Stockholms universitet

Is legislating vaccination neccessary?

An investigation of the combined effect of vaccine and quarantine of school classes using stochastic modelling

Fredrika Lundahl

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2019:15
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 **2019:15**
http://www.math.su.se

Is legislating vaccination necessary?

An investigation of the combined effect of vaccine and quarantine of school classes using stochastic modelling

Fredrika Lundahl*

June 2019

Abstract

In this thesis we investigate an epidemic outbreak of a childhood disease in a population where many but far from everybody are vaccinated. This is a common situation and it is feared to be even more common as vaccine hesitancy and scepticism are on the rise.

Here we are interested in what effect an active response to an outbreak, taking the form of quarantine the unprotected population, might have on the spread of the disease. We consider the population to be divided into school classes, each individual makes contacts inside and outside the class at different rates. When an infection is detected the individual and all unvaccinated class mates get sent to quarantine. A recovered individual attains immunity, which gives our model similarities with a classical SIR model with household structure, but with an additional state of quarantine.

We derive expressions for the values of R_0 for a school class, which is independent of how many infected individuals there are inside it, the expected final proportion school classes that have had at least one infected individual and with the help of a result from a paper by Trapman and Bootsma from 2009 we get an approximate expression for the expected number infected individuals in a school class which leads us to an expression for the expected proportion infected individuals in the population.

Simulations show that the strategy is powerful, the disease reaches much fewer than when no quarantine is used, for example we have a situation where 40% are vaccinated in which 99% of the susceptible population get infected when we do not use quarantine of the school classes and when we use the quarantine strategy 1% get infected.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: ffredrikaa@gmail.com. Supervisor: Pieter Trapman & Abhishek Pal Majumder .

Sammanfattning

I den här uppsatsen undersöker vi ett epidemiutbrott av en barndomssjukdom i en population där många men långt från alla är vaccinerade. Detta är en vanlig situation och den är befarad att bli allt vanligare som en följd av att skeptisicm och tveksamhet gentemot vaccin är på uppgång.

Vi är här intresserade av vilken effekt karantän av den oskyddade populationen kan ha på spridningen av sjukdomen. Vi betraktar populationen som indelad i skolklasser, varje individ kommer i kontakt med individer inom och utanför klassen med olika intensiteter. När en infektion upptäcks sänds individen och samtliga av dennes ovaccinerade klasskamrater i karantän. En tillfrisknad individ uppnår immunitet, vilket ger vår modell likheter med den klassiska SIR modellen med hushållsstruktur, men här med karantän som ett ytterligare tillstånd.

Vi härleder uttryck för värdet på R_0 för en skolklass, vilket är oberoende av hur många smittade individer det är i klassen, den förväntade slutliga proportionen av skolklasser som har haft åtminstone en smittad individ och med hjälp av ett resultat från en artikel av Trapman och Bootsma från 2009 får vi ett uttryck för det förväntade antalet smittade individer i en skolklass vilket leder oss till ett uttryck för den förvändade proportionen smittade individer i populationen.

Simuleringar visar att strategin är kraftfull, sjukdomen når många färre när karantän används, till exempel har vi en situation där 40% är vaccinerade i vilken 99% av den mottagliga populationen blir infekterad när vi inte använder karantän av skolklasserna och 1% när vi använder karantän.

Acknowledgements

I wish to direct a big thank you to my supervisors Pieter Trapman and Abhishek Pal Majumder for their valuable advice, great and humble inputs and big understanding in my situation. I would not have managed this all by my own.

Contents

1 Introduction **1**

2 The Model **2**
2.1 Clarifications and generalizations 2
2.2 States and possible transitions 3
2.3 Parameters and variables 4

3 Theory **5**
3.1 Branching processes 5
3.2 The Standard SIR model 7
3.3 The household model 8

4 Computations **9**
4.1 R_0: The basic reproduction number 9
4.1.1 A base case: no quarantine 9
4.1.2 R_0 for a school class: $\tilde{R}_0$ 10
4.2 Final size equation 15
4.2.1 A macro perspective: Affected school classes 15
4.2.2 A micro perspective: Affected individuals within a school class 17
4.2.3 A combined perspective: infected individuals in the population 20

5 Simulations **21**
5.1 Method 21
5.2 Results 22

6 Conclusions **27**

7 Discussion **28**

8 References **30**

1 Introduction

Vaccination is a way to give immunity or an increased protection against some kinds of infectious diseases (WHO), and it has probably prevented many large outbreaks and saved a tremendous amount of lives. The World Health Organization, WHO, estimates that vaccine saves 2-3 million lives each year (WHO, 2019). Even so, some parents are hesitant to vaccinate their children. The fear of by-effects seem for many of them bigger than the fear of the disease. Maybe because these are, as a consequence of successful (in the aspect of disease prevention) vaccinations, more frequently occurring than epidemic outbreaks. However, keeping infectious diseases under control is often a nationwide, and sometimes worldwide, collaboration and officials are afraid that vaccine scepticism endangers this work. WHO even mentioned vaccine hesitancy as one of the ten threats to global health 2019 (WHO, 2019).

One way to respond to this upcoming problem by legislation. The New Scientist (2009) points out that in some, although few, countries vaccination is mandatory for attending school. Some of these countries are Germany, France and the US. They also mention that after the measles outbreak in Italy a similar policy has come up. Unvaccinated children under 6 years are not welcome to kindergartens and parents of unvaccinated school children are punished with €500 fees.

In this thesis we are trying another approach and investigate how well we can prevent large outbreaks by using quarantine of the unvaccinated population when someone is diagnosed in the close surroundings.

We are doing this by measuring the expected final size and come up with expressions for both the outbreak within a class and the outbreak among classes. Combining these gives the final size counted in proportion of initially susceptible individuals who ever got infected. We are also interested in what kind of role parameters such as infectivity levels, time to diagnosis and percentage vaccinated plays, and if for some combinations this strategy is a good enough alternative to forcing vaccination upon the entire population. This is investigated with simulations.

The kind of diseases we are to investigate are all airborne kinds and the vaccine is considered to give perfect protection. The diseases are considered to mainly spread among children and we allow for the children to get the disease from their classmates or the society as a whole with different probabilities. There is also an exponentially distributed time between the individual getting the disease and being diagnosed, implying that infectious individuals

might have infected their surroundings and unvaccinated classmates before being put into quarantine, which is why the unvaccinated classmates are put into quarantine as well.

The spread among school classes is considered to take the form of a branching process in the beginning, and the epidemic model is a special case of the SIR (Susceptible, Infectious, Recovered) model, where we add the state Quarantine. This thesis begins with a closer description of our model to then in the next section go through some theory about the population structure and the epidemic spread. We then go on with the computations of R_0 for a school class and the final proportion of school classes to get affected to finish with simulations of the spread for different rates and percentages vaccinated.

2 The Model

2.1 Clarifications and generalizations

One must always remember that a mathematical epidemic model is only just that; a model of something happening and not a fully accurate description, however, often accurate enough. There will always be a trade off between realism and mathematical convenience and compromises need to me made. In this thesis we are using the following sometimes somewhat unrealistic generalizations to make the computations easier.

- The population is considered to be homogeneous, meaning that all individuals that are not immune are just as likely to get infected and the disease will evolve the same way for everyone. We also consider the population to be closed so that no individuals leave or enters the population.
- We assume homogeneous mixing in the society, meaning that the probability of meeting someone from the society more than once during the time of the outbreak is considered to be very low.
- The vaccine is assumed to give a 100% protection.
- All school classes are assumed to be of the same size.
- The vaccinated are supposed to be distributed equally among all school classes, i.e. all school classes consist of the same number of vaccinated and unvaccinated children.

- The time between infection and diagnose for an individual is exponentially distributed.
- An infection is always detected and the individual diagnosed at some time point prior to recovery. This is not too unrealistic, considering the distinguishable characteristics of many childhood diseases.
- An individual recovering from the disease is considered immune for further infections. The immunity is considered to be lifelong.
- An individual coming back from quarantine is considered immune, even if she never caught the disease. However, we expect that the probability of the class to get reinfected is very low, low enough for the generalization not to cause any problems.

2.2 States and possible transitions

We model the spread of the infectious disease with the help of an SIR model with an additional state of quarantine to emphasize that the individuals do not spend their whole infectious period among others and that the length of the stay in the infectious state depend on when the individual gets sent to quarantine.

The SIR model is a simple epidemic model assuming that an individual in the case of being infected goes from being susceptible(S) to infected(I) and finally recovers/gets removed(R). A recovered individual attains lifelong immunity. The standard SIR model is further described in section 3.2.

As mentioned, we add the state quarantine(Q) to the model. To make things more clear we also add the state diagnosed(D). An individual is considered to be infective to others the moment she becomes infected, but she can not be put in quarantine until she or someone else in her class has been diagnosed. When someone is diagnosed, all unvaccinated classmates of the person are preventively put in quarantine, that is, the unvaccinated classmates of the diagnosed individual who have not showed any symptoms yet but maybe will later.

The possible transfers between the states are showed below in Figure 1 and 2.

Figure 1: Transitions between the states for the first diagnosed individual in a class

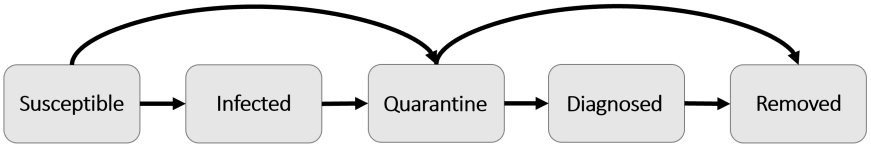

Figure 2: Transitions between the states for individuals preventively put in quarantine. Two possible chains of events exist since individuals might or might not have been infected. However, an individual who has been in quarantine is always considered to attain immunity (recover/get removed).

2.3 Parameters and variables

Here we introduce the following parameters and variables that figure in the thesis, these will be reintroduced upon usage.

n : population size

v : percentage vaccinated

λ_G : global (in society) daily infection rate. Per pair of individuals: λ_G/n

λ_L : local (in class) daily infection rate per pair of individuals

s : size of the school classes (deterministic)

M : size of an outbreak in a school class

D : time to diagnosis for an infected individual. $D \sim Exp(\delta)$, where δ is the rate of detection

T^* : time to the first detection in a school class, given that there is at least one infected in the class. $T^* < \infty$ since detections always happen first.

$I(s)$: infected individuals since the first infection in a school class at time t

$A(t)$: total infectivity of a class at time t, $A(t) = \int_0^t I(s)ds$

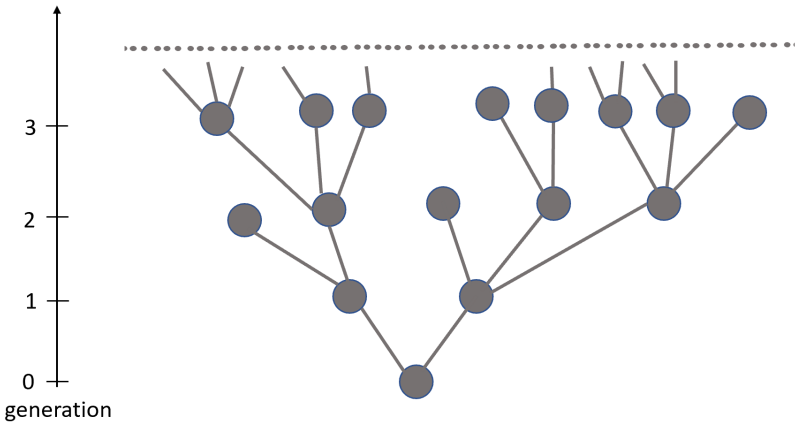

Figure 3: The first four generations of a branching process with one initial individual.

3 Theory

3.1 Branching processes

Early stages of epidemic outbreaks, where the disease has only reached a small part of the population and the probability for an infective to meet another infective or immune individual is small, can often be described with a branching process. Following Sheldon Ross (2014 p 234-), we describe the process by considering a population of individuals with the ability to independently produce offspring of the same kind according to some distribution, and while doing so creating a new generation. The offspring distribution is the same for all individuals. This can easily be translated into epidemics, we then consider a large population with mostly susceptible but also some infective individuals who independently infect new individuals during the infectious period according to some distribution. Figure 3 is an illustration of how the first four generation of a branching process may look.

We let p_j be the probability that an individual has infected j new individuals during her infectious period and then get

$$\mu = \sum_{j=0}^{\infty} j p_j$$

for the mean number of new infected from a single individual. All new

infectives form the next generation of infectives. Letting X_n be the number of individuals in the nth generation and $Z_{n-1,i}$ the number of new infectives from the ith individual in the $(n-1)$th generation we get

$$X_n = \sum_{i=1}^{X_{n-1}} Z_{n-1,i}.$$

For the expected value $E[X_n]$ we condition on X_{n-1} and obtain that

$$\begin{aligned} E[X_n] &= E[E[X_n|X_{n-1}]] = E\left[E\left[\sum_{i=1}^{X_{n-1}} Z_{n-1,i}|X_{n-1}\right]\right] = E[\mu X_{n-1}] \\ &= \mu E[X_{n-1}], \end{aligned}$$

$E[X_{n-1}] = \mu E[X_{n-2}]$,

...

$E[X_2] = \mu E[X_1]$ and

$E[X_1] = \mu E[X_0]$,

where X_0 is a deterministic number describing the number of initially infected individuals. For one initial infected we have $E[X_0] = E[1] = 1$. We then get

$E[X_1] = \mu$

$E[X_2] = \mu^2$

...

$E[X_n] = \mu^n$.

An important observation is that for $\mu \leq 1$ the infectious population goes to zero as $n \to \infty$ with probability 1. Since diseases with $\mu \leq 1$ will not create large outbreaks they are of little interest when it comes to epidemic modeling and we are in this thesis only considering epidemics with $\mu > 1$. In standard SIR epidemics in a homogeneously mixed population μ often coincide with R_0, which is described in section 4.1.

The situation with $X_0 = m$ for $m > 1$ can be seen as m independent branching processes, it then follows that we get $E[X_n] = m\mu^n$.

3.2 The Standard SIR model

The standard SIR model is one of the most simple epidemic models. With the help of what is described by Andersson and Britton (2000, p. 11, 22) we here describe the model, which is of use when setting up an equation describing the proportion of classes that have been affected by the epidemic at some time point, which is done in section 4.1.2.

Individuals are considered to be either susceptible(S), infected(I) or removed(R), meaning that an individual has recovered and therefore is of no further interest for the spread of the disease since lifelong immunity is assumed. The possible transitions between the states are displayed in Figure 4.

Figure 4: Transitions between the states in the SIR model.

In the model an infected individual contacts a given individual according to a Poisson process with rate γ/n during the infected period, where $\gamma > 0$ and n denotes the number of initially susceptible in the population. The population is considered to be homogeneously mixed and closed. A contacted susceptible individual immediately becomes infected. The Poisson processes for different individuals are considered to be independent and identically distributed. The time an individual remains infected is according a fixed but specified distribution and all infectious periods are identically distributed and independent of one another as well.

Since throughout the infectious period of an infective individual each of the other individuals are contacted according to a Poisson process with rate γ/n the infective individual contact other individuals according to a Poisson process with rate γ. This because a sum of independent Poisson processes another Poisson process with rate the sum of the intensities.

Early phases of an epidemic outbreak, where most people are still suscep-

tible, can be well described with a branching process where an individual during her infectious period infects new individuals at the time points of a Poisson process with rate γ.

3.3 The household model

Assuming homogeneous mixing in a population is often of great mathematical convenience, but is often far from the truth and sometimes too far. Andersson and Britton (2000, p. 55) points out the importance of paying attention to the small social structures in a society. These can for example be households, workplaces and schools, in this thesis we look at school classes. This is because children often play a major role in the spread of infections diseases, much due to their more physical behavior, sometimes lack of proper hygiene and less developed immune system.

Like Andersson and Britton do, we consider the population (of children) to be divided into school classes of the same size. Within them diseases easily get a foothold and due to the big mixing of people from different places the disease can quickly spread across the society if adequate actions are not taken.

During a day a child always meet his or her classmates and random people from the society, this may for example be people on the public transport system. The individuals a child meets from the society differ from day to day and the probability that two random individuals from the society meet again is considered to be low enough to be neglected. We define the following rates for the spread of the disease from an infectious individual:

$$\lambda_L : \text{Local(in class) infection rate}$$

$$\lambda_G : \text{Global(in society) infection rate}$$

where most likely $\lambda_L > \lambda_G$ due to the nature of the contacts. The rate of which an infected individual contact a given individual in the society is λ_G/n where n is the number of initially susceptible individuals in the population. If the contacted individual is susceptible she immediately becomes infected. This also means that an infectious individual make in-society contacts with her classmates at that rate, but this rate can according to Andersson and Britton (2000, p. 55) be neglected compared to the local rate λ_L for large n, meaning that most individuals get the infection from someone in their close surroundings. The local rate λ_L is the rate at which an infected individual

contacts a given susceptible classmate, which unlike the global rate is not scaled.

4 Computations

4.1 R_0: The basic reproduction number

To understand and prevent outbreaks it is of great interest to find the average number of new infectives caused by a typical infected in an early stage of the epidemic. This is described by the basic reproduction number, denoted by R_0. According to the threshold limit theorem (described in section 4.2) a large outbreak can not happen for $R_0 \leq 1$ (Andersson and Britton, 2000, p. 6), and infectious diseases which such values of R_0 should therefore not be of big concern. For $R_0 > 1$ a large outbreak occurs with positive probability.

R_0 in a way describes the harm caused by a typical individual. This is information of great importance when deciding which actions should be taken in early stages of an epidemic. R_0 is of great importance when it comes to understanding the development of the spread and is a way of comparing infectious diseases with one another.

4.1.1 A base case: no quarantine

This section is not used, but we present it anyway since it gives insight in the idea of this thesis. Also it might be of interest for further development.

Andersson and Britton (2000, p. 56) presents the following way for obtaining R_0 when having a household model:

$$R_0 = E[M]\lambda E[L]$$

where $E[M]$ is the expected number of infected individuals in a school class with one initially infected individual and λ is the global infection rate and L is the length of the infectious period. The expression can be motivated in the following way: an infectious individual infects in average λ individuals from the society, these are from distinct households with high probability due to the homogeneous mixing and a big enough population. Each of these from the society newly infected start a "subepidemic" in their own

school class comprising in average $E[M]$ individuals. Note that the number of individuals considered infected from a single infectious individual is not just the number of individuals who got infected from a contact with the infectious individual in question. We also count the "indirect infections", i.e. if an individual, say individual a, infect someone globally, who then infect others locally, those local infected individuals are seen as infected by the first individual, individual a.

In our model we chose to define the rate to be daily in order to follow the impact of the number of days until diagnose. We therefore let λ_G be the daily global rate and D the days between infection an diagnose. We assume that the infection always is detected, implying that the individual becomes diagnosed, at a time point prior to recovery and that detection happens within a finite time, that is, $D < \infty$. We have that D is exponentially distributed with parameter δ.

The child is of course still infectious for some more time after the detection of the disease but we assume common sense of the parents, meaning that they let the child stay home until recovery and thus the child does not infect other children any more. Because of this and the fact that an infected individual instantly becomes infectious D is also equal to the number of days the individual infects her surroundings. Introducing R_0^* as the value of R_0 when no quarantine is used and using that $\lambda = \lambda_G E[D]$ and that $E[D] = 1/\delta$ we get

$$R_0^* = E[M]\lambda_G/\delta.$$

The expected value of infectives in a school class with one initially infected, $E[M]$, gets rather complicated since we, unlike with the global contacts, can not assume that two individuals do not meet again in a foreseeable future. We get quite the opposite situation, individuals in the same class meet each other at daily basis which leads to dependencies. This can be solved with simulations, or very approximately with the final proportion technique from section 4.2.1. Something that we however are not doing in this thesis.

4.1.2 R_0 for a school class: $\tilde{R_0}$

In this section we consider the R_0 of a school class in the case where quarantine of all unvaccinated individuals in the school class is used to prevent spread of the disease. Upon detection of an infectious individual in a school

class all unvaccinated students in that class are preventively sent into quarantine. These students may already be infected but not yet diagnosed or may still be susceptible.

Unlike in the previous section we can not say that a typical individual infects $\lambda_G E[D]$ individuals from the society. The variable D is not of interest here since the time spans during which individuals are infecting other students are not identically distributed. This because the detection may happen while the individual is in quarantine and thus the individual then do not spread the disease to others during D time units.

Instead, we here interest ourselves in the value of R_0 for a typical affected class, $\tilde{R_0}$. For that we use the time to the first detection in a school class, which we call T^*.

We introduce some new notation:

$I(t) =$ number of infected individuals at time t

$A(t) = \int_0^t I(s)ds$

We can look upon $A(t)$ as the total infectivity up to time t, which you get if you add up all infectious periods of the infectious individuals in a school class up to time t. $A(T^*)$ then becomes the total infection up to the moment of the first detection. A visualization of $A(T^*)$ is showed in Figrue 5.

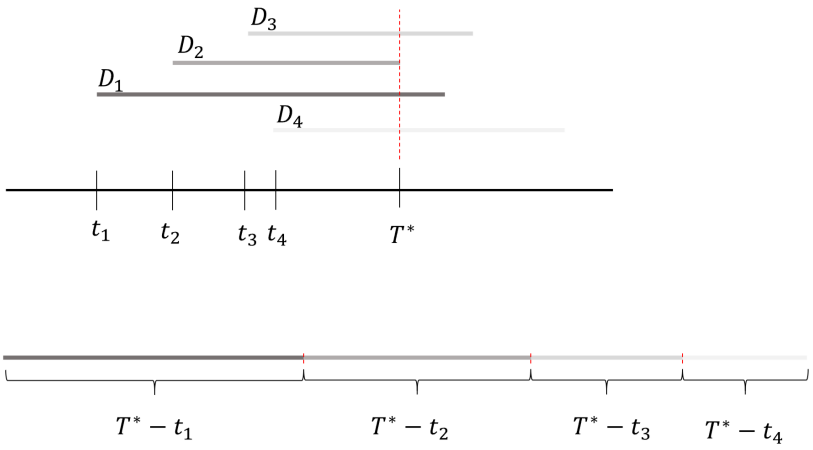

Figure 5: The total infectivity up to time T^* is the sum of the infectious periods up to time T^*, here for the case with 4 infected at the time of detection. Their total infectivity $A(T^*)$ is visualized as one long infectious period which length is the sum of each individual's infectious period up to time T^*.

Recalling that each infected individual independently get detected after an exponentially distributed time with rate δ we get the following important result:

$$A(T^*) \sim Exp(\delta)$$

Proof: Assume that we at time T^* have j infected individuals who consequently got infected prior to T^*. We know that $j \geq 1$ since if we had the opposite situation a detection would not have been possible, i.e. we would have no T^*. We denote the time between infection and diagnosis for the ith infected individual by D_i and the time point when the i:th infected individual got infected by t_i. Note that $t_i < T^*$ for all $i \leq j$ due to our assumption.

All D_i are considered to be i.i.d. $Exp(\delta)$ variables. We recall the following about the exponential distribution:

For the random variable $D \sim Exp(\delta)$
density function: $f_D(t) = \delta e^{-\delta t}$
distribution function: $F_D(t) = 1 - e^{-\delta t} \Rightarrow P(D > t) = e^{-\delta t}$

memoryless property: $P(D > a + b|D > b) = P(D > a)$

minimum of independent exponential variables:

$min\{D_1, D_2, ..., D_n\} \sim Exp(\delta_1 + \delta_2 + ... + \delta_n)$

We begin with using the fact that $A(T^*)$ can be interpreted as the lengths of all infectious periods up to time T^*, the time of the first detection. Therefore

$$\begin{aligned} P(A(T^*) > a) &= P((T^* - t_1) + (T^* - t_2) + ... + (T^* - t_j) > a) \\ &= P\left(jT^* - \sum_{i=1}^{j} t_i > a\right) \\ &= P\left(T^* > \frac{a + \sum_{i=1}^{j} t_i}{j}\right). \end{aligned}$$

We make the substitution $k = \frac{a+\sum_{i=1}^{j} t_i}{j}$. Repeated conditioning leads us to

$$\begin{aligned} &P\left(T^* > \frac{a + \sum_{i=1}^{j} t_i}{j}\right) \\ &= P(T^* > k) \\ &= P(T^* \notin [0, k]) \\ &= P(T^* \notin (t_j, k]|T^* > t_j)P(T^* > t_j) \\ &= P(T^* \notin (t_j, k]|T^* > t_j)P(T^* \notin (t_{j-1}, t_j]|T^* > t_{j-1})P(T^* > t_{j-1}) \\ &= P(T^* \notin (t_j, k]|T^* > t_j) \cdot ... \cdot P(T^* \notin (t_1, t_2]|T^* > t_1)P(T^* > t_1). \end{aligned}$$

Note that $P(T^* > t_1) = 1$ since the first detection can not happen before the first infection. We then only need an expression for the conditional probability that the first detection does not happen in a given interval, constituting of the time between two succeeding time points of infections, given that it has not happened yet.

The probability that a detection will not happen in an interval is the probability that neither of the infected in the interval get detected. Each of the individuals has time to detection distributed as $Exp(\delta)$, for each of them the memoryless property can be used which makes it unnecessary to know who have been infected for the longest, them both have the same probability not to get detected in the interval, given that we know neither of them has been

detected yet. The memoryless property gives us that, given that we have not been infected yet, the probability not to get infected in the interval is the probability not to get infected in a time period the length of the interval, no matter where on the timeline the interval is placed.

Now note that the scenario with none of them getting detected in the interval is the same as the scenario where the one to get detected first does not get detected in the interval. The independence of the detections of the individuals makes us able to use that the distribution of the minimum of exponentials is exponentially distributed with the sum of the rates. Hence

$$
\begin{aligned}
&P(T^* \notin (t_j, k] | T^* > t_j) \cdot ... \cdot P(T^* \notin (t_1, t_2] | T^* > t_1) P(T^* > t_1) \\
&= P(min\{D_1, ...D_j\} > k - t_j) \cdot ... \cdot P(min\{D_1, D_2\} > t_3 - t_2) P(D_1 > t_2 - t_1) \\
&= e^{-j\delta(k-t_j)} e^{-(j-1)\delta(t_j - t_{j-1})} \cdot ... \cdot e^{-2\delta(t_3-t_2)} e^{-\delta(t_2-t_1)} \\
&= e^{-\delta(jk - jt_j + (j-1)t_j - (j-1)t_{j-1} + ... + 2t_3 - 2t_2 + t_2 - t_1))} \\
&= e^{-\delta\left(jk - \sum_{i=1}^{j} t_i\right)}.
\end{aligned}
$$

Recalling that $k = \frac{a + \sum_{i=1}^{j} t_i}{j}$ we get

$$
\begin{aligned}
e^{-\delta\left(jk - \sum_{i=1}^{j} t_i\right)} &= e^{-\delta\left(j\frac{a+\sum_{i=1}^{j} t_i}{j} - \sum_{i=1}^{j} t_i\right)} \\
&= e^{-\delta(a)} \\
&= P(D > a)
\end{aligned}
$$

and since $D \sim Exp(\delta)$ so is $A(T^*)$■

We have now proven that the total infectivity of a school class, from first infection to first detection, is distributed as an exponential distribution with rate δ. The interpretation of the total infectivity as the total lengths of all infectious periods will prove useful now as we introduce the expression for R_0 for a school class when quarantine is issued at first detection.

As in Figure 5, we "translate" the infectious periods up until time T^* of all individuals to one long infectious period with length the sum of each infectious individual's infectious period up to time T^*. We imagine this long infectious period to belong to one individual who each day with rate λ_G contact new individuals, of which the proportion $(1 - v)$ should be with

susceptible individuals in early stages of the outbreak due to vaccination. We, with the help of a branching process approximation, get the basic reproduction number for a school class to be

$$\tilde{R_0} = E[A(T^*)]\lambda_G(1-v) = \lambda_G(1-v)/\delta$$

where we in the last equality used that $A(T^*) \sim Exp(\delta)$.

To compute the total infectivity in a class when recovery may happen before detection one might find a paper from 2015 by Ball et al. of interest.

4.2 Final size equation

When looking at epidemics in large populations the threshold limit theorem can be applied, which divides epidemic outbreaks into minor and major such. In the case of a small outbreak a small, barely noticeable, number of individuals get infected before the disease dies out. When the outbreak is a major one a deterministic proportion can be decided for the amount of people having been infected by the disease before it stops. A major outbreak is according to the theorem only possible when $R_0 > 1$ (more about R_0 can be found in section 4.1). With the help of the threshold limit theorem, a final size equation can be set up (Andersson and Britton 2000 p.6).

4.2.1 A macro perspective: Affected school classes

If we "zoom out" and regard the classes as macro individuals we end up with a homogeneously mixed population, that is with no internal structures. This section is only about the final size counted in macro individuals (school classes), for the final size of an outbreak inside a class we refer to section 4.2.2 and for a combination of the two, yielding the final size counted in infected individuals in total we refer to section 4.2.3.

We do not need to go into the mechanism of quarantine when regarding macro individuals, the time between the first infection and the first detection in a class is to be interpreted as the infectious period of a macro individual and when the unvaccinated individuals in the class enter the state of quarantine the macro individual is entering the state of removal. Consequently, we now deal with a standard SIR (Susceptible, Infected, Removed) model, which is described in section 2.2.

In section 2.2 we mentioned that an infected individual contact others according to a Poisson process with rate γ. Here we have λ_G corresponding to γ. In section 4.1.2 we found that for a school class (here called macro individual) $\tilde{R}_0 = \lambda_G(1-v)/\delta$.

To set up the final size equation we also need the variable $\tilde{\tau}$, which we set to represent the proportion of macro individuals who have been infected at some moment during the outbreak.

In order to *not* get infected you would need to avoid infection from all infected individuals. Since there exists n/s macro individuals the probability that a given infectious such infects you, as a susceptible macro individual, is $\frac{\lambda_G(1-v)/\delta}{n/s} = s\lambda_G(1-v)/(n\delta)$ and thus the probability of avoiding to get the infection from that macro individual is $1 - s\lambda_G(1-v)/(n\delta)$. And if you are to not ever catch the disease you need to avoid infection from every macro individual who has ever been infected. Macro individuals contact each other independently of one another and therefore the probability that you never get infected is $(1 - s\lambda_G(1-v)/(n\delta))^{\tilde{\tau}n/s}$, which converges to $e^{-\tilde{\tau}\lambda_G(1-v)/\delta}$ as $n \longrightarrow \infty$.

Asymptotically the proportion individuals who never got infected should be equal to the probability of avoiding infection, and therefore the final size equation is

$$1 - \tilde{\tau} = e^{-\tilde{\tau}\lambda_G(1-v)/\delta}.$$

We use numerical methods to find the solution to this equation, these leads us to

$$\tilde{\tau} = \frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta}) + \lambda_G(1-v)/\delta}{\lambda_G(1-v)/\delta}$$

where $W(\cdot)$ is the product log function, also known as Lambert W function.

Corless et al. (1996) describes the Lambert W function as the function $W(z)$ satisfying $W(z)e^{W(z)} = z$. Plugging in the proposed solution (2) into the equation (1) we get

$$
\begin{aligned}
1-\tilde{\tau} = e^{-\tilde{\tau}\lambda_G(1-v)/\delta} &\iff 1 - \frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta}) + \lambda_G(1-v)/\delta}{\lambda_G(1-v)/\delta} \\
&= e^{-\frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})+\lambda_G(1-v)/\delta}{\lambda_G(1-v)/\delta}\cdot\lambda_G(1-v)/\delta} \\
&\iff -\frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})}{\lambda_G(1-v)/\delta} \\
&= e^{-W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})-\lambda_G(1-v)/\delta} \\
&\iff -\frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})e^{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})}}{\lambda_G(1-v)/\delta} \\
&= e^{-\lambda_G(1-v)/\delta} \\
&\iff W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})e^{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta})} \\
&= -\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta}
\end{aligned}
$$

which is the same as the definition for the Lambert W function with $z = -\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta}$. It is thus the right solution.

The result is visually presented in Figure 6 for a range of $\lambda_G(1-v)/\delta$ between 1 and 6. We can see that for $\tilde{R_0} = \lambda_G(1-v)/\delta > 1$ the proportion of infected rapidly increases and already for $\lambda_G(1-v)/\delta \approx 2.5$ about 90% of the initially susceptible macro population are to be infected.

4.2.2 A micro perspective: Affected individuals within a school class

If we want to know the proportion of individuals who ever got the disease we need to know the expected proportion of infected individuals in a school class that have had an epidemic outbreak. Knowing that, we can multiply with the proportion of school classes that are expected to get affected and get the proportion of children expected to have caught the disease, which is done in section 4.2.3.

However, computing the proportion of infected children in a class is not as simple as computing the proportion of classes, as we did in the previous section. In a school class we find great dependencies, the probability that two individuals do *not* meet again is really small. On the contrary they are to be situated close to each other for a couple of hours each day for more than a hundred days a year.

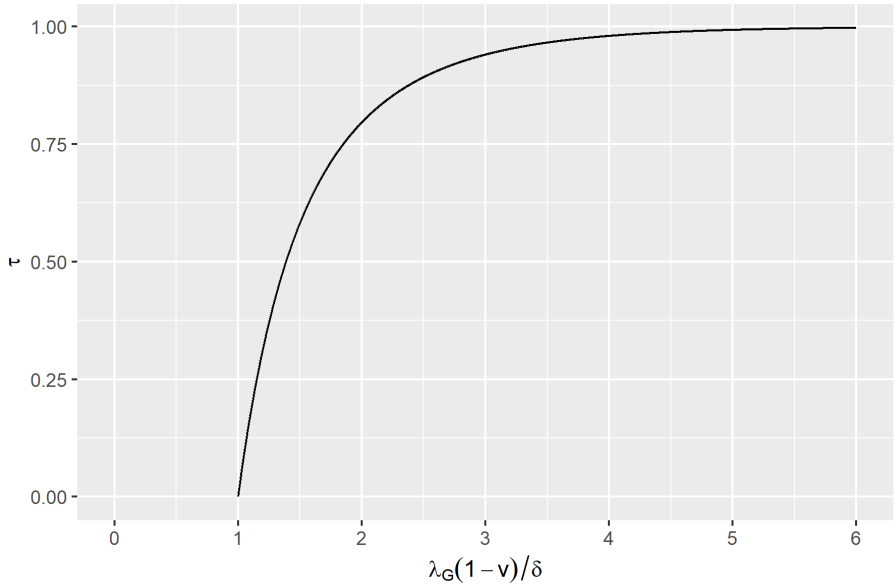

Figure 6: Final proportion of macro individuals who at some point get infected ($\tilde{\tau}$) as a function of $\lambda_G(1-v)/\delta$ ($\tilde{R}_0$).

The close contacts affect both the number of people an infected individual can infect and the probability for a susceptible to get infected and we get a chain-binomial model close the the one described in Andersson and Britton(2000, p. 4-6), which generates expressions very complicated to compute.

If we however could assume that infections occur in a constant rate according to a Poisson process and that we do not experience the depletion of susceptible, something that might be possible in the cases with big school classes, low infection rates, a low percentage vaccinated and a fast detection rate, we could use a result from 2009 by Trapman and Bootsma. The paper goes into the relation between queuing theory and epidemiology and derives the distribution of the number of infectives at the moment of the first detection by establishing a link between continuous time branching processes and the so-called M/G/1 queues with processor sharing, which is M/G/1 queue where we have a processor instead of a clerk and therefore imagine that we serve everyone simultaneously. Understanding the mechanism will however not be of importance here, but for the interested it is well described in the paper.

We however, on the contrary to what is done in the paper, do not approximate our spread to take the form of a continuous time branching process but rather a discrete one. This makes the time aspect incorrect, but the

number of individuals in a generation should be the same.

To be able to assume that the infections inside a school class occur according to the corresponding Poisson process in the paper we instead of the per-pair rate λ_L use a per individual rate λ_L^*.

Very approximately we, with the help of the main result of the paper, get that

$$M \sim Geo(p)$$
$$\text{i.e } P(M = k) = p(1-p)^{k-1}$$

where $p = \delta/(\delta+(1-\pi)\lambda_L^*)$ and π is the smallest root of $\pi = E[e^{-(\delta+(1-\pi)\lambda_L^*)L}]$.

Recall that M denotes the size of the outbreak inside a school class, and since we neglect the small probability of classes getting reinfected this also denotes the final size of the outbreak inside a school class. λ_L^* is the infection rate, once someone is infected in a school class, we neglect the rate of infections from the society, λ_G, since once an infectious individual is in the class it is assumed to start a "subepidemic" rather quickly. L is in the paper the distribution of the infectious period, since we always assume a detection to take place prior to a recovery we here have $L = \infty$. Thus,

$$\pi = E[e^{-\infty(\delta+(1-\pi)\lambda_L^*)}] = e^{-\infty(\delta+(1-\pi)\lambda_L^*)} \to 0$$

and we consequently have $\pi = 0$ and therefore $p = \delta/(\delta + \lambda_L^*)$. Since the expected value of a geometric distribution is $1/p$,

$$E[M] = (\delta + \lambda_L^*)/\delta.$$

This is very intuitive since for two exponential variables, say A and B, with rates α and β we have that

$$P(A < B) = \frac{\alpha}{\alpha + \beta},$$

a proof can be found in Ross, 2014, p. 287. The geometric distribution can therefore be interpreted as the number of times we get to do things before an

event of interest happen, where p is the probability that an event of interest happens. In our situation, the event of interest is the detection, and the "number of times we get to do things" is the number of infections. This means that we can use the geometric distribution to describe the amount of infections that will happen before a detection.

Due to the intuitive nature of the distribution when having $L = \infty$, the result from the paper is not necessarily needed to reach this conclusion, however it is of big use in situations where a recovery prior to a detection is allowed. Allowing it in this paper would nevertheless violate our result in section 4.1.2 about the distribution of the total infectivity in a class. To compute the total infectivity in a class when recovery may happen before detection one might find a paper from 2015 by Ball et al. of interest.

4.2.3 A combined perspective: infected individuals in the population

With the help of the previous sections 4.2.1 and 4.2.2 we here present the expected proportion of individuals who ever get infected. We get this by multiplying the proportion of school classes that would be affected with the proportion of a school class that would be affected. However, due to the approximate nature of the result from section 4.2.2, the affected individuals in a class, the result from this section must also be considered to be approximate.

From section 4.2.1 we got that $\tilde{\tau}$, the proportion of infected school classes is given by

$$\tilde{\tau} = \frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta}) + \lambda_G(1-v)/\delta}{\lambda_G(1-v)/\delta}.$$

From section 4.2.2 we got that

$$E[M] = (\delta + \lambda_L^*)/\delta$$

for the expected number of affected individuals. As a proportion, which we can denote by τ^*, we get

$$\tau^* = E[M]/(s(1-v)) = \frac{\delta + s\lambda_L^*}{\delta s(1-v)}$$

where $s(1-v)$ is the number of initially susceptible in the school class.

The proportion of initially susceptible infected individuals in the population, τ, is therefore

$$\tau = \tilde{\tau}\tau^* = \frac{W(-\lambda_G(1-v)/\delta e^{-\lambda_G(1-v)/\delta}) + \lambda_G(1-v)/\delta}{\lambda_G(1-v)/\delta} \cdot \frac{\delta + \lambda_L^*}{\delta s(1-v)}.$$

5 Simulations

5.1 Method

The simulation can be produced in R in the following way:

Have the following variables, numbers used in this thesis within parenthesis:

pop : size of the population (10000)
m : number of initial infected (1)
v : proportion vaccinated $(0, 0.4, 0.6, 0.8)$
s : size of school classes (25)
rateG : global contact rate $(1/2, 1, 2)$
rateL : local contact rate $(1/5, 1/10, 1/15, 1/20)$
rateD : detection rate $(1/7, 1/3)$
time : time the simulation should go on for, counted in days (100)

(1) Create a data frame with one row for each individual and columns containing the following information:

- Unique number of the individual $(1, 2, ..., pop)$
- Number of the school class the individuals can belong to $(1, 2, ..., pop/s)$
- The current state of the individual. We allow the individual to be in S (for susceptible),R (for removed/recovered, implying immune) or a

number, describing the time of infection. The case of the individual being in a state which is a number thus implies that the individual is infective and got infective at that time point.

- Whether the individual has ever been infected (True/False).
- An exponentially distributed number (with rate $rateD$) describing the time from infection to detection, if the individual gets infected.

(2) Simulate who of the $pop(1 - v)$ susceptible individuals who should be among the m initial infected.

(3) Simulate the infections caused in a day. Do this by, for each infected individual, simulating a binomially distributed number with size parameter the number of susceptible in the class and probability parameter $rateL$. This is the number of newly infected individuals within the class. Randomly choose who to be infected among the susceptible and change their state to the current time. Also choose, for each infected individual, $rateS$ individuals from the population who are contacted. These individuals get infected if their state is S, if so change their state to the current time.

(4) For the infected individuals, check if the current time subtracted with state number (time of infection) exceeds the individual's time to detection. If so, change the state of the individual and all individuals in the same class to R. Increase the current time by 1 and repeat from step (3) while the current time has not yet reached $time$ and the number of susceptible or infected is not 0.

5.2 Results

In Figure 7 we can see how the outbreak evolves for a very aggressive disease when different percentages of the population are vaccinated. We can see that the bigger percentages vaccinated the longer the outbreak will last. When many are vaccinated in a class few are susceptible as a result. The depletion of susceptible classmates then goes then very fast since there are very few of them and the contact rate is per infective-susceptible pair and thus the spread does never gain exponential speed. The high amount of vaccinated however makes the depletion of susceptible individuals outside the school class go slower since many contacts do not result in infections and we see the outbreaks last for longer instead. The disease is too infectious for the vaccinated population to put an end to the spread.

In Figure 8 we have another situation where the time to detection is shorter. Here we can see that for 80% vaccinated the disease has problems getting a foot hold and dies out much quicker than when we have 40 and 60% vaccinated. Note that this disease is even more infectious but it is expected to be detected after 3 days instead of 7.

We also see that the graphs are a bit edgy at places, this has to do with the fact that when a class with many infected get sent to quarantine many infected disappear from the society at the same time. This effect gets even bigger since we simulate in discrete time, meaning that many classes may be sent to quarantine at the same time unit.

We compute the final proportion infected in case of a major outbreak for both the proportion of all initially susceptible, which we denote τ as before, and the proportion in the corresponding case when no quarantine is used, denoted $\hat{\tau}$. It is important to note that even if a major outbreak has a positive probability whenever $R_0 > 1$ different values of R_0 yield different probabilities of a big outbreak. We have given each situation 10 simulations to see if an outbreak happens and see where the proportion converges. In some situations we have not observed any outbreaks, which does not necessarily mean that a big outbreak is impossible, it may just be unlikely to happen and we may have observed a major outbreak if we had done a few simulations more.

We have simulated for the population to consist of $0\%, 40\%, 60\%$ and 80% vaccinated individuals, the results can be found in Table 1,2,3 and 4. In each table we have also investigated the size of the outbreaks for different values of δ, λ_G and λ_L. The values $\delta = 1/7$, $\lambda_G = 2$ and $\lambda_L = 15$ can all be considered to be quite extreme, especially in combination. In a population where neither quarantine nor vaccination is used these parameter values together would mean that an individual is expected to infect close to 4 other individuals in only one day, and is not expected to get detected until 6 more days! This is a situation far from what is normal in real life. We have even used $1/10$ and $1/5$ as values of λ_L, these might be unrealistic but we use them to discover situations where the different percentages vaccinated make a difference.

Reading from the tables we can see that using quarantine makes a big difference for all levels of vaccinated, even none. Many times pretty much the whole susceptible population get infected when no quarantine is used and when we douse quarantine we get the result down with at least two thirds.

When the detection rate δ is $1/7$ the vaccinated population does not always manage to stop the spread of the disease. For the 0%, 40% and 60% vac-

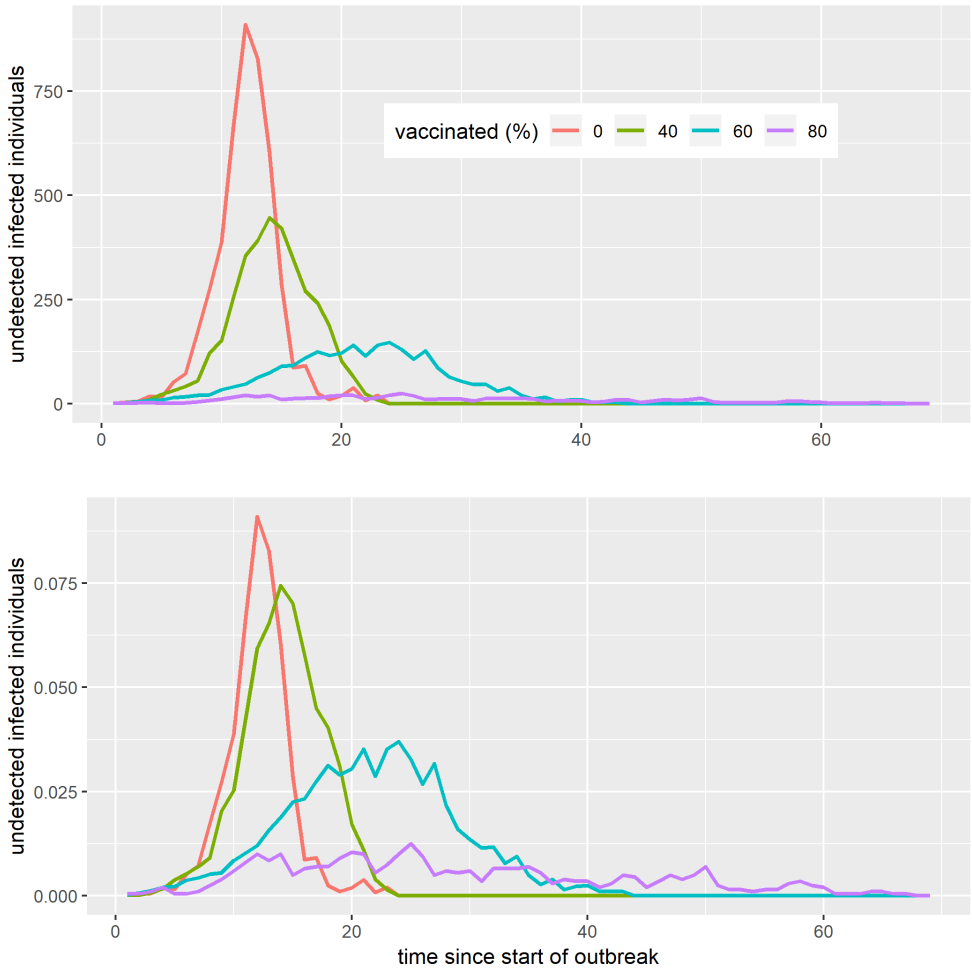

Figure 7: Undetected infected individuals during the outbreak. $\delta = 1/7$, $\lambda_G = 1$ and $\lambda_L = 1/15$, a really infectious and long lasting disease. In the bottom picture is the number of infected as a percentage of initially susceptible individuals. When a big percentage of a population is vaccinated the outbreak lasts for longer. When we deal with really serious diseases like this one the amount of vaccinated, if it is not extremely high, does not stop the spread of the disease, it just makes it take more time.

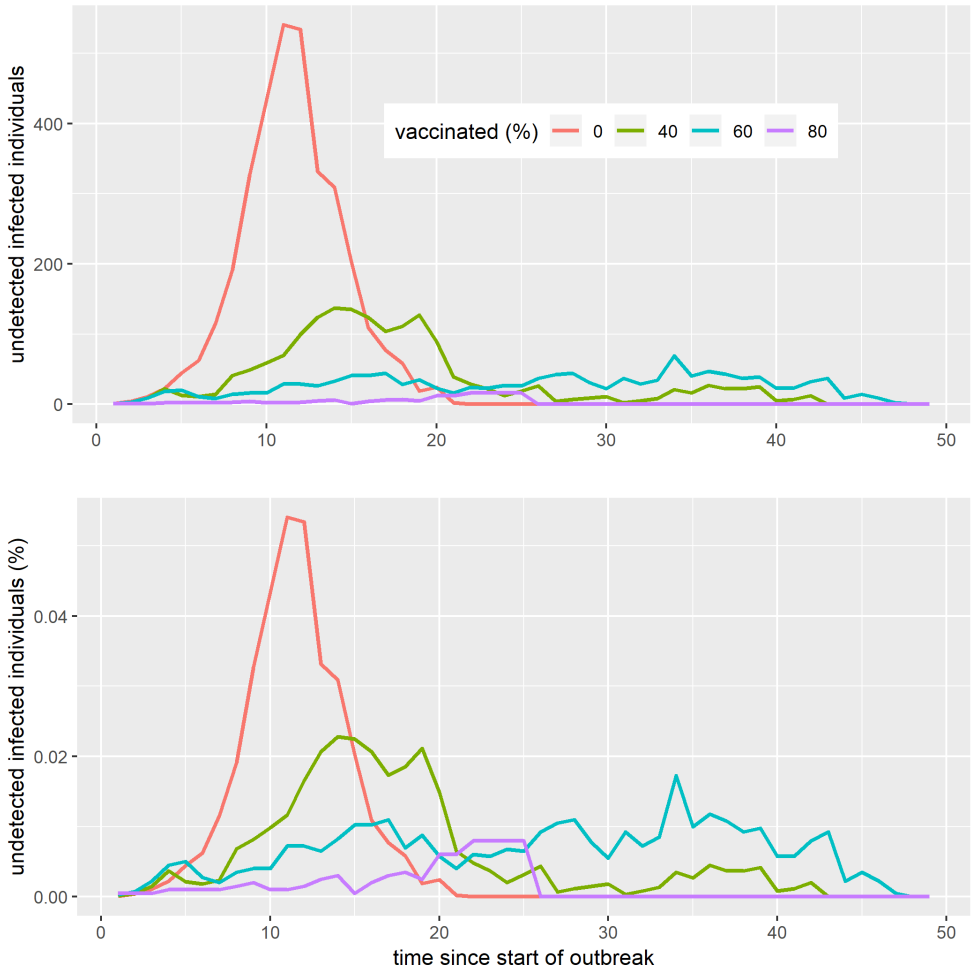

Figure 8: Undetected infected individuals during the outbreak. $\delta = 1/5$, $\lambda_G = 1$ and $\lambda_L = 1/10$, a quite serious disease but not as bad as the one in Figure 7. Here we can see that if we have 80% vaccinated we can stop the spreading before it has affected many. In the bottom picture is the number of infected as a percentage of initially susceptible individuals.

cinated there is not much of a difference regarding the proportion infected among the initially susceptible when we use $\lambda_G = 1$ and $\lambda_L = 1/10$ or $1/15$. For 80% vaccinated we see a considerable decrease, but still many become infected. When $\lambda_G = 1/2$ however the spread quickly stops with only 40% vaccinated.

When we have $\delta = 1/3$ the disease does not get a foothold in the same way. For an extremely infectious disease with $\lambda_G = 2$ and $\lambda_L = 1/5$ vaccination of less than 80% of all individuals does not make much difference. However, if $\lambda_G = 1$ and $\lambda_L = 1/10$ we see a strong effect of vaccination. For 40% vaccinated we can expect 2% of the susceptible population to get infected and for more than 60% vaccinated we have less than 1%.

δ	λ_G	λ_L	τ	$\hat{\tau}$
1/7	1	1/15	0.28	>0.99
1/7	1	1/10	0.33	>0.99
1/7	1/2	1/20	0.20	>0.99
1/7	1/2	1/10	0.26	>0.99
1/3	2	1/5	0.23	>0.99
1/3	1	1/10	0.10	>0.99

Table 1: Simulated final proportion of initially susceptible individuals affected by the disease for 0% vaccinated.

δ	λ_G	λ_L	τ	$\hat{\tau}$
1/7	1	1/15	0.31	>0.99
1/7	1	1/10	0.36	>0.99
1/7	1/2	1/20	<0.1	0.98
1/7	1/2	1/10	0.01	0.99
1/3	2	1/5	0.2	>0.99
1/3	1	1/10	0.02	>0.99

Table 2: Simulated final proportion of initially susceptible individuals affected by the disease for 40% vaccinated.

δ	λ_G	λ_L	τ	$\hat{\tau}$
1/7	1	1/15	0.27	0.98
1/7	1	1/10	0.36	>0.99
1/3	2	1/5	0.18	0.98
1/3	1	1/10	0.01	0.90

Table 3: Simulated final proportion of initially susceptible individuals affected by the disease for 60% vaccinated.

δ	λ_G	λ_L	τ	$\hat{\tau}$
1/7	1	1/15	0.20	0.86
1/7	1	1/10	0.25	0.88
1/3	2	1/5	0.03	0.81

Table 4: Simulated final proportion of initially susceptible individuals affected by the disease for 80% vaccinated.

6 Conclusions

When we deal with *really* infective diseases, vaccination of 40%, 60% and sometimes even 80% of the population is of little help for more than the vaccinated. For very infective diseases those are not enough to stop the disease since the infected makes so many contacts.

Diseases that can go undetected for long are of particular danger. If they additionally are very infective most individuals in a school class may already have caught the disease when a quarantine is issued, and then the quarantine loses some of its effect. For those situations a high percentage must be vaccinated. How high depends on the worst scenario one is willing to accept. If the disease is a little less infectious the strategy works very well.

For diseases that can be detected fast the method with combined quarantine and vaccination is very powerful. We have observed situations where no quarantine is used and the disease has spread to more than 99% of the susceptible population, and for the corresponding situation where quarantine is used the disease has reached less than 10% of all susceptible. In the case with $\delta = 1/3$, $\lambda_G = 1$ and $\lambda_L = 1/10$ we observe that for every 20% vaccinated the proportion of infected susceptible seem to decrease with roughly 50%, which is a very good effect.

To summarize we have seen that often an active response to infections, here in the form of quarantine, may be more powerful and prevent the spread more than an increased level of vaccination, if the proportion vaccinated is already far from 1.

7 Discussion

It is important to note that what is done in section 4.2.2 when the number of infected upon detection was computed is really approximate, especially since we assumed a constant rate of infection. What happens when an outbreak starts in a class is rather a spread of exponential speed. For the approximation to hold the infection rate must be very slow and the detection rate quite fast so that a constant infection rate is reasonable. In this thesis we on the contrary investigated very infectious diseases, inspired by measles. This might also lead to another problem: an over depletion of susceptible. If the results from this theses are to be applied these aspects should be investigated closer.

The strength of this thesis is how it describes the spreading of a disease among classes in a way that is independent of class sizes and local infection rates. The simulations also give a good indication of the strength of the strategy.

In a deeper study it might be of interest to generalize the results. This by allowing there to be a different number of vaccinated in each class, different class sizes, self-recovery and vaccines not working perfectly. That situation could also be used for populations where the populations where people get infected more or less easily because of other aspects.

Also, to really be able to see the strength of combined vaccination and quarantine theoretically more work should be dedicated to section 4.1.1 to work out an expression for R_0 for the situation where quarantine is not used. One should also try to express the constant infection rate λ_L^* from 4.2.1 in order to be able to use the final size expression from section 4.2.2. The conclusions about the effect of the combined vaccine and quarantine strategy in this thesis are only based on simulations.

In this thesis we in a way have tried to see how few we can vaccinate and still save many from infection by using quarantine. However we have no interest in getting fewer people vaccinated. Herd immunity, vaccinating a big enough proportion for R_0 to equal 1 (Andersson and Britton, 2000, p. 188), is the best way to prevent the spread of a disease. It may also be that it is cheaper, quarantines cost a lot; we need people to diagnose the children and also we need parents to stay home from their jobs to take of their children. For long lasting diseases this might have a high cost. The cost aspect should be considered before the strategy in this thesis is considered to better than something else. What we have investigated in this thesis is rather an alternative for when it is problematic or not possible to vaccinate

enough individuals to obtain herd immunity or what to do in situations where it is too late to vaccinate people, like when an outbreak is already ongoing.

8 References

Alm and Britton. 2017. Stokastik. First edition. Liber AB.

Andersson and Britton. 2000. Stochastic Epidemic Models and their Statistical Analysis. Springer.

Ball, Knock and O'Neill. 2015. Stochastic epidemic models featuring contact tracing with delays. Mathematical Biosciences. Volume 266. Pages 23-35.

Corless et al. 1996. About the Lambert W function. Advances in Computational Mathematics. Volume 5. Pages 329-359.

European Centre of Disease Control (ECDC). 2019. Measles. https://-ecdc.europa.eu/en/measles (retrieved 150419)

The New Scientist. 2019. Italy bans unvaccinated children from schools after measles outbreak. https://www.newscientist.com/article/2196534-italy-bans-unvaccinated-children-from-schools-after-measles-outbreaks/ (retrieved 150419)

Ross. 2014. Introduction to Probability Models. Eleventh edition. Elsevier.

Trapman and Bootsma. 2009. A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection. Mathematical Biosciences. Volume 219. Issue 1. p 15-22.

The World Health Organization (WHO). 2019. Rada Akbar. Ten threats to glabal health. https://www.who.int/emergencies/ten-threats-to-global-health-in-2019 (retrieved 150419)

The World Health Organization (WHO). Vaccines. https://www.who.int/-topics/vaccines/en/ (retrieved 150419)