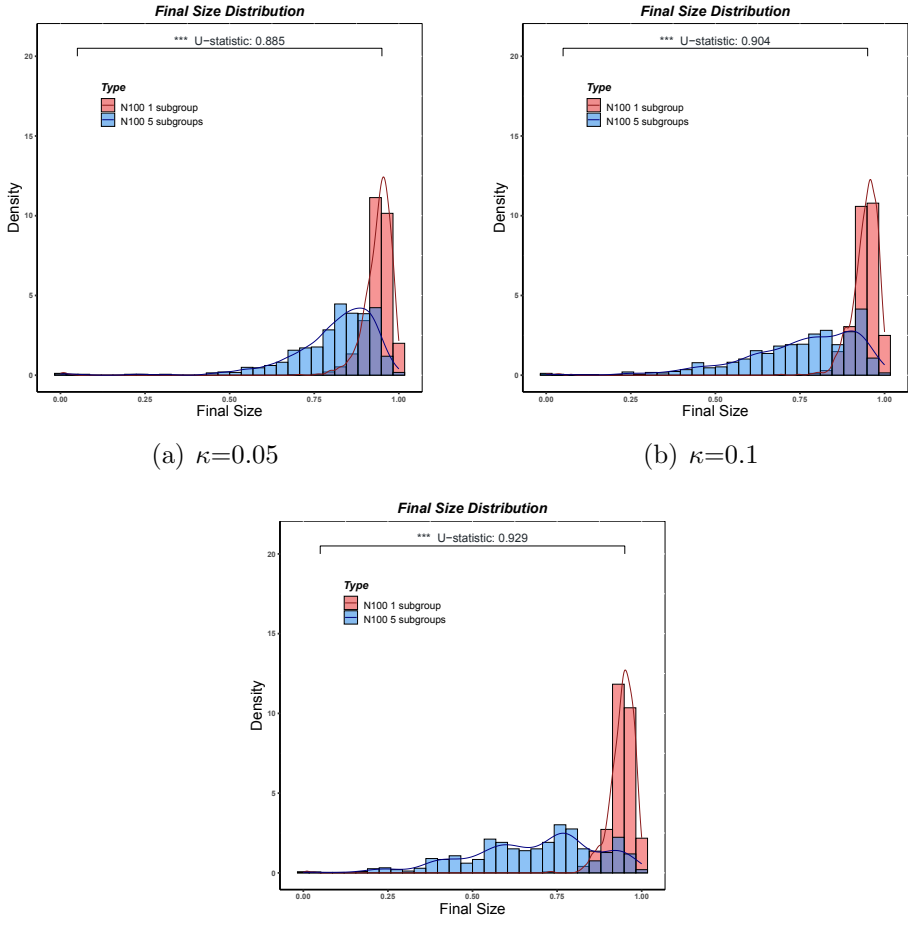


(e) β=7.76, γ=2.48

Figure 4.11: Duration distribution plots for simulations with $\kappa = 0.25$, r=0.01, R_0=3.12. We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

4.3 The latency rate κ

Now we look at the effect of changing κ (**Scenario C**). The latency rate, denoted as κ, signifies the rate of transition from the exposed compartment to the infected compartment within the epidemic model. Furthermore, the $\frac{1}{\kappa}$ signifies the expected time taken for the transition from exposed to infected compartments. In other words, the value of κ is not influenced by the number of the individuals, but depends on the nature of the virus. Changing the value of κ has no effect on the R_0 value, so we change the value of κ by setting the number of days from the exposed compartment to the infected compartment. Here we choose κ values of 0.05, 0.1 and 0.5, corresponding to 20, 10 and 2 days respectively.



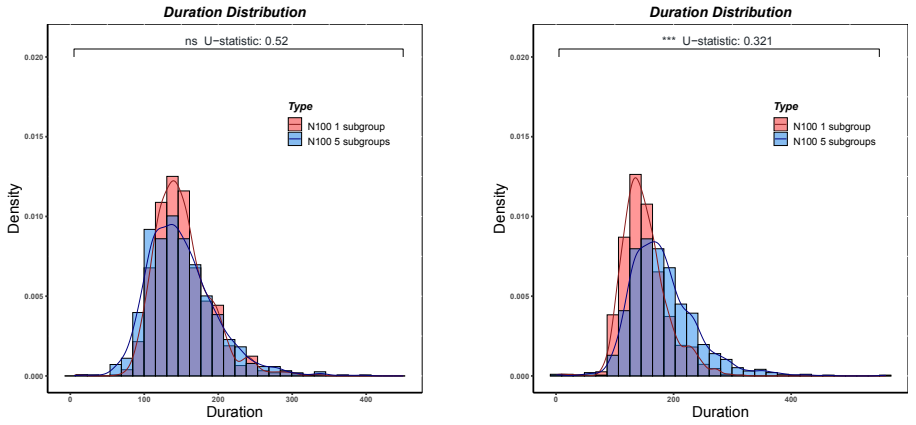
(a) κ=0.05

(b) κ=0.1

(c) κ=0.5

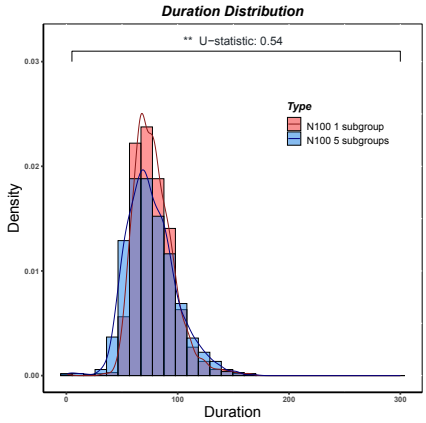
Figure 4.12: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\gamma = \frac{1}{1.61}$, r=0.01.We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

Results for the final size distributions are given in Figure 4.12. The R_0 and r values are constant, the final size distribution is left-skewed and the difference between the distribution of population without and with bubbles is significant. The final size distribution of bubbled population has longer tails than the distribution of population without bubbles, and the shorter the latency period, the longer the tails for the bubbled population.

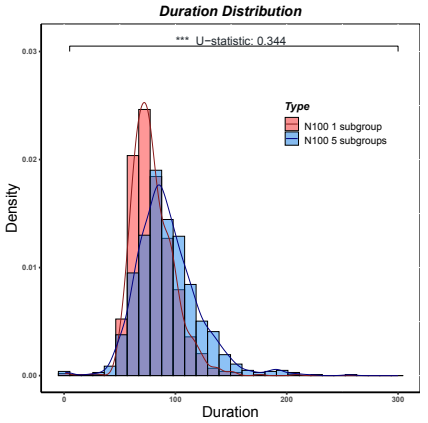


(a) κ=0.05, r=0

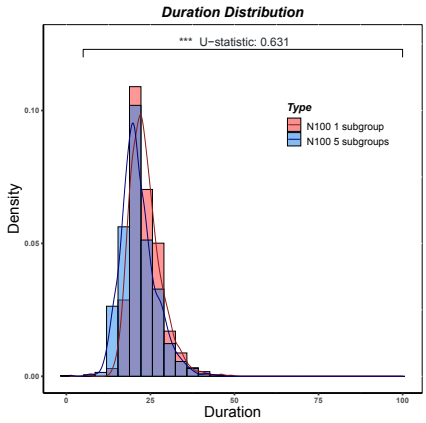
(b) κ=0.05, r=0.01



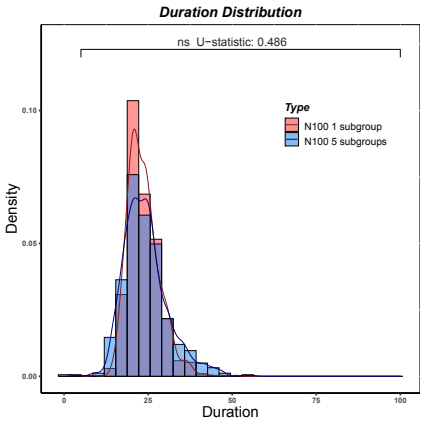
(c) κ=0.1, r=0



(d) κ=0.1, r=0.01



(e) κ=0.5, r=0

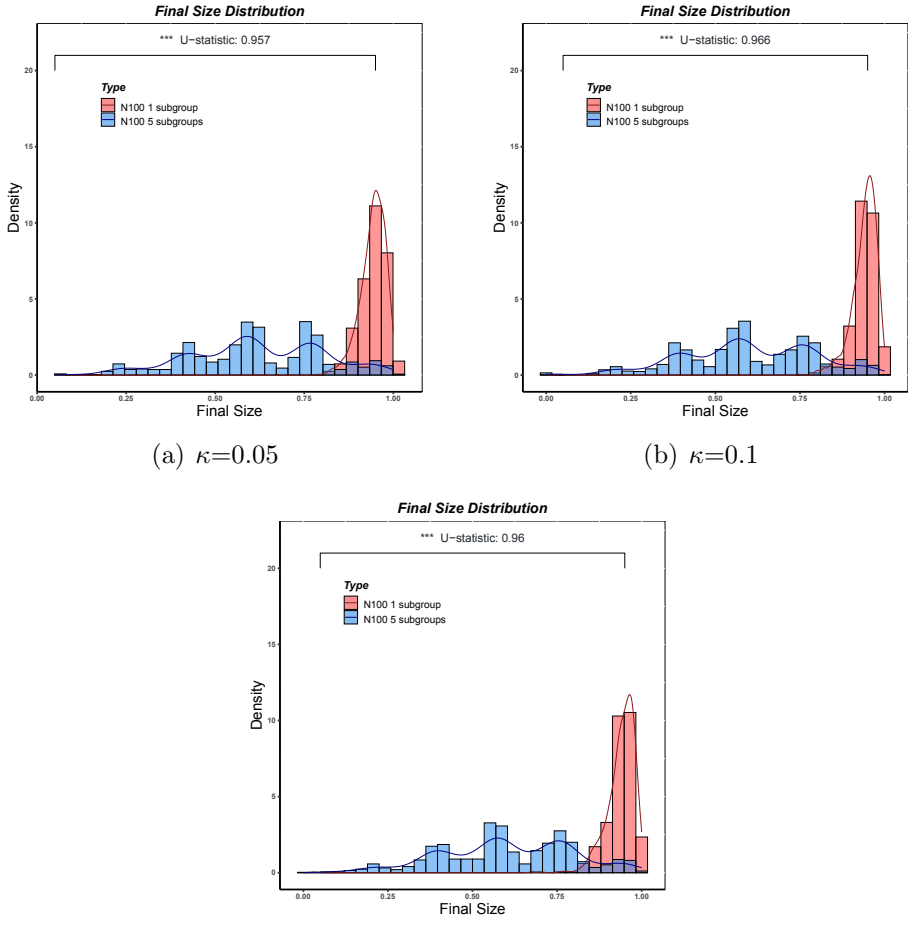


(f) κ=0.5, r=0.01

Figure 4.13: Duration distribution plots for simulations with $\beta = 1.94$, $\gamma = \frac{1}{1.61}$, $R_0 = 3.12$.We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

It is apparent from Figure 4.13 that the greater the value of κ, the shorter the relative duration of the epidemic. It is evident that the duration of the epidemic significantly extends with a κ value of 0.05, spanning approximately 420 days.. An increase in the κ value means that the time from exposure to infection is decreasing, so the duration of the epidemic is also decreasing. Similarly as in the case of fixed R_0, shorter durations lead to stochasticity playing a greater role and that affects populations that have been bubbled more than those that have not.

In this case we can compare the rate of transmission r to see if the presence or absence of transmission can have an impact when κ changes.



(a) κ=0.05

(b) κ=0.1

(c) κ=0.5

Figure 4.14: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\gamma = \frac{1}{1.61}$, r=0.We let I_n=5 for both populations and arrange that each bubble in the bubbled population has 1 infected individual.

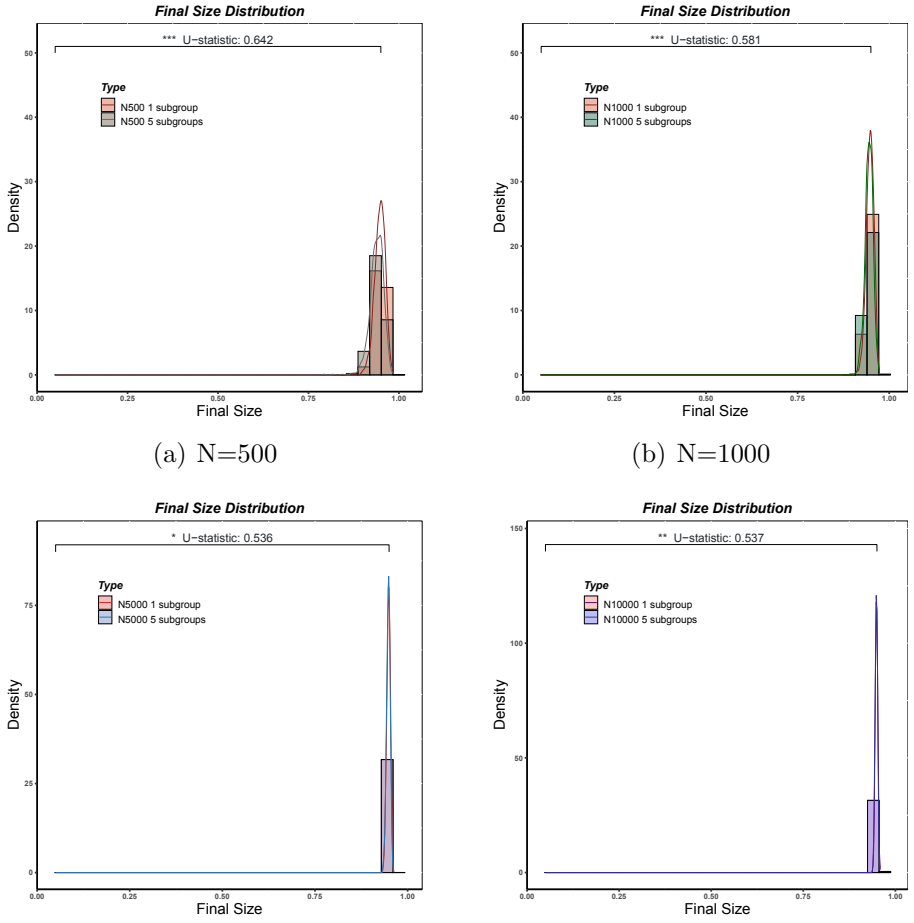
Figure 4.14 shows the final size distributions with $r = 0$. When there is no intergroup transmission, the final size distribution remains largely unaffected by changes in the value of κ. However, as shown in Figure 4.13, duration distributions appear similar for any given κ whether $r = 0$ or $r = 0.01$. Results of the Mann-Whitney test indicate, however, that there are significant differences. In the presence of intergroup transmission, the duration distributions of populations without and with bubbles become increasingly similar until there is no significant difference between the two distributions. However, in the absence of intergroup transmission, the duration distributions of the two populations become significantly different as the latency period decreases. The observed phenomenon suggests that when the latency period is prolonged and there is mobility between groups, it allows individuals more time to migrate between groups. As a result, the chance that an infected individual from one group will carry the epidemic to other groups increases that otherwise would have seen it drop out.

4.4 Populations and Initial Values

Having examined the impact of various transmission parameter values in the stochastic SEIR model in earlier sections, we will now investigate how different populations and initial values influence the epidemic. We will discuss **Scenario D** and **Scenario E** together. Our analysis will rely on the same parameter values from the literature, and we will represent the simulated data visually. Whereas previously we utilized a total population of 100 for the simulations, we will now investigate the impact of population size by considering total populations of 500, 1000, 5000 and 10000. To enable comparisons across the different total populations, we will explore two scenarios: one in which each population is divided into 5 bubbles, and another in which each population is divided into varying numbers of bubbles, each comprising 20 individuals. As the population size increases, we will also establish the initial value of infected individuals, limiting our analysis to one initial infected individual, as well as 5% and 10% of the total population.

4.4.1 Equal number of bubbles

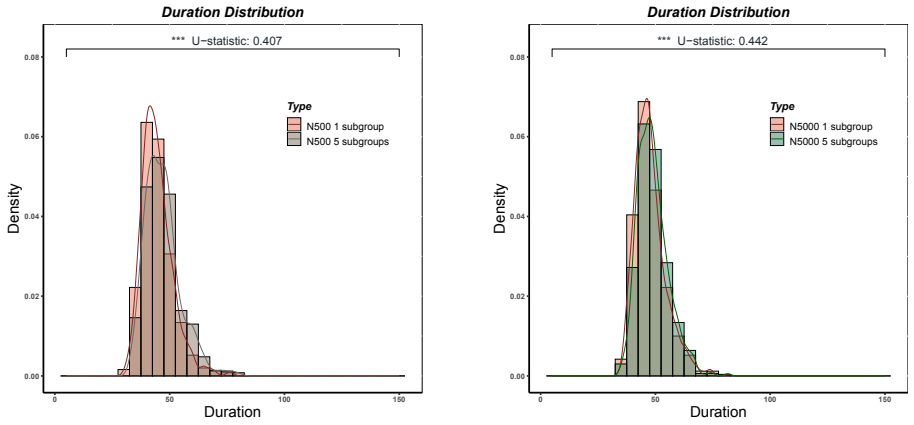
If the total population increases while bubbles remain unchanged, the number of individuals in each group will increase. This, in turn, leads to a shorter and more concentrated distribution in the final size distribution. In other words, a larger total population, with constant transmission parameters, results in more people becoming infected. Thus, the final size distribution observed in the Figure 4.15 is consistent with this trend.



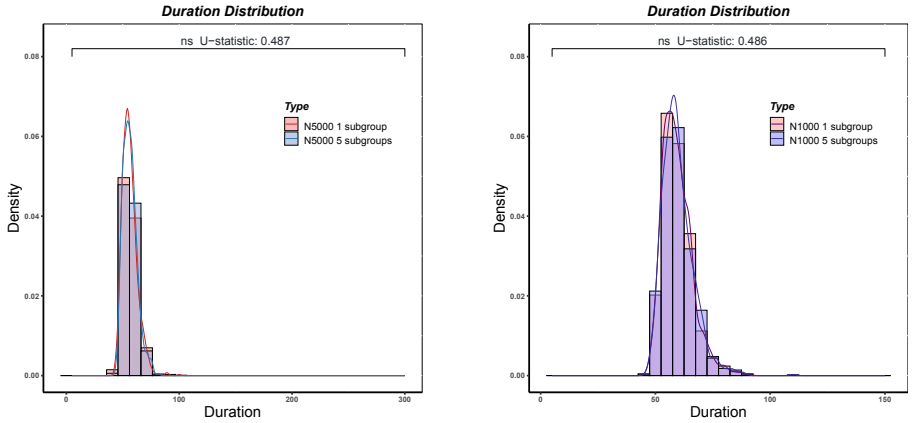
(a) N=500 (b) N=1000

(c) N=5000 (d) N=10000

Figure 4.15: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into 5 bubbles and with 5% of initial infected individuals for each group.



(a) N=500 (b) N=1000



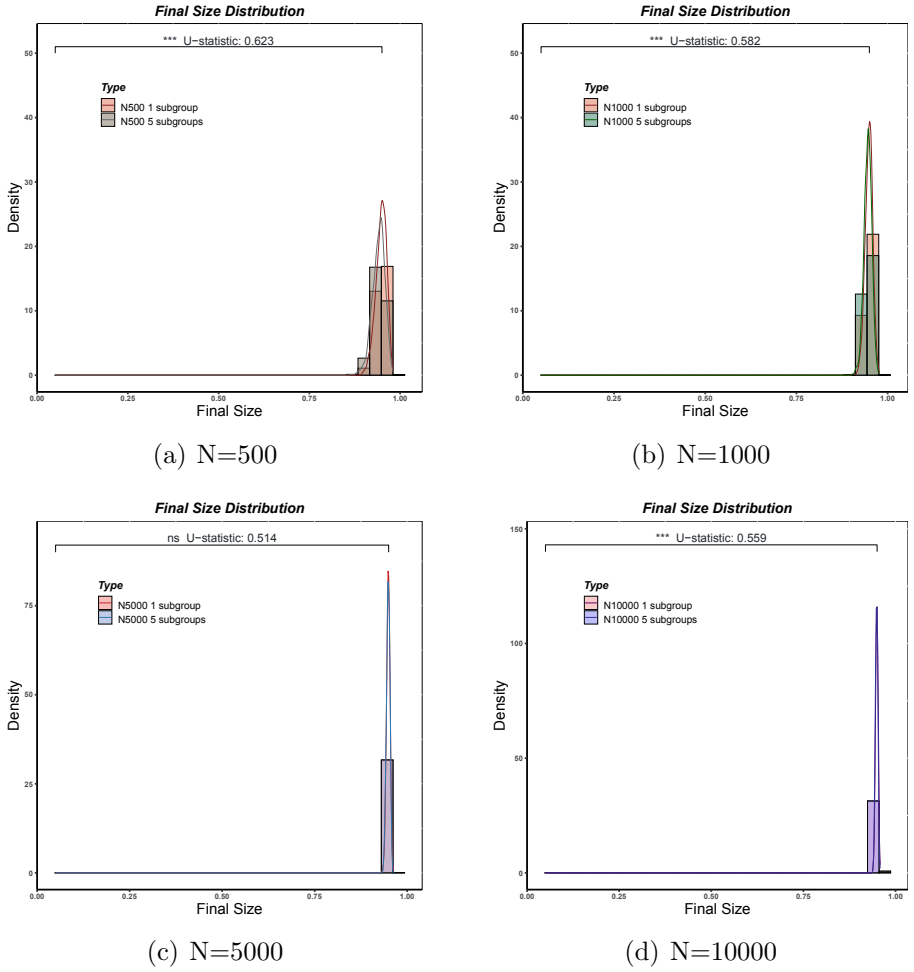
(c) N=5000 (d) N=10000

Figure 4.16: Duration distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into 5 groups and with 5% of initial infected individuals for each group.

As the population size increases, the duration distribution in Figure 4.16 becomes more concentrated and stable, with around 50 days. When comparing the final size distribution of large populations with and without bubbles, both distributions lack a trailing tail, and the corresponding U-statistics are approximately 0.5. As a result, the duration distribution comparisons are not statistically significant. This implies that dividing a larger population into only five groups is less effective in controlling an epidemic.

When the proportion of initial infection values increases, it is possible that the trailing tail of the final size distribution shifts to the right. However, the overall distribution trend with increasing population size remains similar to that observed

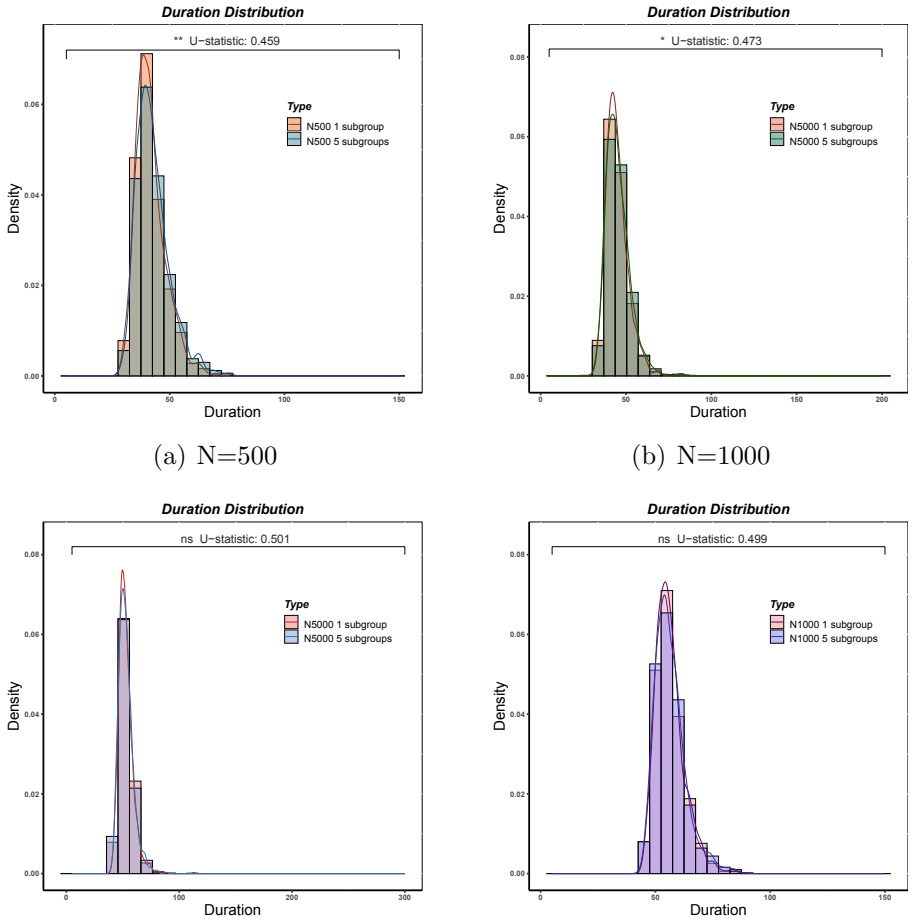
with 5% initial infection values. Figure 4.17 supports this speculation, and we can see that there is no substantial difference in the percentage of initial infections increasing from 5% to 10% as the population size grows. Additionally, the expected duration distributions remain highly similar.



(a) N=500 (b) N=1000

(c) N=5000 (d) N=10000

Figure 4.17: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into 5 groups and with 10% of initial infected individuals for each group.

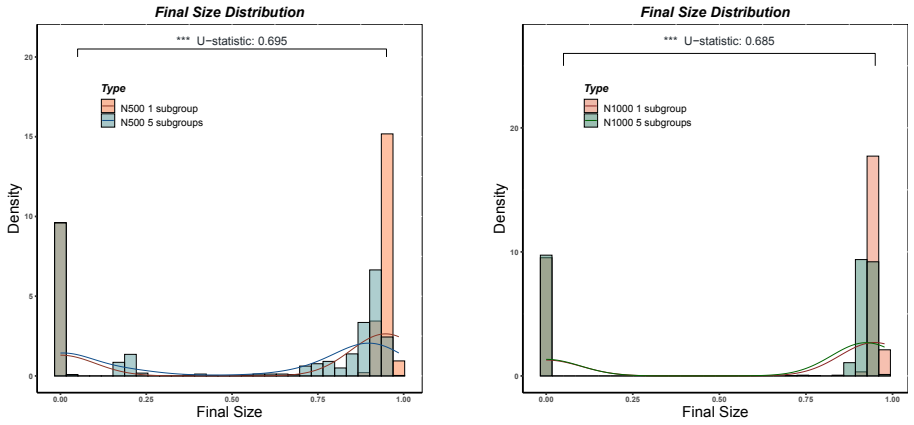


(a) N=500 (b) N=1000

(c) N=5000 (d) N=10000

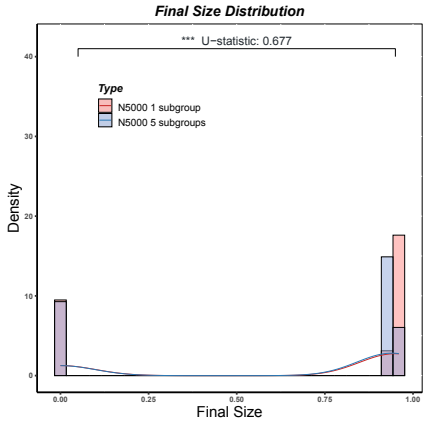
Figure 4.18: Duration distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into 5 groups and with 10% of initial infected individuals for each group.

If we set the initial number of infected individuals to 1, this implies that only one group will have an infected person while none of the other groups will have any infected individuals. It's important to note that there will be no individuals infected during the simulation if the first transition is from the infected compartment to the removed compartment. As a result, no one will transition from the susceptible compartment in the next step.

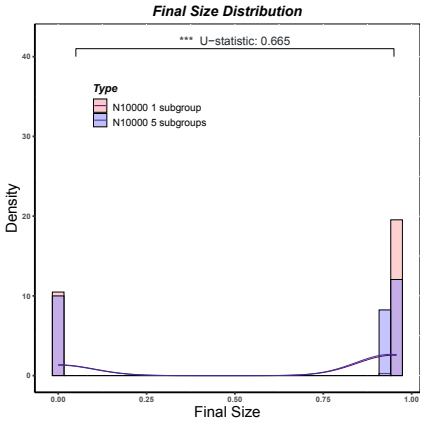


(a) N=500

(b) N=1000



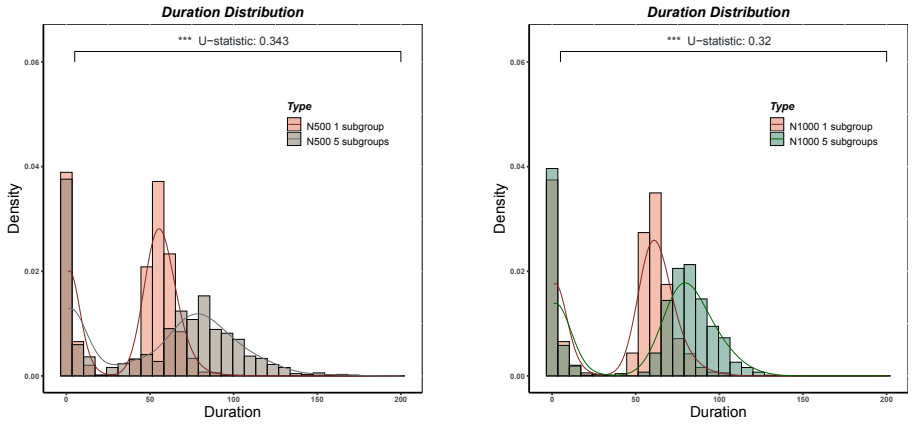
(c) N=5000



(d) N=10000

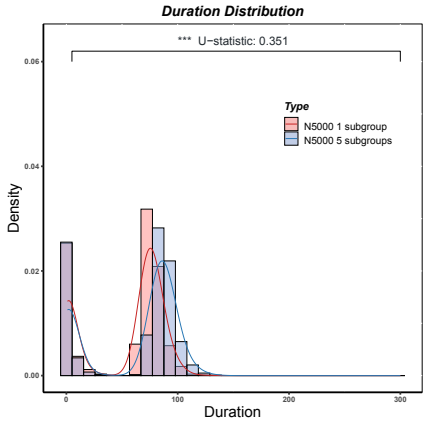
Figure 4.19: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into 5 groups and with only 1 infected individual for each population.

In cases where there is only one group, there is no movement between bubbles, and the final size distribution is typically concentrated at 0 and 0.9. However, when bubbles are introduced, the infected individual can move both within and between groups, leading to more diverse outcomes. Figure 4.19 (a) displays the final size distribution for a total population of 500. Since each group has a small population and a low initial infection rate, the infection rates fluctuate at all stages of the simulation, but it is rare for almost everyone to become infected. As the total population increases, the final size distribution of the bubbled population gradually approaches that of the population without bubbles.

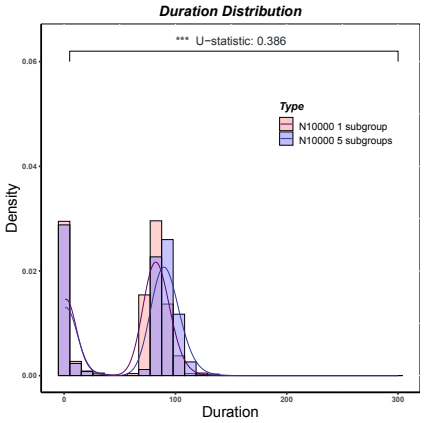


(a) N=500

(b) N=1000



(c) N=5000



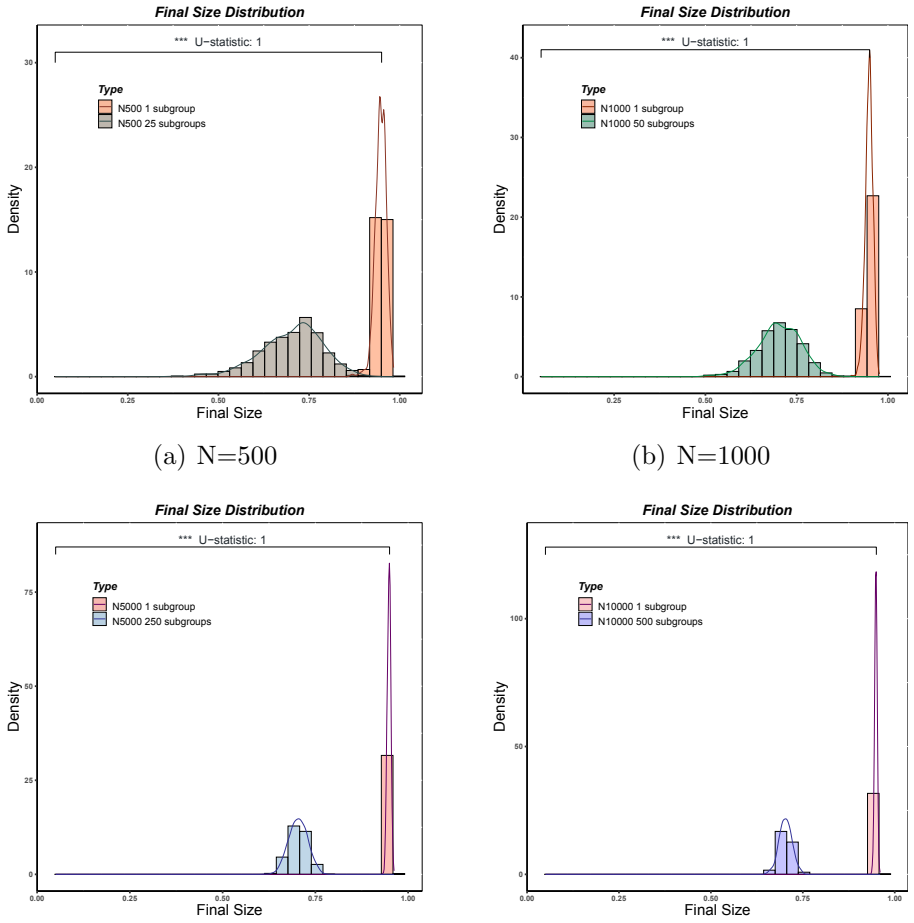
(d) N=10000

Figure 4.20: Duration distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into 5 groups and with only 1 infected individual for each population.

In cases where there is only one infected individual in the simulation, the duration of the epidemic will only be 0 days. This leads to a bimodal distribution of duration. When we compare the increase in initial infection values over time, we can observe that smaller initial infection values provide greater assistance in controlling the epidemic when we divide the population into bubbles. This indicates that we should begin dividing the population at the outset of the epidemic, and the earlier we do so, the better.

4.4.2 Equal size of bubbles

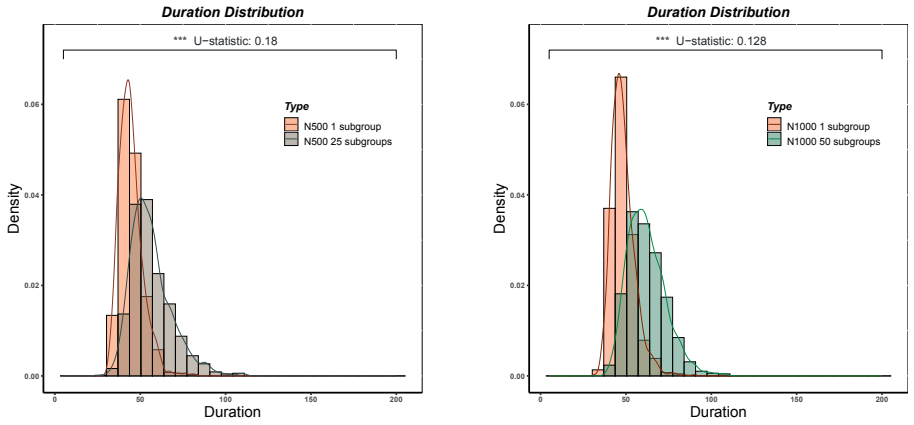
The simulations of populations with equal group size exhibit a contrasting pattern to the simulations of equal group numbers. As the population increases, the final size distribution in Figure 4.21 shifts towards the left and the trailing tail diminishes, although it remains wider than the distribution without bubble. Notably, when the total population exceeds 1000, the final size distributions of the two comparison populations on the same plot do not overlap, resulting in a U-statistic value of 1. This is accompanied by a more distinct difference in the duration distributions as demonstrated in the Figure 4.22.



(a) N=500 (b) N=1000

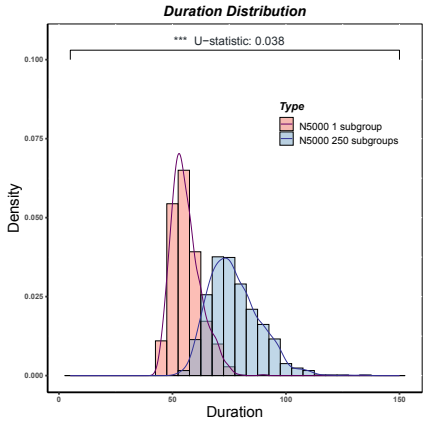
(c) N=5000 (d) N=10000

Figure 4.21: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into different numbers of groups and each group has 20 individuals with 1 initial infected individual. The initial infection value is therefore a 5% of the total population.

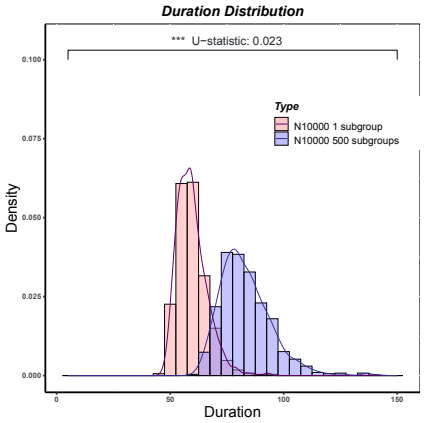


(a) N=500

(b) N=1000



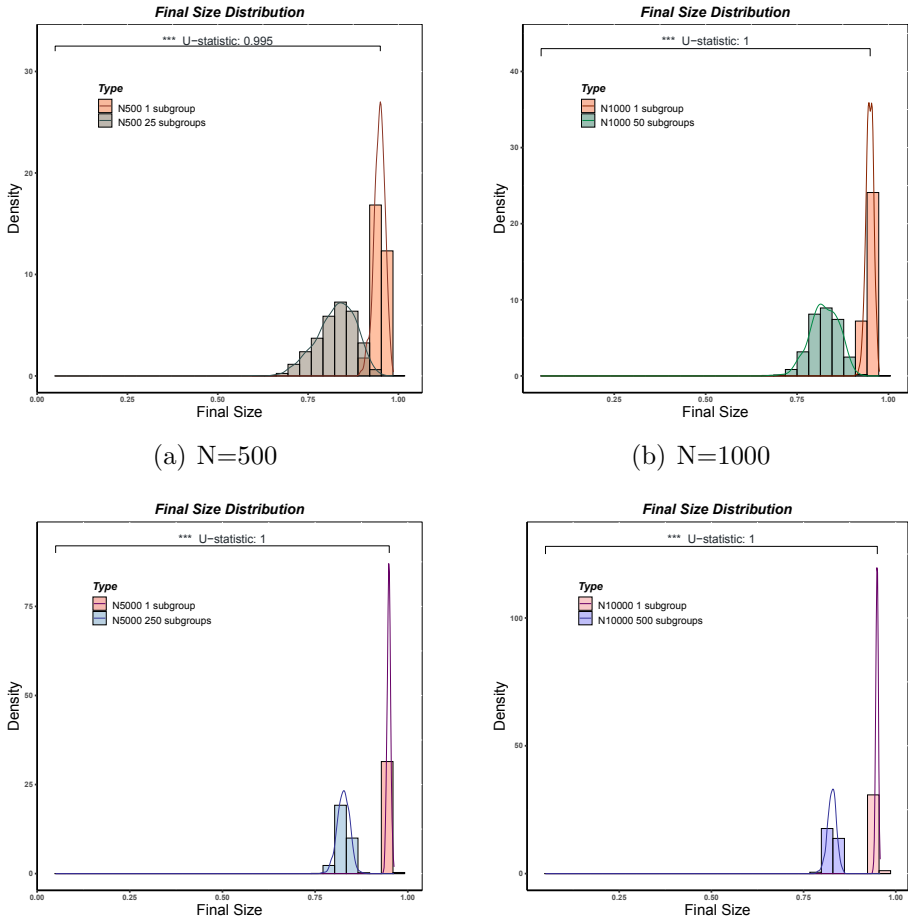
(c) N=5000



(d) N=10000

Figure 4.22: Duration distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into different numbers of groups and each group has 20 individuals with 1 initial infected individual. The initial infection value is therefore a 5% of the total population.

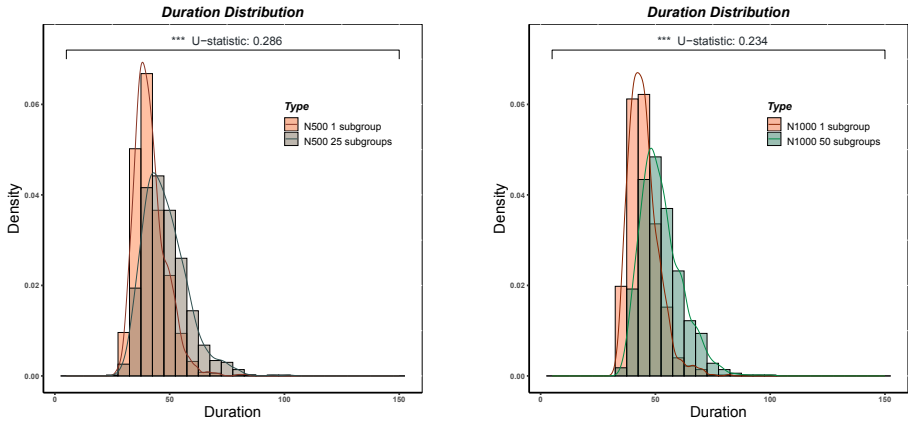
From the final size distribution in Figure 4.21 and duration distribution in Figure 4.22, we can observe that as the population size and subgroup size increase, the number of final infections becomes more concentrated and the duration of the epidemic increases. This suggests that bubbling can be effective, and in fact, increasingly effective, with large populations if the number of individuals per bubble can be kept fairly constant. Furthermore, if we increase the percentage of initial infected individuals, we can expect to observe the same trend.



(a) N=500 (b) N=1000

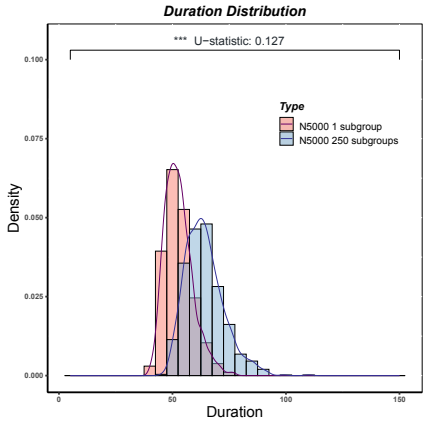
(c) N=5000 (d) N=10000

Figure 4.23: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into different numbers of groups and each group has 20 individuals with 2 initial infected individuals. The initial infection value is therefore a 10% of the total population.

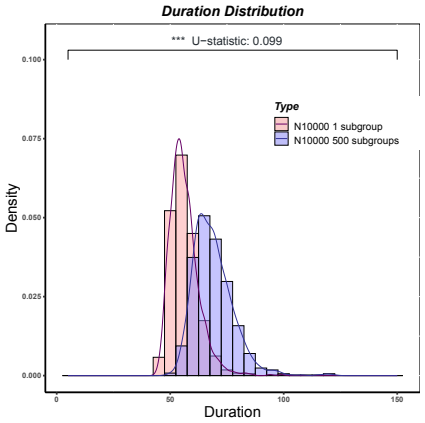


(a) N=500

(b) N=1000



(c) N=5000

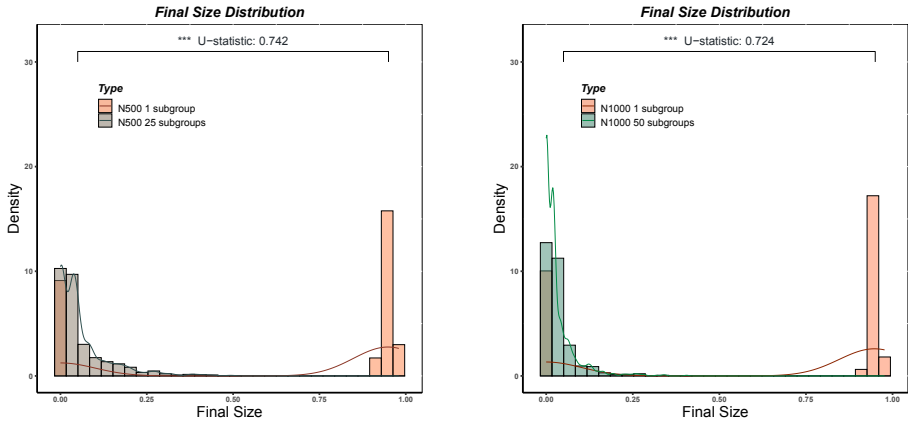


(d) N=10000

Figure 4.24: Duration distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into different numbers of groups and each group has 20 individuals with 2 initial infected individual. The initial infection value is therefore a 10% of the total population.

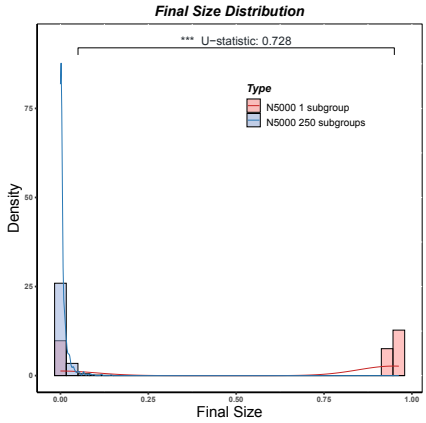
In Section 4.4.1, the expected duration distribution (Figure 4.18) showed less significant differences with increasing population size. However, in this section, the opposite is true. As we divide the population into more groups, the differences in the expected duration distributions (Figure 4.24) become more significant.

Similarly to the Section 4.4.1, when we increase the number of groups while keeping the size of each group fixed and only have 1 initial infected individual, the final size distributions (Figure 4.25) still shifts to the left. Additionally, there is no scenario where everyone ends up infected.

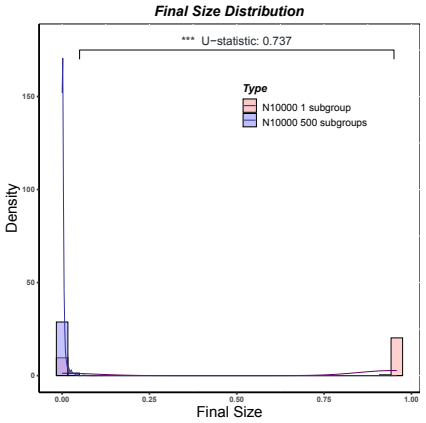


(a) N=500

(b) N=1000



(c) N=5000



(d) N=10000

Figure 4.25: Final size distribution plots for simulations with $R_0 = 3.12$, $\beta = 1.94$, $\kappa = 0.25$, $\gamma = \frac{1}{1.61}$, r=0.01. We divide each population into different numbers of groups and each group has 20 individuals. There is only 1 initial infected individual in total.

The analysis of various population and initial value scenarios indicates that early implementation of population division strategies is preferable when aiming to control epidemics. Furthermore, for larger populations, a more granular approach to population division is recommended for greater effectiveness.

Chapter 5

Conclusion

Within the framework of this paper, firstly, we reviewed Markov chains, the CTMC SIS model, the CTMC SEIR model, the DTMC SEIR model, as well as the connection between stochastic and deterministic models. A stochastic CTMC bubble SEIR model was then constructed that allowed for the division of the population and the mobility of individuals between bubbles. The main objectives were twofold:to assess the effectiveness of epidemic containment using bubble strategies, and to analyse scenarios by systematically varying each parameter in the model. Finally we explored the feasibility of the bubble strategies by analysing the final size distributions and duration distributions of the epidemics in different scenarios.

The results and analyses elucidated in Chapter 4's simulations study yield substantial insights. In Section 4.1, where distinct movement rate values were assigned to baseline parameters, the outcomes establish the efficacy of the bubble strategies in controlling disease propagation. Shifting our focus to Section 4.2, we proceeded to change the values of β and γ so that R_0 changed or remained consistent. The final size distributions shown in Figures 4.4 and 4.6 indicate the mode of the percentage of infections is approximately 10% for infections with smaller R_0 values, whereas a higher R_0 prompts infection of the majority of the population. At the same time, Figure 4.10 establishes that changes in β and γ are not as important as R_0, suggesting that a significant proportion of the population will eventually become infected at a constant R_0 of 3.12. Changes in the transmission parameters β and γ emphasize the key role of the basic reproduction number R_0 in determining the final epidemic distribution. In Section 4.3 an exploration into the impact of the latency rate κ was undertaken. With the extension of the latency period, the mobility across bubbles increases, providing individuals with raised opportunities for movement between bubbles. Consequently, the duration of the epidemic is prolonged.

Another factor that was investigated was the most effective bubbling strategy

with larger populations. When insisting on dividing the population into the same number of bubbles, the final size distribution of the bubbled population will gradually become extremely similar to the final size distribution of the population without bubbles. However, holding the number of individuals within each bubble constant, the bubbled populations show that about 70% of individuals are inevitably infected with the virus. Due to stochastic effects that lead to premature extinction of epidemics in some populations, in small populations, infections may not happen. This is the reason that 70% of individuals are eventually infected and with 30% of individuals are not infected. The final size of the bubbled population will be much smaller compared to the population without bubbles, which for large population sizes and the value of $R_0 = 3.12$ considered here, almost always results in 100% of the population getting infected. As population size grows, the final size gradually concentrates around the 70% threshold. Bubbling is an effective strategy, and its effectiveness is particularly acute in dividing larger population sizes into more bubbles. In addition, the adjustment of initial infection values highlights the crucial role of early population division. By comparing the same population at different values of initial infection, regardless of whether the number of bubbles or the size of the bubbles remains constant, a single initial infected individual results in a smaller final epidemic size than if the initial infection was 5% or 10% of the population.

In essence, this paper enhances our understanding of the dynamics of infectious disease transmission and provides valuable insights for the development of effective control strategies. However, it is important to acknowledge the limitations inherent in the current study. The study in this paper is based on the SEIR model. In actuality, epidemic dynamics consist of more compartments and involve time-varying parameters due to a range of interventions including quarantine policies and vaccination. As demonstrated by Gillis et al. (2021), utilizing a more complex SEIR model in conjunction with genetic algorithms can produce better policy budget. In addition, the application of the bubble SEIR model developed in this thesis to the practical data is a worthwhile direction for continued research. We hope that our work will inspire further research in this area and contribute to the effort to control and prevent the spread of infectious diseases.

Bibliography

Allen, L. J. (2008). An introduction to stochastic epidemic models. *Mathematical epidemiology*, pages 81–130.

Anderson, R. M. and May, R. M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford university press.

B. Roth, A. Payne, H. U. (2021). Timeline: COVID-19 in Nova Scotia. *The Signal*. Available at: `https://signalhfx.ca/timeline-covid-19-in-nova-scotia` (Accessed: October 13th, 2021).

Britton, T., Pardoux, E., Ball, F., Laredo, C., Sirl, D., and Tran, V. C. (2019). *Stochastic epidemic models with inference*, volume 2255. Springer.

Cousins, S. (2020). New Zealand eliminates COVID-19. *The Lancet*, 395(10235):1474.

Danon, L., Lacasa, L., and Brooks-Pollock, E. (2021). Household bubbles and COVID-19 transmission: insights from percolation theory. *Philosophical Transactions of the Royal Society B*, 376(1829):20200284.

Foster, F. (1955). A note on Bailey's and Whittles treatment of a general stochastic epidemic. *Biometrika*, 42(1/2):123–125.

Gagniuc, P. A. (2017). *Markov chains: from theory to implementation and experimentation*. John Wiley & Sons.

Gillis, M., Urban, R., Saif, A., Kamal, N., and Murphy, M. (2021). A simulation–optimization framework for optimizing response strategies to epidemics. *Operations Research Perspectives*, 8:100210.

Hethcote, H. W. (1976). Qualitative analyses of communicable disease models. *Mathematical biosciences*, 28(3-4):335–356.

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM review*, 42(4):599–653.

Hill, E. M. (2023). Modelling the epidemiological implications for SARS-CoV-2 of Christmas household bubbles in England. *Journal of Theoretical Biology*, 557:111331.

Kearns, N., Shortt, N., Kearns, C., Eathorne, A., Holliday, M., Mackle, D., Martindale, J., Semprini, A., Weatherall, M., Beasley, R., et al. (2021). How big is your bubble? characteristics of self-isolating household units ('bubbles') during the COVID-19 Alert Level 4 period in New Zealand: a cross-sectional survey. *BMJ open*, 11(1):e042464.

Pinsky, M. and Karlin, S. (2010). *An introduction to stochastic modeling*. Academic press.

Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A., and Jewell, C. P. (2021). Novel coronavirus 2019-nCoV (COVID-19): early estimation of epidemiological parameters and epidemic size estimates. *Philosophical Transactions of the Royal Society B*, 376(1829):20200265.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Willem, L., Abrams, S., Libin, P. J., Coletti, P., Kuylen, E., Petrof, O., Møgelmose, S., Wambua, J., Herzog, S. A., Faes, C., et al. (2021). The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. *Nature communications*, 12(1):1524.

Appendix A

CTMC Stochastic Epidemic Model

The individual level assumptions that lead to SIS and SEIR models are often not indicated in text book treatments such as Allen (2008). In this appendix we derive the form of the small time interval transition probabilities and, in doing so, indicate the assumptions that are being made. To calculate the probability that the next event is an increase in the number of infecteds, we let $\mathcal{I}$ and $\mathcal{S}$ denote the current set of infected and susceptible individuals. We let $\mathcal{I}_{mn}$ denote event that an infected individual, m, interacts with a susceptible individual n and $\widetilde{\mathcal{I}}$ be the event that the interaction between the two individuals results in an infection. Then

$$\begin{aligned} P\{I(t+\Delta t) = i+1 \big| I(t) = i\} &= \sum_{m\in\mathcal{I}}\sum_{n\in\mathcal{S}} P\{\mathcal{I}_{mn}\widetilde{\mathcal{I}}\} + o(\Delta t) \\ &= \sum_{m\in\mathcal{I}}\sum_{n\in\mathcal{S}} P\{\widetilde{\mathcal{I}}\big|\mathcal{I}_{mn}\}P\{\mathcal{I}_{mn}\} + o(\Delta t) \end{aligned} \tag{A.1}$$

where the $o(\Delta t)$ is a term that represents probabilities that involve two or more events. These probabilities are assumed small. Recall that $o(\Delta t)$ represents a function $g(\Delta t)$ satisfying that $\lim_{\Delta t\to 0}\frac{g(\Delta t)}{\Delta t} = 0$. We assume that the probability that an interaction results in an infection, $P(\widetilde{\mathcal{I}}_n\big|\mathcal{I}_{mn})$, is δ. We also assume that the probability of an interaction between m and n, $P(\mathcal{I}_{mn})$ is $\frac{\eta}{N}\Delta t + o(\Delta t)$. Note that we are assuming that the probability of interaction between any pair of individuals tends to decrease as population size increases. With these assumptions,

$$P\{I(t+\Delta t) = i+1\big|I(t) = i\} = i(N-i)\delta(\frac{\eta}{N}\Delta t) + o(\Delta t). \tag{A.2}$$

We let β be the rate of infection per contact between individuals. The rate depends on δ and η through $\beta := \delta\eta$ alone. Thus,

$$P\{I(t+\Delta t) = i+1\big|I(t) = i\} = i(N-i)\frac{\beta}{N}\Delta t + o(\Delta t). \tag{A.3}$$

To calculate the probability that the next event is an decrease in the number of infecteds, we let $\widetilde{\mathcal{S}}$ be the event that the infected individual m removed, then

$$P\{I(t+\Delta t)=i-1\Big|I(t)=i\}=\sum_{m\in\mathcal{I}}P\{\widetilde{\mathcal{S}}_m\}+o(\Delta t). \tag{A.4}$$

We assume that $P(\widetilde{\mathcal{S}}_m)$ is $\gamma dt+o(\Delta t)$. Then

$$P\{I(t+\Delta t)=i-1\Big|I(t)=i\}=i\gamma\Delta t+o(\Delta t). \tag{A.5}$$

In the SEIR model, once a susceptible individual comes into contact with an infected individual, then the susceptible individual will become exposed. We consider the probability that an susceptible individual moves to the exposed compartment e is $si\frac{\beta}{N}\Delta t+o(\Delta t)$. The probability that an infected individual moves to the removed compartment i is the same as the one calculated for the SIS model and is $i\gamma\Delta t+o(\Delta t)$. Let $\mathcal{E}$ denote the set of exposed individuals and let $\widetilde{\mathcal{E}}_l$ denote the event that individual l moves from exposed compartment to infected. We assume that $P(\widetilde{\mathcal{E}}_l)=\kappa\Delta t+o(\Delta t)$. Then

$$\begin{aligned} p_{(s,e,i),(s,e-1,i+1)}(\Delta t) &= \sum_{l\in\mathcal{E}}P\{\widetilde{\mathcal{E}}_l\}+o(\Delta t) \\ &= e\kappa\Delta t+o(\Delta t). \end{aligned} \tag{A.6}$$