Stockholms universitet

# The Effects of Quarantines on Epidemics in a Kindergarten Setting

Anton Stråhle

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2019:18
Matematisk statistik
Juni 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2019:18**
http://www.math.su.se

# The Effects of Quarantines on Epidemics in a Kindergarten Setting

Anton Stråhle*

June 2019

**Abstract**

In this thesis we examine how a so called SIQS epidemic spreads in a kindergarten setting with a multitype structure. When using the SIQS model we can model the flow of individuals between the subgroups as **S**usceptible → **I**nfectious → **Q**uarantined → **S**usceptible. By doing this we can examine how the diseases spread and how to combat them. We also implement a system of forced withdrawal to examine how such a method would affect the spread of the disease and the total number of days in quarantine.

Every day a child spends in quarantine, or at home as in this case, implies that a working adult has to take care of them. From an economic point of view, which is the primary view of this thesis, it is therefore important to minimize the total number of sick days during an epidemic. Furthermore we also wish to examine how the optimal number of days of forced withdrawal varies for epidemics with different degrees of infectiousness.

By using a forced withdrawal model we come to the conclusion that the spread and total number of days in quarantine can be reduced compared to the reference SIQS model. The method which leads to the minimization of quarantine days, and as a result the minimization of economic loss, is one where the number of days of forced withdrawal are large. By choosing a larger number of days of forced withdrawal we are able to isolate all initially infected individuals such that they can recover without infecting anyone else at the kindergarten. If the number of days is large enough we can be relatively sure that all quarantined children will have recovered upon return, leading to an early termination of the epidemic. Through simulation we are able to derive the number of days of forced withdrawal which minimizes quarantine days for a variety of diseases.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: anton.t.strahle@gmail.com. Supervisor: Pieter Trapman.

# Glossary

**Initially Quarantined**: Denoted by $\eta$. The number of people in quarantine after the initial stage of the epidemic.

**Initial Stage**: The period until the number of infectious individuals reaches approximately zero (and the number of quarantined is $\eta$) in the case of a forced withdrawal model.

**Reignition**: The epidemic does not terminate after the initial stage, instead it could be seen as restarting with some new number of initially infected. The "new" initially infected are the individuals that do not recover during the initial quarantine period.

$q$: The number of days we let the children stay in quarantine in the forced withdrawal model.

**Optimal Value of** $q$: The optimal value or $q$, sometimes referred to as $q_{\min}$, is the value of $q$ which minimizes the total number of sick days.

**ODE**: Ordinary differential equation.

**Local Infection**: When an infectious individual infects someone within their own group (a so called local individual)

**Global Infection**: When an infectious individual infects someone outside their own group (a so called global individual)

**External/Exterior Infection**: When a susceptible individuals gets infected by someone outside of the multitype structure.

**Memorylessness**: Memorylessness is a property of certain distributions indicating that they are independent of the time elapsed. This means that the probability of an event of interest occurring is the same at any trial. The geometric distribution is an example of such a memoryless distribution.

**Reproduction Number**: A quantity, denoted by $R_0$, used in epidemic modelling to indicate the infectiousness of a disease. It can also be related to whether or not a large outbreak can occur.

**Early Termination**: The epidemic terminates after the initial stage of the epidemic, meaning that none of the initially quarantined return susceptible (or in rare cases that some do return infectious but fail to spread the disease further).

# Contents

# Acknowledgements

I would like to thank my supervisor Pieter Trapman as well as Abhishek Majumder for their valuable feedback and ideas.

# 1 Introduction

An epidemic running rampant in a kindergarten could be seen as a problem for everyone involved since it affects both the children, their parents and their teachers. It is therefore of interest to examine different methods on how to most effectively combat epidemics, minimizing sickness and time at home. From an economic standpoint every day a child spends at home results in a working adult having to take care of them which implies some kind of economic loss to both their employer and to the society as a whole. There are of course also the physical and psychological effects of illness and disease which serves as a further incentive to minimize the number of infections and eventually rid the kindergarten of the epidemic as fast as possible.

The main focus of this paper is the economic standpoint since it is more reasonable to quantify economic loss rather than the physical and psychological well being of an individual. Furthermore we could also see the economic loss as a proxy for decrease in well being due to the clear relationship between the two. As previously mentioned, every day a child is sick implies some kind of economic loss. It is therefore crucial to minimize the amount of these days to minimize the economic loss generated by the epidemic. Since it is difficult to determine how much loss is incurred by a missed day we can instead choose to observe the number of days of absence. This is due to the evident positive relationship between absent days and economic loss.

The beginning of this paper includes an introduction to epidemic modelling whilst also bringing up some specific concepts that are used throughout the paper. After that we present the kindergarten setting in its entirety and start implementing the different models we wish to examine based on the theory and concepts introduced in the beginning.

Throughout the paper we attempt to introduce theoretical concepts and derive them when possible. Due to the nature of our models this might be difficult and we might have to rely on simulations or calculations under strict and unrealistic assumptions. The main goal is however to combine both theory and simulation and derive results from their respective outcomes.

# 2 Epidemic Modelling

In statistical infectious disease epidemiology we are generally concerned with the mathematics behind how an epidemic spreads. The spread is usually seen in how different subgroups of a population change in number over time. Consider a disease such as the influenza and assume that no individuals have been vaccinated priorly against it or gained partial immunity from it during a prior outbreak. In the early stages of the epidemic the majority of the population is susceptible to the disease, this subgroup of the population is henceforth denoted by $S$. A smaller subgroup of the population is infected with the disease and

can spread it to those that are susceptible. We denote this subgroup by $I$. After being infectious for some random period of time an infected individual will eventually recover from the influenza and thus enter the final subgroup named recovered which is denoted by $R$. After an individual recovers they are in the case of the influenza seen as immune to the disease and can not be infected again for the remaining period of the epidemic. This type of epidemic model is usually referred to as an SIR model which behaves as follows [2].

$$S \rightarrow I \rightarrow R$$

To observe the spread of a disease such as the influenza it is therefore important to get an understanding of how the individuals move between each subgroup over time.

We will now briefly introduce two different schools of epidemic modelling. The first being compartmental modelling which uses ordinary differential equations to explain the spread of infectious diseases. The other type of modelling, which is more closely related to this thesis as a whole, is stochastic modelling which explains the spread of epidemics through random events. We choose to introduced the compartmental model briefly to shed light on the fact that the methods we use in this thesis are by no means the only ones available. The model is also described to further introduce the subject of statistical epidemiology as a whole. It also proves meaningful to have introduced the concept since it closely relates to certain derivations later on in the paper.

Finally it should be noted that there is no universal framework when it comes to representing time in epidemic modelling. When it comes to compartmental models they generally follow a continuous time frame whilst stochastic models have been developed in both discrete and continuous time.

## 2.1 Compartmental Modelling

A generalized case of the compartmental model that we examine was first introduced in 1927 by Kermack and McKendrick in their paper *A Contribution to the Mathematical Theory of Epidemics* [5] which details how the spread of infectious diseases can be explained by a set of ordinary differential equations. In the equations below we showcase how their general theory can be applied to an SIR model whilst using different notations compared to the authors in an effort to stay consistent throughout the paper.

Let $\beta$ define the rate of infection (i.e the rate at which an infectious individual infects a susceptible individual) and $\gamma$ the rate of recovery (i.e the rate at which and infectious individual recovers). We also let $S(t), I(t)$ and $R(t)$ denote the number of susceptible, infected and recovered individuals at time $t$. In the model developed by Kermack and McKendrick the demographics of the population were not included which leads to the total population, $N = S(t) + I(t) + R(t)$, being constant over time.

The ODE Kermack McKendrick model can be described by the following set of differential equations.

$$\begin{aligned}\frac{S(t)}{dt} &= -\beta S(t)I(t)\\ \frac{I(t)}{dt} &= \beta S(t)I(t) - \gamma I(t)\\ \frac{R(t)}{dt} &= \gamma I(t)\end{aligned}$$

The equations above can be solved given an initial state $(S(0), I(0), R(0)) = (a, b, 0)$ such that $a + b = N$ given that no individuals start of as immune due to for example vaccination.

## 2.2 Stochastic Modelling

When using stochastic modelling we let stochastic events determine how the number of susceptible, infected and recovered individuals change over time rather than having a constant flow of individuals between the different subgroups which occurs in the compartmental model. This type of model seems inherently more intuitive since one could with some certainty assume that infections do not happen at a constant rate but rather at random with some probability. Furthermore the size of each subgroup is also integer valued compared to the deterministic model where the sizes are real valued.

### 2.2.1 The Reed-Frost Model

The type of stochastic epidemic model that we are mainly observing in this thesis is a so called Reed-Frost model. This type of model is often referred to as a chain binomial model since the flow of individuals between subgroups of the population can be explained by different conditioned binomial distributions. Like the aforementioned compartmental model, the model developed by Reed and Frost is also an SIR model, however it usually assumes discrete and not continuous time [2] [1]. When moving from continuous to discrete time we have to redefine our notations. Primarily $t \in \mathbb{N}$ compared to $t \in \mathbb{R}$ as in the continuous case. We also let the number of susceptible individuals at time $t$, $S(t)$, denote the number of susceptible individuals at the start of the $t$:th step in time. For the other subgroups, $I(t)$, $R(t)$ and $Q(t)$ which will be introduced later on, we make the same changes notation. Later on in the paper we introduce some distributions $Y(t)$, $X(t)$, $V(t)$ and $Z(t)$ which indicate the number of events occurring during the $t$:th step in time.

The probability for an individual to go from susceptible to infectious can be explained as the probability that the individuals fails to avoid infection from all other infectious individuals. As such each susceptible individual performs a Bernoulli trial with some probability each step in time to determine if they were

infected or not. From this the number of new infections during the $t$:th step in time can be explained by a binomial distribution under the assumption that infections happen independently of each other. This is an assumption that we make throughout the paper due to mathematical convenience. Let $Y(t)$ denote the number of infections made during the $t$:th step in time conditioned on both the number of infectious and susceptible individuals at the start of that step.

$$Y(t)|S(t), I(t) \sim \text{Bin}\left(S(t), 1 - (1 - \pi_I)^{I(t)}\right)$$

Here $\pi_I$ is the probability of infection and as such $1-(1-\pi_I)^{I(t)}$ is the probability that a susceptible individual does not avoid getting infected by any of the $I(t)$ infectious.

In the SIR model we do not only have the transition from susceptible to infectious but also that of infectious to recovered. Assuming that recoveries happen independently with some probability $\pi_R$ and that the individual recovery times are memoryless the number of recoveries during the $t$:th step in time can also be described by a different binomial distribution. Let $X(t)$ denote the number of recoveries during the $t$:th step in time conditioned on the number of infectious individuals at the start of that step.

$$X(t)|I(t) \sim \text{Bin}\left(I(t), \pi_R\right)$$

Using $X(t)$ and $Y(t)$ we can construct a system similar to the set of differential equations for the compartmental model which shows the random flow of individuals between the groups over time. The equations below are conditioned on the sizes of the subgroups during the start of the previous step in time, $S(t), I(t)$ and $R(t)$.

$$\begin{aligned} S(t+1) &= S(t) - Y(t) \\ I(t+1) &= I(t) + Y(t) - X(t) \\ R(t+1) &= R(t) + X(t) \end{aligned}$$

For the sake of our model we only need to introduce the basic functionality of the a Reed-Frost model since we only some of the basic properties in this thesis. Additional information regarding the Reed-Frost model can be found in section 1.2 of [2] or in [1].

# 3 Modelling

After having introduced the different kinds of epidemic modelling with a focus on the basic SIR model we now wish to examine and explain the models that are used throughout the paper. We begin by explaining the core of the model and then introduce the additional elements as we go forward.

### 3.1 The SIS Model

The primary epidemic model of this paper is an SIS variant of the Reed-Frost model in discrete time. The SIS model is slightly different compared to the SIR model which we previously introduced. After an individual recovers from infection they do not enter the subgroup recovered, which no longer exists, but rather the subgroup of susceptible, meaning that they can be infected again. The model behaves as follows.

$$S \to I \to S$$

The reasoning for implementing the model in discrete time relates back to the scope of the paper which is to examine how quarantines affect epidemics in a kindergarten setting. It seems intuitive that a child would enter and leave the quarantine on a daily basis rather than in continuous time. It is not unreasonable to assume that infections could be said to happen in discrete time since a child will not get infected whilst at home. It is however difficult to make a case for that recoveries happen in discrete time but for the sake of mathematical convenience we still choose to see the model in discrete time where every jump in time $t-1 \to t$ indicates that one day has passed.

Since we have chosen to ignore the demographics of the population like Kermack and McKendrick the total population, $N = S(t) + I(t)$, is still constant through time. If we still assume that all infections happen independently of one another, during each day every susceptible individual performs a Bernoulli trial, with some probability depending on the number of infected, to determine if they get infected during the day or not.

As previously defined, $\pi_I$ is the probability that an infectious individual infects a susceptible individual during a day. From the previous section on Reed-Frost models we can redefine $Y(t)$ since it now only depends on $S(t)$ or $I(t)$ due to the constancy of the total population.

$$Y(t)|S(t) \sim \text{Bin}\left(S(t), (1-\pi_I)^{N-S(t)}\right)$$

We previously defined $\pi_R$ as the probability of recovery each day. The time until recovery should for the sake of mathematical convenience be seen as a geometric distribution due to its inherent memorylessness [2]. This might not seem intuitive at all but is in some regard a necessary assumption to make due to the memorylessness. Let $I$ denote the time an infectious individual stays infectious which we will henceforth call the infectious period.

$$I \sim \text{Geo}(\pi_R)$$

Since the probability for an individual to recover during each passing day is $\pi_R$ due to the memorylessness of the geometric distribution we can redefine the number of recoveries during the $t$:th day, $X(t)$, which now only depends on either $S(t)$ or $I(t)$.

$$X(t)|S(t) \sim \text{Bin}(N - S(t), \pi_R)$$

The change in the number of susceptible and infectious individuals each day can then be defined by the following equations conditioned on $S(t)$ and $I(t)$.

$$S(t+1) = S(t) + X(t) - Y(t)$$
$$I(t+1) = I(t) - X(t) + Y(t)$$

These equations can not be solved like the ordinary differential equations presented in the compartmental model due to the dependence on the outcome of several conditional stochastic variables. We can however solve them in a similar way to the differential equations by observing the expected values of the stochastic variables, rather than the random outcomes. As such we can redefine the equations as follows where we still condition on $S(t)$ and $I(t)$.

$$S(t+1) = S(t) + \mathbb{E}[X(t)] - \mathbb{E}[Y(t)]$$
$$I(t+1) = I(t) - \mathbb{E}[X(t)] + \mathbb{E}[Y(t)]$$

Instead of looking at the population as a whole we can examine how individuals transition from one state to another. To do this we introduce the stochastic process $\{X_j(t) : t \geq 0\}$ which indicates the state of individual $j$ at time $t$ such that $X_3(5) = S$ indicates that individual 3 is susceptible at the start of the 5:th day. Define $\Omega$ as the state space of the process such that $\Omega = \{S, I\}$. Since the transition of an individual only depends on its current state, that is $\mathbb{P}\big(X_j(t+1) = S|X_j(t) = S\big) = \mathbb{P}\big(X_j(t+1) = S|X_j(t) = S, ..., X_j(0) = S\big)$ we have the the stochastic process is a Markov process in discrete time since it satisfies the Markov property. From this we can establish the stochastic matrix of the process.

$$P_j(t) = \begin{pmatrix} (1-\pi_I)^{I(t)} & 1-(1-\pi_I)^{I(t)} \\ \pi_R & 1-\pi_R \end{pmatrix}$$

In this paper we are not be using the idea of the individual Markov processes to derive theoretical proofs. The idea of observing the states of the individuals as a Markov process is however essential for the simulations we perform.

### 3.1.1 Reproduction Number of an SIS Model

An important quantity for all kinds of epidemic models is the basic reproduction number which is usually denoted by $R_0$. The reproduction number indicates if there is a probability that a large outbreak may occur. It is defined by the number of individuals an initial infected infects during their infectious period given a large amount of available susceptible [2].

Assuming that we have $N-1$ initial susceptible individuals and one infected we let $Z$ denote the number of individuals the infected individual infects during its infectious period $I$. To determine the distribution of $Z$ we need to calculate the probability that a susceptible individual gets infected by the initial infected. To do this we begin by calculating the probability that a susceptible does not get infected during the initial infecteds infectious period conditioned on the length of that period.

$$\begin{aligned}
\mathbb{P}(\text{Not infected}|I=1) &= 1-\pi_I \\
\mathbb{P}(\text{Not infected}|I=2) &= (1-\pi_I)^2 \\
&\vdots \\
\mathbb{P}(\text{Not infected}|I=n) &= (1-\pi_I)^n
\end{aligned}$$

Through induction as can be shown that $\mathbb{P}(\text{Not infected}|I) = (1-\pi_I)^I$ which can be seen in section [10.1.2]. Using the law of total probability we can derive the overall probability to not get infected during the initial stage of the epidemic.

$$\begin{aligned}
\mathbb{P}(\text{Not infected}) &= \sum_{i=1}^{\infty}(1-\pi_I)^i\mathbb{P}(I=i) \\
&= \sum_{i=1}^{\infty}(1-\pi_I)^i(1-\pi_R)^{i-1}\pi_R \\
&= (1-\pi_I)\pi_R\sum_{i=1}^{\infty}((1-\pi_I)(1-\pi_R))^{i-1} \\
&= \frac{(1-\pi_I)\pi_R}{\pi_I+\pi_R-\pi_I\pi_R}\underbrace{\sum_{i=1}^{\infty}(1-\pi_I-\pi_R+\pi_I\pi_R)^{i-1}(\pi_I+\pi_R-\pi_I\pi_R)}_{=1} \\
&= \frac{(1-\pi_I)\pi_R}{\pi_I+\pi_R-\pi_I\pi_R}
\end{aligned}$$

The second part of the equation above is equal to 1 since it is the probability function of a geometric distribution with parameter $\pi_I+\pi_R-\pi_I\pi_R$ summed over its complete support $[1,\infty)$. From this we get that the probability that the initial infected infects a susceptible is simply the complement to the above probability.

$$\mathbb{P}(\text{Infected}) = 1-\mathbb{P}(\text{Not Infected}) = 1-\frac{(1-\pi_I)\pi_R}{\pi_I+\pi_R-\pi_I\pi_R}$$

Thus $Z$ takes the distribution below.

$$Z \sim \text{Bin}\left(N-1, 1-\frac{(1-\pi_I)\pi_R}{\pi_I+\pi_R-\pi_I\pi_R}\right)$$

Since $R_0 = \mathbb{E}[Z]$ we get the following.

$$R_0 = \mathbb{E}[Z] = (N-1)\left(1 - \frac{(1-\pi_I)\pi_R}{\pi_I + \pi_R - \pi_I\pi_R}\right)$$

The probability of a large outbreak is 0 if $(N-1)\left(1 - \frac{(1-\pi_I)\pi_R}{\pi_I+\pi_R-\pi_I\pi_R}\right) \leq 1$. If $(N-1)\left(1 - \frac{(1-\pi_I)\pi_R}{\pi_I+\pi_R-\pi_I\pi_R}\right) > 1$ the probability is non-zero and a large outbreak may occur [2].

## 3.2 The SIQS Model

After briefly introducing the SIS core of our model we now wish to introduce a method which is generally used when trying to suppress an epidemic. In cases of extremely contagious epidemics infectious individuals might be put into so called quarantines. Whilst in the quarantine they are temporarily separated from the susceptible population in the hopes that they will recover without infecting anyone else. If the quarantine is successful they will recover without infecting anyone else and thus terminating the epidemic. Define the new subgroup $Q(t)$ as the number of individuals in a quarantine at the start of the $t$:th day. The new model which we are introducing is a so called SIQS model and behaves as follows.

$$S \to I \to Q \to S$$

In this case an individual gets infected and then enters a quarantine after some random period of time. The individual then spends some other random period of time in the quarantine, unable to infect any of the remaining susceptible individuals, until they recover. It is possible to include the possibility that an infectious individual might recover before they are sent to quarantine. This possibility will not be included in the thesis due to the effects it has on some theoretical derivations. This was noted specifically in section [6.1.1] where the inclusion of such a possibility made the derivation difficult.

Let $\pi_Q$ denote the probability to enter the quarantine while infectious. As with the time until recovery for the recovery process in the SIS model we let the time until an infectious individual is put in to quarantine follow a geometric distribution. This is as previously mentioned due to mathematical convenience and the nice properties of the geometric distribution. By implementing the availability of a quarantine we have in some regard redefined the infectious period $I$. Since an individual always return susceptible from the quarantine their infectious period now essentially ends once they enter the quarantine. Therefore $I$ takes the following distribution.

$$I \sim \text{Geo}(\pi_Q)$$

Since the individuals in quarantine are separated from the main population (infected and susceptible) it seems intuitive that the probability that a specific infectious individuals infects a specific susceptible should increase as $Q(t)$ increases. This is simply due to the fact that the number of individuals not in quarantine, $N-Q(t)$, is lower, meaning that there are fewer individuals to interact with. As a result the probability of interactions and consequently infections should increase. The inclusion of such a quarantine adjusted probability would certainly make the model more realistic. The actual application of the quarantine adjusted probability does however complicate the theoretical aspects of the paper. Due to this we will not be incorporating the effect in our models.

Define $Z(t)$ as the number of infected individuals that enters a quarantine during the $t$:th day. Assuming that entries into the quarantine happen independently of each other we can derive the distribution of $Z(t)$ conditioned on $I(t)$ in the same way as we derived $X(t)$.

$$Z(t)|I(t) \sim \text{Bin}(I(t), \pi_Q)$$

The time an individual spends in a quarantine follows another geometric distribution as noted below. This follows directly from the previous definition of the infectious period, $I$, mentioned in section [3.1].

$$Q \sim \text{Geo}(\pi_R)$$

As with the SIS model we can set up the system of equations of stochastic variables with determine the change between the groups. To be able to do this we do however need to define the number of individuals returning from quarantine during the $t$:th day.

Let $V(t)$ be the number of individuals returning from the quarantine during the $t$:th day. Due to the independence of the recoveries and the memorylessness of the recovery period whilst in quarantine, $V(t)$ takes the following distribution.

$$V(t)|Q(t) \sim \text{Bin}(Q(t), \pi_R)$$

The equations below are conditioned on $S(t), I(t)$ and $Q(t)$.

$$\begin{aligned} S(t+1) &= S(t) - Y(t) + V(t) \\ I(t+1) &= I(t) + Y(t) - Z(t) \\ Q(t+1) &= Q(t) - V(t) + Z(t) \end{aligned}$$

As in the previous case in section [3.1] we can examine the expected values of each stochastic event and as such get an understanding of how the disease would spread in expectation.

If we once again look at the individual Markov chain, $\{X_j(t) : t \geq 0\}$, of an arbitrary individual $j$ we get that the stochastic matrix looks as follows.

$$P_j(t) = \begin{pmatrix} (1-\pi_I)^{I(t)} & 1-(1-\pi_I)^{I(t)} & 0 \\ 0 & 1-\pi_Q & \pi_Q \\ \pi_R & 0 & 1-\pi_R \end{pmatrix}$$

Note that the state space of the Markov chain is now $\Omega = \{S, I, Q\}$.

### 3.2.1 Reproduction Number of an SIQS Model

The reproduction number is not the same in an SIQS model as in a general SIS model. When not accounting for a quarantine adjusted probability of infection the only difference from from the SIS model is that the infectious period changes. In the case of the SIS epidemic the probability to leave the infectious state was $\pi_R$, but as can be observed in the stochastic matrix $P_j(t)$ above the probability to leave is now $\pi_Q$. This leads to a difference in $R_0$, assuming that $\pi_Q \neq \pi_R$. Performing the same calculations as in the section [3.1] we are left with the following whilst still assuming one initial infected.

$$R_0 = (N-1)\left(1 - \frac{(1-\pi_I)\pi_Q}{\pi_I + \pi_Q - \pi_I\pi_Q}\right)$$

## 3.3 Multitype Epidemics

The last concept we wish to introduce before applying everything to our setting is the idea of a multitype epidemic. A multitype epidemic is characterized by dividing the total population into different types $i = 1, ..., k$. When doing that we have to redefine the aforementioned subgroups of the population, $S, I$ and $Q$. Define $S_i(t), I_i(t)$ and $Q_i(t)$ as the number of susceptible, infected and quarantined individuals of type $i$ at the start of the $t$:th day. The reasoning behind these different types is to refine the model further.

It is unlikely that every individual infects every other individual with the same probability. It is however reasonable to assume that an individual might belong to some group that they interact with more frequently. Due to the more frequent interactions there is also higher probability to infect individuals within an individuals own group compared to the rest of the population. This specific type of multitype epidemic model is referred to as a household model [2].

As the name suggests a household model is usually used to model the spread of disease with some kind of household structure. In theory however the household model is characterized by dividing the population into many small homogeneously mixing groups. The grouping of the population creates two levels of mixing, within and between the groups [3]. This means that an individual infects people within their group or household with some probability and the rest of the population with another.

If we now assume that these small groups represent some kind of predefined social structure we can get a more realistic understanding of the spread of an

epidemic. In our case this structure is the distribution of children into the different groups or classes at the kindergarten. It seems reasonable to assume that since a child spends more time with the other children within their own group than those in others. The probability to infect those within the same group should therefore be higher compared to the probability to infect someone in another group.

Define $\pi_L$ as the probability to infect a local individual (i.e someone within their own group) and $\pi_G$ to infect a global individual (i.e someone outside their own group). We also let $n_i$ and $m_i$ denote the number of susceptible and infected of each type at $t = 0$ such that $N_i = n_i + m_i$.

$$S(t) = \sum_{i=1}^{k} S_i(t) \quad I(t) = \sum_{i=1}^{k} I_i(t) \quad Q(t) = \sum_{i=1}^{k} Q_i(t)$$
$$S(0) = \sum_{i=1}^{k} n_i \quad I(0) = \sum_{i=1}^{k} m_i \quad N = \sum_{i=1}^{k} N_i$$

Define $Y_i(t)$ as the number of new infected individuals of type $i$ during the $t$:th day. The probability to get infected has now changed compared to the previous models and does not only depend on the number of total infected but rather on the number of infected within each type. For a susceptible individual of type $i$ we can thus derive the following probability.

$$\mathbb{P}(\text{Not infected during } t\text{:th day} | I_1(t), ..., I_k(t)) = (1-\pi_L)^{I_i(t)}(1-\pi_G)^{\sum_{j\neq i} I_j(t)}$$

From this, whilst still assuming independence of infections, we get that $Y_i(t)$ takes the following distribution.

$$Y_i(t)|S_i(t), I_1(t), ..., I_k(t) \sim \text{Bin}\left(S_i(t), 1-(1-\pi_L)^{I_i(t)}(1-\pi_G)^{\sum_{j\neq i} I_j(t)}\right)$$

We can derive $V_i(t)$ and $Z_i(t)$ in the same way as $V(t)$ and $Z(t)$. Then we get the following equations for conditioned on $S_1(t), ...S_k(t)$, $I_1(t), ..., I_k(t)$ and $Q_1(t), ...Q_k(t)$.

$$\begin{aligned}
S_1(t+1) &= S_1(t) - Y_1(t) + V_1(t) \\
I_1(t+1) &= I_1(t) + Y_1(t) - Z_1(t) \\
Q_1(t+1) &= Q_1(t) - V_1(t) + Z_1(t) \\
&\vdots \\
S_k(t+1) &= S_k(t) - Y_k(t) + V_k(t) \\
I_k(t+1) &= I_k(t) + Y_k(t) - Z_k(t) \\
Q_k(t+1) &= Q_k(t) - V_k(t) + Z_k(t)
\end{aligned}$$

This can also be expressed on vector form as can in some regard be seen in section [6.3.2] which makes it a bit nicer.

### 3.3.1 Reproduction Number of a Multitype SIQS Model

According to Britton & Andersson [2] the reproduction number of a multitype epidemic can be obtained from the dominant eigenvalue of a certain matrix $A_K = (\mu_{ij})$ $i, j = 1, ..., k$. Define $\mu_{ij}$ as the expected number of individuals of type $j$ an initially infectious individual of type $i$ infects during its infectious period. To be able to determine the eigenvalues we first have to get an understanding of $\mu_{ij}$.

Throughout the paper we let $m_i = 1$ and as such $n_i = N_i - 1$ for all $i = 1, ..., k$, meaning that the total number of initially infected are $k$. The number of individuals of type $j$ an initially infected individual of type $i$ infects during its infectious period follows some distribution $Z_{ij}$. Thus $\mu_{ij} = \mathbb{E}(Z_{ij})$. To determine the distribution of $Z_{ij}$ we need to calculate the probability that an individual gets infected during the initial infecteds infectious period.

Let $I$ be the infectious period of the initial susceptible of type $i$. As defined in section [3.2], $I$ takes the following distribution.

$$I \sim \text{Geo}(\pi_Q)$$

The probability that the initial infected individual of type $i$ does not infect an initial susceptible of type $i$ can be derived as follows.

$$\begin{aligned}
\mathbb{P}(\text{Not infected by initial infected}|I = 1) &= 1 - \pi_L \\
\mathbb{P}(\text{Not infected by initial infected}|I = 2) &= (1 - \pi_L)^2 \\
&\vdots \\
\mathbb{P}(\text{Not infected by initial infected}|I = n) &= (1 - \pi_L)^n
\end{aligned}$$

Through induction we can derive that $\mathbb{P}(\text{Not infected by initial infected}|I) = (1 - \pi_L)^I$ as can be seen in section [10.1.3] using a more general probability of infection. The probability that an individual gets infected is as such the complement to that probability. The distribution of $Z_{ii}$ is therefore the following.

$$Z_{ii}|I \sim \text{Bin}\left(n_i, 1 - (1 - \pi_L)^I\right)$$

From this we can derive $\mu_{ii}$ using the law of total expectation.

$$\begin{aligned}
\mu_{ii} = \mathbb{E}[\mathbb{E}|Z_{ii}|I]] &= \mathbb{E}\left[n_i \left(1 - (1 - \pi_L)\right)^I\right] \\
&= n_i \left(1 - \mathbb{E}\left[(1 - \pi_L)^I\right]\right) \\
&= n_i \left(1 - \sum_{i=1}^{\infty} (1 - \pi_L)^i \mathbb{P}(I = i)\right) \\
&= n_i \left(1 - \sum_{i=1}^{\infty} (1 - \pi_L)^i (1 - \pi_Q)^{i-1} \pi_Q\right)
\end{aligned}$$

$$\mu_{ii} = \mathbb{E}[\mathbb{E}[Z_{ii}|I]] = n_i \left(1 - \pi_Q(1-\pi_L)\sum_{i=1}^{\infty}((1-\pi_L)(1-\pi_Q))^{i-1}\right)$$

$$= n_i \left(1 - \frac{\pi_Q(1-\pi_L)}{\pi_L + \pi_Q - \pi_L\pi_Q} \underbrace{\sum_{i=1}^{\infty}(1-\pi_L)^I(1-\pi_Q)^{i-1}\pi_Q(\pi_L + \pi_Q - \pi_L\pi_Q)}_{1}\right)$$

$$= n_i \left(1 - \frac{\pi_Q(1-\pi_L)}{\pi_L + \pi_Q - \pi_L\pi_Q}\right)$$

Based on the previous calculations we can derive $\mu_{ij}$ for all possible $i$ and $j$ in an analogous manner. The only quantity that would change is the probability of infection, from $\pi_L$ to $\pi_G$. We can thus express $\mu_{ij}$ as follows.

$$\mu_{ij} = \begin{cases} n_i \left(1 - \frac{\pi_Q(1-\pi_L)}{\pi_L+\pi_Q-\pi_L\pi_Q}\right) & i = j \\ n_i \left(1 - \frac{\pi_Q(1-\pi_G)}{\pi_G+\pi_Q-\pi_G\pi_Q}\right) & \text{else} \end{cases}$$

Given $\mu_{ij}$ we can construct the matrix $A_K$ and thus derive the reproduction number. If we let $\lambda_1, ..., \lambda_k$ be the solutions to the following equation we get $R_0$ as well.

$$|A_K - I\lambda| = 0 \qquad R_0 = \max(\lambda_1, ..., \lambda_k)$$

It is very important to note that these theoretical results depend on having sufficiently large populations which might not always be the case when applying the model to a real life scenario [2]. In our multitype models this implies that sufficiently large groups are needed and not sufficiently many groups.

# 4 Setting

The scope of this paper is to examine how the different quarantine methods affect the total number of quarantine days in a kindergarten setting. As previously mentioned the number of sick days can be related to some kind of economic loss which we wish to minimize. In the case of our kindergarten setting the idea of a quarantine translates to the child staying home until they recover and then returning immediately. In theory this type of setting is well described by the SIQS model which we introduced in section [3.2].

There are however still some issues we need to deal with. Primarily kindergartens are not made of a homogeneously mixing population of children. Children could be assumed to have different social groups and the kindergartens could be assumed to be separating children into different groups or classes. To solve the latter issue it seems applicable to introduce a household effect to our SIQS model, which we introduced in section [3.3]. By doing this we include the

effect of children being divided into separate groups, partially limiting which individuals they contact and possibly infect.

To include the social structure of the children one would have to implement an underlying social graph essentially describing how the children interact, within, and between groups, which is not something we introduce in this paper due to the theoretical complexity of such a model. In theory we implement our model on a so called complete graph with weighted edges and differing weights between global and local vertices. This is by no means ideal but for now we consider to be sufficient for the questions we wish to answer.

Another issue with quarantines in a kindergarten setting is the fact that children might not return as susceptible as they do in the ideal scenario in which the theoretical SIQS epidemic takes place. It is possible that a child might return from their time at home whilst still being infectious, most likely due to the recklessness of their guardian.

Since we can not change or model for human behaviour we will attempt to implement other types of quarantines, which directly forces children to stay at home for a certain numbers of days before returning. We wish to observe how such a forced withdrawal affects the spread of epidemics, but also how it affects the total number of days in quarantine. The primary scope is to compare these two types of models, primarily through simulations, to get an understanding of how they perform against one another.

After having established the main framework which our models will operate in we have to decide for how long we wish to observe our models. This is mainly relevant to our simulations since varying the length of the observations will drastically impact our results. Due to wanting to observe the total number of sick days, or the proportion of sick days, we have to be able to account for the fact that an epidemic might reach the endemic state. As a result we could theoretically have an infinite number of sick days if let the number of days of observation go to infinity. This would however not be the case if the epidemic were to terminate early since the number of sick days would be finite if the number of days of observation would go to infinity.

The length of our simulations does therefore have to be chosen in a way such that we are able to notice the effect of an epidemic reaching the endemic stage. Moreover we also want to choose a number of reasonable scale to reduce the computations needed in our simulations. The length which we will henceforth use is one calender year, or 365 days, since the value is large enough to account for the effect of an endemic disease but also small enough to be manageable in simulation. If this chosen value was different the results would surely differ.

# 5 SIQS

The first model we will examine is a multitype SIQS model in an ideal scenario. In this case we assume that infectious children are sent home after some ran-

domly distributed time $I$. This time can be seen as the period when the child is infectious but shows no visible symptoms of the disease. Once a child is sent home they stay at home, in the quarantine, until the day when they recover and are then immediately sent back to kindergarten. This is the exact model which we described previously in section [3.3] and we can therefore use the theoretical results we derived.

This model is based on the idealistic assumption that children return from the quarantine only when no longer infectious and exactly after recovery. It is however of importance to observe how the epidemic spreads in such an ideal scenario to showcase that an epidemic might turn endemic even under conditions which should limit its spread. The SIQS model mainly serves as a reference model to compare our other models to.

## 5.1 Theory

To start of we have to assign numeric values to the probabilities we defined previously. The values in Table [1] below have no real substance behind them and are chosen simply showcase the methods we have derived.

**Table 1:** SIQS Values

| Notation | Definition | Value |
|---|---|---|
| $\pi_L$ | Probability to infect locally | 0.03 |
| $\pi_G$ | Probability to infect globally | 0.003 |
| $\pi_R$ | Probability to recover | 0.14 |
| $\pi_Q$ | Probability to be sent home | 0.4 |
| $k$ | Number of classes | 20 |
| $n_i$ | Number of initially susceptible of group $i$ | 19 |
| $m_i$ | Number of initially infected of group $i$ | 1 |

We let $n_i = 19$ and $m_i = 1$ for all $i$ meaning that we have $N_i = 20$ children in each class where 19 start of as susceptible and 1 as infected. If we now substitute in the values into the formula from section [3.3.1] we get the following matrix $A_K$.

$$A_K = (\mu_{ij}) \quad i, j = 1, ..., k$$

$$\mu_{ij} = \begin{cases} n_j \left(1 - \frac{\pi_Q(1-\pi_L)}{\pi_L + \pi_Q - \pi_L \pi_Q}\right) & i = j \\ n_j \left(1 - \frac{\pi_Q(1-\pi_G)}{\pi_G + \pi_Q - \pi_G \pi_Q}\right) & \text{else} \end{cases}$$

$$A_K = \begin{cases} 19\left(1 - \frac{0.4(1-0.03)}{0.03+0.4-0.012}\right) & i = j \\ 19\left(1 - \frac{0.4(1-0.003)}{0.003+0.4-0.0012}\right) & \text{else} \end{cases}$$
$$= \begin{cases} 1.363 & i = j \\ 0.142 & \text{else} \end{cases}$$

The matrix then looks as follows.

$$A_K = \begin{pmatrix} 1.363 & 0.142 & \dots & 0.142 \\ 0.142 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0.142 & \dots & \dots & 1.363 \end{pmatrix}$$

From this matrix we can calculate the reproduction number when we do not take the quarantine adjusted probability into account. This is done, as previously mentioned in section [3.3.1] by determining the dominant eigenvalue of the positive matrix. This gives us that $R_0 = 4.061$, which indicates that there is indeed a probability for a large outbreak. Since we know that there exists a probability for a large outbreak given our currently assigned probabilities we wish to simulate this epidemic to see how it actually develops.

## 5.2 Simulation

Using the probabilities from the table in the previous section we can simulate this multitype SIQS epidemic using **R**. After running the simulation we are left with the results in Figure [1] in the Appendix. In the figure we see that a large proportion of the population is in quarantine in the so called endemic state. The endemic state in the case of SIQS model is the state to which the model subgroups converge to as $t \to \infty$.

$$\big(S(t), I(t), Q(t)\big) \to \big(S^*(t), I^*(t), Q^*(t)\big) \quad \text{as } t \to \infty$$

We also note that the number of susceptible children present at the kindergarten is below 50 percent which is not sustainable in the long run. In the case of an infectious disease such as this one a kindergarten would have to implement additional measures to combat the epidemic, since simply removing infectious children until they recover does not seem to be giving the desired effect. This can be done by for example introducing mandatory vaccinations or closing down the kindergarten for some time and hoping that everyone recovers. The focus of this paper is however to implement and compare the method of forced withdrawal for a fixed number of days to see how the spread of the epidemic and the number of total quarantine days are affected.

# 6 Forced Withdrawal

In this section we discuss one method of how a kindergarten could go about solving the issues presented in the previous section. In this case the time spent in quarantine is not a random variable like before. Let $Q = q$ where $q \in \mathbb{N}_+$ indicating that the number of days an individual spends in a quarantine is fixed at $q$. In this case it could be argued that the theory from section [3.3] still holds for $R_0$ since the distribution of the time in quarantine is not accounted for in the calculations. This is however based on the assumption that the infectious period ends once an infectious individual enters the quarantine.

## 6.1 Theory

Compared to an SIQS model the forced withdrawal model is usually not a Markov process. This is due to the fact that the process does usually not satisfy the Markov property, since the probability to leave the quarantine at time $t$ depends on the state of process at time $t - q$. As such if $q \neq 1$ the Markov property is not satisfied. An example of this can be seen below.

$$\mathbb{P}\big(X(t) = S | X(t-1) = Q, X(t-q) = Q\big) = 1 - (1 - \pi_R)^q$$
$$\mathbb{P}\big(X(t) = S | X(t-1) = Q, X(t-q) \neq Q\big) = 0$$

If the individual Markov process is in state $Q$ at time $t - 1$ and was in state $Q$ at time $t - q$ we know that the process will leave $Q$ during the next day. The process then enters state $S$ with probability $1-(1-\pi_R)^q$ and $I$ with $(1-\pi_R)^q$. If the process is not in state $Q$ at time $t-q$ we know that process will not leave state $Q$ during the next day. As such the probability to enter state $S$ is 0. Since the probability to enter state $S$ at time $t$ does not only depend on the state at time $t - 1$, but also the state at time $t - q$, it is evident that the Markov property is not satisfied. This can be shown analogously with the probability to enter state $I$ whilst having been in quarantine.

### 6.1.1 Another Definition of $R_0$

In the SIQS model we defined the infectious period, $I$, as the time until and individual enters the quarantine. In this case, where we allow for the fact that individuals might return infected this is in theory no longer the case. Let $\xi_q$ denote the random variable that indicates the number of times an initial infectious individual is sent to quarantine. Since the time for recovery is memoryless as previously mentioned the probability to not recover during the $q$ days in quarantine is therefore $(1 - \pi_R)^q$. The probability to recover is then the complement of that probability. Since the recovery process and the infectious period is memoryless the number of times an individual is sent into quarantine, $\xi_q$, is also memoryless and can therefore be described by the following geometric distribution

$$\xi_q \sim \text{Geo}(1-(1-\pi_R)^q)$$

Let $I'$ denote the infectious period in this case. Let $I^{(1)}, ...$ be independent and identically distributed $I$-variables.

$$\begin{aligned} I'|\xi_q &= \sum_{l=1}^{\xi_q} I^{(l)} \\ \mathbb{E}[I'] &= \mathbb{E}[\xi_q]\mathbb{E}[I] \\ &= \left(\frac{(1-\pi_R)^q}{1-(1-\pi_R)^q}\right)\frac{1}{\pi_Q} \end{aligned}$$

The reason why we can define $I^{(1)}, ...$ as independent is due to the memorylessness of $I$. Let $Z'_{ii}|I'$ denote the number of individuals of type $i$ and initial infectious of type $i$ infects during its new infectious period.

$$Z'_{ii}|I' \sim \text{Bin}\left(n_i, 1-(1-\pi_L)^{I'}\right)$$

Since we want to calculate the expected value of infections, $\mu'_{ii}$, we can do this by using the definition of the probability generating function. We know that the probability generating function is defined as $\Pi_W(z) = \mathbb{E}[z^W]$ where $W$ is some non-negative discrete random variable and $z \in (0, 1]$. By using the law of total expectation once again we are left with the following expression for $\mu'_{ii}$.

$$\begin{aligned} \mu'_{ii} = \mathbb{E}[\mathbb{E}|Z'_{ii}|I']] &= \mathbb{E}\left[n_i\left(1-(1-\pi_L)\right)^{I'}\right] \\ &= n_i\left(1-\underbrace{\mathbb{E}\left[(1-\pi_L)^{I'}\right]}_{\Pi_{I'}(1-\pi_L)}\right) \end{aligned}$$

Since $\pi_L$ is a probability we know that $1-\pi_L \in (0, 1]$ as long as $\pi_L \neq 1$. To continue we have to derive the probability generating function of $I'$ to complete our calculations. From the theorem in section [10.1.1] we know that the probability generating function of a stochastic sum of stochastic variables can be expressed as follows where we let the sum be defined as $S_N = \sum_{i=1}^{N} X_i$.

$$\Pi_{S_N}(z) = \Pi_N(\Pi_{X_1}(z))$$

For the theorem to hold we assume that $X_1, ...X_N$ are independent and identically distributed non-negative discrete random variables and that they are all independent of the non-negative discrete random variable $N$.

As such $\Pi_{I'}(z) = \Pi_{\xi_q}(\Pi_I(z))$. Since both $I$ and $\xi_q$ follow different geometric distributions on the support $\{1, 2, 3, ...\}$ we ought to define the probability generating function of such a distribution. If $G \sim \text{Geo}(p)$ then $\Pi_G(z) = \frac{zp}{1-z(1-p)}$.

Using this we can calculate $\Pi_{I'}(z)$ as follows.

$$\begin{aligned}
\Pi_{I'}(z) &= \Pi_{\xi_q}(\Pi_I(z)) \\
&= \frac{\frac{z\pi_Q}{1-z(1-\pi_Q)}(1-(1-\pi_R)^q)}{1-\frac{z\pi_Q}{1-z(1-\pi_Q)}(1-(1-(1-\pi_R)^q))} \\
&= \frac{\frac{z(1-(1-\pi_R)^q)\pi_Q}{1-z(1-\pi_Q)}}{\frac{1-z(1-\pi_Q)-z\pi_Q(1-(1-(1-\pi_R)^q))}{1-z(1-\pi_Q)}} \\
&= \frac{z(1-(1-\pi_R)^q)\pi_Q}{1-z(1-\pi_Q)-z\pi_Q(1-(1-(1-\pi_R)^q))} \\
&= \frac{z(1-(1-\pi_R)^q)\pi_Q}{1-z(1-(1-(1-\pi_R)^q)\pi_Q}
\end{aligned}$$

Since $z = 1 - \pi_L$ in our case we are left with the following expression of $\mu'_{ii}$.

$$\mu'_{ii} = n_i \left(1 - \frac{(1-\pi_L)(1-(1-\pi_R)^q)\pi_Q}{1-(1-\pi_L)(1-(1-(1-\pi_R)^q)\pi_Q}\right)$$

We can derive $\mu'_{ij}$ analogously for $i \neq j$ such that the following holds.

$$\mu'_{ij} = \begin{cases} n_j \left(1 - \frac{(1-\pi_L)(1-(1-\pi_R)^q)\pi_Q}{1-(1-\pi_L)(1-(1-(1-\pi_R)^q)\pi_Q}\right) & i = j \\ n_j \left(1 - \frac{(1-\pi_G)(1-(1-\pi_R)^q)\pi_Q}{1-(1-\pi_G)(1-(1-(1-\pi_R)^q)\pi_Q}\right) & \text{else} \end{cases}$$

From this we can construct the matrix $A'_K$ and derive the alternative reproduction number, $R'_0$ from its dominant eigenvalue. Using the values in Table [2] below we derive $R'_0$ as follows. It should be noted that as with the previous introduction of actual probabilities for infection, recovery and others the chosen value of $q = 10$ does not have any reasoning behind it and is simply chosen to showcase the alternate definition.

**Table 2:** Forced Withdrawal Probabilities

| Notation | Definition | Value |
|---|---|---|
| $\pi_L$ | Probability to infect locally | 0.03 |
| $\pi_G$ | Probability to infect globally | 0.003 |
| $\pi_R$ | Probability to recover | 0.14 |
| $\pi_Q$ | Probability to be sent home | 0.4 |
| $q$ | Time in quarantine | 10 |

$$\mu'_{ij} = \begin{cases} n_j \left(1 - \frac{(1-\pi_L)\pi_Q(1-(1-\pi_R)^q)}{1-(1-\pi_L)(1-\pi_Q(1-(1-\pi_R)^q))}\right) & i = j \\ n_j \left(1 - \frac{(1-\pi_G)\pi_Q(1-(1-\pi_R)^q)}{1-(1-\pi_G)(1-\pi_Q(1-(1-\pi_R)^q))}\right) & \text{else} \end{cases}$$

$$= \begin{cases} 1.716 & i = j \\ 0.182 & \text{else} \end{cases}$$

$$A'_K = \begin{pmatrix} 1.716 & 0.182 & \dots & 0.182 \\ 0.182 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0.182 & \dots & \dots & 1.716 \end{pmatrix}$$

After computing the eigenvalues we get that $R'_0 = 5.174$ which is a lot larger compared to the other definition which left us with $R_0 = 4.061$. This is however to be expected due to increasing the average length of the infectious period in the alternative definition.

## 6.2 Simulation

In the simulation we use the values from Table [2] in the previous section. Performing this simulation in the same way as the previous one over one year in **R** we are left with some very interesting results.

We see in Figure [2] in the Appendix that the number of children in quarantine in the endemic state is actually lower than in the case of the SIQS model in Figure [1]. This would be a vast improvement if it was not for the fact that the number of infectious children still attending kindergarten is a lot higher, which of course is not to be desired. We also see a higher spike in the number of initially quarantined compared to the SIQS model. This is due to the fact that individuals are forced to stay in the quarantine for 10 days whilst they can return in less than that in the SIQS model. As such there is a 10 day period in the initial stage of the epidemic where individuals will enter the quarantine whilst none leave. Since individuals leave the quarantine daily in the SIQS model the maximum number of initially quarantined individuals will be lower since the expected time in quarantine is $\frac{1}{\pi_R} \approx 7.15 < 10$.

When comparing Figure [1] and Figure [2] we have made improvements concerning the economical aspect of an epidemic by reducing the number of quarantined individuals in the endemic state. By doing this we have however created an undesirable situation where a majority of the children are sick whilst still attending kindergarten. It is clear that this approach has not had the desired effects. Due to this we wish to find a more optimal value of $q$ which leads to a minimization of the number of total sick days and hopefully a termination of the epidemic.

### 6.2.1 Letting $q$ vary

For the previous case where we let $q = 10$ we did not achieve the desired effects. We now wish to simulate the epidemic presented in Table [2] whilst letting $q$ vary. This is done in a similar fashion to the previous simulations. In Figure [3] in the Appendix we can observe the optimal value of $q$ for two different methods of taking the average, the mean and the median.

The reason why we choose to observe both methods is due to the sometimes large differences between them. In the case of taking the mean a single larger outbreak might change the results drastically whilst the median in some respect ignores such extremities. If the probability of such an extremity were quite large it does however seem unreasonable to ignore them completely and as such we could argue for the usage of the mean instead. Since the results vary a lot between a large and a small outbreak it does however seems unintuitive to use the mean since the results will be somewhere in the middle of both extremities. As for the median we simply observe the 50th percentile which tells us very little. Due to this it seems reasonable to observe a higher quantile such that the $q$ which minimizes the proportion of sick days does so with in some percentage of the cases.

**Table 3:** Optimal Values of $q$ for Different Methods

| Method | Median | 0.75 | 0.9 | 0.95 | Mean |
|---|---|---|---|---|---|
| $q$ | 40 | 44 | 48 | 48 | 48 |

The lower values for the quantile methods are caused by the exclusion of the worst outbreaks which are present in the usage of the mean. Using the quantile method and a large sample of simulations one could choose $q$ such that it satisfies some probability of minimization. Going forward we mainly utilize the results from the mean since we wish to account for the effects of larger outbreak, no matter their probability of occurrence, whilst keeping the other methods in mind.

In the case of the infectious disease which we have discussed throughout the paper this far we notice that the $q$ which minimizes the proportion of sick days in mean during a year is 48 [3]. This is an extremely large value which in practice means that once a child is sent home they have to stay at home for up to one and a half months. Even though it could be argued that minimizing the total number of sick days is for the greater good of the society as a whole, it seems difficult to argue that children should be forced to stay at home for up to such a long time. This becomes even more problematic when the children will spend a majority of the time not being infectious. This would of course also imply that the guardian or guardians of that child would miss almost a combined one and a half months of work which would hardly be appreciated by them or their employers.

However, the scope of this thesis is not to examine how the individuals would react to the proposed solutions and as such the optimal number of days a child

should stay at home is 48. The model also operates under unrealistic assumptions which makes the results inapplicable to a real world scenario.

This large number of days does not seem very reasonable at first. It can however be explained by the logic of the reproduction number. If we assume that $q$ is somewhat large and that the epidemic is infectious enough then a vast majority of initially infected and the susceptible they infect will be in quarantine at roughly the same time. This can be seen in Figure [2] in the Appendix. This is due to children being infected early and being sent into quarantine for a long time. As such children will enter the quarantine whilst none return for the first $q + 1$ days. Thus if the disease is infectious enough most children will have gotten infected and been sent to the quarantine before the first individual returns. For extremely infectious disease the expected number of individuals sent to quarantine before time $q + 1$ should be close to $N$.

In this scenario there is now a case of infectious individuals recovering whilst at home, if however at least one of them were to return and still be infected we are left with a similar scenario as to that in the calculation of $R_0$. The scenario being a lot of susceptible individuals and some random number of infected which will most likely result in the epidemic starting over or reigniting as we will henceforth call such a restart.

The minimization of the proportion of sick days should therefore be achieved by getting the disease to terminate as early as possible, rather than letting it reignite one or multiple times. Even though this might seem intuitive it should be taken into account that for the termination to occur for very infectious disease $q$ has to be very large. This will lead to a large influx of quarantined in the early stages of the disease but none in the later stages due to the epidemic having terminated.

### 6.3 Early Termination and Minimization of Sick Days

The reason why the optimal value of $q$ was so large in the previous section can then be described by observing the number of initially quarantined. Define $\eta$ as the maximum number of initially quarantined individuals. As can be seen in Figure [4] in the Appendix all individuals are put into quarantine at approximately the same time meaning that $\eta \approx N$ in this case. If almost the entire population enters the quarantine for $q$ days they are to return at approximately the same time. If as previously mentioned one quarantined individual returns and is still infectious the epidemic will very likely reignite due to the large number of recently recovered and returned susceptible. Since all $\eta$ recoveries happen independently have the following probability.

$$\mathbb{P}(\text{All } \eta \text{ individuals recover}) = (1 - (1 - \pi_R)^q)^\eta \tag{1}$$

We can see that for the probability of full recovery to be high, either $q$ has to be large or $\eta$ must be small. In the current case where $\eta \approx N$ we have no other option rather than increasing $q$ which is what we were trying to show.

Furthermore we define the distribution of the number of returning infectious of the initial $\eta$ quarantined as follows.

$$V_\eta|\eta \sim \text{Bin}\left(\eta, (1-\pi_R)^q\right)$$

### 6.3.1 Simulated Values of $\eta$

Since $\eta$ depends on outcomes of a large number of different conditional stochastic events it is difficult to derive an explicit expression for it. We can however derive an estimate for it by observing the results from a large number of simulations. In Figure [1(a)] we observe the estimate of $\eta$, $\hat{\eta}$ for a variety of diseases with different $R_0$ and their respective $q_{\min}$. As one would assume the estimated number of initially quarantined individuals grows as the infectiousness of the disease increases.

As we saw in Figure [4] in the Appendix the number of initially quarantined is approximately the entire population for the more infectious disease. As such we can only minimize equation (1) by increasing $q$. Moreover we can now roughly calculate the probability of an early termination for the disease which we have been observing throughout the paper.

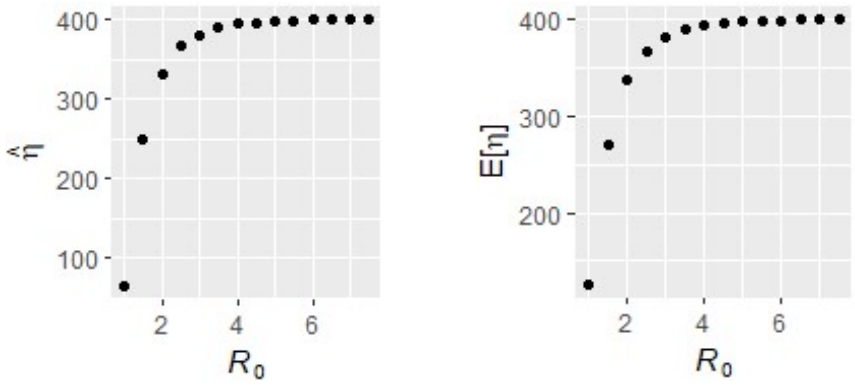
Previously we determined that $q_{\min} = 48$ and recently that $\hat{\eta} \approx 395$ for $R_0 \approx 4$. By substituting these values into equation (1) we get the following probability of an early termination.

$$\begin{aligned}
\mathbb{P}(\text{All } \eta \text{ individuals recover}) &= \prod_{j=1}^{395} 1 - (1-\pi_R)^{48} \\
&= (1-(1-\pi_R)^{48})^{395} \\
&= (1-(1-0.14)^{48})^{395} \\
&= 0.75
\end{aligned}$$

We should also take into account that this can be seen as lower bound since even if $V_\eta|\eta = \hat{\eta} > 0$ there is always some probability that the returning infectious individuals fail to spread the disease further which leads to an early termination anyway. That probability should however tend to zero for larger values of $V_\eta|\eta$ or very infectious disease.

### 6.3.2 Expected Value of $\eta$

We can calculate the expected value of $\eta$ by setting up a system of equations as we did for the SIS and SIQS model. Since we included a multitype structure we do however need to introduce some new notations since it becomes quite tedious




(a) Simulated Values of $\eta$ (b) Expected Values of $\eta$

**Figure 1:** Comparison of $\eta$

to write up the equations for each subgroup.

$$\mathbf{S}(t) = \begin{pmatrix} S_1(t) \\ \vdots \\ S_k(t) \end{pmatrix} \quad \mathbf{I}(t) = \begin{pmatrix} I_1(t) \\ \vdots \\ I_k(t) \end{pmatrix} \quad \mathbf{Q}(t) = \begin{pmatrix} Q_1(t) \\ \vdots \\ Q_k(t) \end{pmatrix}$$

Furthermore we let $\mathbf{N} = \begin{pmatrix} n_1 + m_1 \\ \vdots \\ n_k + m_k \end{pmatrix}$ such that $\mathbf{N} = \mathbf{S}(t) + \mathbf{I}(t) + \mathbf{Q}(t)$.

We also introduce the vectors $\mathbf{Y}(t)$ which holds all information regarding the number of new infectious of each type, $\mathbf{Z}(t)$ denotes the number of individuals sent to quarantine of each type.

$$\mathbf{Y}(t) = \begin{pmatrix} Y_1(t) \\ \vdots \\ Y_k(t) \end{pmatrix} \quad \mathbf{Z}(t) = \begin{pmatrix} Z_1(t) \\ \vdots \\ Z_k(t) \end{pmatrix}$$

Since individuals leave the quarantine after $q$ days we know that the number of individuals returning from the quarantine at time $t$, $\mathbf{V}(t)$, is equal to the number of individuals sent to quarantine, $\mathbf{Z}(t)$, at time time $t - q$. Since we account for the fact that individuals can return infectious we do however have to split up the the returning individuals into two groups. Let $\mathbf{V}^{(1)}(t)$ denote the vector of returning susceptible individuals of each type and $\mathbf{V}^{(2)}(t)$ the vector of

returning infectious. Under the assumption that $Q = q$ we have the following.

$$\mathbf{V}(t) = \begin{pmatrix} V_1(t) \\ \vdots \\ V_k(t) \end{pmatrix}$$

$$\mathbf{V}^{(1)}(t) = \begin{pmatrix} V_1^{(1)}(t) \\ \vdots \\ V_k^{(1)}(t) \end{pmatrix} \quad \mathbf{V}^{(2)}(t) = \begin{pmatrix} V_1^{(2)}(t) \\ \vdots \\ V_k^{(2)}(t) \end{pmatrix}$$

We are yet to formally define $V_i(t)$, $V_i^{(1)}(t)$ and $V_i^{(2)}(t)$ which we do as follows.

$$V_i(t)|Z_i(t-q) = \begin{cases} Z_i(t-q) & t > q \\ 0 & \text{else} \end{cases}$$

$$V_i^{(1)}(t)|V_i(t) \sim \text{Bin}\big(V_i(t), 1-(1-\pi_R)^q\big)$$

$$V_i^{(2)}(t)|V_i^{(1)}(t), V_i(t) = V_i(t) - V_i^{(1)}(t)$$

The final expression is based on the fact that the number of returning individuals is always the sum of the returning number of infectious and susceptible individuals. Using the vectors which we have defined we can establish the following system of equations in which we can observe the mean of the number of individuals of each type and each subgroup at time $t$. The equations are conditioned on $\mathbf{S}(t), \mathbf{I}(t)$ and $\mathbf{Q}(t)$.

$$\begin{aligned} \mathbf{S}(t+1) &= \mathbf{S}(t) - \mathbb{E}[\mathbf{Y}(t)] + \mathbb{E}[\mathbf{V}^{(1)}(t)] \\ \mathbf{I}(t+1) &= \mathbf{I}(t) + \mathbb{E}[\mathbf{Y}(t)] - \mathbb{E}[\mathbf{Z}(t)] + \mathbb{E}[\mathbf{V}^{(2)}(t)] \\ \mathbf{Q}(t+1) &= \mathbf{Q}(t) + \mathbb{E}[\mathbf{Z}(t)] - \underbrace{\mathbb{E}[\mathbf{V}^{(1)}(t)] - \mathbb{E}[\mathbf{V}^{(2)}(t)]}_{\mathbb{E}[\mathbf{V}(t)] = \mathbb{E}[\mathbf{Z}(t-q)]} \end{aligned}$$

This system can be solved with some effort which therefore makes us able to derive $\mathbb{E}[\eta]$. It should be noted that in this system we no longer see individuals as discrete units, meaning that we might have fractions of new infections during a day. Since $Q(t)$ increase monotonously for $t < q$ the time at which $Q(t)$ reaches its peak should be at or in relative proximity to $q$. The function is monotonously increasing since we have no individuals leaving $Q(t)$ when $t < q$ whilst fractions of individuals continue to enter the state. We define $\mathbb{E}[\eta]$ as follows.

$$\mathbb{E}[\eta] = \max \mathbf{1}^T \mathbf{Q}(t)$$

Performing this calculation manually is extremely tedious and thus we resort to performing it in **R**. After performing the calculations we are left with the results in Figure [1(b)] to the right. By comparing Figure [1(a)] and Figure

[1(b)] we can see that the $\eta$ we obtained from the mean of the simulations do not differ noticeably from the expect for $R_0 = 1$. This is due to the outcome of a limited number of simulations which in this case seems to differ from the expected value. The large deviation in the expected $\eta$ is also more noticeable for lower values of $R_0$ since the probability of a small outbreak is a lot larger than for higher values of $R_0$.

After having derived the expected value of $\eta$ we can continue by examining the probability that the epidemic reignites. The distribution of returning infected individuals, $V_\eta$, which conditioned on having $\mathbb{E}[\eta]$ initially quarantined, can be expressed as follows.

$$V_\eta \sim \text{Bin}(\mathbb{E}[\eta], (1-\pi_R)^q) \tag{2}$$

$$p_{V_\eta}(k) = \binom{\mathbb{E}[\eta]}{k}((1-\pi_R)^q)^k(1-(1-\pi_R)^q)^{(\mathbb{E}[\eta]-k)} \tag{3}$$

From equation (3) above we can derive a lower bound for the probability that the epidemic terminates after the initial stage for a certain value of $q$. This probability will however be conditioned on $\eta$. If we assume $\eta = \mathbb{E}[\eta]$ then we can derive such a lower bound for the probability of an early termination. Furthermore we can also assume the worst case scenario, that $\eta = N$ and from this derive a value of $q$ which satisfy the probability an early termination even under the worst circumstances.

To be able get a better understanding of the probability of an early termination beyond a lower bound created under certain assumptions, we would have to derive the probability that the epidemic terminates conditioned on $V_\eta > 0$, which in turn has to be conditioned on $\mathbb{E}[\eta]$ to be calculated. Under the assumption that every individual had the same distribution for the numbers of individuals they would infect during their infectious period this could be done using a branching processes. This is however not the case for our models since we have a finite population where the number of infections during an individuals infectious period depend on the number of susceptible individuals which vary throughout the epidemic.

If we were to make some changes and further assumptions to account for this we would be presented with the problem that the returning $V_\eta$ infectious individuals would return to different groups at different times which present further issues. Due to this the true probability of termination after the initial stage of the epidemic under the assumption that $\eta = \mathbb{E}[\eta]$ as well as others is very difficult to derive.

A way to refine the lower bound could be to simply use brute force the probability that the process would terminate even though some individuals returns whilst still infectious. For larger values of $V_\eta$ it seems intuitive that the probability of termination would tend to zero, if the disease is infectious enough and as such we could most likely derive a second lower bound which should be closer to the true value than the first. To do this we would still have to make assumptions

regarding the time of return and the distribution of returning infected between the groups. Due to the number of assumptions this would require we choose to leave the lower bound of the probability as it is.

## 6.4 Varying $q$ for Different Diseases

In the previous section we focused on the case of a very infectious disease where $R_0 = 4.061$. In this section we would like to simulate how $q$ would vary if we varied the infectiousness of the epidemic. To do this we would tweak to global, and local probabilities of infection so that $R_0$ changes and then observe which $q$ produces the lowest total proportion of sick days for the different diseases.

### 6.4.1 A Less Infectious Disease

In this case we let $\pi_L = 0.0073$ and $\pi_G = 0.00073$. From this we can calculate $R_0$ just as we did in section [5.1]. The dominant eigenvalue of the matrix $A_K = (\mu_{ij})$ is 1.01 which is the $R_0$ of this less infectious disease.

Since the probability of a large outbreak is non-zero due to $R_0 > 1$ we observe the development of the epidemic through simulation. However, since $R_0$ is very close to one we can assume that the optimal value of $q$ for a disease such as this should be significantly smaller compared to the previous model. Since $R_0 = 1.01$ an initial infectious infects an expected total of 1.01 other individuals under ideal circumstances during its infectious period when there are $n_i = 19$ susceptible individuals of each group.

Since we do not adjust our probabilities of infection based on the individuals currently in quarantine the occurrence of a reignition of the epidemic is most likely zero. This is due to the fact that any returning infectious do not have a large enough pool of susceptible individuals since some will be in quarantine and thus the expected number of individuals returning infectious would infect is less than one, meaning that the disease would terminate. If we however were to account for the change in probability brought by putting children in quarantine the results would surely differ.

Performing the same simulations in the same ways as for the previous disease we end up with the optimal values of $q$ presented in Table [4] below.

**Table 4:** Optimal Values of $q$ for Different Methods

| Method | Median | 0.75 | 0.9 | 0.95 | Mean |
|---|---|---|---|---|---|
| $q$ | 23 | 27 | 29 | 38 | 29 |

As we expected the optimal values of $q$ are significantly smaller for this disease than its previous, more infectious, counterpart. In Figure [5] in the Appendix we can observe the spread of the epidemic and it is very clear that the disease is not infectious enough to spread beyond the initial stage. Once the number of quarantined individuals grows there is not enough susceptible individuals to

infect and as a result the number of infectious individuals drops to zero and the epidemic terminates.

### 6.4.2 A More Infectious Disease

In this case we will examine an even more infectious disease compared to the original one. We now let $\pi_L = 0.045$ and $\pi_G = 0.0045$ and perform the same calculations as before.

This results in the dominant eigenvalue of $A_K = (\mu_{ij})$, and the $R_0$ for this disease being 6.03. Due the extremely infectious nature of the disease even a small number of available susceptible individuals would bring the expected number of infections during an infectious period to a value above one. Thus the disease will rarely terminate due to a lack of available susceptible individuals like in the previous, less infectious, case. It is reasonable to assume that the optimal value of $q$ might be very similar to the value which we obtained in the original epidemic from Table [1].

As can be seen in Figure [1(a)], $\hat{\eta} \approx N$ as in the case of the original disease. Thus the epidemic would likely reignite if any of the initially quarantined individuals were to return and still be infectious. This is the same argument as we made in the original model. Since the value of $\pi_R$ has not changed the optimal $q$ should be very similar since we can not allow for any individual to return or the epidemic will likely reignite. The only difference being that the probability of a termination under the assumption that some initially quarantined return infectious is lower for this disease compared to the original one. As such the optimal value of $q$ should be higher.

Performing the same simulations as last time we end up with the results presented in Figure [6] in the Appendix. As we can see the development looks extremely similar to that of the original epidemic. The optimal values of $q$ can be chosen according to Table [5] below.

**Table 5:** Optimal Values of $q$ for Different Methods

| Method | Median | 0.75 | 0.9 | 0.95 | Mean |
|---|---|---|---|---|---|
| $q$ | 43 | 45 | 52 | 53 | 57 |

### 6.4.3 For Other Values of $R_0$

Lastly we can observe how $q$ behaves when we let $R_0$ vary between a vast number of values. To begin with we need to make some assumptions of how our disease spreads. As has been the case in the three previous diseases $\pi_L = 10\pi_G$. If we assume that the probability of infection solely depends on how much individuals interact it seems plausible to also assume that a child would interact a lot more with children in their own group compared to those in others meaning that there should some constant factor relating $\pi_L$ to $\pi_G$. If we are finally to assume

that this factor is equal to 10 meaning that an child spends 10 times as much time with a arbitrary child in their own group compared to an arbitrary child in another group. Limiting the diseases in this way allows us to easily calculate $R_0$ for all possible probabilities of infection.

$$\mu_{ij} = \begin{cases} 19\left(1 - \frac{0.4(1-10\pi_G)}{10\pi_G+0.4-0.4\cdot 10\pi_G}\right) & i = j \\ 19\left(1 - \frac{0.4(1-\pi_G)}{\pi_G+0.4-0.4\pi_G}\right) & \text{else} \end{cases}$$
$$= \begin{cases} 19\frac{10\pi_G}{6\pi_G+0.4} & i = j \\ 19\frac{\pi_G}{0.6\pi_G+0.4} & \text{else} \end{cases}$$

This would give us the following matrix.

$$A_K = 19\begin{pmatrix} \frac{10\pi_G}{6\pi_G+0.4} & \frac{\pi_G}{0.6\pi_G+0.4} & \cdots & \frac{\pi_G}{0.6\pi_G+0.4} \\ \frac{\pi_G}{6\pi_G+0.4} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\pi_G}{6\pi_G+0.4} & \cdots & \cdots & \frac{10\pi_G}{6\pi_G+0.4} \end{pmatrix}$$

The dominant eigenvalue of $A_K$ can after some tedious calculations and a lot of matrix algebra be expressed on the following form. The matrix calculations can be seen in [10.1.3].

$$R_0 = \frac{10\pi_G}{6\pi_G + 0.4} + 19\frac{\pi_G}{0.6\pi_G + 0.4} \tag{4}$$

Solving (4) for $\pi_G$ we are left with one positive and one negative solution. Since $\pi_G$ is a probability the solution which we are looking for is the positive.

$$\pi_G = \frac{-2755 + 33R_0 + \sqrt{7590025 + 46170R_0 + 729R_0^2}}{57000 - 90R_0} \tag{5}$$

By using (5) we recalculate $\pi_G$ for different values of $R_0$ and therefore simulate for a variety of different hypothetical epidemics with varying degrees of infectiousness.

We could perform an enormous number of simulations for all different values of $R_0$ by increasing by very small increments. This would be extremely time consuming, but would give us a graph indicating the optimal value of $q$ given an epidemic with a certain $R_0$.

Due to the time constraints we can instead choose to observe how $q$ changes for larger increments giving us Figure [2]. In the figure we see very clearly that $q_{\min}$ is dependant on the value of $R_0$. When we reach the very infectious diseases this increase seems to slow down. This is likely due to the fact that $\eta \approx N$ for

all disease over a certain infectious threshold as can be seen in Figure [1(a)]. Furthermore the more infectious diseases are likely to reignite if one individual returns from the initial quarantine without recovering.

To be able to minimize the total number of sick days we must therefore allow for no infectious individuals to return. Due to this the main value that would affect $q$ in these cases would be the probability of recovery since equation (1) can only be maximized by either increasing $q$ or decreasing $\pi_R$ for diseases with $\eta \approx N$. If we for example were to increase the average recovery time the values of $q_{\min}$ would increase, and vice versa.

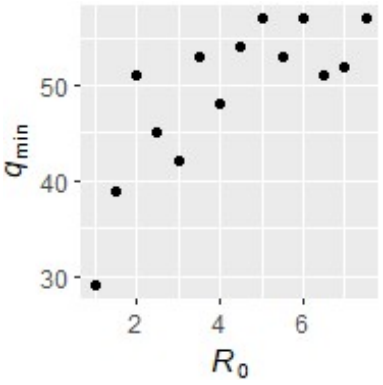



**Figure 2:** Different $q_{\min}$

The reason to why the values of $q_{\min}$ do not increase monotonously with $R_0$ in Figure [2] is due to the number of simulations being somewhat low. Therefore some stochastic outcomes could shift the result significantly leading to larger values of $q_{\min}$ for less infectious disease compared to others. In theory the increase should however be seen as monotonous.

# 7 Renewal Theory

It is possible to further refine the model by implementing the possibility of a restart of the epidemic when there are no infectious individuals present. Since we are dealing with SIS diseases it is not unlikely that such a disease would reinfect the kindergarten some time after the initial infection. A child present at the kindergarten could possibly be infected in a variety of different scenarios or places outside of the kindergarten, and as such restart the spread in the kindergarten.

To include this we can add the possibility of a general probability of an exterior infection, meaning that there is always a probability of infection even though there are no infectious individuals present at the kindergarten. By doing this the epidemic would terminate and then start over once a susceptible individual suffers an exterior infection. This could then be used in combination with renewal theory to examine the average proportion of children in quarantine each day during a longer period of time. It could also be of interest to observe the long run number of infections per day in the model.

## 7.1 Renewal Reward Process

Consider that we have an external probability of infection $\pi_E$. If we then assume that the epidemic is truly terminated once the entire population have returned

to being susceptible we can determine the distribution of the idle period. Due to the independence of infections, which we also assume for the external infections, we know that the probability that all $N$ individuals avoid infection during a day is $(1-\pi_E)^N$ which can be proved through induction in a similar way to [10.1.2]. The length of the idle period which we denote $D$ then takes the following distribution.

$$D \sim \text{Geo}\left(1-(1-\pi_E)^N\right)$$

The length of a busy period, which we will denote by $B$, is a bit tougher to calculate but it can be estimated using simulations. The cycle length, which we denote as $T$, is thus $D+B$. Using the same simulations we can also estimate the number of days in quarantine during each busy period, and as a result the total number of quarantine days during each cycle. Note that a complete cycle is the idle period combined with the busy period. By using a renewal reward process we can then calculate the long run proportion of individuals in the quarantine per day.

In the mentioned renewal reward process which we henceforth denote $\{R(t), t \geq 0\}$ the total reward during a cycle, which we will denote $R$, is the total number of days in quarantine during the cycle. Note however that all quarantine days occur during the busy period. We let $\{N(t), t \geq 0\}$ denote the underlying renewal process where a renewal occurs at time $t$ when $S(t) = N$ and $S(t-1) \neq N$. Using these notation we have the following where $R_i \sim R$.

$$R(t) = \sum_{i=1}^{N(t)} R_i$$

The renewal reward process "hits stationarity" after the first renewal since we have $k$ starting infectious individuals at the beginning but only 1 after each subsequent renewal. Since the starting conditions of each cycle vary between the first and the subsequent cycles we only wish to observe the process from the first renewal until the last renewal. The reason why we stop at the last renewal is because the potentially last busy period will end early which means that its length and number of quarantine days will be inaccurate.

We wish to examine the long run number and proportion of individuals in quarantine during a single day. As such we use Theorem [10.1.4] from the Appendix.

We henceforth denote $\mathbb{E}[R]$ by $\delta$. Since we know that $\frac{R(t)}{t}$ converges almost surely to $\frac{\delta}{\mu}$ we can estimate both quantities using simulation to get an approximate result.

We know that $\mu = \mathbb{E}[T] = \mathbb{E}[B] + \mathbb{E}[D] = \mathbb{E}[B] + \frac{1}{1-(1-\pi_E)^N}$. The remaining expected value, $\mathbb{E}[B]$, is difficult if not impossible to derive analytically and as such it has to be approximated through simulation.

To showcase this method we choose to examine an epidemic with $R_0 = 1.5$. From this we can use equation (5) and get that $\pi_G \approx 0.0011, \pi_L = 0.011$. We also have to choose a suitable value for $\pi_E$ which we set to 0.000011. We use the previous values for $k, m_i, n_i$ which leads to $N = 400$. Using these values we get that the average length of an idle period is $\frac{1}{1-(1-0.00001)^{400}} \approx 228$. A reason that we choose $\pi_E$ to be so small is because we do not want it to have too large of an impact on the spread during the busy period.

When introducing the possibility of an external infection the previously derived theory regarding $R_0$ has to be re-examined. In the case of a multitype epidemic there now exists a new group which infects the other groups but can not be infected itself. By introducing such a group the theory centred around that $R_0$ can be expressed as the dominant eigenvalue of $A_K$ fails due to the groups definition.

It should also be noted that we can derive the long run number of infections per busy period by simply dividing $\frac{R(t)}{t}$ by $q$. We denote these number of infections by $\nu = \frac{\delta}{q}$. Since $q$ is constant the almost sure convergence presented above still holds. Therefore the number of infections per day over an entire cycle converges almost surely to $\frac{\delta}{\mu q} = \frac{\nu}{\mu}$.

## 7.2 Simulation

Through some simulation where each value of $q$ is simulated over 100000 days we get the following estimates.

**Table 6:** Outcomes of Renewal Reward Process Simulations

| $q$ | $\hat{\mu}$ | $\hat{D}$ | $\hat{B}$ | $\hat{\delta}$ | $\hat{\nu}$ |
|---|---|---|---|---|---|
| 21 | 1970.62 | 220.02 | 1750.59 | 229185.50 | 10913.60 |
| 26 | 530.37 | 244.94 | 285.43 | 32881.52 | 1264.674 |
| 31 | 324.96 | 210.53 | 114.43 | 10657.72 | 343.79 |
| 36 | 306.45 | 212.45 | 93.99 | 8065.22 | 224.03 |
| 41 | 312.79 | 226.27 | 86.53 | 7241.53 | 176.62 |
| 46 | 318.87 | 236.53 | 82.34 | 6827.69 | 148.43 |
| 51 | 316.81 | 228.86 | 87.95 | 7403.09 | 145.16 |
| 56 | 303.55 | 213.12 | 90.44 | 7688.85 | 137.30 |
| 61 | 313.08 | 215.06 | 98.03 | 8722.04 | 142.98 |
| 66 | 328.36 | 224.24 | 104.12 | 9338.35 | 141.49 |
| 71 | 341.52 | 233.01 | 108.51 | 10452.80 | 147.22 |
| 76 | 317.08 | 199.22 | 117.86 | 11017.10 | 144.96 |

We note that the smaller values of $q$ have higher estimates for $\delta$. This is due to the increase in length of the busy period as can be seen in $\hat{B}$. This does in turn

occur due to the increasing probability of a reignition which lower values of $q$ bring. For values of $q < 21$ we observed no renewals in our simulations which is likely a result of the aforementioned increase in the probability of a reignition. For some lower values of $q$ it is also not unlikely that individuals returning from a quarantine will return to a kindergarten where the disease is still spreading due to the low infectivity of the disease in question. We note from Figure [2] that the optimal value of $q$ for $R_0 = 1.5$ is approximately 40. When including the external probability of infection this $q$ should increase slightly and that is why $\hat{\delta}$ is the smallest for $q$:s around 40 to 50.

Using the estimates from Table [6] in combination with Theorem [10.1.4] we can derive the long run number of children in quarantine per day and the number of infections per day.

$$\frac{R(t)}{t} \xrightarrow{a.s} \frac{\delta}{\mu} \approx \frac{\hat{\delta}}{\hat{\mu}} \quad \text{as } t \to \infty$$
$$\frac{R(t)}{tq} \xrightarrow{a.s} \frac{\nu}{\mu} \approx \frac{\hat{\nu}}{\hat{\mu}} \quad \text{as } t \to \infty$$

**Table 7:** Long Run Estimations

| $q$ | $\frac{\hat{\delta}}{\hat{\mu}}$ | $\frac{\hat{\delta}}{\hat{\mu}N}$ | $\frac{\hat{\nu}}{\hat{\mu}}$ |
|---|---|---|---|
| 21 | 116.30 | 0.291 | 5.54 |
| 26 | 62.00 | 0.155 | 2.38 |
| 31 | 32.80 | 0.082 | 1.06 |
| 36 | 26.32 | 0.066 | 0.73 |
| 41 | 23.15 | 0.06 | 0.56 |
| 46 | 21.55 | 0.058 | 0.47 |
| 51 | 23.37 | 0.058 | 0.45 |
| 56 | 25.33 | 0.063 | 0.45 |
| 61 | 27.85 | 0.070 | 0.45 |
| 66 | 28.44 | 0.071 | 0.43 |
| 71 | 30.60 | 0.076 | 0.43 |
| 76 | 34.75 | 0.086 | 0.46 |

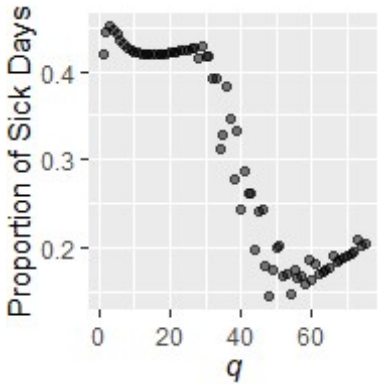
From Table [7] we see a decrease in the number of infections per day as $q$ increases. This number stops decreasing at around $q \approx 46$ since the epidemics start behaving similarly. In these cases we can see as in Figure [8] that most busy periods are brief, with some exceptions, and with to reignitions. Furthermore we also note a decrease in the proportion of children in quarantine when increasing $q$. At some value of $q \approx 46$ however we seem to be getting similar outcomes which as previously mentioned are shorter busy periods with no reignitions. As

a result an increase in $q$ in these cases would not reduce the number of infections and only increase the number of days an already recovered children spend at home.

# 8 Conclusion

Throughout this thesis we have examined how we could reduce the number of total days in quarantine and control the spread of epidemics of varying infectiousness in a kindergarten setting. When applying the basic SIQS model we found that the total number of sick days was very high and therefore came to the conclusion that some other method would have to be implemented to combat the disease.

**Figure 3:** Sick Days




In the case of the original disease the proportion of sick days is approximately 0.53 as can be seen in Figure [1] in the Appendix. To reduce this number we implemented a forced withdrawal model where the quarantine time was no longer stochastic but fixed. By doing this we where able to reduce the proportion of sick days significantly in the original disease as can be seen in Figure [3] above. In the case of when we choose the optimal value of $q$ we are even able to reduce the proportion to approximately 0.145.

We where also able to find optimal values of $q$ for a number of diseases with varying infectiousness as can be seen in Figure [2]. The values which have been called the optimal values of $q$ or $q_{\min}$ are based on taking the mean of the epidemic. Since the mean is heavily influenced by the larger deviations in quarantine days that are present in larger outbreaks or extremely early terminations it is interesting to also observe the median as we have done in Table [3], [4] and [5]. From the tables we observed that the medians where always lower compared to the mean, due to the influence of the larger outbreaks. We also observed what happened when we took different quantiles of the simulation and as such only ignored the largest deviations.

We are also able to obtain a lower bound for the probability of an early termination by using the expected value of $\eta$ can be seen in Figure [1(b)] and different values of $q$. The lower bound can be seen in Figure [7] in the Appendix for different values of $q_{\min}$ and $R_0$. Under the assumption that the number of initially quarantined is approximately the entire population, that is for $R_0 \gtrapprox 4$, the lower bound is simply (1) with $\eta = N$.

When comparing different disease the true probability of an early termination will depend on the infectiousness since the probability of termination given some

random number of returning infectious is smaller for more infectious diseases. Using this lower bound we could choose a value of $q$ in a scenario where we want the probability of a large outbreak to be at least some value.

Furthermore, we also implemented the possibility of an external infection where children could get infected outside of the kindergarten. We used this in combination with renewal theory to derive some interesting results in the long rung which in some regard looked very similar to those from our other simulations.

Lastly it should be reiterated that the results and conclusions drawn in this thesis are by no means relatable to a real world scenario since the assumptions which our models are based on are unrealistic. As previously mentioned we have not accounted for the adjusted probability of infection that would occur when individuals leave for the quarantine. We have also made some unrealistic assumptions regarding the distribution of the infectious, quarantine and recovery period. It seems more probable that the the recovery period of a disease should be a fixed value or some uniform distribution at that value instead of being geometrically distributed.

We have also chosen to ignore certain implications of extended quarantine times in regards to the setting. A child which spends longer time at home might infect their family and as such creating a second household where the epidemic spreads. The aim of this thesis was never to make any real life reflections but instead to showcase different techniques to which a theoretical epidemic could be limited under the assumptions presented.

# 9 Further Developments

The ideas in this thesis can be developed further in a variety of ways. An important aspect which we have chosen to exclude is that kindergarten groups, and certainly not kindergartens as a whole, might not fulfil the homogeneous mixing present in the household model which we have assumed. It is actually highly unlikely that a child in a class interacts with each other child the same way. To deal with this issue one could implement an underlying, possible weighted, social graph and letting the probabilities of infection depend on the edges in said graph.

Throughout the thesis we have not chosen to adjust the probabilities of infection based on the fact that some individuals stay in quarantine. In reality this is not the case since if the number of individuals in a group of potential contacts decrease the probability to interact, and thus infect, each individual should increase. By including this adjustment the theoretical parts of this paper would be very difficult do derive whilst the simulations would contribute with more realistic results.

The lower bound which was calculated in section [6.3] and shown in Figure [7] could most likely be improved by either using brute force calculations or the introduction of other techniques. This would certainly shed more light on the

difference between the simulated and theoretical results. In the simulated case as previously mentioned the optimal value of $q$ is calculated by observing a large amount of simulations where one large outbreak might change the results drastically. As such it would be very interesting to observe how a more precise probability of an early termination could be used to justify different values of $q$. It could for example be possible to recommend a lower number of days of forced withdrawal until the probability of early termination reaches a certain threshold.

We could also perform our simulations over different period lengths. As mentioned and seen previously the number of sick days peaks in the early stages of our forced withdrawal model for the respective $q_{\min}$ of each disease. Since the epidemics terminate after the initial stage the proportion of total sick days decrease as we increase the number of days in our simulation. If we however were to decrease the number of days in the simulation, the optimal value of $q$ would likely change. Since the results and the optimal values of $q$ would differ it could be of interest to observe different lengths of the simulations.

Lastly we could always increase the scale of our simulations in order to obtain even more precise results.

# 10 Appendix

## 10.1 Calculations

### 10.1.1 Probability Generating Function of a Random Sum

This theorem and proof can also be found in [4].

*Let $X_1$, $X_2, \ldots$ be independent and identically distributed non-negative, discrete random variables and let $N$ be a non-negative discrete random variable which is independent of $X_1$, $X_2, \ldots$. Let $S_0 = 0$ and $S_n = X_1 + X_2 + \ldots + X_n$, for $n \geq 1$. Then the following holds.*

$$\Pi_{S_N}(z) = \Pi_N\Big(\Pi_X(Z)\Big).$$

*Proof.*

$$\begin{aligned}
\Pi_{S_N}(z) = E[z^{S_N}] &= \sum_{n=0}^{\infty} E[z^{S_N}|N=n] \cdot \mathbb{P}(N=n) \\
&= \sum_{n=0}^{\infty} E[z^{S_n}|N=n] \cdot \mathbb{P}(N=n) \\
&= \sum_{n=0}^{\infty} E[z^{S_n}] \cdot \mathbb{P}(N=n) \\
&= \sum_{n=0}^{\infty} \Big(\Pi_X(z)\Big)^n \cdot \mathbb{P}(N=n) = \Pi_N\Big(\Pi_X(z)\Big)
\end{aligned}$$

□

### 10.1.2 Probability to Avoid Infection

Using the definitions and notations introduced in this thesis the probability to not be infected by an initial infected conditioned on their infectious period is the following.

$$\mathbb{P}(\text{Not Infected}|I) = (1-\pi_I)^I$$

*Proof.*

We prove this using induction. We know that the probability to not get infected by an infectious individual during a jump in time is $1 - \pi_I$. To reduce the notation let $P(i) = \mathbb{P}(\text{Not Infected}|I = i)$.

*Base case.*

$P(n)$ holds for the base case where $n = 1$.

$$P(1) = 1 - \pi_I = (1-\pi_I)^1$$

*Inductive step.*

Assuming that $P(n)$ holds $P(n+1)$ also holds due to the independence of the attempted infections. The probability to no be infected during $n+1$ jumps in time is as such the probability to no be infected during the first $n$ jumps in time as well as not being infected during the last jump in time.

$$P(n)(1-\pi_I) = (1-\pi_I)^{n+1} = P(n+1) \qquad \square$$

### 10.1.3 Analytical Eigenvalue of $A_K$

Let $a = \frac{10\pi_G}{6\pi_G+0.4}$ and $b = \frac{\pi_G}{0.6\pi_G+0.4}$ then $A_K$ looks as follows.

$$A_K = \begin{pmatrix} a & b & \dots & b \\ b & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ b & \dots & \dots & a \end{pmatrix} \quad |A_K - \lambda I_K| = \begin{vmatrix} a-\lambda & b & \dots & b \\ b & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ b & \dots & \dots & a-\lambda \end{vmatrix}$$

We now add all rows to the first row.

$$\begin{vmatrix} a-\lambda+(k-1)b & a-\lambda+(k-1)b & \dots & a-\lambda+(k-1)b \\ b & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ b & \dots & \dots & a-\lambda \end{vmatrix}$$

Then we subtract the first column from the other columns.

$$\begin{vmatrix} a-\lambda+(k-1)b & 0 & \dots & 0 \\ b & a-\lambda-b & & \vdots \\ \vdots & & \ddots & \vdots \\ b & \dots & \dots & a-\lambda-b \end{vmatrix}$$

Then we divide the 2nd until the $k$th column by $a-\lambda-b$ and assuming that $\lambda \neq a-b$.

$$(a-\lambda-b)^{k-1} \begin{vmatrix} a-\lambda+(k-1)b & 0 & \dots & 0 \\ b & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ b & \dots & \dots & 1 \end{vmatrix}$$

If we then compute the remaining determinant using Laplace expansion we get the following.

$$\begin{aligned}|A_K - \lambda I_K| &= (a-\lambda-b)^{k-1}(a-\lambda+(k-1)b) \\ &= \{\text{Substituting } a \text{ and } b\} \\ &= \left(\frac{10\pi_G}{6\pi_G+0.4} - \lambda - \frac{\pi_G}{0.6\pi_G+0.4}\right)^{k-1}\left(\frac{10\pi_G}{6\pi_G+0.4} - \lambda + (k-1)\frac{\pi_G}{0.6\pi_G+0.4}\right)\end{aligned}$$

The reproduction number is then on the following form.

$$\begin{aligned}R_0 &= \max\left\{\lambda : \left(\frac{10\pi_G}{6\pi_G+0.4} - \lambda - \frac{\pi_G}{0.6\pi_G+0.4}\right)^{k-1}\left(\frac{10\pi_G}{6\pi_G+0.4} - \lambda + (k-1)\frac{\pi_G}{0.6\pi_G+0.4}\right) = 0\right\} \\ &= \max\left\{\lambda : \left(\frac{10\pi_G}{6\pi_G+0.4} - \lambda - \frac{\pi_G}{0.6\pi_G+0.4}\right) = 0 \vee \left(\frac{10\pi_G}{6\pi_G+0.4} - \lambda + (k-1)\frac{\pi_G}{0.6\pi_G+0.4}\right) = 0\right\} \\ &= \max\left\{\frac{10\pi_G}{6\pi_G+0.4} - \frac{\pi_G}{0.6\pi_G+0.4}, \frac{10\pi_G}{6\pi_G+0.4} + (k-1)\frac{\pi_G}{0.6\pi_G+0.4}\right\} \\ &= \frac{10\pi_G}{6\pi_G+0.4} + (k-1)\frac{\pi_G}{0.6\pi_G+0.4}\end{aligned}$$

The final equality holds for all $k \geq 1$ and $\pi_G \in [0,1]$.

### 10.1.4 Convergence of a Renewal Reward Process

This proof can also be found in [6].

Let $R_i \sim R$. If $\mathbb{E}[R] < \infty$ and $\mathbb{E}[T] = \mu < \infty$ then as $t \to \infty$.

$$\frac{R(t)}{t} \xrightarrow{a.s} \frac{\mathbb{E}[R]}{\mu}$$

*Proof.*

$$\frac{R(t)}{t} = \frac{\sum_{i=1}^{N(t)} R_i}{t} = \frac{\sum_{i=1}^{N(t)} R_i}{N(t)}\frac{N(t)}{t}$$

$\frac{\sum_{i=1}^{N(t)} R_i}{N(t)} \xrightarrow{a.s} \mathbb{E}[R]$ by the Strong Law of Large Numbers

$\frac{N(t)}{t} \xrightarrow{a.s} \frac{1}{\mu}$ From Proposition 7.1 in [6] □

## 10.2 Figures

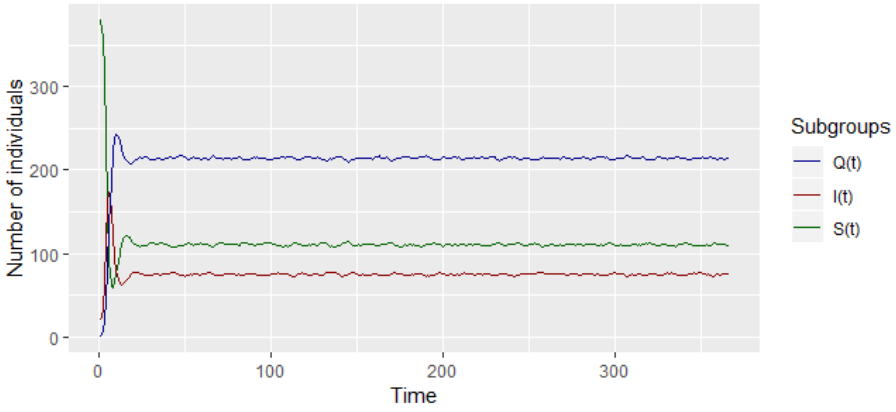



**Figure 1:** A multitype SIQS model with the mean of 100 simulations over 365 days for the probabilities defined in Table [1].

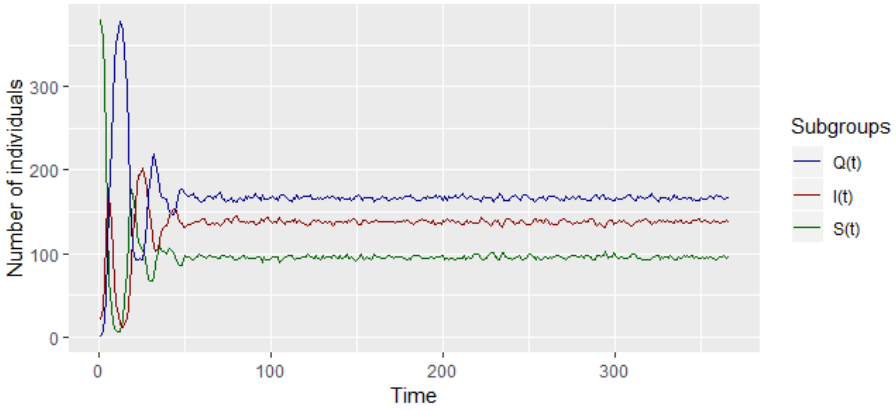



**Figure 2:** A multitype Forced Withdrawal model with the mean of 100 simulations over 365 days for the probabilities defined in Table [2].

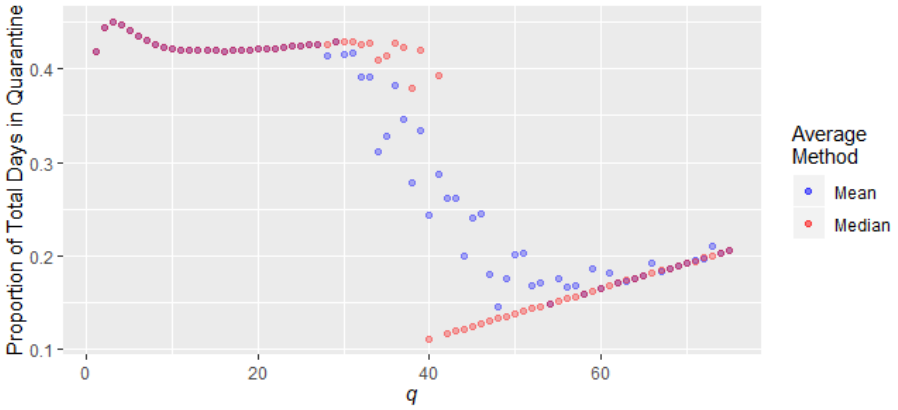



**Figure 3:** The proportion of total days in quarantine for a Forced Withdrawal model with the mean and median of 25 simulations over 365 days for $q \in [1, 75]$ with $(\pi_L, \pi_G) = (0.03, 0.003)$.

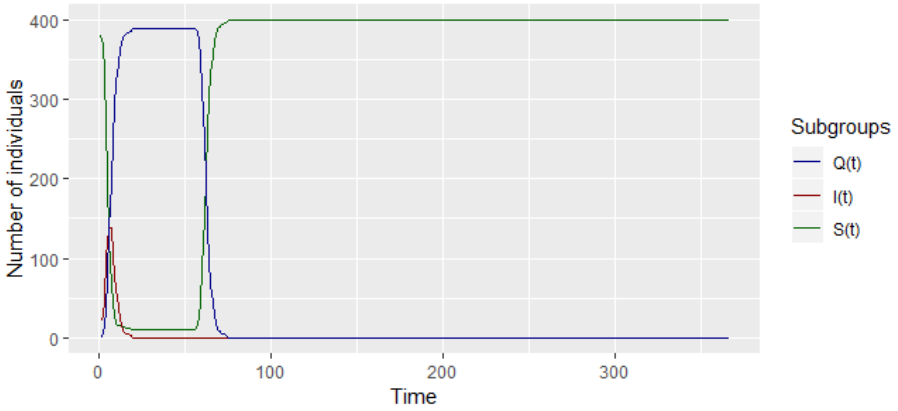



**Figure 4:** The number of total days in quarantine for a Forced Withdrawal model with the mean of 25 simulations over 365 days for $q = 48$ with $(\pi_L, \pi_G) = (0.03, 0.003)$.

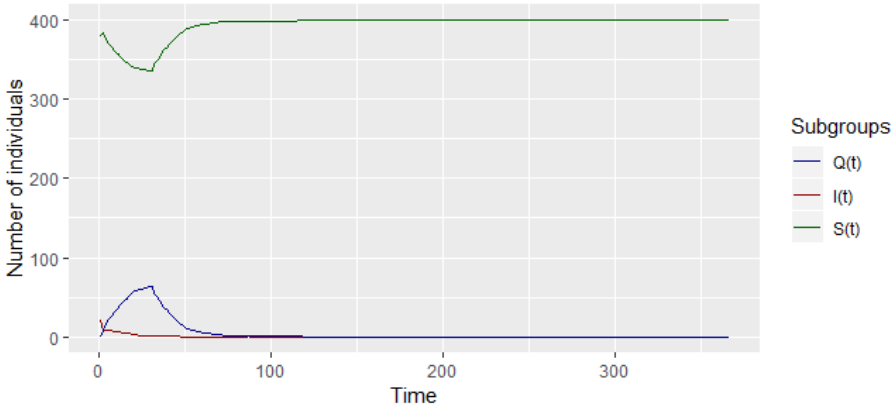



**Figure 5:** The number of total days in quarantine for a Forced Withdrawal model with the mean of 25 simulations over 365 days for $q = 29$ with $(\pi_L, \pi_G) = (0.0073, 0.00073)$.

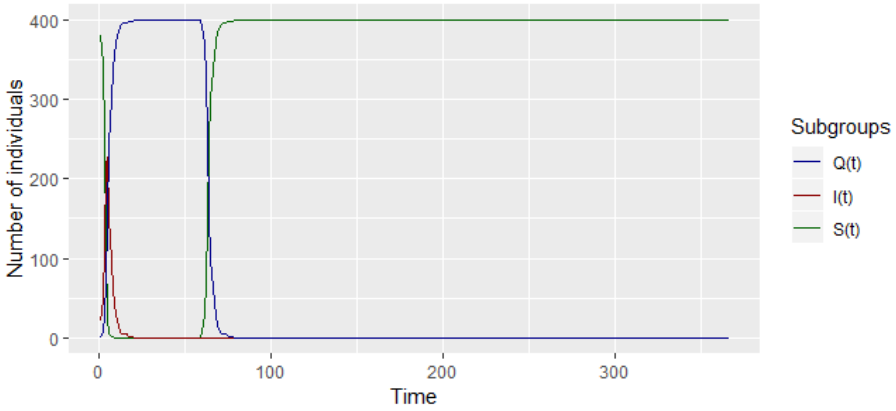



**Figure 6:** The number of total days in quarantine for a Forced Withdrawal model with the mean of 25 simulations over 365 days for $q = 57$ with $(\pi_L, \pi_G) = (0.045, 0.0045)$.

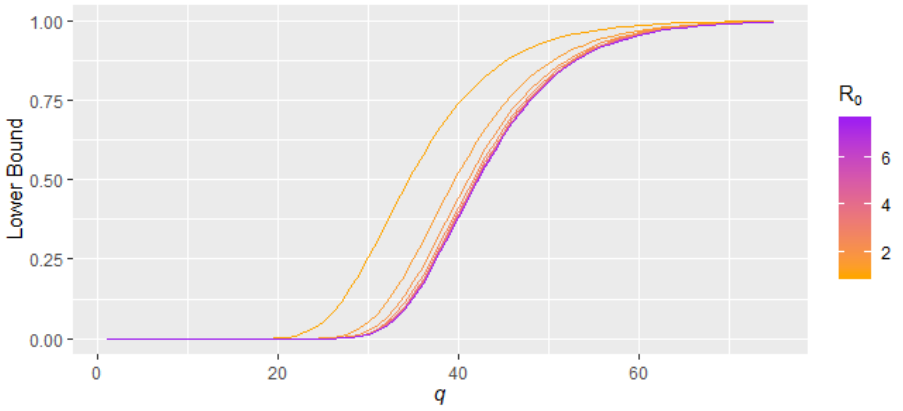



**Figure 7:** A lower bound for the probability of an early termination for epidemics with varying infectiousness conditioned on $\eta = \mathbb{E}[\eta]$.

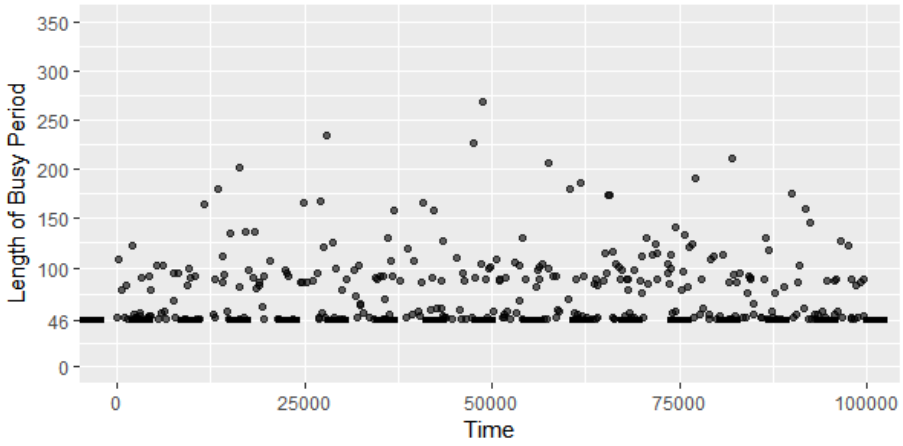



**Figure 8:** The length of each busy period for an epidemic with $R_0 = 1.5$, $q = 46$, $\pi_E = 0.000011$ simulated over 100 000 days.

# References


[1] H. Abbey, "An examination of the reed-frost theory of epidemics", *Human Biology*, vol. 24, no. 3, pp. 201–233, 1952.

[2] H. Andersson and T. Britton, "Stochastic epidemic models and their statistical analysis, volume 151 of lecture notes in statistics", *Lecture Notes in Statistics*, vol. 151, Jan. 2000.

[3] F. Ball, L. Pellis, and P. Trapman, "Reproduction numbers for epidemic models with households and other social structures ii: Comparisons and implications for vaccination", *Mathematical biosciences*, vol. 274, pp. 108–139, 2016.

[4] A. Gut, *An Intermediate Course in Probability*. Springer New York, Jan. 2009.

[5] W. Kermack and A. McKendrick, "A contribution to the mathematical theory of epidemics", *Proc. Roy. Soc. Edinburgh*, vol. 115, pp. 700–721, Jan. 1927.

[6] S. M. Ross, *Introduction to probability models*. Academic press, 2014.