Stockholms universitet

Some schools impose measles vaccination as admission requirement. Is it for the better or worse?

Lai Mei Yip Lundström

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2020:12
Matematisk statistik
Juni 2020

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2020:12**
http://www.math.su.se

Some schools impose measles vaccination as admission requirement. Is it for the better or worse?

Lai Mei Yip Lundström*

June 2020

Abstract

This thesis analyses the epidemic of the highly infectious measles in a large school children population under the premise that some schools impose measles vaccination as an admission criteria. We develop a model framework for which we can use to analyse the effects of such a measure, and its extent, on important parameters of an epidemic such as the basic reproduction number and the final outcome.

Based on the standard SIR (Susceptible, Infected, Removal) model but with two levels of mixing, a school children population is divided into classes of the same size and a proportion of these classes admit ONLY vaccinated children, leaving the burden of unvaccinated children to be shared by those classes that do not have this admission criteria, according to some distribution. These classes, in turn, can have varying numbers of unvaccinated students. Each student, regardless of vaccination status, makes contact with students who are in the same class (*local* contacts) and also with those who are not in the same class (*global* contacts), which give rise to the two levels of mixing. However, only an unvaccinated student who is contacted by an infectious student can be infected and the infection is immediately. All infectious students who recover (or die) from the virus attain lifetime immunity and can never be infected again.

Results of our analysis show that, in general, not having vaccination admission criteria or increasing the proportion of classes that admit unvaccinated students helps to bring down the basic reproduction number and the proportion of ultimately infected. The extent of this positive effect, however, is contingent on the level of vaccination coverage, how virulent the strain of virus is and how often students make global contacts.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: lalu6398@student.su.se. Supervisor: Pieter Trapman.

Acknowledgement

I would like to express my immense gratitude to my supervisor, Pieter Trapman, for being the kindest and most patient thesis supervisor one can ever wish for. No question is ever too trivial for Pieter to answer.

I dedicate this thesis to my loving and supportive family of $1 + 3$.

Finally, spread love, not virus.

Sammanfattning

Det här examensarbetet analyserar en epidemi i en stor befolkning skolbarn med det mycket infektiösa mässlingviruset, under förutsättningen att vissa skolor inför mässlingsvaccination som antagningskrav. Vi utvecklar ett modellramverk som kan användas för att analysera effekterna av en sådan åtgärd, och dess omfattning, utifrån viktiga parametrar för en epidemi såsom grundläggande reproduktionsnummer och antalet smittade.

Baserat på standard SIR-modellen (mottaglig, infekterad, borttagen), men med två nivåer av blandning, delas en skolbarnspopulation in i klasser av samma storlek. Och en del av dessa klasser tillåter ENDAST vaccinerade barn, vilket lämnar bördan för ovaccinerade barn till de klasserna som inte har detta antagningskrav, enligt en viss fördelning. Dessa klasser kan i sin tur ha olika antal ovaccinerade studenter. Varje student, oavsett vaccinationsstatus, tar kontakt med elever som är i samma klass (*local* kontakter) och även de som inte är i samma klass (*global* kontakter), vilket ger upphov till två nivåer av blandning. Men bara en ovaccinerad student som kontaktas av en smittsam student kan smittas och infektionen är omedelbar. Alla smittsamma studenter som återhämtar sig (eller dör) av viruset uppnår livstidsimmunitet och kan aldrig smittas igen.

Resultaten av vår analys visar att inte ha detta antagningskrav eller öka andelen klasser som tillåter ovaccinerade studenter kan minska det grundläggande reproduktionsantalet och andelen slutligen infekterade. Omfånget av dess positiva effekt beror emellertid på vaccinationstäckningsnivån, hur smittsam den nuvarande virusstammen är och hur ofta elever tar globala kontakter.

Contents

1 Introduction

Many potentially life-threatening infectious diseases, such as measles, polio and tetanus, that have historically erupted into major epidemics and killed many were brought under control, thanks to concerted vaccination programmes. According to World Health Organization (WHO) [21] [22], before widespread vaccination, measles, in particular, have had major outbreaks approximately every 2 to 3 years and caused an estimated 2.6 million deaths each year. Between 2000 and 2018, measles vaccination has prevented an estimated 23.2 million deaths. However, in recent years, measles has seen resurgence in some developed countries, such as France, Italy and the U.S.. The Center for Disease Control and Prevention in the U.S. [19] reported that 2019 registered the greatest number of cases in the country since 1992 and most of the people infected were not vaccinated against measles. Why do people not want to vaccinate themselves or their children against potentially deadly viruses, such as measles, when there is one readily available? Vaccine hesitancy.

Vaccine hesitancy is on the rise. According to an article published in The Lancet Child & Adolescent Health [20], vaccine hesitancy is reported in more than 90% of the countries in the world. In Europe, vaccination coverage for measles is below 95% in many countries. There is a genuine concern among health authorities that vaccine hesitancy is leading to loss of herd-immunity, which was built up through mass immunisation.

In order to reverse the loss of herd-immunity, some countries, like Italy, Germany and the U.S., have legislated vaccination and children can be refused school placements if their parents have chosen voluntarily not to vaccinate them. Yet in other countries, like Sweden, vaccination is only strongly encouraged but not compulsory. In the case where legislation is not in place or where exemptions are liberally applied, schools may decide to take things into their own hands and impose vaccination as a school admission requirement. This means that those schools that do not have the vaccination admission criteria will have a high proportion of unvaccinated students. How will this affect the spread of a highly infectious disease, such as measles, in a large population of school children? What is the probability of an outbreak? What is the proportion of unvaccinated students who will be infected? Modelling a measles outbreak in case of such a situation is the motivation of this thesis.

Hence, this thesis attempts to answer the questions above by developing a model framework to analyse a measles outbreak in a large population of school children, under two scenarios. One, no schools impose vaccination as an admission criteria and all schools take in unvaccinated children; two, some schools have imposed vaccination as an admission criteria while the rest do not, leading to all unvaccinated children distributed only amongst the latter.

There is a rich body of literature in mathematical modelling of the spread of infectious diseases. One of the most cited literature is the deterministic model proposed by Kermack and McKendrick [11]. In this model, the population is assumed to be homogeneous mixing. Another famous work is a book written by Bailey [5], which includes his earlier work on the theory of epidemics. Bailey's work is also based on a homogeneous mixing population. Nowadays, however, the focus is on models that incorporate heterogeneities. Andersson and Britton [2] have written a set of comprehensive lecture notes on stochastic epidemic models and their statistical analysis, which include a very useful chapter on models that allow heterogeneity in the population. Further, Ball, Mollison and Scalia-Tomba [7] have developed a comprehensive model framework for epidemics with removal in populations that mix at two levels. In addition, a paper written by Britton, Kypraios and O'Neil [8] is also of interest as they construct a stochastic epidemic model with three levels of mixing and apply their methods on a real measles outbreak dataset. There is, in fact, a plethora of research work done on mathematical modelling of infectious diseases from different perspectives and angles, some of which are applied on the case of measles. This thesis finds a fresh angle by specifically studying the case where a proportion of schools impose vaccination as an admission criteria, as opposed to the current real-life situation where either all schools have vaccination criteria or none has it.

In order to be as close to reality as possible and yet still keep the mathematics tractable, we use a stochastic model that allows two levels of mixing in the population. The model we choose to work with is a generalisation of the standard SIR model with two levels of mixing put forth by Ball *et al* [7]. The standard SIR model is a simple stochastic model for epidemics where S.I.R stands for "Susceptible"(S), "Infectious"(I) and "Removed"(R). One of the assumptions of this model is that there is homogeneous mixing in the population, that is, the social structure is flat. Ball *et al* [7] generalises this model by allowing populations to mix at two levels, namely *local* and *global*. It is also known as the *household model*. It is intuitive that the rate at which individuals come into contact with one another locally will be higher than the rate at which they make contact with individuals outside their social settings. A *local* infection refers to the event of an infectious individual infecting someone within the same household, while a *global* infection refers to the event of an infectious individual infecting an individual outside the household. A detailed description of the model and the assumed social structure in the population will be provided in Section 3.

In Section 3, we delve into the theoretical aspects of how an epidemic spreads like a branching process at the initial stages of an outbreak. We also give detailed explanations of both the standard SIR model and the household model.

Section 4 is dedicated to deriving expressions for the basic reproduction number and the final outcome of an outbreak with two levels of mixing.

More significantly, we take into consideration the proportion of classes that admit unvaccinated students and see how this plays a role in these quantities.

In Section 5, we evaluate the expressions derived in Section 4 by varying the main model parameter, which is the proportion of classes that admit unvaccinated individuals, as well as other parameters such as the level of vaccination coverage in the population and global contact rates.

We conclude our findings in Section 6 and discuss the limitations of this thesis in Section 7.

2 Definitions of main parameters and quantities

For convenience of reference, we collect the main parameters and quantities in this thesis into a list here and provide their definitions.

- Basic reproduction number : average number of initially susceptible individuals infected by an infectious individual
- Local epidemic: epidemic within a household
- Global epidemic: epidemic of all households, which is the same as one involving the whole population
- Final outcome: proportion of initially susceptible individuals who are ultimately infected in the global epidemic
- Local contact rate: contact rate between two given individuals within the same household
- Global contact rate: contact rate between two given individuals belonging to different households
- Super individual: a household or in the context of this thesis, a class of students
- N: population size
- λ_L: local contact rate
- $\lambda_G/(N-1)$: global contact rate
- R_*: basic reproduction number of an epidemic with two levels of mixing.
- α: proportion of school classes that admit unvaccinated students
- γ: level of vaccination coverage
- τ: final outcome

3 Theoretical aspects

3.1 Branching process and epidemic

In the initial stages of an epidemic outbreak, it is reasonable to expect that majority (if not the entire) of the population is susceptible to the virus/disease and hence it is highly likely that the individuals whom the initial infectious individuals make contact with have not been infected yet. This motivates the branching process approximation of an epidemic in its early stages.

Using Ross [18] as the main reference, we define the branching process. Consider an initial population of a finite number of individuals with independent and identically distributed life spans I. They form the ancestor generation. During the life span of an ancestor individual, he produces offsprings following a Poisson process at the rate of λ. We call his offsprings the next generation and these offsprings, in turn, have also the same independent and identically distributed life spans and produces offsprings of their own following the same Poisson process, thereby creating yet another generation, and so on. We assume that all individuals make independent decisions of how many offsprings they produce regardless of their life spans, that is, all Poisson processes are independent of one another and also independent of the life spans. See Figure 1 for an illustration of the branching process.

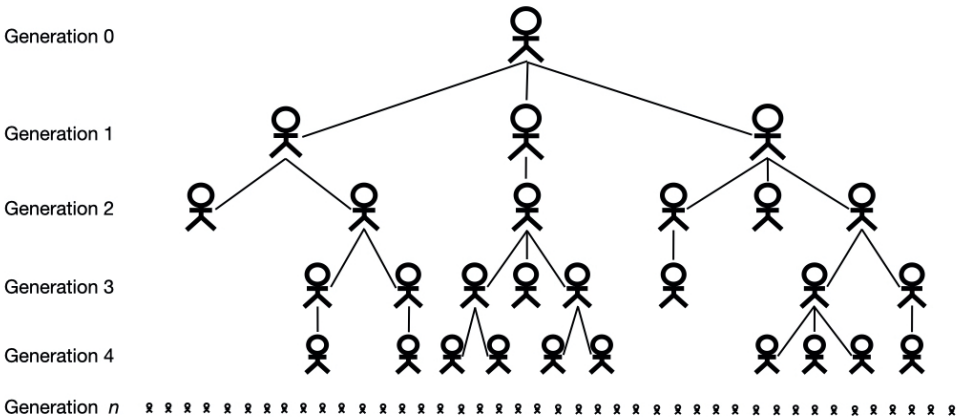

Figure 1: Illustration of the branching process of a single ancestor

Relating the above to an epidemic process, consider a finite number of initial infectious individuals in a large population of individuals where majority of them are susceptible to the virus/disease. These *ancestor infectious individuals* have independent and identically distributed infectious periods I. During their infectious periods, they *produce* the next generation of in-

fectious individuals following a Poisson process at rate λ by making contact with them. This new generation of infectious individuals, who inherit the same infectious period distribution, then go on and produce the next generation of infectious individuals through the same Poisson process by making contact with the rest of the susceptible individuals in the population, and so on.

To illustrate, let X_n denote the number of infectious individuals in the nth generation of the epidemic. It is easy to see that the branching process gives us a Markov Chain of $\{X_n, n = 0, 1, 2, ...; X_n \in \mathbb{N}\}$, since the number of infectious individuals in a new generation depends only on the number of infectious individuals in the previous one.

Suppose we have 1 initial infectious individual, i.e. $X_0 = 1$, we want to know what is the expected number of infectious individuals at the nth generation, i.e. $E[X_n]$. We begin by first taking note that the the average number of offsprings produced by an infectious individual is $\lambda * \iota$, where $E[I] = \iota$. This is the case since $\lambda * \iota$ is the average number of infectious individuals infected by an infectious individual during his infectious period, given that the Poisson process of offspring reproduction and infectious period are mutually independent. Hence

$$\begin{aligned} E[X_n] &= E[E[X_n|X_{n-1}]] \\ &= E[X_{n-1} * \lambda * \iota] \\ &= \lambda * \iota * E[X_{n-1}]. \end{aligned}$$

Given $E[X_0] = 1$, we get

$$\begin{aligned} E[X_1] &= \lambda * \iota \\ E[X_2] &= \lambda * \iota * E[X_1] = (\lambda * \iota)^2 \\ &\vdots \\ E[X_n] &= \lambda * \iota E[X_{n-1}] = (\lambda * \iota)^n \end{aligned}$$

So the average number of infectious individuals in the nth generation is directly dependent on the average number of infectious individuals produced by a given infectious individual and it is easy to see that $(\lambda * \iota)^n \to 0$ when $\lambda * \iota < 1$. Let us call this quantity the reproduction number of an infectious individual. In other words, the expected number of infectious individuals will tend towards zero as $n \to \infty$ if the reproduction number of an infectious individual is less than one.

We shall see how this reproduction number is tied to the probability that the epidemic dies out or explodes.

Let τ denotes the probability that an epidemic started by one initial infectious individual explodes into an outbreak, which means we have $1-\tau$ as

the probability that the number of infectious individuals in such an epidemic diminishes to zero, that is, the epidemic dies out. In the case where $\lambda * \iota < 1$, $1 - \tau = 1$. We illustrate this by using the argument in Ross [18], Chapter 4.7, which essentially is derived using Markov's Inequality [18][Proposition 2.6],

$$\begin{aligned}(\lambda * \iota)^n = E[X_n] &= \sum_{j=1}^{\infty} jP(X_n = j) \\ &\geq \sum_{j=1}^{\infty} 1 * P(X_n = j) \\ &= P(X_n \geq 1), \quad j \geq 1.\end{aligned}$$

So if $\lambda * \iota < 1$, then $E[X_n] \to 0$, which implies $P(X_n \geq 1) \to 0$. This, in turn, implies $P(X_n = 0) = 1 - \tau \to 1$. It can be shown that $P(X_n = 0) = 1$ even when $\lambda * \iota = 1$.

In the case where $\lambda * \iota > 1$, then the epidemic will either die out or explode into an outbreak. Then the probability that the epidemic started by the initial infectious individual dies out (i.e. the branching process of the ancestor infectious individual goes extinct) is,

$$\begin{aligned}1 - \tau &= \sum_{j=0}^{\infty} P(X_n = 0 | X_1 = j) * P_j \\ &= \sum_{j=0}^{\infty} (1 - \tau)^j * P_j,\end{aligned}$$

where j is the number of infectious individuals *produced* by him and P_j denotes the probability that an infectious individual will infect j susceptible individuals during his infectious period. In other words, P_j is the mixed Poisson distribution of the number of infectious individuals produced by an infectious individual. We can see that $1 - \tau = 0$ is always a solution to the equation above. However, one can show that when $\lambda * \iota > 1$, besides zero, the other $1 - \tau$ value is the smallest positive solution that satisfy the equation above. We refer interested readers to Jagers [10] for its proof.

Take note that $P(X_n = 0 | X_1 = j) = (1-\tau)^j$, which follows the argument that the epidemic dies out if and only if all of the j branches started by the j infectious individuals in the first generation eventually die out. By the branching process analogy of an epidemic, every infectious individual *produces* the next generation of infectious individuals independently of one another and every infectious individual is the *ancestor* infectious individual of the branch of infectious individuals started by him. And recall that

we have $(1-\tau)$ as the probability that an epidemic started by an initial infectious individual dies out. Hence, we get $P(X_n = 0|X_1 = j) = (1-\tau)^j$.

Rearranging the equation above, we get

$$P(\text{large outbreak}) = \tau = 1 - \sum_{j=0}^{\infty}(1-\tau)^j * P_j,$$

given 1 initial infective. If there are m initial infectives, then the probabilities that the epidemic will die out and explode are $(1-\tau)^m$ and $1-(1-\tau)^m$, respectively.

We use the random graph representation of an epidemic in Section 3.2.3 to show that τ is also the proportion of initially susceptible individuals who are ultimately infected in the epidemic outbreak.

3.2 Epidemic models

In this section, we present detailed descriptions of the standard SIR model and in particular, the household model. In addition, we analyse an epidemic within the framework of random graph theory, and see how certain parameters of interest can be easily derived, especially with a constant infectious period.

3.2.1 The Standard SIR Model

As mentioned in the introduction, a simple model that describes the spread of an infectious disease is the standard SIR model. Using Andersson and Britton [2] as the main reference, we provide a definition of this model. Without loss of generality, consider an epidemic that begins in a population consisting of 1 infectious individual and $N-1$ susceptible individuals. Every infectious individual has independent and identically distributed infectious period I that follows a certain distribution, with mean ι. It is assumed that the population is closed, homogeneous as well as homogeneous mixing. Under the assumption of homogeneous mixing, each infectious individual independently makes contact with a given individual following a homogeneous Poisson process at a rate of $\lambda/(N-1)$. Also, since the population is assumed to be homogeneous, they do not have varying susceptibility, and so if a given individual is contacted by an infectious individual and he is still susceptible, he becomes immediately infected and is able to infect other remaining susceptible individuals. Once an infected individual passes his infectious period (or die), he is assumed to attain lifetime immunity from the virus, that is, considered to have "recovered" and therefore, exit the epidemic. The epidemic ends when all infectious individuals have "recovered" and exited the epidemic.

We can see that the basic reproduction number, which is the average number of initially susceptible individuals infected by an infectious individual is

E[number of initially susceptible individuals infected by an infectious individual]
$=E$[number of contacts made by an infectious individual during his infectious period]

[independence between contact process and infectious period]

$$=\frac{\lambda}{N-1} * (N-1) * \iota = \lambda * \iota.$$

The standard SIR model, while mathematically convenient to model and easy to understand, is quite distant from reality due to many of its underlying assumptions. The model we discuss next is called the *household model*, which is built upon the SIR model with violation to one of its assumptions, that of homogeneous mixing in the population. This is also the model that we use throughout the rest of this thesis.

3.2.2 The Household Model

As mentioned in the introduction and Section 3.2.1, the household model allows two levels of mixing in the population, as opposed to homogeneous mixing in the standard SIR model. In real-life, our society has a complex social structure and the extent and speed of the spread of an infectious diseases depend very much on it. People belong to various social groups, such as families, schools, workplaces, hobby groups and so on, and individuals in the same social group make contact with one another more often than with individuals outside the social group. In the context of this thesis, it is intuitive that students make more frequent contacts with their classmates than with schoolmates as well as students from other schools.

Hence, an infectious student makes contact with other students in the school children population following two independent homogeneous Poisson processes:

- λ_L: local contact rate (per day), that is, the rate at which an infectious student contacts a given classmate in a day;
- $\lambda_G/(N-1)$: global contact rate (per day), that is, the rate at which a student contacts another student in the population of school children that consists of N students, regardless of vaccination status. We express the global contact rate in this form in order to keep λ_G, the mean number of global contacts, constant even as $N \to \infty$. While most literature on epidemic modelling regards $\lambda_G/(N-1)$ as the global contact rate between a given infectious individual and a given susceptible individual and N as the number of initial susceptible individuals, we do

not do so here because one of our investigation parameters is the level of vaccination coverage in the school children population. Hence, by defining λ_G and N our way, we are saying that the average number of global contacts made by a given student is kept constant, even as the population changes, but of these global contacts, how many of them are unvaccinated is controlled by the level of vaccination coverage in the population.

By the definition of global contact rate above, an infectious student's global contacts refer not only to students outside of his class, but also his classmates. Following the rationale of Andersson and Britton [2], however, the global contact rate of an infectious student with his classmates is negligible, since it is intuitively much smaller than the local contact rate. Hence, if an infectious student were to infect his classmate, it is be due to local contact and not global contact. We treat schoolmates as global contacts by assuming that students from different classes rarely interact. Although this is rather unlikely in reality, we do this for mathematical convenience.

We also assume that:

- infectious period I is fixed
  According to WHO [22], a measles infectious individual is generally infectious 4 days before and 4 days after onset of rashes, and is most likely to infect others before the onset of rashes. A logical explanation is that after onset of rashes, it becomes a clear sign of disease and the infectious individual will highly likely avoid contact with other individuals. In the context of this thesis, we assume students will stop going to school once rashes develop and they will stay at home until recovery, essentially giving an infectious student 4 days to infect other students.

- the size of a class is the same across all schools in the population and assumed to be 20 students. While the size of a class is kept fixed at 20, the number of unvaccinated students can vary between 0 to 20 and we take note that an epidemic does not involve those classes that do not have any unvaccinated students, which follows from the next assumption.

- vaccinated students have 100% protection from measles.

The approach in this thesis is to first look at the spread of measles within a class of students through local contact amongst classmates. The infected students then make global contact with students in other classes, thereby creating local epidemics in other classes. And then we "zoom out" and look at the epidemic of school classes.

During the early stages of the epidemic, all of the global contacts made by infectious students will highly likely belong to different classes when the total number of school classes is large. Each of these global contacts will then start an epidemic within his own class, he and his infected classmates then pass on the virus to other classes in the same way. When we regard every class as a super individual, the initial stages of the epidemic amongst the super individuals can be approximated by the branching process described in Section 3.1 and the epidemic between these individuals follows the standard SIR model with homogeneous mixing. From now on, we will use super individual and a class of students interchangably.

Let μ denote the average number of infected students in a class (including the initial infectious student) and γ the proportion of school children who are unvaccinated. These μ students then, on average, collectively make global contact with $\mu * \lambda_G * \gamma$ unvaccinated students on a daily basis. Since the infectious period is fixed at 4 days, this means that the basic reproduction number of an infectious super individual is

$$R_* = 4 * \mu * \lambda_G * \gamma.$$

By the branching process approximation of an epidemic that we have presented in Section 3.1, a large outbreak is possible if and only if $R_* > 1$.

3.2.3 Random graph interpretation of an epidemic

In order to relate to the context of this thesis, but without loss of generality, we consider a population made up of a total of N unvaccinated and vaccinated individuals. Then, consider the standard SIR epidemic with one level of mixing in such a population and a contact rate of $\lambda_G/(N-1)$ between two individuals.

What we have here is the same as a household epidemic where the household size is 1. We represent all the individuals in the population by vertices of a graph. However, only $N\gamma$ of them are "active", in the sense that they are part of the epidemic since those who are vaccinated will never be infected. With a stochastic infectious period I, say a given initially susceptible individual i becomes infected, then a given susceptible individual j will be infected by him with probability $\lambda_G I_i/(N-1) \approx 1-\exp[-\lambda_G I_i/(N-1)]$, which follows from the first order Taylor approximation of $\exp[-\lambda_G I_i/(N-1)]$. Now, let us represent this potential contagion from i to j by a directed arrow, which means we draw an arrow from i to j with probability $1-\exp[-\lambda_G I_i/(N-1)]$. Hence, as soon as an arrow has been established, it means that i will contact j during his infectious period, if the former ever gets infected, and if j is still susceptible, he will be infected. By the same analogy, if an arrow is drawn from j to i, this means that if j ever gets infected, then he will infect i during his infectious period. We see that these arrows are not drawn independently due to the random

variable I. But if I is constant, that is, infectious period is fixed at some constant a, then the random graph representation becomes much simpler and can be explained by the Bernoulli random graph $\mathcal{G}(\tilde{N}, p)$. A simple and easy to understand explanation of Bernoulli random graph is given in Andersson and Britton [2], Section 7.2.

Essentially, this means that we can now draw an edge, not arrows, between two vertices of the graph with the same probability $p = 1 - \exp[-\lambda_G a/(N-1)]$. This is because the arrows can now be drawn independently of each other and since both arrows have the same probability, it suffices to have one undirected edge between the two vertices. Also, any two vertices are said to belong to the same component if and only if there exists a path of edges between them. It is easy to see that the standard SIR epidemic with one level of mixing consisting of $N\gamma - 1$ initial susceptible individuals and 1 infectious individual will then have one component where all the ultimately infected individuals are linked to the initial infectious individual by a path of edges between pairs of unvaccinated individuals. See Figure 2 for an illustration of such a contact graph. This component is made up of active vertices connected by solid edges, originating from the large dark circle.

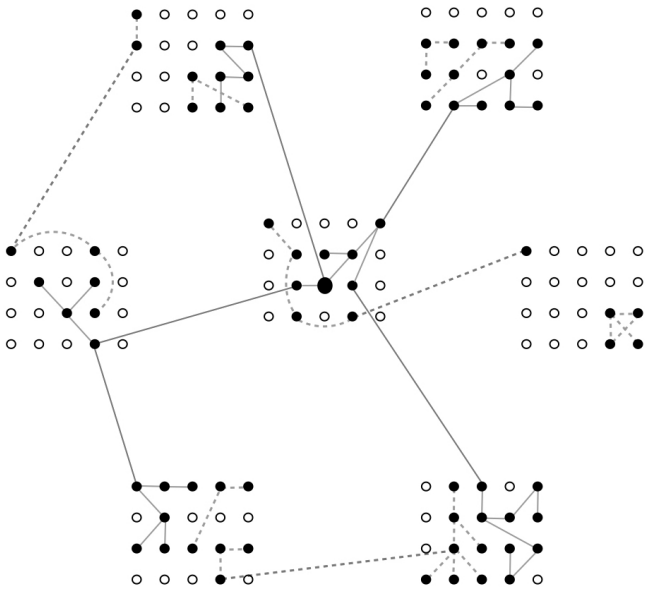

Figure 2: Contact graph with one level of mixing. The black circles are unvaccinated students while the white ones are. The large black circle is the initial infectious individual. The solid edges denote contacts that were activated, that is, virus transmission has occurred, while the dashed lines denote potential contacts but were not activated because neither the initial infectious individual nor those infected by him and his offsprings have edges connected to them.

We can now apply a well-known result in random graph theory, encapsulated as Theorem 7.1 in Andersson and Britton [2], Section 7.2:

Consider the $\mathcal{G}(\tilde{N}, p)$ *random graph model. Assume* $p = \beta/\tilde{N}$*, as* $\tilde{N} \to \infty$*. If* $\beta \leq 1$*, then a vertex chosen at random will belong to a component of size* $O(1)$*. On the other hand, if* $\beta > 1$*, then the relative size of the largest component converges in probability to some constant* C *strictly between* 0 *and* 1*, as* $\tilde{N} \to \infty$*. Also, a randomly chosen vertex will belong to this large component with probability* C *and it will belong to a component of size* $O(1)$ *with probability* $1 - C$*.*

Applying this theorem to the epidemic above and taking note that $p = 1 - \exp[-\lambda_G a/(N-1)] \approx \lambda_G a/(N-1)$, which again follows from first-order approximation of $\exp[-\lambda_G a/(N-1)]$, and $\tilde{N} = N\gamma$, this gives us $\beta \approx \lambda_G a\gamma$. Recall earlier in Section 3.2.1 that $\lambda * \iota$ is the basic reproduction number in a standard SIR epidemic with independent and identically distributed infectious periods I that has expectation ι. Replacing ι with the constant a and taking into account that γ of the population is unvaccinated, the basic

reproduction number becomes $\lambda_G a\gamma$, which means β is also the basic reproduction number. Then, following the theorem, if the basic reproduction number $\lambda_G a\gamma$ is greater than 1, the relative size of the largest component, which is the proportion of initially susceptible individuals ultimately becoming infected, is the probability that a uniformly chosen susceptible will belong to this component, which in turn, is the asymptotic probability of a large outbreak.

We go back now to the epidemic with two levels of mixing. Assuming that there is initially one infectious student, using the same notation as in Section 3.1, let τ be the probability of a large outbreak started by this uniformly chosen unvaccinated student. See Figure 3 for a graphical illustration of the epidemic. The configuration is exactly the same as Figure 2 except now that the black circles are unvaccinated students being put into classes of the same size demarcated by the grey circles. For example, the circle on the right (middle) is a class with five unvaccinated students. Take note that with two levels of mixing, there are two types of edges for virus transmission: the blue ones denote transmission through local contact while the red ones denote transmission through global contact.

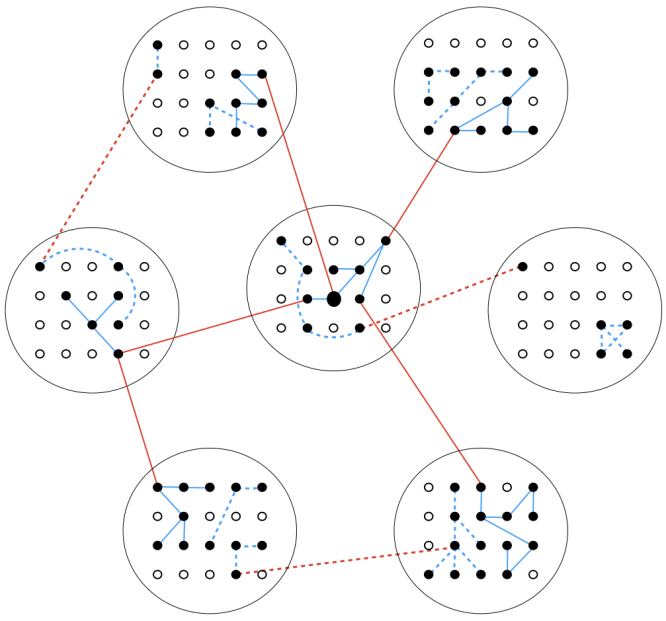

Figure 3: Contact graph with two levels of mixing. The large black circle is the initial infectious student. The grey circles denote boundaries of classes. Although class size is fixed, every class can have a different number of unvaccinated students. Blue edges denote local contact; red edges denote global contact.

We see that even though there are two levels of mixing, the component remains unchanged and is still made up of initially unvaccinated students linked to the initial infectious student by their respective paths of edges. Take note that there are now two different probabilities of infection tied to the edges, with the red ones having probability of global infection while the blue ones have probability of local infection.

However, we also see that the red edges, in fact, denote infectivity emanating from the class of the initial infectious student, which is analogous to infectivity emanating from an infectious individual, and we recall that the epidemic of the super individuals (i.e. classes of students) is assumed to follow the standard SIR model. This means that, under the assumption that we have a large school population and hence a large number of classes, if the average number of initially susceptible classes infected by the initial infectious class is greater than 1, then using the random graph interpretation, the relative size of the largest component is the probability that a uniformly chosen unvaccinated student belongs to it, which in turn, is the asymptotic probability of a large outbreak τ. This average number of initially susceptible classes infected by the initial infectious class is precisely R_*.

But computation of R_* is a little more cumbersome since different classes can have a different number of unvaccinated students, that is, we have to take into consideration the size-biased distribution. In the next section, we show how we derive expressions for R_* and τ. We also see that with the random graph interpretation, it becomes relatively easy to compute the risk of infection, that is, the probability that a unvaccinated student who belongs to a class with u unvaccinated students will become infected.

4 Basic Reproduction Number and Final Outcome

This section is devoted to deriving expressions for the important quantities of an epidemic, namely the basic reproduction number of an infectious super individual R_* and the proportion of initially unvaccinated students who are ultimately infected in the global epidemic τ. As mentioned earlier, the methods we use to derive these quantities give us a relatively easy way to derive the risk of infection of a given unvaccinated student who belongs to a class with u unvaccinated students, which can be useful to help identify the maximum number of unvaccinated students a class should have given an acceptable risk level.

4.1 Scenario I: No vaccination admission criteria

As mentioned in the introduction, due to some schools imposing vaccination as an admission criteria, unvaccinated children can only be admitted to schools that do not have this criteria. Let α denote the proportion of schools that do not have the said admission criteria and γ is the proportion of unvaccinated students in the population of school children. Assuming that all schools have the same number of classes, α is also the proportion of total school classes that do not have the admission criteria. Hence, these classes will bear the burden of unvaccinated students according to some distribution, which we discuss later in this section.

In this scenario, all schools admit both vaccinated and unvaccinated children, which means we have $\alpha = 1$.

In order to compute R_*, we need to know μ, which is the average number of infected students in a class including the initial infectious student(s). To begin, let $P_k^{n,m}$ denote the probability that k unvaccinated students (excluding the initial infectious student(s)) in a class with n initial unvaccinated students and m initial infectious students will ultimately be infected, that is, the $P_k^{n,m}$ is the distribution of the final size of the local epidemic.

The distribution of the final size of the local epidemic is, in general, difficult to compute even for the simple Reed-Frost epidemic model, as shown in Andersson and Britton [2], Section 1.2. Ball *et al* [7] give a version using results derived by Lefèvre and Picard [12] who use Gontcharoff polynomials, as well as a version by Addy, Longini and Haber [1] where the final size probabilities can be determined recursively from a triangular system of linear equations. Andersson and Britton [2] has also given a version that is similar to Addy *et al*'s, but uses the Wald's identity for epidemics derived by Ball [6]. We decided to use the version derived by Andersson and Britton, whose proof is relatively easy to follow and with a constant infectious period, becomes very straightforward to solve. Encapsulated as Theorem 2.2 in Andersson

and Britton [2], Section 2.4, we compute the $P_k^{n,m}$s recursively as follows:

$$\sum_{k=0}^{l} \binom{n-k}{l-k} \frac{P_k^{n,m}}{E[\exp\{-4\lambda_L(n-l)\}]^{k+m}} = \binom{n}{l}$$
$$\iff \sum_{k=0}^{l} \binom{n-k}{l-k} \frac{P_k^{n,m}}{\exp\{-4\lambda_L(n-l)\}^{k+m}} = \binom{n}{l},$$

for $0 \geq l \geq n, 0 \geq k \geq n$. The expression $E[\exp\{-4\lambda_L(n-l)\}]$ is the Laplace transform of the constant infectious period of 4 days with argument $\lambda_L(n-l)$.

Recall that the number of unvaccinated students in such a class follows a certain distribution, which we will now derive. Assume that the number of school classes is large (which is implied by the large school children population), let s_u denote the asymptotic proportion of classes that have u unvaccinated students. Since all classes are of the same size (recall that each class has 20 students), this implies that s_u has the following asymptotic distribution, that is, the probability that a randomly chosen class is of size u is

$$s_u = \frac{\binom{20}{u} \gamma^u (1-\gamma)^{20-u}}{\sum_{u=1}^{20} \binom{20}{u} \gamma^u (1-\gamma)^{20-u}}, \quad 1 \leq u \leq 20,$$

where the probability of getting an unvaccinated student is the proportion of unvaccinated students at large because all classes accept unvaccinated students. We need to have a normalizing constant since the support set starts from 1 and not 0. Let us denote it by c and we will soon see that this normalizing constant does not contribute to the computations.

Now we initiate the epidemic with a uniformly chosen unvaccinated student. By the law of large numbers, the probability that this student belongs to a class with u unvaccinated students is

$$\pi_u = \frac{u * s_u}{\sum_{u=1}^{20} u * s_u}.$$

Recall that $s_u = \frac{1}{c} \binom{20}{u} \gamma^u (1-\gamma)^{20-u}$, this gives the distribution of π_u

as

$$\pi_u = \frac{u * \frac{1}{c}\binom{20}{u}\gamma^u(1-\gamma)^{20-u}}{\sum_{u=1}^{20} u * \frac{1}{c}\binom{20}{u}\gamma^u(1-\gamma)^{20-u}}$$
$$= \frac{u * \binom{20}{u}\gamma^u(1-\gamma)^{20-u}}{\sum_{u=1}^{20} u * \binom{20}{u}\gamma^u(1-\gamma)^{20-u}}$$

[We can start the summation in the denominator from zero, since $0*\binom{20}{u}\gamma^u(1-\gamma)^{20-u}$ does not contribute to the sum. But then this becomes the expectation of a binomial distribution with parameters 20 and γ!]

$$= \frac{u * \frac{20!}{(u!)(20-u)!}\gamma^u(1-\gamma)^{20-u}}{20*\gamma}$$
$$= \frac{(20-1)!}{(u-1)!(20-u)!}\gamma^{u-1}(1-\gamma)^{20-u}$$
$$= \binom{20-1}{u-1}\gamma^{u-1}(1-\gamma)^{20-u}, \qquad 0 \le u-1 \le 20-1.$$

It is easy to see that this is also the distribution of the number of unvaccinated students left in a class after one of them has been globally infected. In other words, $u-1$ is a realisation of the binomially distributed random variable $U-1$ that denotes the number of unvaccinated students left in a class after one of them is newly infected from outside.

Combining this with the distribution of the final size of the local epidemic, we are now ready to give an expression for the average number of infectious students in a local epidemic (i.e. including the initial infectious student),

$$\mu = \sum_{u-1=0}^{20-1} \pi_u \Big(1 + \sum_{k=0}^{u-1} k * P_k^{u-1,1}\Big).$$

Before we give the final expression for R_*, recall that $\lambda_G/(N-1)$ is the global contact rate between two given students, regardless of vaccination status. Since γ is the proportion of unvaccinated student, then the global contact rate between a given infectious student and a given unvaccinated student is $\lambda_G * \gamma/(N-1)$.

Finally, the basic reproduction number of a super individual, given 1

initial infective, is

$$R_* = 4 * \lambda_G * \gamma * \sum_{u-1=0}^{20-1} \pi_u \Big(1 + \sum_{k=0}^{u-1} k * P_k^{u-1,1}\Big).$$

4.1.1 Probability of a large outbreak and final outcome of a global epidemic

Using the same notation as the branching process approximation of an epidemic in Section 3.1, let τ denote the probability of a large outbreak when there is initially one infectious class with one initial infectious student, that is, τ is the probability of a large outbreak started by a uniformly chosen unvaccinated student. Then

$$P(\text{no large outbreak}) = 1 - \tau.$$

Let Z denote the size of the first generation of susceptible super individuals infected by the initial infectious super individual. It is easy to see that this is the number of global infections emanating from the initial infectious super individual (see Figure 3). This is the case because every infectious student in this initially infectious class is assumed to make global contacts that belong to distinct classes, based on the assumption that the number of classes is large.

By the branching process approximation of an epidemic described in Section 3.1, the epidemic dies out if and only if every branch of the first generation of infectious super individuals dies out, that is,

$$1 - \tau = \sum_{z=0}^{\infty} (1 - \tau)^z * P(Z = z),$$

where $P(Z = z)$ is the distribution of the number of initially susceptible super individuals infected by the initial infectious super individual.

To find the distribution of Z, we can condition on the final size of the local epidemic of the initially infectious class plus the initial infectious student. Let K denote the final size of the local epidemic. Now, since the number of global contacts made by a given infectious student during his infectious period of 4 days is Poisson distributed with mean $4 * \lambda_G * \gamma$, then the total number of global contacts made by the infectious students (including the initial infectious student) in the initially infectious class is a sum of Poisson random variables with mean $4 * \lambda_G * \gamma(1 + K)$, that is, we have

$$Z|K \sim Po(4 * \lambda_G * \gamma(1 + K)).$$

Applying this to the expression for τ, we get

$$\begin{aligned}
1-\tau &= \sum_{z=0}^{\infty}(1-\tau)^z * P(Z=z) \\
&= E[(1-\tau)^Z] \\
&= E\Big[E[(1-\tau)^Z|K]\Big] \\
&= E\Big[\exp\{-4\lambda_G\gamma(1+K)\tau\}\Big] \\
&= E\Big[E\big[\exp\{-4\lambda_G\gamma(1+K)\tau\}|U\big]\Big]
\end{aligned}$$

[The inner expectation is the moment generating function of the non-negative, integer-valued variable $1+K|U$, with parameter $-4\lambda_G\gamma\tau$ and U is the number of unvaccinated students in a class. Recall that $U-1$ is the number of unvaccinated students left in the class after one of them has been globally infected.]

$$\begin{aligned}
&= E\Big[\sum_{k=0}^{U-1} e^{-4\lambda_G\gamma(1+k)\tau} * P_k^{U-1,1}\Big] \\
&= \sum_{u-1=0}^{20-1} \pi_u * \Big(\sum_{k=0}^{u-1} e^{-4\lambda_G\gamma(1+k)\tau} * P_k^{u-1,1}\Big) \\
&\iff \\
\tau &= 1 - \sum_{u-1=0}^{20-1} \pi_u * \Big(\sum_{k=0}^{u-1} e^{-4\lambda_G\gamma(1+k)\tau} * P_k^{u-1,1}\Big).
\end{aligned}$$

There are a number of things we can observe here:

- $E[Z] = R_*$

  We can see this by:

$$\begin{aligned}
E[Z] &= E\Big[E[Z|K]\Big] \\
&= E[4\lambda_G\gamma(1+K)] \\
&= 4\lambda_G\gamma E\Big[E[(1+K)|U]\Big] \\
&= 4\lambda_G\gamma E\Big[1+\sum_{k=0}^{U-1} k * P_k^{U-1,1}\Big] \\
&= 4\lambda_G\gamma \sum_{u-1=0}^{20-1} \pi_u\Big(1+\sum_{k=0}^{u-1} k * P_k^{u-1,1}\Big)
\end{aligned}$$

- $1-\tau$ is the probability of a uniformly chosen unvaccinated student completely escapes infection in the global epidemic.

  Consider a class of u unvaccinated students with no initial infectious student. From the global contact rate of $\frac{\lambda_G}{N-1}$, the probability that a given unvaccinated student in this class is infected globally by a given infectious student is approximately $\frac{4\lambda_G\gamma}{\gamma N-1}$ (for a given γ), which gives us $1-\frac{4\lambda_G\gamma}{\gamma N-1} = \exp(\frac{-4\lambda_G\gamma}{\gamma N-1})$ as the probability that a given unvaccinated student escapes global infection from a given infectious student, as $N \to \infty$. Recall that the proportion of initially unvaccinated students being infected in the global epidemic is τ, which means in order for a given unvaccinated student to escape global infection, he needs to escape infection from these $\tau(\gamma N-1)$ infected students. So the probability that a given unvaccinated student escapes global infection is

  $$\exp\left[\frac{-4\lambda_G\gamma}{\gamma N-1}\right]^{\tau(\gamma N-1)} = \exp(-4\lambda_G\gamma\tau).$$

  But this is not enough for him to completely escape infection because if any of his potential contacts gets globally infected, then he will also be infected. We use the random graph representation of an epidemic in Figure 3 to illustrate this point. Let us focus on the grey circle in the upper left corner. The unvaccinated student who was globally infected by the initial infectious student in another class causes four other unvaccinated classmates to be infected because these five students have local contact with one another. Conversely, if it had been another student out of these five who was globally infected, the outcome would have been the same. Hence, in order for a given unvaccinated student to completely escape infection, all of his unvaccinated classmates whom he has contact with must also escape infection. This number is precisely the outcome of the local epidemic k when one of the u unvaccinated students has been globally infected. Hence, the probability that a given unvaccinated student completely escapes infection is $\exp(-4\lambda_G\gamma\tau)^{1+k}$. Averaging it out with the distribution of the final size of the local epidemic and the size-biased distribution, we arrive at the said probability for a uniformly chosen unvaccinated student.

4.1.2 Risk of infection

The task at hand here is to compute the risk of infection of a given unvaccinated student who belongs to a class with u unvaccinated students, after which we can compare it with the overall risk of infection τ or compare it with the risk of infection of a given student who belongs to a class with a different u.

We denote a given unvaccinated student who belongs to a class with u unvaccinated students as a type u student. Consider an epidemic started by a given type u student and we are interested to know the probability of a large outbreak started by him. Hence let τ_u denote the probability of a large outbreak when there is initially one type u infectious student. So we have

$$\text{P(no large outbreak started by type } u \text{ student)} = 1 - \tau_u.$$

Analogous to the computation for τ in Section 4.1.1, let Z_u denote the size of the first generation of susceptible super individuals infected by the initial infectious type u super individual. The epidemic dies out if and only if every branch of this first generation of infectious super individuals dies out. Recall that $1 - \tau$ is the probability that a branch dies out, this gives us the probability that the epidemic dies out, in the event it is started by a type u student, as

$$1 - \tau_u = \sum_{z_u=0}^{\infty} (1-\tau)^{z_u} * P(Z_u = z_u).$$

To find the distribution of Z_u, we condition again on the final size of the local epidemic of the initially infectious type u class plus the initial infectious student. Let K_u denote the final size of this local epidemic. So now the conditional Poisson process of the total number of global contacts made by this initial infectious type u class (or type u super individual) has the following distribution:

$$Z_u | K_u \sim Po(4 * \lambda_G * \gamma(1 + K_u)).$$

Using the results in Section 4.1.1, it is easy to see that this gives us

$$\begin{aligned}
1 - \tau_u &= \sum_{z_u=0}^{\infty} (1-\tau)^{z_u} * P(Z_u = z_u) \\
&= E\Big[\exp\{-4\lambda_G \gamma(1+K_u)\tau\} \Big] \\
&= \sum_{k_u=0}^{u-1} e^{-4\lambda_G\gamma(1+k_u)\tau} * P_{k_u}^{u-1,1} \\
&\iff \\
\tau_u &= 1 - \sum_{k_u=0}^{u-1} e^{-4\lambda_G\gamma(1+k_u)\tau} * P_{k_u}^{u-1,1}.
\end{aligned}$$

Applying again the random graph interpretation of an epidemic, if the basic reproduction number of a type u super individual is greater than 1,

then τ_u is also the probability that a given type u student belongs to the largest component, which then is equivalent to the probability of a type u student becoming infected in the global epidemic.

It is now possible to quantify the risk of infection of an unvaccinated student who belongs to a class with u unvaccinated students against the overall risk of infection (or against $\tau_{\tilde{u}}$) by constructing the risk ratio

$$\frac{\tau_u}{\tau}.$$

If it is more than 1, then the risk of infection is higher when one belongs to a class with u unvaccinated students as compared to a uniformly chosen unvaccinated student and the opposite if it is less than 1.

4.2 Scenario II: Vaccination admission criteria in place

In this scenario, we have α, which is the proportion of classes that admit unvaccinated students, no longer equals to one. So α is any real number in the open interval of $(0, 1)$. In the case of $\alpha = 1$, we simply have scenario I.

Applying all that we have learnt in scenario I, the asymptotic proportion s_u^α of classes that have u unvaccinated students, given α, now becomes

$$s_u^\alpha = \binom{20}{u} \left(\frac{\gamma}{\alpha}\right)^u \left(1 - \frac{\gamma}{\alpha}\right)^{s-u}, \quad 1 \leq u \leq 20,$$

taking note that γ/α is the probability that a random student in this subpopulation of school classes is unvaccinated. Observe that $\gamma < \alpha$, for otherwise it means there are more unvaccinated students than the number of classes that can accommodate them. In the case where $\gamma = \alpha$, it means every student in such a class is unvaccinated, that is, it is deterministic.

We also update the size-biased distribution as follows:

$$\pi_u^\alpha = \binom{20-1}{u-1} \left(\frac{\gamma}{\alpha}\right)^{u-1} \left(1 - \frac{\gamma}{\alpha}\right)^{20-u}.$$

The basic reproduction number of an infectious super individual, given 1 initial infective, is now

$$R_* = 4 * \lambda_G * \gamma * \sum_{u-1=0}^{20-1} \pi_u^\alpha \Big(1 + \sum_{k=0}^{u-1} k * P_k^{u-1,1}\Big).$$

The expression for τ now becomes

$$\tau = 1 - \sum_{u-1=0}^{20-1} \pi_u^\alpha * \Big(\sum_{k=0}^{u-1} e^{-4\lambda_G \gamma \tau k} * P_k^{u-1,1}\Big).$$

Take note that the risk of infection remains unchanged since its computation does not involve the size-biased distribution.

5 Computations

In this section, we evaluate and show the results of the quantities of interest in Section 4. They are computed using programming software `R` [17].

5.1 Methods

Bearing in mind that the main objective of this thesis is to compare the effects between some schools impose vaccination as an admission criteria and no schools impose such a criteria on a measles outbreak, α is our main parameter of interest and we center our investigation around how α interacts with other parameters that made up the expressions of the various quantities of interest in Section 4. We use United Kingdom as a reference country for our choice of proportion of unvaccinated children, since this is the most prominent country in Europe to have lost its measles eradication status [14].

- γ

  In 2018 to 2019, 90.3% of children in UK have received their first dose of the MMR (Measles, Mumps, Rubella) vaccine by 24 months of age and 94.5% by 5 years old. But only 86.4% received their second dose by the time they turned 5 years old [16]. So we find it reasonable to focus our investigation on $\gamma = 0.20$ and $\gamma = 0.05$.

- α

  This parameter controls the proportion of school classes that admit unvaccinated children, which is an important parameter to investigate in the context of this thesis. For every value of γ, we let α vary in the interval $[\gamma, 1]$ for some choices of λ_L and λ_G values and derive the combinations of these parameters for which $R_* > 1$. We are also going to investigate α's relationship with τ and τ_u. We use the `uniroot` function in software `R` [17] to solve for τ.

- λ_L

  The local contact rate between a given infectious student and a given classmate is λ_L. This gives the probability of an unvaccinated student getting infected locally as approximately

  $$\frac{4 * \lambda_L * \text{number of classmates in the class}}{\text{number of classmates in the class}} = 4\lambda_L,$$

  which, in turn, is approximately $1 - \exp(-4\lambda_L)$ following first-order Taylor approximation of $\exp(-4\lambda_L)$.

For a highly infectious disease like measles, we find it reasonable to investigate $\lambda_L = (0.1, 0.175, 0.6)$, since this gives us the probability of infection by an infectious classmate as $33\%, 50\%$, and 90%, respectively. We do this so that we have the possibility to analyse the situation when a particular strain of measles is less virulent.

- λ_G

  λ_G is chosen in such a way that if the whole school population is unvaccinated (i.e. if $\gamma = 1$), R_* is roughly in the range of 10 to 20. This is to attain the basic reproduction number (R_0) of measles, which is often cited to be between 12 to 18 as reported by Anderson and May [4] . On the other hand, Guerra *et al* [14] have written a paper analysing the various published R_0 estimates for measles and found that the range of values are wider than 12 to 18. We understand that R_* and R_0 are different quantities, but we use the latter as a reference value. Interestingly, Lorenzo, Ball and Trapman [15] have written a paper that shows how R_0 for models with social structure can be computed. Using their work, we could have computed estimates of R_0 for measles in the case of a SIR epidemic with two levels of mixing, but this is beyond the scope of this thesis. Take note that if $\gamma = 1$, it implies that every class has ONLY unvaccinated students.

5.2 Results

We begin by investigating the three-prong relationship of R_*, α and γ. See Figure 4. By trial and error, we derive three λ_G values such that R_* is in the range of 10 to 20 when $\gamma = 1$, that is, when the entire school population is unvaccinated.

In general, with everything else being constant, R_* decreases as α increases, as shown across all subfigures in Figure 4. In Figure 4(a), the range of R_* drops from approximately between 7 to 14 to between 5 to 10. So increasing the proportion of classes that admit unvaccinated students helps to reduce R_*. We see this same positive effect in Figure 4(b) to Figure 4(d) as well, albeit to varying degrees. We also observe how the rate of reduction changes as γ decreases. It progresses from approximately linear when $\gamma = 0.7$ to clearly exponential when γ eventually reaches 0.05. This implies that for higher levels of vaccination coverage, the positive effect of an increasing α in bringing down R_* loses steam faster than in the case of a lower level of vaccination coverage. We see the tails of the green lines flattens out as γ decreases. This is an effect of high vaccination coverage, where the meagre number of unvaccinated students will quickly spread thin amongst the school classes that admit unvaccinated students as α increases, up to a point where the increase in the number of such classes barely has an effect on R_*.

We focus now on Figure 4(c) and Figure 4(d) since these two levels vaccination coverage are of main interest in our analysis. In Figure 4(c), we first remark the range of R_* is now between 2 to 4, in case of a deterministic size-biased distribution. When $\alpha = 1$, this range is approximately between 0.5 to 1, which means when there is no vaccination admission criteria, there will not be a large outbreak, regardless of how infectious the strain of measles is. But in the event if $\alpha < 1$, we see that there are certain values of α that will cause R_* to cross the threshold. Another observation is a larger λ_L needs a larger α in order to keep its corresponding R_* under threshold, which is not surprising. To illustrate, when $\lambda_G = 0.255$, the only way to keep R_* below 1 is to have $\alpha = 1$. On the other hand, for $\lambda_G = 0.185$, α needs only to be in the range of 0.55 to 0.65 in order to keep R_* under the threshold. Figure 4(d) shows the case when the level of vaccination coverage is at 95%, which is the level stipulated by WHO in order to stop measles from spreading. Here, we see that R_* is really close to zero for a wide range of α even when both λ_L and λ_G are large. The only instance when R_* crosses the threshold (and even so, it is barely above 1) is when $\lambda_G = 0.255$ and the size-biased distribution is deterministic, that is, when every class that admit unvaccinated students has only unvaccinated students. Clearly, if the level of vaccination coverage is as high as 95%, there is barely any remarkable impact on R_* if some classes impose vaccination admission criteria.

Then we observe that all three lines in every plot pivot from the same point. This is due to the size-biased distribution being deterministic, that is, $\gamma/\alpha = 1$. In this case, $\pi_1 = \pi_2 = ... = \pi_{19} = 0$ except for $\pi_{20} = 1$. Hence

$$R_* = 4 * \lambda_G * \gamma * 1 * (1 + \sum_{k=0}^{20-1} k * P_k^{20-1,1}),$$

and from the recursive formula for $P_k^{u-1,1}$, we can see (after some computations) that the distribution of $P_k^{20-1,1}$ is concentrated at the end of the outcome domain $k = 20-1$ even for small λ_L. Alternatively, we can see this by approximating the probability that the final outcome of the local epidemic is 18, given 1 initial infectious student, that is, 1 student survived the local epidemic. The probability that just one student escape infection is approximately $\exp(-4\lambda_L)^{19}$, since he has to escape infection from the initial infectious student and 18 other classmates who were ultimately infected. Even for $\lambda_L = 0.1$, this probability is approximately $\binom{19}{1} * \exp(-4*0.1)^{19} \approx 0.00942$, which means the probability that this student does not escape infection is $1 - 0.00942 = 0.99058$. But this is also the probability that all 19 of the initial infectious student's unvaccinated classmates are ultimately infected in the local epidemic. This probability will be even higher for bigger λ_L. Hence this explains why R_* for all three values of λ_L are very close to one another when γ/α is very close to 1.

In Figure 4(b), the first thing we observe is R_* drops to a lower range as compared to Figure 4(a). With everything else kept unchanged except for γ, this is surely the positive effect of vaccination and we see a continuous drop in the range of R_* as γ decreases as shown in Figure 4(c) and Figure 4(d). Another remarkable observation in Figure 4(a) is the three green lines are very close to one another, which is a clear indication that regardless of how infectious this strain of measles is, R_* is essentially the same. We can see this again from the expression for R_*. Now we have the presence of the size-biased distribution. When γ is big, the size-biased distribution "favours" classes with more unvaccinated students since γ/α is big. Further, recall that we have earlier found out that the distribution of $P_k^{u-1,1}$ for big values of u is concentrated at the end of the outcome domain, for all λ_L. This explains why R_* is almost the same for all λ_L. Intuitively, when 70% of the school population is unvaccinated, even in the case where there is no vaccination admission criteria (i.e. $\gamma = 1$), this means that every class has $0.7 * 20 = 14$ unvaccinated students and we know from the distribution of $P_k^{14-1,1}$ that all 14 of these unvaccinated students will almost surely be infected, regardless of how infectious the strain of measles is. Next, we see λ_L starts to make a difference in R_* as γ decreases, as shown in how the three lines progressively separate from one another from Figure 4(a) to Figure 4(d). This tells us that when the level of vaccination coverage is low, R_* does not depend on how virulent the strain of measles is. On the other extreme, when the level of vaccination coverage is high, it is very difficult for the virus to take hold even if it is a virulent strain plus the constraint of some classes impose admission criteria, as shown in Figure 4(d).

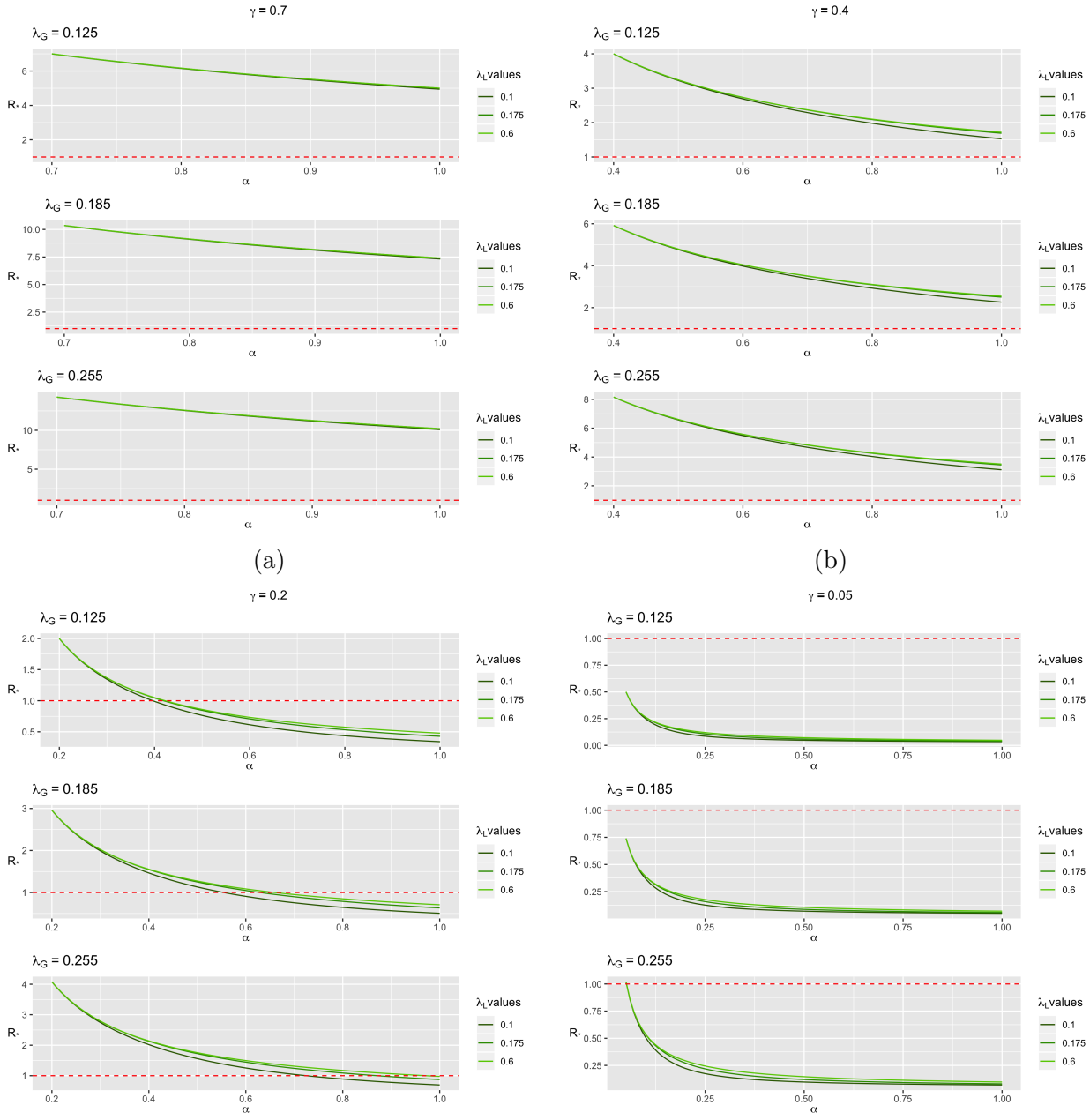

Figure 4: Relationship between average number of global contacts, proportion of schools that admit unvaccinated students and vaccination coverage. The red dotted line is drawn at $R_* = 1$. From (a) to (d), the level of vaccination coverage in the population is 30%, 60%, 80% and 95%, respectively, which are derived from $1 - \gamma$.

In Figure 5, we see the critical values of α and λ_G for which $R_* = 1$, for the same four levels of vaccination coverage as in Figure 4.

Clearly, these two parameters have a positive correlation. In other words, the average number of global contacts increases as the proportion of classes that admit unvaccinated students increases. The logical explanation is that since the unvaccinated students are spread across a wider base as the number of classes that admit them increases, thus allowing the average number of global contacts to increase to get to $R_* = 1$.

Our first observation is how little it takes for λ_G to be in order to make R_* exceed 1 when the level of vaccination coverage is low, as shown in the first two plots. This means that in order to prevent a measles outbreak from happening, we need to keep the average number of global contacts made per day by a given student to below these values, which is quite unrealistic.

In each plot, we take note of how all three lines are merged when $\gamma/\alpha = 1$, which again is due to the size-biased distribution effect we have explained above. Another remarkable observation is how the rate of change of λ_G changes from approximately linear to logarithmic as γ decreases, which implies that a positive unit change in α leads to a smaller change in λ_G as we move to higher levels of vaccination coverage. This means that as the level of vaccination coverage increases, an increase in the proportion of classes that admit unvaccinated students has less of an effect on the average number of global contacts allowed. This coincides with our observation in Figure 4 where the positive effect of an increasing α on R_* flattens out as the level of vaccination coverage increases. To illustrate, let us zoom into the middle green line of $\gamma = 0.05$ in Figure 5, that is, when $\lambda_L = 0.175$. When α is 0.5, λ_G is approximately 2.1, which is a reasonable average number of global contacts made by a student per day, say a neighbour whom he plays with occasionally after school or a schoolmate who is in the same football club. But this means that when the level of vaccination coverage is high, there is room for some school classes to refuse admission of unvaccinated student and not make R_* exceed 1. On the other hand, when $\gamma = 0.2$, we see that even when α is 1, the biggest that λ_G can be is only ≈ 0.36 for the smallest λ_L.

We also observe that the three lines progressively separate as γ decreases, which coincides with our observation in Figure 4. This implies that when the level of vaccination coverage is low, all λ_L share a similar set of (λ_G, α) critical values for which $R_* = 1$.

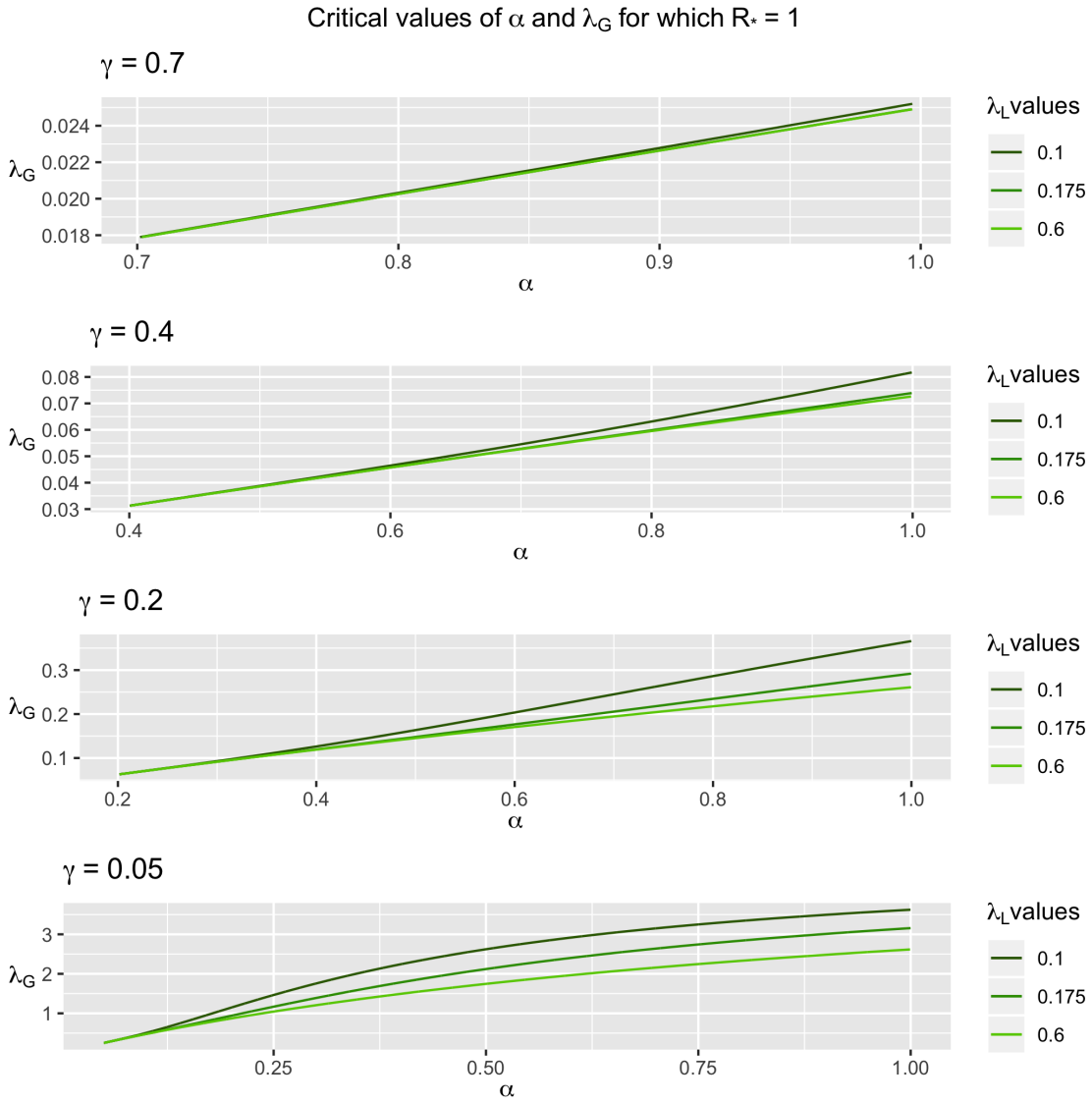

Figure 5: Critical values of α and λ_G for which $R_* = 1$. Take note that each α axis starts from its respective γ.

Figure 6(a) to Figure 6(d) show the connection amongst τ, α, λ_G and γ for the same levels of vaccination coverage as in the two figures above. Here, we present the τ values for a set of three λ_G values that were chosen such that its range is 1.5 times higher than the range of critical λ_G values in Figure 5. We do so in order to make sure $R_* > 1$ and we know there is a positive probability of a large outbreak.

In general, increasing α has a positive effect of bringing down τ across the board, albeit to varying degrees depending on both λ_G and γ.

We first look at Figure 6(a) and Figure 6(b), where the levels of vaccination coverage is at 30% and 60% respectively. In Figure 6(a), it is clear that even if there is no vaccination admission criteria (i.e. $\alpha = 1$), we cannot bring τ down to zero even with unrealistically low λ_G values, regardless of the virality of the strain of measles. But we must bear in mind here that if the school population is large, even a 1% reduction in τ can be big in absolute numbers. The situation becomes quite different when the level of vaccination coverage goes up to 60%, as shown in Figure 6(b). In the case when $\lambda_G = 0.12$, we see that it is possible to bring τ down from 1 to ≈ 0.63 as $\alpha \to 1$ for the most virulent strain of measles.

We focus now on Figures 6(c) and Figure 6(d). The effect of the virality of the strain of virus becomes more apparent here. In the middle plot of both figures, we see that for $\lambda_L = 0.1$, it is possible for $\tau = 0$ even though α is not one, implying that there is room for some school classes to impose vaccination criteria and not cause an outbreak. However, it is important to take note that this is valid only for the λ_G values used in the computation. Given the same level of vaccination level but a different λ_G, the situation becomes very different, as shown in the last plot of both figures.

Similar to Figure 4 and Figure 5, we observed that all three lines "pivot" from the same point, which again is due to the effect of a deterministic size-biased distribution. Recall earlier that we have found out that in this case, all of the students in those classes that admit unvaccinated students are unvaccinated and the distribution of $P_k^{20-1,1}$ is heavily concentrated at $P_{19}^{20-1,1}$, for all λ_L. Hence, all three λ_L values give almost identical τ values when $\gamma/\alpha = 1$. Also, the three lines more or less overlap one another when the level of vaccination coverage is low, which coincides with our observations of Figure 4 and Figure 5. This is not surprising since they share a very similar set of R_* and critical (λ_G, α) values. For each level of γ, the lines begin to separate after α crosses a certain value, with the separation becoming more apparent as γ decreases. Our interpretation is that in the event there is a large outbreak (i.e. $R_* > 1$), when the level of vaccination coverage is low, the proportion of ultimately infected does not really depend of the virality of the strain of measles. Conversely, when the level of vaccination coverage is higher, a weaker strain of the virus will result in a smaller τ in comparison to a more virulent strain, especially as

$\alpha \to 1$. A logical explanation is that when there is a bigger proportion of classes that admit unvaccinated students, there is a wider base to shoulder the burden of unvaccinated students, which makes it relatively more difficult for a weaker strain of virus to take hold.

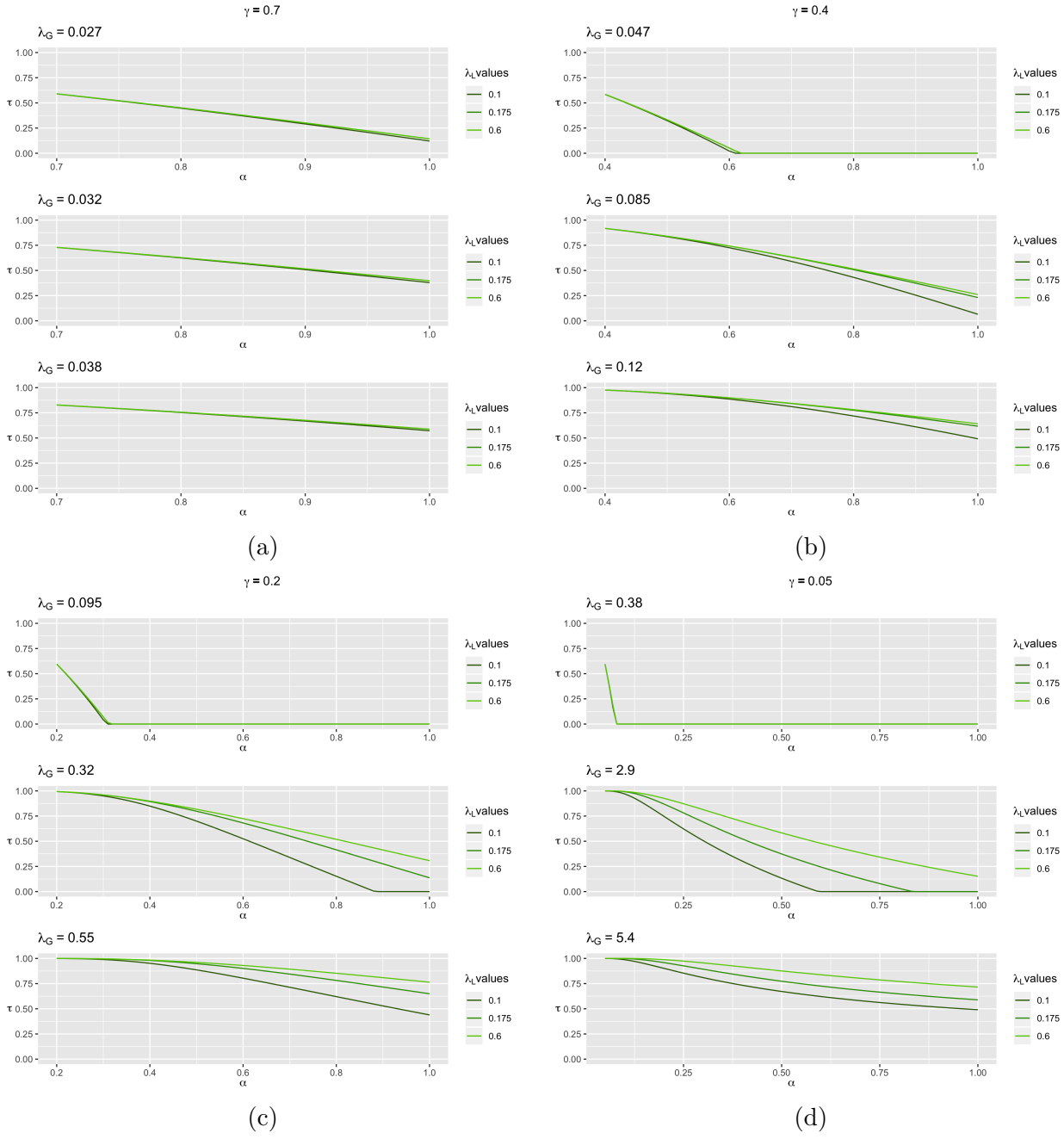

Figure 6: Relationship between τ and α at vaccination coverage 30%, 60%, 80% and 95% using λ_G values that are 1.5 times higher than the range of critical values in Figure 5.

Finally, we look at the results for risk of infection. The question that we want to answer here is: How does the presence of vaccination admission criteria affect the probability of a type u student becoming infected and consequently the relative risk of infection? If we study the expression for τ_u, we see that its dependence on α is embedded in τ. Hence, the effect of α on τ_u is manifested through τ.

In order to make the analysis tractable, we center our investigation around a set of selected parameter values. This is the setup of our investigation:

- choose two α values

  We look at $\alpha = 0.7$ and $\alpha = 1$. Since we model the vaccination criteria being imposed on an ad-hoc basis, we assume that only some schools will go to this extent.

- focus on $\gamma = 0.2$ and $\gamma = 0.05$

  We focus on these two values as we have earlier identified them as the levels of vaccination coverage that are of main interest in this thesis.

- focus on $\lambda_L = 0.1$ and $\lambda_L = 0.6$

- choose a λ_G value such that there is a positive probability of a large outbreak

  With no specific preference, we choose to work with the last λ_G value shown in Figure 6 that corresponds to the chosen γ value.

- compute τ and τ_u, for $1 \leq u \leq 20$, for every α and present all risk of infection relative to the τ value

  Interested readers can derive risk ratios $\tau_u/\tau_{\tilde{u}}$ by computing $\frac{\tau_u/\tau}{\tau_{\tilde{u}}/\tau}$.

Results of our computations are shown in Table 1 and Table 2.

Across the board, for a larger α, the risk ratios are larger compared to those when α is smaller, with the difference between the two risk ratios becoming bigger as u becomes larger. This is an interesting observation because when we derive all τ_u from the risk ratios and make horizontal comparisons, we see that $\tau_u(\text{smaller } \alpha) \geq \tau_u(\text{bigger } \alpha)$ for all u. This is telling us that although the probability of a type u student getting infected is lower (if not the same) when the proportion of classes that admit unvaccinated students is larger, this reduction (or non-reduction) is much smaller relative to the overall reduction in probability of getting infected. In other words, while everyone, in general, benefits from a larger proportion of classes that admit unvaccinated students, this benefit is not shared equally amongst all types of students. This coincides with an increasing trend in the risk ratios as

u increases, which means that those unvaccinated students in classes with more unvaccinated classmates are in a relatively worse position. Further, in both tables, we see that the increase tapers off after reaching a certain u and for some values of u, they have the same risk ratio. Since the denominator is kept constant, this implies that some types of students share the same risk of infection, that is, the probability of a type u student getting infected is the same as a type $u+1$ student, which means that this type u student is actually not much worse off than type $u+1$ student. This is especially apparent for the bigger λ_L value such that when the strain of measles is highly virulent, many types of students have more or less the same risk of infection.

$\gamma = 0.2, \lambda_G = 0.55$				
	$\lambda_L = 0.1$		$\lambda_L = 0.6$	
Risk ratio	$\alpha = 0.7$ $\tau = 0.71$	$\alpha = 1$ $\tau = 0.43$	$\alpha = 0.7$ $\tau = 0.89$	$\alpha = 1$ $\tau = 0.76$
$\frac{\tau_1}{\tau}$	0.36	0.39	0.35	0.36
$\frac{\tau_2}{\tau}$	0.46	0.48	0.58	0.60
$\frac{\tau_3}{\tau}$	0.57	0.65	0.76	0.81
$\frac{\tau_4}{\tau}$	0.70	0.81	0.88	0.96
$\frac{\tau_5}{\tau}$	0.84	1.02	0.95	1.06
$\frac{\tau_6}{\tau}$	0.97	1.23	1.01	1.13
$\frac{\tau_7}{\tau}$	1.09	1.44	1.04	1.18
$\frac{\tau_8}{\tau}$	1.18	1.62	1.06	1.22
$\frac{\tau_9}{\tau}$	1.25	1.76	1.08	1.25
$\frac{\tau_{10}}{\tau}$	1.29	1.88	1.10	1.26
$\frac{\tau_{11}}{\tau}$	1.33	1.97	1.10	1.27
$\frac{\tau_{12}}{\tau}$	1.35	2.04	1.11	1.28
$\frac{\tau_{13}}{\tau}$	1.36	2.09	1.11	1.28
$\frac{\tau_{14}}{\tau}$	1.38	2.13	1.11	1.30
$\frac{\tau_{15}}{\tau}$	1.38	2.16	1.11	1.30
$\frac{\tau_{16}}{\tau}$	1.39	2.18	1.11	1.30
$\frac{\tau_{17}}{\tau}$	1.39	2.20	1.11	1.30
$\frac{\tau_{18}}{\tau}$	1.39	2.23	1.11	1.30
$\frac{\tau_{19}}{\tau}$	1.39	2.25	1.11	1.30
$\frac{\tau_{20}}{\tau}$	1.39	2.25	1.11	1.30

Table 1: Risk ratios for 80% vaccination coverage and average number of global contacts per day is 0.55.

$\gamma = 0.05, \lambda_G = 5.4$				
	$\lambda_L = 0.1$		$\lambda_L = 0.6$	
Risk ratio	$\alpha = 0.7$ $\tau = 0.58$	$\alpha = 1$ $\tau = 0.49$	$\alpha = 0.7$ $\tau = 0.80$	$\alpha = 1$ $\tau = 0.71$
$\frac{\tau_1}{\tau}$	0.79	0.83	0.71	0.74
$\frac{\tau_2}{\tau}$	0.93	1.00	1.00	1.07
$\frac{\tau_3}{\tau}$	1.08	1.18	1.15	1.25
$\frac{\tau_4}{\tau}$	1.24	1.36	1.20	1.33
$\frac{\tau_5}{\tau}$	1.37	1.55	1.22	1.36
$\frac{\tau_6}{\tau}$	1.50	1.69	1.23	1.38
$\frac{\tau_7}{\tau}$	1.56	1.81	1.23	1.39
$\frac{\tau_8}{\tau}$	1.62	1.89	1.23	1.39
$\frac{\tau_9}{\tau}$	1.65	1.93	1.23	1.39
$\frac{\tau_{10}}{\tau}$	1.68	1.97	1.23	1.39
$\frac{\tau_{11}}{\tau}$	1.68	2.00	1.23	1.39
$\frac{\tau_{12}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{13}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{14}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{15}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{16}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{17}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{18}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{19}}{\tau}$	1.70	2.02	1.23	1.39
$\frac{\tau_{20}}{\tau}$	1.70	2.02	1.23	1.39

Table 2: Risk ratios for 95% vaccination coverage and average number of global contacts per day is 5.4.

6 Conclusion

It is clear from the results presented in Section 5.2 that not having vaccination admission criteria or an increase in the proportion of classes that do not have this admission criteria helps to bring down the basic reproduction number of the epidemic of school classes and the proportion of initially unvaccinated who ultimately gets infected in the global epidemic. The extent of this positive effect, however, is contingent on the level of vaccination coverage, how virulent the strain of measles is and how often students contact one another outside their classes.

In the case where the level of vaccination coverage is at 80%, increasing α can help to bring the basic reproduction number down to below 1 for those λ_G values used in the computation. On the other hand, at 60% and below, the same act does not bring the basic reproduction number down to below 1, given the same set of λ_G values. We have also seen that increasing α helps to bring τ down across the board. But again, the extent of its positive effect is not consistent. The positive effect of an increasing α on reducing τ loses steam when vaccination coverage goes down and/or average number of global contacts goes up. Further, although τ_u is smaller for all u when α is larger, one's relative risk of infection (compared with the general probability of getting infected) becomes worse off for larger α. In other words, a larger proportion of classes that admit unvaccinated students benefits the unvaccinated students as a whole (and hence the society), but not quite the case on an individual level. The reason being although now one has a higher chance of escaping global infection, once infected (or once one of the unvaccinated classmates gets infected), one is still going to get the same outcome for the local epidemic. And it is worse for those in the larger u classes, up to a certain point though, after which they are more or less the same "worse off". What is apparent too is there is less room for α to have an effect on τ and consequently relative risk of infection when the strain of measles is highly virulent as well as when the level of vaccination coverage is high.

Concluding, if some schools decide to impose vaccination as an admission criteria, the extent of its consequences is very much dependent on the level of vaccination coverage, the virality of the strain of measles and how often students make global contacts, which is arguably partly dependent on how urbanised and densely populated the city/region is. The positive effect of not having such an admission criteria is most apparent when vaccination coverage is high. When vaccination coverage goes down, this positive effect is less apparent. However, we should also bear in mind we are talking about proportion here. If the school population is large, a seemingly small reduction in the proportion of ultimately infected may still be big in absolute numbers. Hence, since not having such an admission criteria, in general, helps to reduce the number of unvaccinated students who ultimately gets in-

fected, albeit to different extent depending on the underlying circumstances, our recommendation is not to have such a criteria.

7 Limitations

In choosing to use the standard SIR model as the basis for our model, we inevitably inherit its limitations. For one, the simplifying assumptions of a closed and homogeneous population are far from real-life, which render the expressions for various quantities of interest in our model only as good as rough approximations. In particular, varying susceptibility (as opposed to homogeneous susceptibility assumed in the SIR model) will affect the final outcome of an epidemic. Interested readers can find out more in Andersson and Britton [2], Section 6.4 as well as a research paper [3] by the same authors. Further, we have chosen to work with a constant infectious period, which also has its drawbacks. Meester and Trapman [13] have shown that the final outcome will be overestimated with the assumption of a constant infectious period. However, we can argue that an overestimation is certainly more prudent than an underestimation, especially for a highly infectious and potentially deadly virus like measles. Finally, we have simplified λ_G to the average number of global contacts made by a given student. In fact, λ_G is a function of many factors, such as social distance, social behaviour and the inherent virality of the virus (i.e. is it airborne? Or spreads only by contact with bodily fluids, etc.). Taking into account all these heterogeneities would have yielded a model that is much closer to reality, but it would make the mathematics less tractable.

Intuitively, one would think that having vaccination as an admission criteria on an ad-hoc basis is not a good idea and what we have done here is to confirm this intuition with scientific investigation. While it may already be known to policy makers that such an act will worsen a measles outbreak, this thesis can be used as a tool for them to further convince schools why they are not allowed to do so.

One area of further research is whether or not parents who do not want to vaccinate their children would choose to give up schools of their choice or endure longer commute to schools than to vaccinate their children, should there be schools that impose such an admission criteria. This may inadvertently help authorities pull up the level of vaccination coverage. We leave this, however, to the sociologists as this is beyond the scope of mathematical research.

8 References

References

[1] Addy, C. L., Longini, I., M. and Haber, M. 1991. A generalised stochastic model for the analysis of infectious disease final size data. Biometrics. Volume 47. Pages 961 - 974.

[2] Andersson, H. and Britton, T. 2000. Stochastic Epidemic Models and Their Statistical Analysis. Springer.

[3] Andersson, H. and Britton, T. 1998. Heterogeneity in epidemic models and its effect on the spread of infection. Journal of Applied Probability. Volume 35. No. 3. Pages 651 - 661.

[4] Anderson, M. R. and May, M. R. 1982. Directly Transmitted Infectious Diseases: Control by Vaccination. Science. New Series. Volume 215. No. 4536. Pages 1053 - 1060.

[5] Bailey, N. T. J. 1975. The mathematical theory of infectious diseases and its applications. 2nd edition. Griffin.

[6] Ball, F. G. 1986. A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. Advances in Applied Probability. Volume 18. Pages 289 - 310.

[7] Ball, F., Mollison, D. and Scalia-Tomba, G. 1997. Epidemics with two levels of mixing. The Annals of Applied Probability. Volume 7. No. 1. Pages 46 - 89.

[8] Britton, T., Kypraios, T. and O'Neil, D. P. 2011. Inference for Epidemics with Three Levels of Mixing: Methodology and Application to a Measles Outbreak. Scandinavian Journal of Statistics. Volume 38. Pages 578 - 599.

[9] European Centre of Disease Control (ECDC). 2019. Measles. https://www.ecdc.europa.eu/en/measles (retrieved 31 January 2020)

[10] Jagers, P. 1975. Branching processes with biological applications. Wiley.

[11] Kermack, W. O. and McKendrick, A. G. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society A. Volume 115. Pages 700 - 721.

[12] Lefèvre, C. and Picard, PH. 1990. A non-standard family of polynomials and the final size distribution of Reed-Frost epidemic processes. Advances in Applied Probability. Volume 22. Pages 25 - 48.

[13] Meester, R. W. J. and Trapman, J. P. 2011. Bounding basic characteristics of spatial epidemics with a new percolation model. Advances in Applied Probability. Volume 43. Pages 335 - 347.

[14] MGuerra, F., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y. and Crowcroft, N. S.. 2017. The basic reproduction number (R_0) of measles: a systematic review. The Lancet Infectious Disease. Volume 17. Issue 12. Pages e420 - e428.

[15] Pellis, L., Ball, F. and Trapman, P. 2012. Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R_0. Mathematical Biosciences. Volume 235. Pages 85 - 97.

[16] NHS Digital. Childhood Vaccination Coverage Statistics - England 2018-19. https://files.digital.nhs.uk/4C/09214C/child-vacc-stat-eng-2018-19-report.pdf (retrieved 11 April 2020)

[17] R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

[18] Sheldon, M. R. 2014. Introduction to Probability Models. Eleventh Edition. Elsevier.

[19] The Centers for Disease Control and Prevention (CDC). 2019. Measles Cases and Outbreaks: Measles Cases in 2019. https://www.cdc.gov/measles/cases-outbreaks.html (retrieved 31 January 2020)

[20] The Lancet Child & Adolescent Health. 2019. Vaccine hesitancy: a generation at risk. Editorial. Volume 3. Issue 5. Page 281.t

[21] The World Health Organization. 2019. Measles. https://www.who.int/news-room/fact-sheets/detail/measles (retrieved 31 January 2020)

[22] The World Health Organization. Vaccine-Preventable Diseases Surveillance Standards: Measles. https://www.who.int/immunization/monitoring_surveillance/burden/vpd/WHO_SurveillanceVaccinePreventable_11_Measles_R1.pdf?ua=1 (retrieved 4 February 2020)

[23] The World Health Organization, Regional Office for Europe European Region loses ground in effort to eliminate measles. http://www.euro.who.

int/en/media-centre/sections/press-releases/2019/european-region-loses-ground-in-effort-to-eliminate-measles (retrieved 11 April 2020)