Reactivation of infectious diseases
A study on the seroepidemiology of varicella

Roeland Warringa
MSc Thesis

March 21, 2011

Supervisors:

Dr. M. van Boven
Dr. M.C.J. Bootsma
Dr. J. Wallinga

Contents

Colofon

What you are reading is a MSc thesis for the master 'Stochastics and Financial Mathematics' at the University of Utrecht. The thesis was written during an internship at the National Institute for Public Health and the Environment. This thesis discusses the disease phenomenon reactivation. The goal is to obtain insight in the importance of reactivation. We estimate parameters of mathematical models that use serological data of varicella and contact patterns.

1 Introduction

1.1 Reactivating infections

In this thesis we develop models for reactivating infectious diseases. An infectious disease is said to have the reactivation property if, after a primary infection, the pathogen is not completely cleared but remains dormantly present within the host and can give rise to subsequent disease episodes. The way one should think about reactivation is that after the immune system of the host has adapted to the virus and defeated it, the virus retreats and hides in the body. If the immune system becomes less effective against the virus, which may take decades, the virus may emerge from its hiding place and cause another round of disease and/or infectiousness. Prominent examples of reactivating infectious diseases that infect humans are herpes simplex virus 1 and 2, also known as human herpes virus 1 and 2, herpes simplex virus 3, also known as varicella zoster, and cytomegalovirus. Here we shall develop general theory to deal with reactivating pathogens in an age-structured host population, and apply the theory to a data set on varicella zoster virus.

1.2 Symptoms

The symptoms after the initial infection and after reactivation may be similar, but this is not necessarily the case. For instance, for varicella, the first infection causes chickenpox while the reactivation causes herpes zoster. Chickenpox has a duration of 4 to 7 days and causes blisters all over the body, while herpes zoster can last weeks and blisters are only at the section of the body corresponding to the section of the nerve system where varicella retreated to. Another difference is that chickenpox seems to be a lot more contagious than herpes zoster. The first infectious episode caused by varicella will always be chickenpox. It does not matter whether the individual was infected by someone with chickenpox or with herpes zoster, since it is the same virus. However the viral load in the infectious dose may be different which may result in a more or less severe infectious period.

1.3 Boosting of immunity

Although reactivation is not caused by reinfection, reinfection with the virus may influence reactivation in the sense that reinfection can decrease the incidence of reactivation. The idea is that immunity against a particular infection can decrease over time, especially if the immune system does not encounter the virus anymore. Reinfection may trigger the immune system of the host and enhance the immunity against the virus, thus making it harder for the virus to successfully reemerge. The importance of this mechanism is not well known but the incidence

of herpes zoster increased after vaccination against varicella in several countries. This lowered the prevalence of chickenpox and hence the rate of (re)infection, while there were still many individuals with latent varicella. (source: wikipedia: herpes zoster, with references).

1.4 Questions to be addressed

Not all are answered in this thesis, but for most there is a satisfying answer.

- What is the effect of herpes zoster on the prevalence of chickenpox?

- Do infections with varicella due to herpes zoster significantly increase the fraction of the population that is seropositive with varicella?

- Can we disentangle the force of infection caused by chickenpox and herpes zoster?

- Can we estimate the fraction of the population that needs to be vaccinated to eradicate varicella

- Does the fact that someone infected with varicella can develop herpes zoster and so infect new individuals make it harder to eliminate varicella?

- Can we give a definition and an estimate of some sort of reproduction number for varicella?

- Is there a way to divide the reproduction number into a component caused by primary infection and a component by reactivating infections?

Note that varicella can be introduced from the outside when one country eliminates it, so eradication in the sense that the varicella virus itself has completely disappeared is never possible, unless there is a world wide vaccination program to eradicate varicella.

1.5 Contents

In this thesis we have tried to incorporate these considerations in mathematical model(s). We used Mathematica (Wolfram Research Inc.) to run the models. The first chapter of the thesis discusses basic models for infectious disease transmission and can be skipped if the reader is familiar with compartmental models. The second, third and the first half of the forth chapter contain a model we will use to analyse seroprevalence data of chickenpox. Specifically in the third chapter we describe the data. The second half of chapter four presents the results of the

model. The fifth chapter describes a numerical simulation method and results of these simulations. In the sixth chapter we discuss the results and present a way to extend and improve the model. The appendix contains a table with the used notation and a discussion of an article by De Koeijer *et al.* [5]. It will not be used, but is interesting in it's own right to model reactivation epidemics.

2 Transmission models with reactivation

2.1 A simple compartmental model

2.1.1 SIR

2.1.1.1 Setting up the model frame We start with a simple situation where reactivation does not occur. We have a population where a non-lethal disease is introduced for the first time. In this thesis, the population will consist of humans, but it is also possible to consider animals. We assume that no individuals leave or enter the population, so we neglect migration, birth and death (and the disease is non-lethal, so that does not kill individuals either). This is an approximation often used for a disease that lasts very short in comparison to the lifespan of individuals. The total population size is always assumed large, so that a single infectious individual does not notably decrease the number of susceptible individuals. Furthermore we treat all individuals similarly, thus making no distinction between individuals apart from their infection status, this also implies homogeneous mixing of individuals. When an individual gets infected he may infect other individuals and recover after some time after which he is immune to further infections. In this situation we would like to know how many individuals will get infected, how fast the disease spreads, how many individuals escape, etc. To do this we translate our assumptions into a mathematical model.

First we introduce the notation for our individuals: Let $S(t)$ be the number of susceptible individuals at time t, $I(t)$ is the number of infectious individuals at time t and $R(t)$ is the number of recovered or removed individuals at time t. We will take years as the scale for t. To model the change in time of these states we make use of a system ordinary differential equations (ODEs). Note that this means that the states S, I and R will work with fractional individuals. The total population size, which we assume to be constant, is denoted by N and is equal to $S(t) + I(t) + R(t)$ for all times t. The fact that the total population is constant means that $\frac{dN}{dt} = 0$ and hence we have $\frac{dN}{dt} = \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$. We denote the initial number of S, I and R individuals (at time $t = 0$) by $S(0)$, $I(0)$ and $R(0)$. Usually we will not mention them explicitly. A susceptible individual can become infected by meeting an infectious individual. If infection occurs he will go from state S to state I. An infectious individual can recover and when this occurs he goes from state I to state R.

To describe the change in time of the number of individuals in these three states we use differential equations. We have β as a transmission parameter, that is the contact rate per individual times the transmission probability upon contact from an infected individual to a susceptible individual. The rate at which susceptibles get infected at time t will then be described by $\beta S(t)I(t)$. This describes both

an outflow of susceptibles and an inflow of infectious individuals. The recovery of individuals is α per infectious individual. The total recovery rate is then $\alpha I(t)$. This describes an outflow of infectious individuals and an inflow of recovered individuals and implies an exponentially distributed infectious period with parameter α. The SIR model, as it is called, is then described by the following system of differential equations (Diekmann *et al.* [1] chapter 1.3)

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \alpha I \\ \frac{dR}{dt} &= \alpha I. \end{aligned} \qquad (2.1)$$

The starting values for this ODE are $S(0)$, $I(0)$ and $R(0)$.

2.1.1.2 Invasion An important quantity in infectious disease models is the basic reproduction number, denoted by R_0.

Definition 2.1.1 R_0, *the basic reproduction number, is defined as the expected number of susceptibles that one infectious individual infects, during his entire infectious period[1], in an otherwise fully susceptible population.[2]*

When one infectious individual is introduced in an otherwise susceptible population we have that $S(t) \approx N$ for large N and t close to 0. (We assume that the infectious individual is introduced at time $t = 0$.) So the number of individuals he infects is described by a Poisson process with parameter βN. This means that for every unit of time t of his infectious period we expect βN new cases. The rate of recovery is exponentially distributed with parameter α. The expected duration of the infectious period is $\frac{1}{\alpha}$. Hence for this model we have have $R_0 = \frac{\beta N}{\alpha}$. When $R_0 < 1$ the infection will die out. In case of $R_0 > 1$ there will be an epidemic. Note that in reality when stochastic effects are important for $R_0 > 1$, there is a positive probability that after introduction of one infective individual in an entirely susceptible population that there will be no epidemic. This is not taken into account in this deterministic model.

2.1.1.3 Final size When the epidemic is over there will be a fraction of the population who will have escaped infection. One minus that fraction is the final size, which depends on the initial conditions. Let $s(t) = \frac{S(t)}{N}$ denote the

---

[1]The entire infectious period may consist of multiple disjoint infectious periods. In the model of section 2.1.1 there is at most one per individual, but later on there will be multiple.

[2]R_0, the basic reproduction number, is not to be confused with R, the number of recovered individuals.
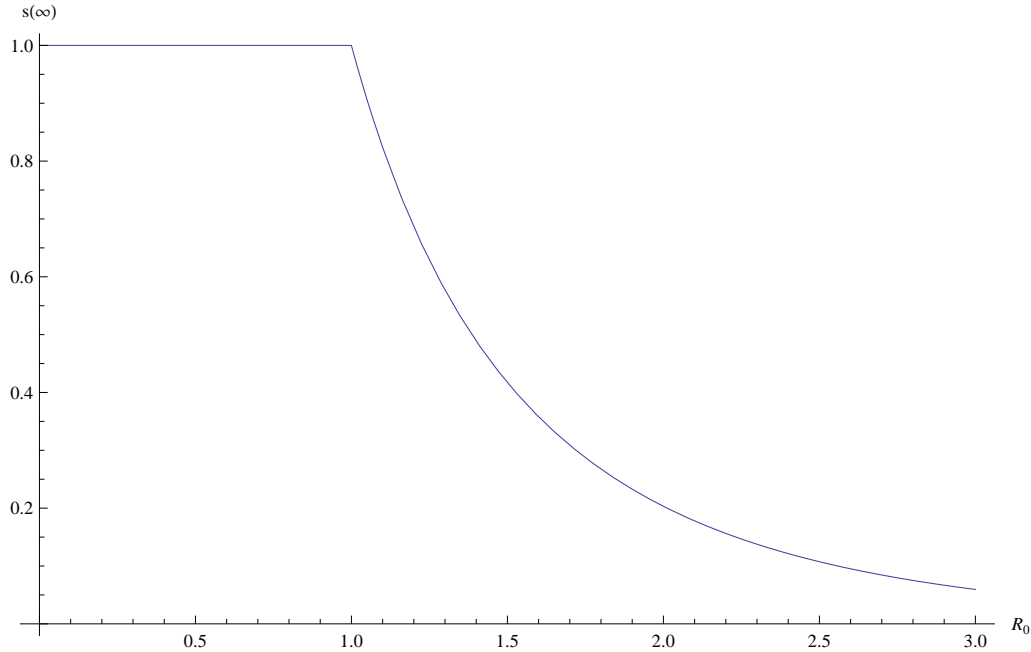
Figure 1: $s(\infty)$ of our model (system (2.1)) as a function of R_0

proportion of susceptibles. $1 - s(\infty) = 1 - \lim_{t\to\infty} s(t)$ will be our final size. We could numerically solve the system (2.1) obtain an approximation for $s(\infty)$ by $s(t)$ for a large t. This is certainly a proper way of estimating our final size, but for this model we have an implicit expression for $s(\infty)$, Diekmann *et al.* [1] chapter 1.3.

$$\log s(\infty) = R_0(s(\infty) - 1)$$

When $R_0 \leq 1$ the relevant root is $s(\infty) = 1$. When $R_0 > 1$ the unique root between 0 and 1 determines the final size. In Figure 1 we display $s(\infty)$ as a function of R_0.

2.1.1.4 Different point of view We see in our model that $R_0 = \frac{\beta N}{\alpha}$ depends on N, the population size. This is because the number of contacts per time unit depends on N. If this dependency is unwanted we can adapt the βSI term in our model to $\beta\frac{SI}{N}$. The model then looks as follows

$$\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{SI}{N} \\
\frac{dI}{dt} &= \beta\frac{SI}{N} - \alpha I \\
\frac{dR}{dt} &= \alpha I.
\end{aligned}$$

9

This variant gives $R_0 = \frac{\beta}{\alpha}$, Diekmann *et al.* [1] chapter 1.3. Now R_0 does not depend on the population size anymore. If we take βN of this model equal to the β of the model we had before, the difference is purely cosmetically. This variant can however be convenient if one wants to look the effect of different population sizes without having to change β. An example is a city that expands by building new homes at its borders, so it grows both geographically as well in population. If we assume that this means the density of the population per area roughly stays the same, the people should then not have more intensive contact despite the increased population and hence having the model independent of population size is useful.

If the model would be expanded such that N can vary during evaluation of the model, something we will not look at, then the models are really different. In general we will use a model where R_0 does not depend on the total population size.

2.1.2 SIR with birth and death

2.1.2.1 Model The model from section 2.1.1 is not suitable to model a disease with reactivation, even if a reactivation term is added to the model. In case that individuals can only reactivate a finite number of time the model may still be used, but if reactivation can occur endlessly, the individuals with dormant infection will reactivate endlessly, since they do not die and hence the disease will never die out and eventually everyone will have encountered infection. Still, from a realistic point of view even if reactivation could happen only once, this model may to simplistic. If the time scale of the disease dynamics is not much smaller than the time scale of demography, for instance because reactivation may occur a long period after the infection, we cannot ignore birth and death anymore. Therefore we add it to our model.

A simple way to incorporate birth and death is to add a parameter μ that determines the rate of both birth and death, so the total population size will remain constant. We still treat all individuals similar and assume that the disease is non-lethal so it does not increase the death rate of individuals. We assume there is no maternal infection or immunity so every newborn individual will be a susceptible. The model then looks as follows

$$
\begin{aligned}
\frac{dS}{dt} &= \mu N - \mu S - \beta \frac{SI}{N} \\
\frac{dI}{dt} &= \beta \frac{SI}{N} - \mu I - \alpha I \\
\frac{dR}{dt} &= \alpha I - \mu R.
\end{aligned}
$$

2.1.2.2   Invasion   The average time an infectious individual spends in state I is now $\frac{1}{\alpha+\mu}$ because an infectious individual can not only recover, but also die. Therefore $R_0 = \frac{\beta}{\alpha+\mu}$. Note there will not be a final size in the sense that we can talk about a number of susceptibles that eventually have been infected since they are replenished by birth and the disease may not die out, and if it does die out we eventually get a fully susceptible population.

2.1.2.3   Equilibria   In this model there will be steady states, also called equilibria. In the case that $R_0 < 1$ there will be one positive and in the case that $R_0 > 1$ there will be two, depending on the starting values. Here positive means that the states S, I and R are ≥ 0 and at least one of them is > 0. This is a natural condition, since there are no negative individuals. The steady states are easily found by putting

$$\begin{aligned}
\frac{dS}{dt} &= 0 \\
\frac{dI}{dt} &= 0 \\
\frac{dR}{dt} &= 0.
\end{aligned}$$

(2.2)

Remember that $N = S + I + R$. This gives us the trivial solution $S^\infty = N$, $I^\infty = R^\infty = 0$ which exists for all $R_0 \geq 0$ and:

$$\begin{aligned}
S^\infty &= \frac{\alpha+\mu}{\beta}N &&= \frac{N}{R_0} \\
I^\infty &= \frac{\beta-\alpha-\mu}{\beta(\alpha+\mu)}\mu N &&= (1 - \frac{1}{R_0})\frac{\mu N}{\alpha+\mu} \\
R^\infty &= \frac{\beta-\alpha-\mu}{\beta(\alpha+\mu)}\alpha N &&= (1 - \frac{1}{R_0})\frac{\alpha N}{\alpha+\mu}
\end{aligned}$$

which is only positive and different from the trivial solution for $R_0 > 1$.

2.1.3   SIR with reactivation

2.1.3.1   Model   Now we are going to add reactivation to our model. Reactivation means that recovered individuals become infectious again at a certain rate, without being reinfected. In the model this means that the individual goes from state R to I. For now reactivation is identical to the first infection in terms of infectious period and infectiousness. We assume that there is no limit to how often an individual can reactivate and the process only stops at death. We use the parameter ρ for the constant rate of reactivation, which can happen immediately

from recovery onward: Note that this implies an exponentially distributed period between two disease episodes. In reality when an individual has recovered from infection he will have some form of immunity preventing him from reactivating on shortly after recovery. With this addition the model becomes (of course with all assumptions from the birth and death model in place)

$$\frac{dS}{dt} = \mu N - \mu S - \beta \frac{SI}{N}$$
$$\frac{dI}{dt} = \beta \frac{SI}{N} + \rho R - \mu I - \alpha I$$
$$\frac{dR}{dt} = \alpha I - (\rho + \mu) R.$$

2.1.3.2   Invasion and extinction   We can now make a distinction between infectious periods where before, without reactivation we had only one. R_0 as we defined it takes all infectious periods into account, but we would like a quantity that gives us the reproduction number of the first period only. We call this reproduction number R_1.

Definition 2.1.2 *We define R_1 to be the expected number of susceptibles that one infectious individual infects in a fully susceptible population during his initial infectious period, not taking into account any infectious periods due to reactivation.*

By this definition R_1 is not influenced by reactivation and hence the same as R_0 in the birth and death model without reactivation in section 2.1.2.1 and so $R_1 = \frac{\beta}{\alpha + \mu}$. This comes with a price, namely that $R_1 < 1$ does not mean that the disease will die out. R_1 does not take into account infections due to a reactivated individual. R_0 still has the property that $R_0 < 1$ means that the disease will die out.

For our SIR model with reactivation we can calculate R_0 explicitly with the so called first-step analysis. In words: R_0 is the expected number of new infections from the first infectious period, hence R_1, plus the probability that the individual will reactivate times R_0 because after reactivation the whole process starts anew. This is because we use the exponential distribution in our models, which is memoryless. The probability that an infectious individual reactivates is the probability that he goes from I to R times the probability of going from R to I without dying. These probabilities are $\frac{\alpha}{\alpha + \mu}$ and $\frac{\rho}{\rho + \mu}$. Hence we get the equation

$$R_0 = R_1 + \frac{\alpha \rho}{(\alpha + \mu)(\rho + \mu)} R_0 = \frac{\beta}{\alpha + \mu} + \frac{\alpha \rho}{(\alpha + \mu)(\rho + \mu)} R_0.$$

Solving this equation gives us $R_0 = \frac{\beta(\rho + \mu)}{\mu(\alpha + \rho + \mu)}$.

2.1.3.3 Equilibria For this variant of the model we can also solve the equations for the steady state (2.2) and besides the trivial solution $S^\infty = N$, $I^\infty = R^\infty = 0$, we find

$$
\begin{aligned}
S^\infty &= \frac{\alpha+\rho+\mu}{\beta(\rho+\mu)}\mu N & &= \frac{N}{R_0} \\
I^\infty &= \left(\frac{\rho+\mu}{\alpha+\rho+\mu} - \frac{\mu}{\beta}\right)N & &= (R_0 - 1)\frac{\mu}{\beta}N \\
R^\infty &= \left(\frac{1}{\alpha+\rho+\mu} - \frac{\mu}{\beta(\rho+\mu)}\right)\alpha N & &= (R_0 - 1)\frac{\alpha\mu}{\beta(\rho+\mu)}N.
\end{aligned}
$$

We see that only for $R_0 > 1$ the equation the non-trivial steady state is valid and that $R_0 < 1$ indeed means that the disease will die out. R_1 may give us an indication of how the epidemic will start and R_0 for if it will last. R_0 does in general not tell us much about the start of an epidemic, since for most diseases $\beta >> \rho$, though by the differential equations reactivated individuals will be present immediately after time $t = 0$.

2.1.4 Different infectivity with reactivation

One can add a different infectious period for the initial infectious period and a reactivated infectious period, or even distinction between reactivations. A model with additional dormant and reactivation states can be more appropriate for for instance varicella, which on first infection causes chickenpox and herpes zoster with reactivation. Both infect people with varicella.

We look at an extension of our model. We assume that reactivation occurs at most once and that the infectious period after reactivation may be different both in duration and infectiousness compared to the first infectious period. The number of states an individual can be in grows to 5. We still have state S for the susceptible individuals. I_1 is for the individuals in their first infectious period. For the individuals that have been infected but have not reactivated jet we use the state D. They are said to have a dormant (sleeping) infection, also called latently infected. An individual that is in his infectious period due to reactivation is said to be in I_2. The state R holds the recovered and immune individuals that can no longer reactivate. Individuals go through the states in this order so $S \to I_1 \to D \to I_2 \to R$ and at all times, in every state an individual can die. To denote different infectivity and different length of infectious period we use two transmission parameters β_1 and β_2, and two recovery parameters α_1 and α_2. β_1 and α_1 are the transmission parameter resp. recovery parameter for the first infectious period and β_2 and α_2 for the second infectious period, which is caused by reactivation. The starting values at time $t = 0$ are given by $S(0)$, $I_1(0)$, $D(0)$, $I_2(0)$ and $R(0)$. We call this the SI$_1$DI$_2$R model. This gives us the following differential equations (with the birth/death model assumptions)

$$\frac{dS}{dt} = \mu N - \mu S - (\beta_1 I_1 + \beta_2 I_2)\frac{S}{N}$$

$$\frac{dI_1}{dt} = (\beta_1 I_1 + \beta_2 I_2)\frac{S}{N} - (\alpha_1 + \mu)I_1$$

$$\frac{dD}{dt} = \alpha_1 I_1 - (\rho + \mu)D$$

$$\frac{dI_2}{dt} = \rho D - (\alpha_2 + \mu)I_2$$

$$\frac{dR}{dt} = \alpha_2 I_2 - \mu R.$$

2.1.4.1   Reproduction numbers   Now we look at the reproduction numbers. R_1 is fairly easy. It looks only at first time infections caused by and individual in his full period in I_1 and hence $R_1 = \frac{\beta_1}{\alpha_1+\mu}$, as before, only now with the index 1 for the β and α. R_0 is a little different since now an individual cannot repeatedly reactivate, furthermore the reactivated infectious period is different. To derive it we first introduce the probability that an individual in state I_1 reaches I_2 before dying: $\mathbb{P}(I_1 \to I_2) = \mathbb{P}(I_1 \to D)\mathbb{P}(D \to I_2) = \frac{\alpha_1}{\alpha_1+\mu}\frac{\rho}{\rho+\mu} = \frac{\alpha_1\rho}{(\alpha_1+\mu)(\rho+\mu)}$. Secondly, let R_2 be the reproduction number of the second infectious period I_2. This is derived in the same way as R_1 only now we use the parameters with subscript 2, giving $R_2 = \frac{\beta_2}{\alpha_2+\mu}$. With this we can express R_0, it is R_1 times the probability of reaching I_2 times R_2. We get $R_0 = R_1 + \mathbb{P}(I_1 \to I_2)R_2$. So $R_0 = \frac{\beta_1}{\alpha_1+\mu} + \frac{\beta_2\alpha_1\rho}{(\alpha_1+\mu)(\alpha_2+\mu)(\rho+\mu)}$.

2.1.4.2   Equilibria   The equilibrium states of this model are $S^\infty = N$, $I_1 = D = I_2 = R = 0$ and the non-trivial solution.[3]

$$S^\infty = \frac{N}{R_0}$$

$$I_1^\infty = \frac{R_0 - 1}{R_0}N\frac{\mu}{\alpha_1 + \mu}$$

$$D^\infty = \frac{R_0 - 1}{R_0}N\frac{\alpha_1}{\alpha_1 + \mu}\frac{\mu}{\rho + \mu}$$

$$I_2^\infty = \frac{R_0 - 1}{R_0}N\frac{\alpha_1}{\alpha_1 + \mu}\frac{\rho}{\alpha_2 + \mu}\frac{\mu}{\rho + \mu}$$

$$R^\infty = \frac{R_0 - 1}{R_0}N\frac{\alpha_1}{\alpha_1 + \mu}\frac{\alpha_2}{\alpha_2 + \mu}\frac{\mu}{\rho + \mu}$$

---

[3]I use a mathematica application to find these expressions. Including R_1 in the application did not result in any R_1 in the expressions

For this variant of the model we also have that $R_0 > 1$ is required for this non-trivial solution to exist.

2.1.5 An age structured model

2.1.5.1 Host demography Since there is usually a long time period between the first disease episode following infection and the disease episode as a result of reactivation, considering an age stratified model is a natural thing to do. We first focus on a model without reactivation and afterwards introduce reactivation to the model. Our presentation and model formulation will be based on Chapter 7 of Diekmann *et al.* [1]. We look at population functions depending on two variables $S(a,t)$, $I(a,t)$ and $R(a,t)$ will be the number of susceptibles, infectious and recovered individuals again, with a interpreted as age and t as time. Throughout this thesis we will assume that age and time are measured in units of years. The state of an individual will be described by both his infection state and his age. When an individual is born at time t, he will be of age a at time $t + a$, provided he is still alive. So an individual born at time t will always be on the one dimensional line $(0+\tau, t+\tau)$ in the positive age/time plane.

We first look at the total population size $N(a,t)$ while we ignore the disease dynamics. The age demography dependents only on birth and death. We assume no migration. The change of N in time is given by $\frac{dN}{dt}(a,t)$. Using the chain rule we see that $\frac{dN}{dt}(a,t)$ is equal to

$$\frac{\partial N}{\partial t}(a,t) + \frac{\partial N}{\partial a}(a,t)\frac{da}{dt} = \frac{\partial N}{\partial t}(a,t) + \frac{\partial N}{\partial a}(a,t),$$

since $\frac{da}{dt} \equiv 1$. We set $N_0(a) = N(a,0)$ as the starting demography function. The death rate is modeled by $\mu(a)$ as a function of the age of an individual. μ determines the probability that an individual lives until age a: $F_d(a) = e^{-\int_0^a \mu(a^*)\,da^*}$. $F_d(a)$ is called the survival function. Type one mortality, which we will use later, is given by $F_d(a) = 1$ if $a < A$ and $F_d = 0$ if $a \geq A$ for a fixed A. For $a > 0$ the differential equation for the population demography is as follows (Diekmann *et al.* [1] chapter 7.1):

$$\frac{dN}{dt}(a,t) = \frac{\partial N}{\partial t}(a,t) + \frac{\partial N}{\partial a}(a,t) = -\mu(a)N(a,t).$$

The birth rate is given by $N(0,t) = \int_0^\infty b(a)N(a,t)\,dt$, where $b(a)$ is a birth rate function dependent of the age of the parent(s). It is also possible to set a constant inflow c of newborns $N(0,t) = c$. This ensures a constant population, both in age and size. We will use a constant inflow throughout this thesis.

When $\frac{\partial N}{\partial t}(a,t) = 0$ we have a stable age distribution. We take $N(a)$ as this

distribution and such that $\int_0^\infty N(a) = 1$ (unlike $N(a,t)$). The stable age distribution of the population is given by

$$\frac{\partial N(a)}{\partial a} = -\mu(a)N(a) \Rightarrow N(a) = N(0)e^{-\int_0^a \mu(\xi)\,d\xi}. \tag{2.3}$$

throughout this thesis we assume a stable age distribution.

2.1.5.2 Infection dynamics The infection dynamics with age stratification are an upgrade from the model without age. We assume the disease is non-lethal and that infectivity and contact patterns vary with age. The part $\frac{dS}{dt}$ becomes $\frac{\partial S}{\partial t}(a,t) + \frac{\partial S}{\partial a}(a,t)$ just like in de demography and the same applies for I and R. For the infection dynamics we introduce a contact function $c(a,a^*)$. It is the rate at which individuals of age a^* make contact with individuals of age a (Note $c(a,a^*) = c(a^*,a)$ need not be true). Furthermore our parameters β, α will be dependent on age: $\beta(a)$, the infectivity rate and $\alpha(a)$, the recovery rate. Note that whereas β also had the contact rate in it, $\beta(a)$ is only the infectivity rate of an individual of age a. $\beta(a)$ together with $c(a,a^*)$ and the entire infectious population at time t, $I(a,t)$ for all age a, determine the infection rate. We obtain the following partial differential equations (PDEs)

$$\begin{aligned}
\frac{\partial S}{\partial t}(a,t) + \frac{\partial S}{\partial a}(a,t) &= -\mu(a)S(a,t) - S(a,t)\int_0^\infty \beta(a^*)c(a,a^*)I(a^*,t)\,da^* \\
\frac{\partial I}{\partial t}(a,t) + \frac{\partial I}{\partial a}(a,t) &= -(\alpha(a) + \mu(a))I(a,t) \\
&\quad + S(a,t)\int_0^\infty \beta(a^*)c(a,a^*)I(a^*,t)\,da^* \\
\frac{\partial R}{\partial t}(a,t) + \frac{\partial R}{\partial a}(a,t) &= -\mu(a)R(a,t) + \alpha(a)I(a,t).
\end{aligned} \tag{2.4}$$

The functions $S_0(a)$, $I_0(a)$ and $R_0(a)$ are the starting functions[4] for the distribution at time $t = 0$. We assume that the state of being infectious of recovered has no influence on the birth rates of an individual and every newborn is susceptible. Then we have simply that $S(0,t) = N(0,t)$ and $I(0,t) = R(0,t) = 0$, for all time t.

2.1.6 The basic reproduction number

We want to determine whether the infection is able to establish itself in the host population in its stable age distribution. To this end we calculate the basic reproduction number R_0 which represents the expected number of secondary cases

---

[4]Do not confuse R_0, the basic reproduction number with $R_0(a)$, the starting age demography for the recovered individuals.

that a typically infected host is expected to produce in an infinitely large (host and pathogen) population in the early stages of an epidemic with respect to the PDEs (2.4). Diekmann *et al.* [1] chapter 7.3 provides a way to determine R_0 assuming separable mixing and no reactivation. We will however not use this particular variant of R_0 since for the case of varicella we cannot make the assumption of separable mixing, but it is given here anyway. Separable mixing means that the contact pattern $c(a, a^*)$ can be split up in a function $f(a)$ for the contacted individual of age a and a function $g(a^*)$ for the infectious individual of age a^*, so $c(a, a^*) = f(a)g(a^*)$. We assume the total population and age distribution constant in time and take $N(a)$ the normalised age distribution ($\int_0^\infty N(a)\,da = 1$) of the total population from equation (2.3). $h(\tau, a)$ is the transmission rate at time τ after infection of an individual infected at age a, given a contact. This $h(\tau, a)$ replaces the $\beta(a)$ and the recovery of an infectious individual that is implied by $I(a, t)$ in the integral in the PDEs (2.4). Normally we would need to multiply $h(\tau, a)$ by some sort of total number of infectious individuals[5], but here we start with one typical infectious individual. $F_d(a)$ is the probability of being the host still being alive at age a as in section 2.1.5.1. Recall that $F_d(a)$ can be expressed in terms of $\mu(a)$ and it is calculated as follows $F_d(a) = e^{-\int_0^a \mu(a^*)\,da^*}$. R_0, taken directly from Diekmann *et al.* [1], is given by

$$R_0 = \int_0^\infty \psi(a)f(a)N(a)\,da \qquad (2.5)$$

with ψ describing the infectious individual

$$\psi(a) = \int_0^\infty h(\tau, a)g(a + \tau)\frac{F_d(a + \tau)}{F_d(a)}\,d\tau. \qquad (2.6)$$

Clearly this is much more complicated than R_0 in the model for an unstructured host population developed in Chapter 2. By using easy functions, like constants and $F_d(a) = 0$ if $a \le a_d$ and 1 if $a > a_d$ for some age a_d, the formulas become easier.

2.1.7 Reactivation in the age stratified model

We expand the system equations (2.4) to contain reactivation as in section 2.1.4. Analogous to section 2.1.4 we still have $S(a, t)$ for the susceptibles, $I_1(a, t)$ for the first infectious period, $D(a, t)$ for the latently infected individuals, $I_2(a, t)$ for the reactivated infectious period and $R(a, t)$ for the immune individuals. $\beta_1(a)$ and $\alpha_1(a)$ are the infection and recovery rate for I_1 and $\beta_2(a)$ and $\alpha(a)$ are that for I_2. $\rho(a)$ is the age dependent reactivation rate. The PDEs with reactivation are:

---

[5]Later on we will see that the incidence of infection plays that role.

$$\frac{\partial S}{\partial t}(a,t) + \frac{\partial S}{\partial a}(a,t) = -\mu(a)S(a,t) - S(a,t)(\int_0^\infty \beta_1(a^*)c(a,a^*)I_1(a^*,t)\,da^*$$

$$+ \int_0^\infty \beta_2(a^*)c(a,a^*)I_2(a^*,t)\,da^*)$$

$$\frac{\partial I_1}{\partial t}(a,t) + \frac{\partial I_1}{\partial a}(a,t) = -(\alpha_1(a) + \mu(a))I_1(a,t)$$

$$+ S(a,t)(\int_0^\infty \beta_1(a^*)c(a,a^*)I_1(a^*,t)\,da^* \qquad (2.7)$$

$$+ \int_0^\infty \beta_2(a^*)c(a,a^*)I_2(a^*,t)\,da^*)$$

$$\frac{\partial D}{\partial t}(a,t) + \frac{\partial D}{\partial a}(a,t) = -(\mu + \rho(a))(a)D(a,t) + \alpha_1(a)I_1(a,t)$$

$$\frac{\partial I_2}{\partial t}(a,t) + \frac{\partial I_2}{\partial a}(a,t) = -(\mu(a) + \alpha_2(a))I_2(a,t) + \rho(a)D(a,t)$$

$$\frac{\partial R}{\partial t}(a,t) + \frac{\partial R}{\partial a}(a,t) = -\mu(a)R(a,t) + \alpha_2(a)I_2(a,t).$$

The starting functions are $S_0(a)$, $I_{1,0}(a)$, $D_0(a)$, $I_{2,0}(a)$, $R_0(a)$.[6] The newborns are given by $S(0,t) = N(0,t)$ and $I_1(0,t) = D(0,t) = I_2(0,t) = R(0,t) = 0$.

2.2 The endemic equilibrium

2.2.1 Setup in general

In this section we introduce an implicit integral equation that gives us the force of infection in a situation where the disease has reached an equilibrium (steady state). The central variables become independent of time, and the PDEs described in section 2.1.5.2 reduce to a set of ODEs in which the variables only depend on age a. The setup of this model is so that it actually does not depend the differential equations. That way we can set the length of stay in the infectious period to 0 for all individuals that become infected. In a DE model this would kill your disease immediately, nevertheless the DE models are a good step up for this one and provide good intuition for what is happening.

For simplicity we start without reactivation. We consider reactivation in section 2.2.3. We have the following assumptions. There is a demography in the population depending on birth and death (no migration), everyone is born susceptible and the total population is large. The disease is non-lethal. First we will define the ingredients. $\lambda(a)$ will be our force of infection, the variable a indicates

---

[6]Do not confuse R_0, the basic reproduction number with $R_0(a)$, the starting age demography for the recovered individuals.

the age dependency of λ. What this force of infection gives us is the rate of infection of a susceptible individual (what $\int_0^\infty \beta(a)c(a,a^*)I(a,t)\,da^*$ was in section 2.1.5.2) and a way to compute the probability to escape infection up to age a, which we name $F_S(a)$. In other words $F_S(a)$ is the probability of being in state S for an individual of age a. We have $F_S(a) = e^{-\int_0^a \lambda(a^*)\,da^*}$. Recall that $h(\tau, a^*)$ is the rate of infection upon contact time τ after infection age a^*. Note that $h(\tau, a^*)$ also contains the recovery rate of an individual, by letting his infectivity go to 0 for relatively[7] large τ. We use the contact function in the following way: $c(a, a^* + \tau)$ where $+\tau$ is used to track the age of the infectious individual during the infectious period, so the contact pattern can change accordingly. $h(\tau, a^*)$ cannot be separated from $c(a, a^* + \tau)$, since individuals need to make contact to infect other individuals. As said, our force of infection $\lambda(a^*)$ describes the rate at which individuals of age a^* get infected. This we multiply by the total susceptible population of age a^*, which is described by $S(a^*) = N(a^*)F_S(a^*)$, with $N(a^*)$ the age distribution of the total population as before and we had defined $F_S(a^*)$ the probability to escape infection up to age a^*. We get $\lambda(a^*)S(a^*)$ as the incidence at age a^* (what $S(a,t)\int_0^\infty \beta(a)c(a,a^*)I(a,t)\,da^*$ was in section 2.1.5.2). This parts dependency on λ is what makes the equation implicit. Lastly an individual needs to be alive to infect other individuals and hence we include the probability to be alive time τ after infection age a^*. We use our $F_d(a^*)$ we had in section 2.1.5 and get, using the Bayes theorem, $\frac{F_d(a^*+\tau)}{F_d(a^*)}$, the survival function since infection. Taking all this together we get, consistent with Diekmann et $al.$ [1] chapter 7.5, the following implicit equation for $\lambda(a)$

$$\lambda(a) = \int_0^\infty \int_0^\infty h(\tau, a^*)c(a, a^* + \tau)\lambda(a^*)S(a^*)\frac{F_d(a^* + \tau)}{F_d(a^*)}\,d\tau\,da^*. \qquad (2.8)$$

The trivial solution $\lambda \equiv 0$ corresponds to a state where there is no disease. This may be the only solution, but we will look at a situation where we find a non-trivial solution.

2.2.2 Simplifications

Equation (2.8) is quite hard to implement in a computer program like Mathematica. Therefore we will make three main simplifications. To begin with we take type one mortality structure, that means that all individuals live to a chosen age a_d and then die, but not before. Secondly we assume that h is independent of the age at infection. This means that we may assume that $h(\tau, a) \equiv h(\tau)$. Thirdly we assume the infectious period T to be fixed and 'short enough' so that we can work with the short disease approximation. We treat $h(\tau)$ as a constant before and after T, because T is 'short enough'. This means that $h(\tau) = \beta'$ if $\tau \leq T$

---

[7]With respect to the disease we model.

and $h(\tau) = 0$ if $\tau > T$ for a fixed constant β'. We let $\beta = \beta'T$. The T is 'short enough' assumption also allows us to drop the mortality risk during infection and such that the contact patterns do not change during infection, so $c(a, a^* + \tau)$ is approximated by $c(a, a^*)$. The inner integral of equation (2.8) reduces to β and the simplified equation becomes

$$\lambda(a) = \beta \int_0^{a_d} c(a, a^*)\lambda(a^*)S(a^*) \, da^*.$$

We let $f_I(a) = \lambda(a)S(a) = \lambda(a)N(a)e^{-\int_0^a \lambda(\eta) \, d\eta}$, be the incidence of infection of individuals of age a. Then the equation becomes

$$\lambda(a) = \beta \int_0^{a_d} c(a, a^*)f_I(a^*) \, da^*. \tag{2.9}$$

2.2.3 Implementing reactivation

Next, we want to extend equation (2.9) to have an endemic equilibrium model with reactivation. Here we choose to implement reactivation directly into the simplified version of the model and not in equation (2.8). We add reactivation by simply adding an integral to the right hand side of (2.9) that determines the force of infection generated by reactivation. In view of our assumption of an infinitely short duration of infection (the short disease approximation), we may safely assume that individuals that have been infected go immediately from state S to D.[8] Recall from section 2.1.4 that an individual in state D is a latent carrier of the pathogen and that reactivation is possible. We add subscripts to distinguish between first infection, subscript 1, and reactivation, subscript 2. The new model contains two transmission parameters β_1 for primary infection and β_2 for individuals infectious due to reactivation.[9] We define $f_{I_2}(a)$ to be the incidence of reactivation, depending on the age of the dormant individual and set $f_{I_1}(a) = \lambda(a)N(a)e^{-\int_0^a \lambda(\eta) \, d\eta}$, the same as $f_I(a)$ for the model without reactivation. Putting it all together, the integral equation describing the equilibrium force of infection in the model with reactivation becomes

$$\lambda(a) = \beta_1 \int_0^{a_d} c(a, a^*)f_{I_1}(a^*) \, da^* + \beta_2 \int_0^{a_d} c(a, a^*)f_{I_2}(a^*) \, da^*. \tag{2.10}$$

The key for this equation is to get an expression for $f_{I_2}(a)$. Therefore we define $\rho(a)$ to be the age dependent rate of reactivation and $F_D(a)$ to be the probability that an individual of age a is in state D. $F_D(a)$ is given by the integral from 0 to

---

[8]this is not immediately clear from the expression, but we can see any reference to a time in state I_1 being absent, or rather being completely ignored.

[9]We do not take into account the possibility that the route of transmission changes so much that we would need to adapt the contact function c for reactivation

a over integration variable ξ of 'the probability of infection' $\lambda(\xi)e^{-\int_0^\xi \lambda(\eta)\,d\eta}$ times 'the probability not having reactivated yet as of age of infection' $e^{-\int_\xi^a \rho(\eta)\,d\eta}$, and thus

$$F_D(a) = \int_0^a \lambda(\xi)e^{-\int_0^\xi \lambda(\eta)\,d\eta}e^{-\int_\xi^a \rho(\eta)\,d\eta}\,d\xi. \tag{2.11}$$

With equation (2.11) at hand we can build an expression for the incidence of reactivation, $f_{I_2}(a)$. We multiply the rate of reactivation $\rho(a)$ at age a with the pool of D individuals of age a, which is $N(a)F_D(a)$, so that

$$f_{I_2}(a) = \rho(a)N(a)F_D(a) = \rho(a)N(a)\int_0^a \lambda(\xi)e^{-\int_0^\xi \lambda(\eta)\,d\eta - \int_\xi^a \rho(\eta)\,d\eta}\,d\xi.$$

Insertion of both f_{I_1} and f_{I_2} into Equation (2.10) completes the derivation of the equation describing the equilibrium force of infection.

In the case of varicella newborn children obtain maternal immunity from seropositive mothers. When we want to include this in the model we use a starting age a_s for our force of infection. This means that the lower integral border, which was 0 will be replaced by a_s. This approach gives all newborns maternal immunity, it can only be used in populations which has almost every individual of fertile age seropositive. Inserting a_s into equation (2.10) gives

$$\lambda(a) = \beta_1 \int_{a_s}^{a_d} c(a,a^*)f_{I_1}(a^*)\,da^* + \beta_2 \int_{a_s}^{a_d} c(a,a^*)f_{I_2}(a^*)\,da^* \tag{2.12}$$

and we set $\lambda(a) = 0$ for $a \leq a_s$. The last part means we don't have to change all other definitions to fit with maternal immunity. We will come back to maternal immunity in section 4.2.3 and 4.2.4.

An interesting quantity is the mean age at infection. We use it as a generalized infection age together with ρ to approximate which percentage of individuals reactivates, so we know which interval of ρ to use in the model. Following Diekmann *et al.* [1], chapter 7.5., it is given by

$$\bar{a}^* = \frac{\int_0^\infty a\lambda(a)F_S(a)F_d(a)\,da}{\int_0^\infty \lambda(a)F_S(a)F_d(a)\,da} \tag{2.13}$$

In this thesis we use type two mortality, so $F_d(a) = 1$ if $0 \leq a < a_d$ and 0 otherwise. Recall that $F_S(a) = e^{\int_0^a \lambda(a^*)\,da^*}$ is the probability for an individual of age a to be in state S. If we take both these into account, and use partial integration the formula for the mean age becomes

$$\bar{a} = \frac{\int_0^{a_d} F_S(a)) \, da}{\int_0^{a_d} \lambda(a) F_S(a) \, da}.$$ (2.14)

2.2.4 Splitting up the force of infection

In equation (2.10) the right hand side consists of the sum of two integrals. The first integral describes the force of infection generated by individuals in their first infectious period, and the second integral describes the force of infection generated by the reactivated individuals. Hence we can split up the force of infection in into a part generated by the first infection and a part generated by reactivation. The part generated by first infection $\lambda_{I_1}(a)$ is given by

$$\lambda_{I_1}(a) = \beta_1 \int_0^{a_d} c(a, a^*) f_{I_1}(a^*) \, da^* = \beta_1 \int_0^{a_d} c(a, a^*) \lambda(a^*) N(a^*) e^{-\int_0^{a^*} \lambda(\eta) \, d\eta} \, da^*.$$ (2.15)

The second equality is shown here to stress the fact that equation (2.15) is not the same as (2.9) because of the distinction between λ_{I_1} and λ.

The part generated by reactivation $\lambda_{I_2}(a)$ is given by

$$\begin{aligned} \lambda_{I_2}(a) &= \beta_2 \int_0^{a_d} c(a, a^*) f_{I_2}(a^*) \, da^* && (2.16) \\ &= \beta_2 \int_0^{a_d} c(a, a^*) \rho(a^*) N(a^*) \int_0^{a^*} \lambda(\xi) e^{-\int_0^{\xi} \lambda(\eta) \, d\eta - \int_\xi^{a^*} \rho(\eta) \, d\eta} \, d\xi \, da^*. \end{aligned}$$

To determine $\lambda_{I_1}(a)$ and $\lambda_{I_2}(a)$ we need to solve (2.10) and then insert $\lambda(a)$ into (2.15) and (2.16). If there is no reactivation we have that $\lambda_{I_1}(a) = \lambda(a)$ and hence, in that particular case, equation (2.15) is the same as (2.9).

2.2.5 Discretisation of the model

Equation (2.10) does not lend itself readily to numerical analysis. Therefore we use a version in which we have a number of discrete age classes. It may seem we are making an error with discretising, but this version has an existence right of its own. For chickenpox most infections occur at a young age (< 18 is a safe upper bound, data will be presented in section 3.1). These individuals attend school and hence have most contact with their classmates, so there is something to be said for putting them in the same discretised age class.

We discretise equation (2.10) by replacing the integrals with sums. As a consequence the parameters have to be discretised as well. The most important thing to explain is that the continuous variables a and a^* that stand for the age

of an individual are to be replaced by variables like a_i. Subscript j will also function as summation index, like a^* was the integration variable. Likewise, the force of infection $\lambda(a)$ will take a finite number of values λ_i, the contact function $c(a, a^*)$ becomes a contact matrix $C = \{c_{ij}\}_{1 \leq i,j \leq A}$, and the reactivation rate $\rho(a^*)$ will be replaced by ρ_i. In order not to unnecessarily complicate the equations we let the a_j represent the length (in years) of the j-th age group, i.e. we use the a_j parameter to replace the integral boundary (and they are <u>not</u> the boundary values of the age classes, the boundary values are of the form $\sum_{j=0}^{i-1} a_j$, with a_0 the starting value, usually 0, for the beginning of the ith age class). The distribution of the population $N(a^*)$ discretises to N_j. Putting it all together, the force of infection in age group i is given by

$$\lambda_i = \beta_1 \sum_{j=1}^{A} (\lambda_j a_j c_{ij} N_j e^{-\sum_{k=1}^{j} \lambda_k a_k}) + \tag{2.17}$$

$$\beta_2 \sum_{j=1}^{A} (\rho_j a_j c_{ij} N_j \sum_{l=1}^{j} \lambda_l a_l e^{-\sum_{k=1}^{l}(\lambda_k a_k) - \sum_{m=l+1}^{j} \rho_m a_m}).$$

With these λ_i we can also compute the mean age at infection, by setting $\lambda(a) = \lambda_i$ if $a \in (\sum_{j=0}^{i-1} a_j, \sum_{j=0}^{i} a_j]$, with $a_0 = 0$ or another starting value (in case of maternal immunity see equation (2.12)) in equation (2.14).

In the same way as when we split up the force of infection, section 2.2.4, β_1 times the first sum of equation (2.17) gives the $\lambda_{I_1 i}$ and the rest gives the $\lambda_{I_2 i}$. So $\lambda_i = \lambda_{I_1 i} + \lambda_{I_2 i}$ and

$$\lambda_{I_1 i} = \beta_1 \sum_{j=1}^{A} (\lambda_j a_j c_{ij} N_j e^{-\sum_{k=1}^{j} \lambda_k a_k})$$

$$\lambda_{I_2 i} = \beta_2 \sum_{j=1}^{A} (\rho_j a_j c_{ij} N_j \sum_{l=1}^{j} \lambda_l a_l e^{-\sum_{k=1}^{l}(\lambda_k a_k) - \sum_{m=l+1}^{j} \rho_m a_m}).$$

The contribution of reactivation towards new infections by contact group i is given by $r_i(\beta_1, \beta_2, \rho) = \frac{\lambda_{I_2 i}}{\lambda_{I_1 i} + \lambda_{I_2 i}}$. Then to calculate the total contribution of reactivation towards new infections we use

$$r(\beta_1, \beta_2, \rho) = \frac{\sum_{j=1}^{A} e^{-\sum_{k=1}^{j-1} \lambda_k a_k}(1 - e^{-\lambda_j a_j})r_j(\beta_1, \beta_2, \rho)}{1 - e^{-\sum_{i=1}^{A} \lambda_i a_i}}, \tag{2.18}$$

where $\sum_{j=1}^{A} e^{-\sum_{k=1}^{j-1} \lambda_k a_k}$ is the fraction of indivuals still seronegative when they enter age class j, $(1 - e^{-\lambda_j a_j})r_j(\beta_1, \beta_2, \rho)$ is the fraction of individuals in age group j infected by a reactivated individual and $1 - e^{-\sum_{i=1}^{A} \lambda_i a_i}$ is the fraction of individuals that eventually get infected.

2.2.6 R_1 of the simplified Force of Infection model

Here we give R_1, defined in section 2.1.3, for the model. This is actually R_1 of a model without reactivation (so actually R_0), but if reactivation has little impact to the initial age distribution of infectious individuals we consider it as an approximation of the true R_1. To determine R_1 we set up the next generation matrix of our model. The next generation matrix is set up in discrete time and will give us the R_1 of the continuous time force of infection model. So it is good to take into account that it is an approximation of R_1 of the simplified model. Our next generation matrix is a multiplication of 3 parts. Firstly it is β_1 times our basic contact matrix $\mathbf{Q}$ (The values of the entries are given in section 3.2) and that matrix multiplied by the matrix that has the fractions of the contact age groups at its diagonal, zeros otherwise. We define[10] $\bar{N}_i$ to be the fraction of the ith contact age group. The next generation matrix $\mathbf{M}$ is then defined as follows.[11]

$$\mathbf{M} = \beta_1 \mathbf{Q} \begin{pmatrix} \bar{N}_1 & 0 & \cdots & 0 \\ 0 & \bar{N}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \bar{N}_5 \end{pmatrix} = \beta_1 \begin{pmatrix} q_{11}\bar{N}_1 & q_{12}\bar{N}_2 & \cdots & q_{15}\bar{N}_5 \\ q_{21}\bar{N}_1 & q_{22}\bar{N}_2 & \cdots & q_{25}\bar{N}_5 \\ \vdots & \vdots & \ddots & \\ q_{51}\bar{N}_1 & q_{52}\bar{N}_2 & & q_{55}\bar{N}_5 \end{pmatrix}.$$

From this matrix we determine the dominant eigenvalue, that is the maximum of all eigenvalues. Then R_1 is this dominant eigenvalue, Diekmann *et al.* [1] chapter 5.2. The corresponding normalised eigenvector tells us the stable age distribution of the infectious individuals in the beginning of the epidemic. We will not give R_0, as it is much more complicated due to reactivation. In the specific case that reactivating infections are not infectious ($\beta_2 = 0$), the basic reproduction number equals the primary reproduction number ($R_0 = R_1$).

2.2.7 Force of infection on the SIR model with reactivation

We will not use the equation given in this section, but we stated it here for completeness. Recall section 2.1.3 where we have a SIR model with a reactivation parameter that could bring individuals from state R back to state I. Here we take the simplified equation (2.9), including all assumptions and add reactivation parameter function $\rho(a)$, now indicating the rate at which an individual of age a goes from state R to I (So unlike in section 2.2.3 $\rho(a)$ does not send individuals to a different state of infectiousness). It looks as follows: S $\overset{\lambda(a)}{\rightarrow}$ I $\overset{\leftarrow \rho(a)}{\longleftrightarrow}$ R until the individuals dies. If we ignore the time spent in state I again we have only

---

[10]$\bar{N}_i$ is different from the N_j in section 2.2.5, as here we talk about fraction sizes of the separate contact groups. In section 2.2.5 we talk about the step size for equation (2.17).

[11]Because of the definition of $\mathbf{Q}$ later in this thesis, it is a 5×5 matrix.

S and R left for an individual to be in. Note that this is only possible if the total time spent in state I is a negligible compared to the life span of the host. The probability for an individual of age a to be in state S is $F_S(a)$ and hence the probability for an individual of age a to be in state R is $1 - F_S(a)$. Then we have $\rho(a)(1 - F_S(a))N(a)$ as the age dependent incidence of reactivation. Recall from section 2.2.3 that the incidence of infection is $\lambda(a)F_S(a)N(a)$. From state R an individual can reactivate multiple times and this is modeled by the convention that an individual does not leave state R upon reactivation, but simply gives an imput into the force of infection $\lambda(a)$. This is actually simpler than in section 2.2.3, where we had only one reactivation per individual and had to make sure in that a seropositive individual had not reactivated yet. Adding the part that reactivation contributes to the force of infection to equation (2.9) we get

$$
\begin{aligned}
\lambda(a) &= \beta \int_0^{a_d} c(a, a^*)\lambda(a^*)F_S(a^*)N(a)\,da^* + \\
&\quad \beta \int_0^{a_d} c(a, a^*)\rho(a^*)(1 - F_S(a^*))N(a)\,da^* \\
&= \beta \int_0^{a_d} c(a, a^*)(\lambda(a^*)F_S(a^*) + \rho(a^*)(1 - F_S(a^*)))N(a)\,da^*.
\end{aligned}
$$

The equation can be used for an infection that has similar infectious period for both initial infection and reactivation and the sum of all infectious periods is negligible compared to the lifespan of a host. We can combine this equation with a different reactivated period compared to the first infectious period. The scheme then looks as S $\overset{\lambda(a)}{\rightarrow}$ I$_1$ $\rightarrow$ D $\overset{\rho(a) \rightarrow}{\longleftrightarrow}$ I$_2$. So an individual is born susceptible, becomes infected at rate $\lambda(a)$ and goes into the dormant state. From there he can reactivate with rate $\rho(a)$ into the reactivated infectious state, from where he returns to the dormant state. We use β_1 and β_2 as in section 2.2.3 and get the equation (which we will not use either)

$$
\begin{aligned}
\lambda(a) &= \beta_1 \int_0^{a_d} c(a, a^*)\lambda(a^*)F_S(a^*)N(a)\,da^* + \\
&\quad \beta_2 \int_0^{a_d} c(a, a^*)\rho(a^*)(1 - F_S(a^*))N(a)\,da^*.
\end{aligned}
$$

3 Data

In this chapter we display and explain the data to our disposal.

3.1 Serological survey

We have data from Van der Klis *et al* [4], containing a cross-sectional population sample containing sera of 1515 persons. These data were analysed to obtain information on the infection status of the persons using a standard test. In addition, for each person the age was recorded. Secondly we have a contact matrix, explained in section 3.2, which we assume to be true. These are the two sets of data at our disposal.

The sera were analysed to determine the varicella zoster virus specific antibody levels, and to give information on the infection status of the individuals. Here we use a simple binary classification based on a fixed cut off in which each person is classified as having encountered the infection if its titre is equal or higher than the cut off, and uninfected otherwise. The precise value of the cut off is based on a mixture model in which the negative (i.e. uninfected) and positive (i.e. infected components) are characterised by certain distributions. The parameters of these distributions are estimated from the data (results not shown). For the current data the analyses show that the cut off at which a measurement has equal probability to belong to the negative or positive component is at 0.25 IU/ml. Figure 2 displays the logarithm of the titre and the age of each of the data points.

3.1.1 Confidence intervals for the data

We can determine per category of age the fraction of seropositive individuals and the corresponding confidence interval (CI) belonging to that age. For this there exist several ways to determine a CI, (p_-, p_+) for the probability to be susceptible p, based on the binomial distribution. We will use the following formula where S is the number of still susceptible individuals, N is the total number of individuals.

$$\sum_{k=0}^{S} \left(\begin{array}{c} N \\ k \end{array} \right) p_-^k (1 - p_-)^{n-k} = 0.025 \tag{3.1}$$

and

$$\sum_{k=0}^{S-1} \left(\begin{array}{c} N \\ k \end{array} \right) p_+^k (1 - p_+)^{n-k} = 0.975 \tag{3.2}$$

As found in Prins *et al.* [8] chapter 7.2.4.1. It is specifically stated that this way of taking a CI is useful for small datasets. Per year our dataset is not big. Equation (3.1) gives the upper bound for p and equation (3.2) gives us the lower
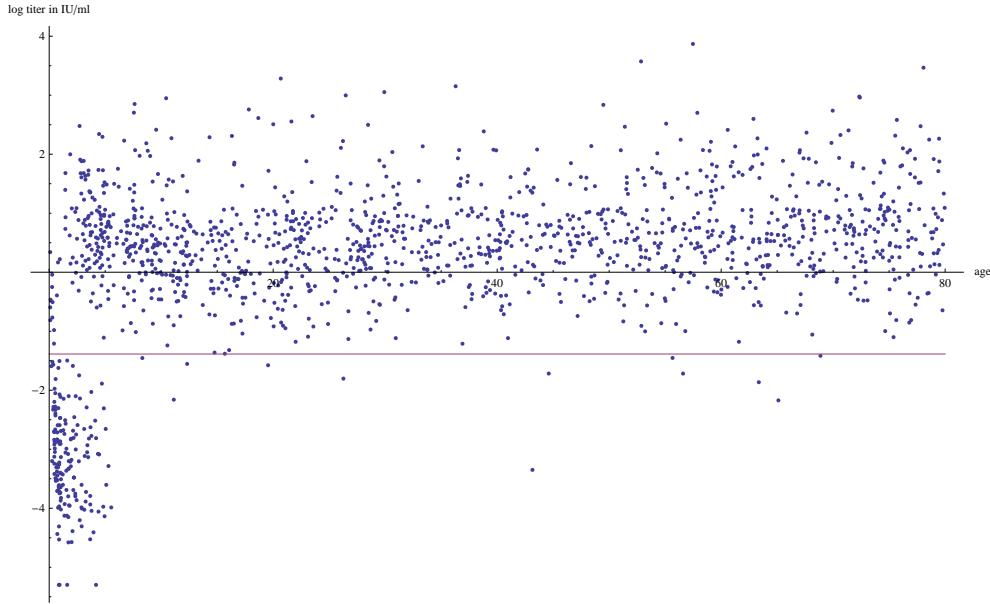
Figure 2: The log-titre of antibodies against varicella zoster of 1515 individuals as a function of age age, as measured in the Netherlands 2006-2007. The dashed red line denotes the cut off value of 0.25 IU/ml.

bound for p of the 95% CI. We want actually the 95% CI of $(1 - p)$, which is the probability for an individual to be seropositive. We take the results of (3.1) and (3.2) and subtract them from 1 (so lower and upper interchange) to get the 95% CI $(1 - p_+, 1 - p_-)$. We take $p_- = 0$ if $S = 0$ which makes (3.2) unusable, and only a lower bound for the probability to be seropositive can be calculated.

Figure 3 displays the fraction of infected individuals per year is displayed together with the calculated 95% confidence intervals. We want the probability for an individual to be seropositive derived from the model, to follow this graph. The red points are the upper boundaries, the black the actual fraction of seropositive individuals in the data (they coincide with the upper bound when the fraction of seropositive individuals in the data is 1) and the blue points are the lower boundaries of the 95% CI.

3.2   Contact matrix

The contact pattern is given by a 5×5 contact matrix $\mathbf{Q}$ from Wallinga *et al.* [9] and Mossong *et al.* [7]. $\mathbf{Q}_{ij}$ denotes the contact rate for the age groups from group j to i. We have the following age groups (in years): 0-4, 5-19, 20-39, 40-59 and 60+. The matrix is given by
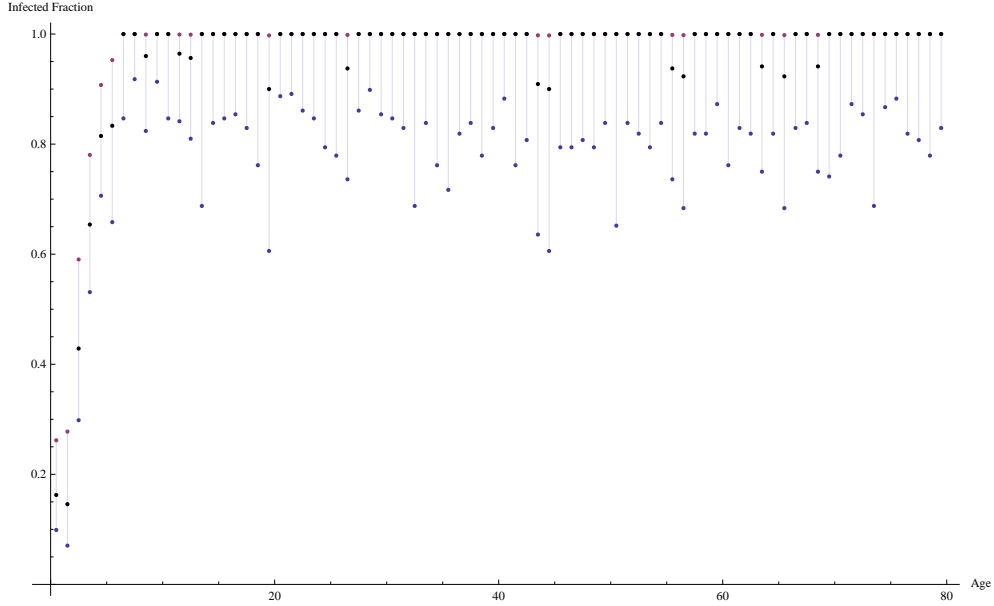
Figure 3: The fraction of seropositive individuals per year according to the data. The black dots indicate the true fraction, the red dots the upper bound and the blue dots the lower bound of the 95% confidence interval described in section 3.1.1.

$$\mathbf{Q} = \begin{pmatrix} 0.742474 & 0.269282 & 0.174178 & 0.953492 & 0.149120 \\ 0.269287 & 0.709787 & 0.120211 & 0.128524 & 0.114687 \\ 0.174186 & 0.120214 & 0.248755 & 0.159513 & 0.110271 \\ 0.953515 & 0.128530 & 0.159516 & 0.193967 & 0.120349 \\ 0.149130 & 0.114687 & 0.110269 & 0.120391 & 0.190162 \end{pmatrix}.$$

It may seem that matrix $\mathbf{Q}$ is equal to matrix C defined in section 2.2.5, but this is not the case. $\mathbf{Q}$ is a 5×5 matrix while C may be any $n \times n$, $(n \geq 5)$ matrix, determined by how we cut up the age classes. An element c_{ij} of C is defined by looking at which age groups i and j correspond to with respect to the matrix $\mathbf{Q}$. c_{ij} is then equal to the corresponding element of $\mathbf{Q}$. This forces us to have at least the ages 5, 20, 40 and 60 as cut offs of age classes for C and that is why we have the $n \geq 5$ for C.

4 Estimation

In this chapter a procedure is introduced to estimate our model parameters from the serological data. Throughout, estimates are obtained by the method of maximum likelihood. Below we derive the likelihood function using the theory developed in chapter 2 (section 4.1), and present specific results for a number of scenarios (section 4.2).

4.1 The likelihood function

We aim to estimate parameters β_1, β_2 and ρ from equation (2.17). The data described in section 3 only contains (relevant to us) the age and the titre of an individual. The titre is a measure for whether an individual has been infected (for a detailed explanation see section 3.1). For our estimation we define a log-likelihood function $l(\beta_1, \beta_2, \rho | \mathbf{a}, \mathbf{u})$ which is the logarithm of the likelihood function $L(\beta_1, \beta_2, \rho | \mathbf{a}, \mathbf{u})$. The likelihood function calculates the probability of the distribution of infectious individuals in the data if we assume the force of infection model with the parameters β_1, β_2 and ρ. The imput of the data is denoted by $\mathbf{a}$ and $\mathbf{u}$, which stand for the age and titre of our individuals, whom are labeled by ι.

Let n denote the total number of data points (in our case $n = 1515$). We use the $\tilde{\iota}$ for the age class of the individual with identifier number ι. We let a_j represent the length of the age class j. Furthermore $A(\iota)$ represents the number that needs to be added to $\sum_{j=1}^{\tilde{\iota}-1} a_j$ to reach his true age.[12] Finally u_ι is the titre of individual ι, and we set the cut off at 0.25. An individual with a titre below this cut off is classified as not infected; otherwise he is classified as infected. With the above notational conventions the log-likelihood is given by

$$l(\beta_1, \beta_2, \rho | \mathbf{a}, \mathbf{u}) = \sum_{\iota=1}^{n} \left\{ \begin{array}{ll} \log(1 - e^{-(\sum_{j=1}^{\tilde{\iota}-1} \lambda_j a_j) - A(\iota)\lambda_{\tilde{\iota}}}) & \text{if } u_\iota \geq 0.25 \\ \log(e^{-(\sum_{j=1}^{\tilde{\iota}-1} \lambda_j a_j) - A(\iota)\lambda_{\tilde{\iota}}}) & \text{if } u_\iota < 0.25 \end{array} \right\}. \quad (4.1)$$

Notice that there is a slight disingenuity, since the parameters to be estimated (β_1, β_2 and ρ) are not directly visibly in the right hand side of (4.1), but implicitly present in the λ_j defined by equation (2.17). The expression $e^{-(\sum_{j=1}^{\tilde{\iota}-1} \lambda_j a_j) - A(\iota)\lambda_{\tilde{\iota}}}$ is the probability to escape infection up till the age of the individual with identifier number ι with respect to the chosen (by the parameters) force of infection λ. So we see that this likelihood determines sums the logs of the probabilities of having or not having escaped infection with respect to our parameters.

---

[12]In other words $\sum_{j=1}^{\tilde{\iota}-1} a_j + A(\iota)$ is the age of the individual.

4.1.1 The maximum likelihood estimate

We define $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\rho}$ such that for all other β_1, β_2 and ρ we have that

$$l(\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}|\mathbf{a}, \mathbf{u}) \geq l(\beta_1, \beta_2, \rho|\mathbf{a}, \mathbf{u}).$$

We call $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\rho}$ a maximum likelihood estimate[13] (MLE). In general this maximum need not be unique for the chosen parameters nor does it need to exist. For instance if the log-likelihood converges to a maximum asymptotically for one of the parameters. In our case at least a local maximum for the force of infection and β_1 will be shown to exist according to the method, but also that the case of β_2 and ρ is different. We will not need to look at asymptotes of the force of infection because of its interpretation. The force of infection cannot become infinitely large, because then no seronegatives would remain.

4.1.2 Confidence intervals

The maximum likelihood method gives us an estimate for our parameters, but we would also like to know if there are other parameters in the neighbourhood that are not much less likely than the ones we found. To do this we calculate the 95%[14] CI using the chi-square approximation of the profile likelihood (e.g., McCullagh *et al.* [6] appendix A).

The general setup takes data $\mathbf{y}$ and estimates parameters ψ and ϕ with dimensions d and $n - d$. Say we want a CI for ψ, regarding ϕ as nuisance parameters. The procedure is as follows. We determine $\hat{\psi}$ and $\hat{\phi}$, the maximum likelihood estimate. We take the difference of the log-likelihood function with the MLE inserted, with a log-likelihood where we take the supremum over ϕ. A value for ψ is in the 95% CI if this difference does not exceed the boundary condition, which is $\frac{1}{2}$ times the χ^2 distribution with d degrees of freedom, cut off at 95%. The CI is then the region of ψ that is allowed by the following equation

$$2l(\hat{\psi}, \hat{\phi}|\mathbf{y}) - 2\sup_{\phi} l(\psi, \phi|\mathbf{y}) \leq \chi^2_{d,95\%}. \tag{4.2}$$

So in our case if we want a CI for say β_1 we have the following equation

$$2l(\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}|\mathbf{a}, \mathbf{u}) - 2\sup_{\beta_2,\rho} l(\beta_1, \beta_2, \rho|\mathbf{a}, \mathbf{u}) \leq \chi^2_{1,95\%} \approx 3.841. \tag{4.3}$$

The MLE may be on a boundary point, so for instance if we estimate a parameter in $[0, \infty)$, and the MLE is 0. If this happens the CI generated by this method

---

[13]There are 3 parameters, but it is only 1 estimate, hence we do not use the plural form of estimate.

[14]One can take any percentage ξ, just replace 95% by ξ everywhere.

may not be entirely correct anymore. Still the error is not too serious and we will remain using this method of determining the CI, assuming it is correct.

4.1.3 Details of the estimation procedure

To process the data we use the computer program Mathematica. Firstly we explain the technical details about how we cut up our age classes and force of infection with respect to our contact matrix $\mathbf{Q}$ in equation (2.17). The general approach with general data and contact matrix is not included, we immediately state our method. We have λ_1 through λ_{27} for the age vector[15] (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 79.99), where elements denote the boundaries of the age classes, so 0,1 means the age class $[0, 1)$, 1,2 the age class $[1, 2)$ and 20,25 the age class $[20, 25)$ and the same goes for all of them. Then for instance λ_1 covers age class $[0, 1)$, λ_2 covers $[1, 2)$ and λ_{20} covers $[20, 25)$. Now we have 27 distinct λ_i but many will have the same value. This is because we only have a 5×5 contact matrix. Of course we need a 27×27 matrix for the equation so we build it by awarding the age classes in the same contact group the same contact pattern. This 27×27 matrix will be our c_{ij}. The definition of c_{ij} results in only 5 distinct values for the λ_i. This can be seen from equation (2.17) where the λ_i in the same contact group have the same right hand side (the right hand side has only the subscript i in the matrix entries c_{ij}, so if $c_{ij} = c_{i'j}$ then $\lambda_i = \lambda_{i'}$). So effectively we have only 5 equations instead of 27, which cuts down the processing time roughly by a factor 10. We denote those 5 values for λ_i by $\lambda_{(1)}$ through $\lambda_{(5)}$. The first age class $[0, 1)$ will have $\lambda_1 = \lambda_{(1)}$, the second age class $[1, 2)$ will have $\lambda_2 = \lambda_{(1)}$ up till age class $[4, 5)$ $\lambda_5 = \lambda_{(1)}$. These age classes correspond to the first age group for our contact matrix. Then the force of infection becomes $\lambda_{(2)}$ so age class $[5, 6)$ will have $\lambda_6 = \lambda_{(2)}$. The next jumps will be at 20, so $\lambda_{20} = \lambda_{(3)}$, then at 24, and lasty at 26.

To implement the data we choose a grid of the parameters β_1, β_2 and ρ. For each of the three parameter we may wish to choose a different step size. Then we let the program solve equation (2.17) giving us the corresponding $\lambda_{(1)}$ through $\lambda_{(5)}$. These $\lambda_{(i)}$ we insert in (4.1) to get a log-likelihood value. So we get a large (depending on the grid) list of parameters with their log-likelihood. We simply take the maximum of all log-likelihoods to find our estimate for the parameters.

For reactivation we have Johnson *et al.* [3] telling us that "It is estimated that 23-30% of the population in Europe will develop herpes zoster during their lifetime; approximately 50% of all people reaching the age of 85 years will have experienced Herpes zoster.". Hence we choose the range for our parameter ρ such that approximately 23-30% reactivates. We use the mean age of infection

---

[15]The 79.99 instead of 80 is for programming reasons

$\bar{a}f$ as infection age from which we calculate the proper range of ρ.

We use (4.2) to calculate confidence intervals, and confidence regions/volumes. For example if we want to have a CI for β_1 we take (4.3). For β_1 We will use only that what is allowed by equation (4.3) plus one step to each side so we are sure the entire 95% confidence interval is contained in the interval we give. This means that the true CI may be smaller at most one step size at each side of the interval and we have at least 95% confidence, determined by the precision of the grid.

We cannot separate β_2 and ρ. This is explained more in detail in section 4.2.1. We will take them together so we then have 2 degrees of freedom for the χ^2 distribution when we take the confidence region. Note that here that we get that the region may be at most one step size larger in the direction of the corresponding parameter than the actual confidence region, since 'adding a point at each side of the interval' like with the β_1 CI does not have much meaning here.

4.2 Results

4.2.1 Base scenario

In this chapter we apply the estimation procedure of section 4.1.3. In this particular section we consider a baseline scenario, and in the subsequent sections we consider variations on the base scenario. Table 1 shows the results.

Table 1: Maximum likelihood estimate of the epidemiological parameters β_1, β_2, and ρ, with associated 95% confidence intervals. Also shown are derived parameters such as the forces of infection in the five age groups at the maximum likelihood estimate, the reproduction number of varicella R_1, and the mean age at infection with varicella $\bar{a}$.

Parameter	MLE	95%CI	Accuracy	Range
β_1	4.49	$(3.99, 5.03)$	steps of size$\frac{1}{100}$	3 - 6
β_2	0	$[0, \infty)$	steps of size $\frac{1}{10}$	0 - 4
ρ	0	$[0, \infty)$	steps of size$\frac{1}{3000}$	$\frac{9}{3000}$ - $\frac{15}{3000}$
$\lambda_{(1)}$	0.308	$(0.259, 0.366)$	by β_1, β_2 and ρ	
$\lambda_{(2)}$	0.174	$(0.162, 0.196)$	by β_1, β_2 and ρ	
$\lambda_{(3)}$	0.079	$(0.069, 0.097)$	by β_1, β_2 and ρ	
$\lambda_{(4)}$	0.050	$(0.045, 0.066)$	by β_1, β_2 and ρ	
$\lambda_{(5)}$	0.069	$(0.060, 0.082)$	by β_1, β_2 and ρ	

R_1	8.3	(7.3,9.3)	by β_1
$\bar{a}$	3.9	(3.1, 4.8)	by the $\lambda_{(i)}$

The maximum likelihood estimate is $(\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}) = (4.49, 0, 0)$, indicating that reactivating infections are not infectious at all ($\hat{\beta}_2 = 0$), and that reactivation does not occur ($\hat{\rho} = 0$).

Table 1 furthermore shows that the transmission parameter β_1 can be estimated with considerable precision, while the parameters β_2 and ρ cannot be estimated separately. In fact, the confidence intervals for these parameters span $\mathbb{R}_+$.

Intuitively, the fact that β_2 and ρ cannot be estimated separately can be explained as follows. More reactivation (high ρ) increases the herpes zoster cases. To maintain the same force of infection, the infectivity of herpes zoster, β_2 must decrease. Less reactivation means that β_2 must increase. Hence we cannot separate β_2 and ρ because the influence each other in such a direct way. In the most extreme case, when there is no reactivation β_2 can be anything because herpes zoster does not occur. Or the other way around if herpes zoster is noninfectious ρ may be everything. This is the reason their confidence intervals span $\mathbb{R}_+$.

Although the parameters β_2 and ρ cannot be estimated separately, it is possible to estimate the product of these parameters with precision, and, more precisely, to determine a credible upper bound for the product. We use our CI equation (4.2) in the following form to establish their combined confidence region.

$$2l(\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}|\mathbf{a}, \mathbf{u}) - 2\sup_{\beta_1} l(\beta_1, \beta_2, \rho|\mathbf{a}, \mathbf{u}) \leq \chi^2_{2,95\%} \approx 5.991 \qquad (4.4)$$

Interestingly, the confidence region (dots) in Figure 4 shows that we cannot discard much of the parameter region that we considered reasonable a priori (red dots), since only the part of parameter space with high reactivation rate (e.g., 0.005 per year) and high transmissibility (e.g., 4) is not contained in the 95% confidence region of β_2 and ρ. This may be due to the fact that many young individuals are already seropositive. Much of the data of the older individuals, which are almost all already seropositive, then gives little information.

For the contribution of herpes zoster to new infections we have in section 2.2.5 the equation (2.18). We apply this equation into the following version of the likelihood function.

$$2l(\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}|\mathbf{a}, \mathbf{u}) - 2\sup_{\beta_1, \beta_2, \rho|r(\beta_1, \beta_2, \rho) \geq r} l(\beta_1, \beta_2, \rho|\mathbf{a}, \mathbf{u}) \leq \chi^2_{1,95\%} \approx 3.841 \qquad (4.5)$$

$r(\beta_1, \beta_2, \rho) \geq r$ is a condition on (β_1, β_2, ρ) such that only the points where the
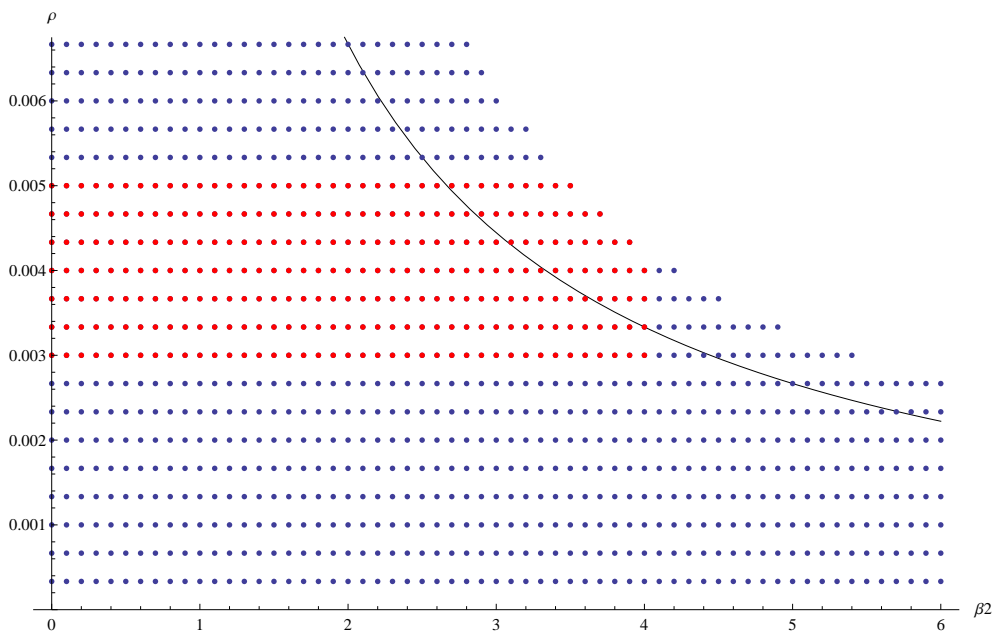
Figure 4: The dots give the 95% confidence region of herpes zoster transmission parameter β_2 together with reactivation parameter ρ per year determined by equation (4.4). The red dots represent parameter combinations that are considered realistic a priori. The black line goes near the points where there exist β_1 such that the contribution of herpes zoster to new infection lies between 7,0% and 7,3%.

contribution of herpes zoster to new infections exceeds r are included. We say that $l(r) = \sup_{\beta_1, \beta_2, \rho | r(\beta_1, \beta_2, \rho) \geq r} l(\beta_1, \beta_2, \rho | \mathbf{a}, \mathbf{u})$, so that it is visible that we have a likelihood of only 1 parameter and this is why use the chi-square distribution with only 1 degree of freedom.

The upper bound of the contribution of herpes zoster to the new infections is 7.3%, based on equation (4.5). The point where this contribution is attained is $(\beta_1, \beta_2, \rho) = (4.11, 6.0, 0.002)$. It is has highest β_2 allowed by the grid and a relatively low ρ. These parameter values indicate that β_2 has a stronger effect on the contribution of herpes zoster to the new infections than ρ has. This is most likely due to the restriction that an individual may reactivate at most once, since a reactivation reduces the pool of individuals with a dormant infection. The black line ($\beta_2 \rho = \frac{1}{75}$) in Figure 4 goes near points ((β_2, ρ) directly left and/or directly below or on the black line) such that there exists a β_1 such that the contribution of herpes zoster exceeds 7,0%. This indicates that it is not problematic that $(\beta_1, \beta_2, \rho) = (4.11, 6.0, 0.002)$ is a point on the upper boundary of the grid for β_2, in the sence that there are points with much lower β_2 that give a similar result for the contribution of herpes zoster. In Table 2 the force of infection, split up force of infection and the contribution of herpes zoster is shown per age group.

Table 2: Table of the force of infection in general, generated by chickepox and generated by herpes zoster and the contribution of herpes zoster to the total of new infections in the upper bound case of $\beta_1 = 4.11$, $\beta_2 = 6.0$ and $\rho = 0.002$. (When the force of infection does not add up precisely it is due to rounding errors.)

Force of infection	by chickenpox	by herpes zoster	Zoster contribution
$\lambda_{(1)} = 0.298$	$\lambda_{I_1(1)} = 0.281$	$\lambda_{I_2(1)} = 0.017$	$r_{(1)}(\beta_1, \beta_2, \rho) = 0.058$
$\lambda_{(2)} = 0.185$	$\lambda_{I_1(2)} = 0.162$	$\lambda_{I_2(2)} = 0.022$	$r_{(2)}(\beta_1, \beta_2, \rho) = 0.120$
$\lambda_{(3)} = 0.088$	$\lambda_{I_1(3)} = 0.073$	$\lambda_{I_2(3)} = 0.015$	$r_{(3)}(\beta_1, \beta_2, \rho) = 0.172$
$\lambda_{(4)} = 0.060$	$\lambda_{I_1(4)} = 0.046$	$\lambda_{I_2(4)} = 0.014$	$r_{(4)}(\beta_1, \beta_2, \rho) = 0.230$
$\lambda_{(5)} = 0.075$	$\lambda_{I_1(5)} = 0.063$	$\lambda_{I_2(5)} = 0.012$	$r_{(5)}(\beta_1, \beta_2, \rho) = 0.165$

Here is visible how quickly the disease spreads. The only $r_{(i)}$ below the total contribution of herpes zoster is $r_{(1)}$, which is the contribution of herpes zoster to the infections in individuals up to 5 years only. That contribution determines most of the total contribution. Also note how the force of infection generated by chickenpox decreases rapidly with age, while the force of infection generated by herpes zoster is relatively stable.

The confidence region (dots) in Figure 4 displays a hyperbola indicating there may a relation between β_2 and ρ of the form $\beta_2 \rho = c$, for some constant c. If we
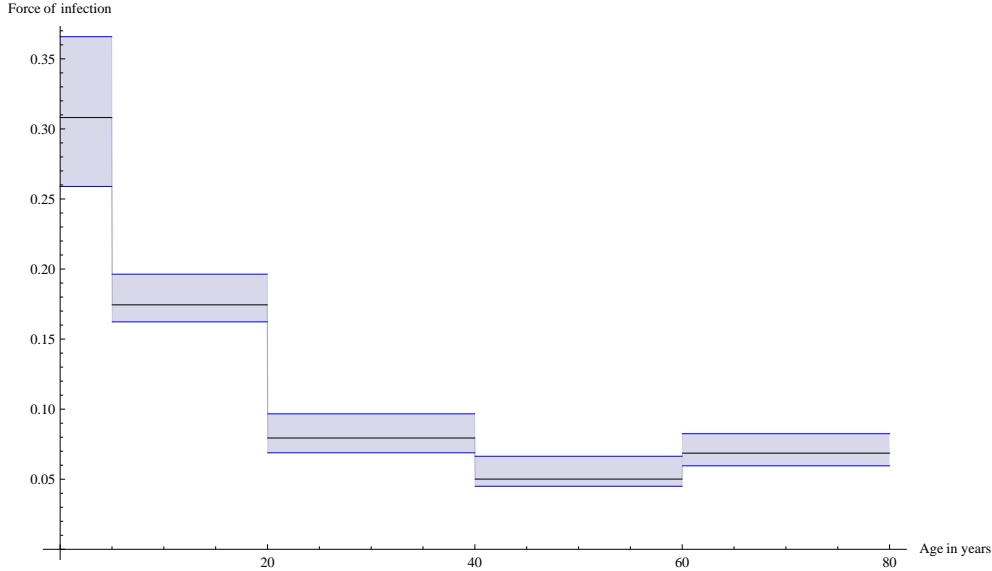
Figure 5: Estimated force of infection per age group (black line), with associated 95% confidence intervals (lila area) for the base scenario.

look at equation (2.17) where now ρ_i becomes ρ, we have

$$\beta_2 \rho \sum_{j=1}^{A} (a_j c_{ij} N_j \sum_{l=1}^{j} \lambda_l a_l e^{-\sum_{k=1}^{l}(\lambda_k a_k) - \rho \sum_{m=l+1}^{j} a_m})$$

for the second part of the right hand side. We see that we have a $\beta_2 \rho$ as a factor before the sum and an $e^{-\rho}$ in the sum, which is there by the restriction that an individual may only reactivate once. The $e^{-\rho}$ contributes little and hence the confidence region for β_2 and ρ with fixed β_1 (β_1 is not fixed, but should not change to much.) should approximately look like a hyperbola. Note that by the $e^{-\rho}$ there is a difference between β_2 and ρ, otherwise only the product of the two would truly matter.

The MLE also gives us an estimate for the values of $\lambda_{(i)}$, displayed in Table 1. The CI for $\lambda_{(i)}$ is actually a little bigger than what we see in table 1 because the algorithm used only takes points that are strictly within the CI. The interpretation of $\lambda_{(1)} = 0.308$ is that it gives an approximate probability of $1 - e^{-0.308} \approx 27\%$ per year for an individual to become seropositive. Figure 5 displays the $\lambda_{(i)}$ with CI in a graph. Here we used the CI equation with all three parameters free.

$$2l(\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}|\mathbf{a}, \mathbf{u}) - 2l(\beta_1, \beta_2, \rho|\mathbf{a}, \mathbf{u}) \leq \chi^2_{3,95\%} \approx 7.815. \tag{4.6}$$

Clearly the force of infection is at its highest for the first 20 years of life. Unlike the product $\beta_2 \rho$ the force of infection can be estimated with substantial precision
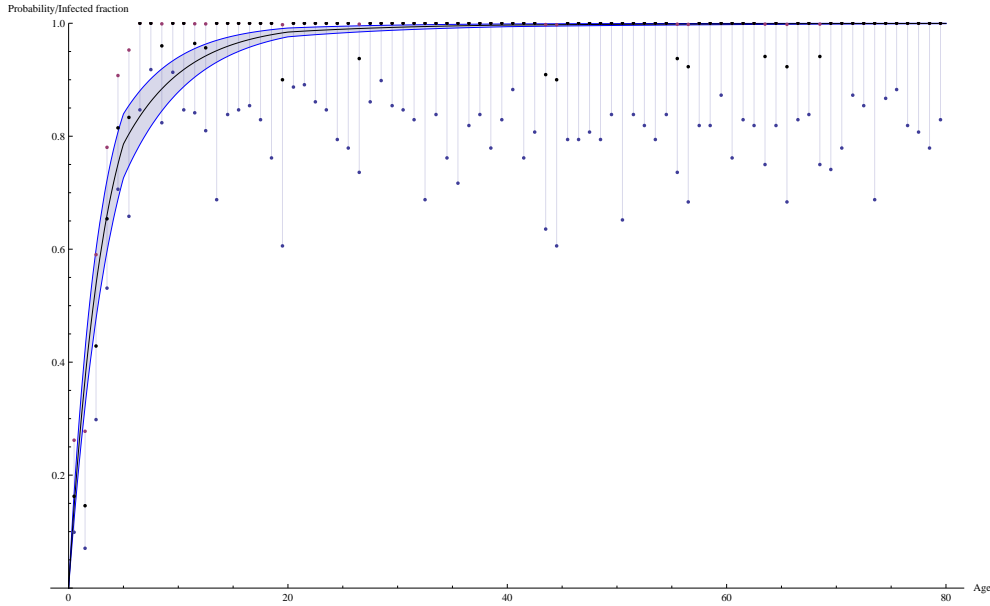
36

Figure 6: Estimated probability of infection (black line) with associated 95% confidence region (lila area) for the base scenario, together with the data (cf Figure 3).

despite the fact that the older individuals in the data give us little information. This is because the contact matrix is assumed fixed, still giving us information about the older individuals.

Figure 6 shows the probability of being seropositive together with the fractions of seropositive individuals in the data with the profile likelihood 95% confidence interval as described in section 3.1.1. The probability of being seropositive is $1 - e^{-(\sum_{j=1}^{\tilde{\iota}-1} \lambda_j a_j) - A(\iota)\lambda_{\tilde{\iota}}}$ as given in section 4.1. If we look from the age of 8 onward, then for every black point not equal to 1, only one individual is seronegative (compared to approximately 10-18 seropositives per year of age for the latter ages). From age 1 to 2 we see a drop. We assume this is caused by maternal immunity of newborn children stimulating high titres for 0-6 month olds. The second year (This is the age range $(1,2]$) profile likelihood does not reach the probability of being seropositive, this may also be explained by the overestimation due to maternal immunity of the first year, but this is not certain. It can also be that the contact pattern or the age step sizes we used in equation (2.17) are not fine enough.

4.2.2   Smaller age steps

Table 3: Maximum likelihood estimate of the epidemiological parameters β_1, β_2, and ρ, with associated 95% confidence intervals. Also shown are derived parameters such as the forces of infection in the five age groups at the maximum likelihood estimate, the reproduction number of varicella R_1, and the mean age at infection with varicella $\bar{a}$. Here age steps of $\frac{1}{2}$ year are used instead of the vector given in 4.1.3.

Parameter	MLE	95%CI	Accuracy	Range
β_1	4,17	$(3.73, 4.62)$	step size of $\frac{1}{100}$	3 - 6
β_2	0	$[0, \infty)$	step size of $\frac{1}{10}$	0 - 4
ρ	0	$[0, \infty)$	step size of $\frac{1}{3000}$	$\frac{9}{3000}$ - $\frac{14}{3000}$
$\lambda_{(1)}$	0.300	$(0.260,0.367)$	by β_1, β_2 and ρ	
$\lambda_{(2)}$	0.173	$(0.161,0.195)$	by β_1, β_2 and ρ	
$\lambda_{(3)}$	0.080	$(0.069,0.097)$	by β_1, β_2 and ρ	
$\lambda_{(4)}$	0.050	$(0.045,0.067)$	by β_1, β_2 and ρ	
$\lambda_{(5)}$	0.069	$(0.060,0.082)$	by β_1, β_2 and ρ	
R_1	7.7	$(6.8, 8.6)$	by β_1	
$\bar{a}$	4.0	$(3.1,4.8)$	by the $\lambda_{(i)}$	

In this section we use age steps of $\frac{1}{2}$ year to compute our estimations. The processing time of one grid point went up greatly hence the smaller age steps gave us the restriction for the fineness of the grid we use throughout. Unfortunately we could not run the entire grid. The results for the β_1, R_1 and the MLE for $\lambda_{(i)}$ are all done correctly, but the CI for $\lambda_{(i)}$ is less precise. Mainly the reactivation part is not taken into account, but that should have less effect than primary infections. The small age steps are implemented in equation (2.17) hence the parameters β_1 (taking R_1 along) and β_2 should change, but $\lambda_{(i)}$, which evaluated directly into the log-likelihood, should not (and ρ neither). Indeed both β_1 and R_1 decrease, indicating that R_1 is overestimated, but the rest stays roughly the same as in section 4.2.1, though $\lambda_{(1)}$ drops slightly

In Figure 7 we have the probability to be seropositive together with the fraction, with CI, in the data that is seropositive. The drop of $\lambda_{(1)}$ is visible (more or less) in Figure 7. The data CI of 2 year olds is closer to the seropositive probability graph, while the data CI for the 8 year olds is further off the force of infection MLE (not further of the force of infection CI).

4.2.3 Maternal immunity

In Figure 6 of the base scenario we see that at the young ages the actual data differs from the estimate of the probability of being seropositive. The first point, corresponding to the first year of life, is good, but then for a year older the fraction of seropositive individuals drop. We expect that this is caused by maternal immunity for varicella. When a child receives breast feeding from his or her mother,
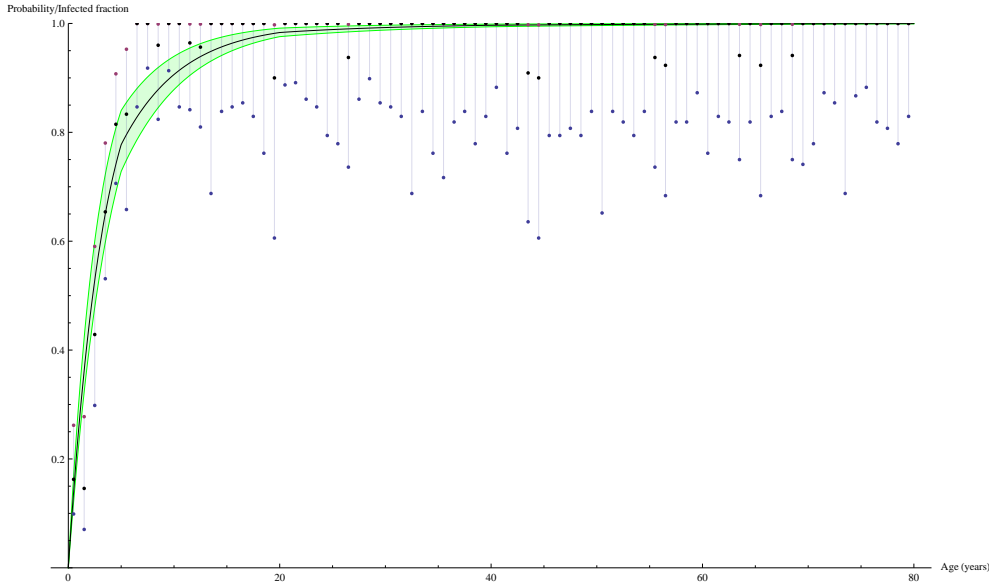
Figure 7: Estimated probability of infection (black line) with associated 95% confidence region (transparent green area) for the small age step scenario, together with the data (cf Figure 3).

the child also receives some immunity from her, giving the child protection for a few months. However this also shows in the titre therefore possibly giving us a false indication. This protection wanes after a few months and hence explaining the lower fraction of seropositive individuals for the second life year.

In this en the next section reevaluations are given without children up to the age of 6 months, with normal age steps again. Other than that the procedure in this section is identical to the original setup. This way we also discard the children that do not get maternal immunity because of a seronegative mother. We assume that they are a negligible minority, because there are very little seronegative adults.

Table 4: Maximum likelihood estimate of the epidemiological parameters β_1, β_2, and ρ, with associated 95% confidence intervals. Also shown are derived parameters such as the forces of infection in the five age groups at the maximum likelihood estimate, the reproduction number of varicella R_1, and the mean age at infection with varicella $\bar{a}$. The data used is simply the part without individuals up to 6 months of age.

Parameter | MLE | 95%CI | accuracy | Calculation range

β_1	4.50	$(3.99, 5.04)$	$\frac{1}{100}$	3 - 6
β_2	0	$[0, \infty)$	$\frac{1}{10}$	0 - 4
ρ	0	$[0, \infty)$	$\frac{1}{3000}$	$\frac{9}{3000}$ - $\frac{15}{3000}$
$\lambda_{(1)}$	0.309	(0.259,0.367)	by β_1, β_2 and ρ	
$\lambda_{(2)}$	0.175	(0.162,0.196)	by β_1, β_2 and ρ	
$\lambda_{(3)}$	0.080	(0.069,0.097)	by β_1, β_2 and ρ	
$\lambda_{(4)}$	0.050	(0.045,0.067)	by β_1, β_2 and ρ	
$\lambda_{(5)}$	0.069	(0.060,0.083)	by β_1, β_2 and ρ	
R_1	8.3	(7.3,9.3)	by β_1	
$\bar{a}$	3.9	(3.1,4.8)	by the $\lambda_{(i)}$	

We get estimates $\hat{\beta}_1 = 4.50$, $\hat{\beta}_2 = 0$ and $\hat{\rho} = 0$. The β_1 confidence interval gives us $(3.99, 5.04)$ and the corresponding force of infection is $\lambda_{(1)} = 0.309$, $\lambda_{(2)} = 0.175$, $\lambda_{(3)} = 0.080$, $\lambda_{(4)} = 0.050$ and $\lambda_{(5)} = 0.069$. The CI are in table 4. $R_1 = 8.3$ with CI $[7.3, 9.3]$. All in all quite similar to the basic result when we included children up to 6 months, actually every minor change is of an increase in the parameters and the force of infection, where we would have expected a decrease.

Figure 8 illustrates result. Not much has changed in comparison with section 4.2.1. The second year the profile likelihood still does not reach the probability of being seropositive 95% confidence interval, indicating that this difference was not generated by the maternal immunity in the way we modeled it in this section. In section 4.2.4 we set things up a little different and expect to get better results for the second year of life. The fraction of infected individuals in the data has decreased for the first life year. This indicates that indeed maternal immunity plays a role in the titre, but it seems to have hardly any effect on the force of infection.

4.2.4 Maternal immunity using equation (2.12)

In this section we have used the alternation to our standard method described with the equation (2.12). The most notory underlying change is that the first half year is not used. So actually $\lambda_{(1)}$ now covers from 0.5 up till 5 years of age. Therefore there is a * next to it in the table and we introduce $\lambda_{(0)} = 0$ to cover up till the age of 6 months.
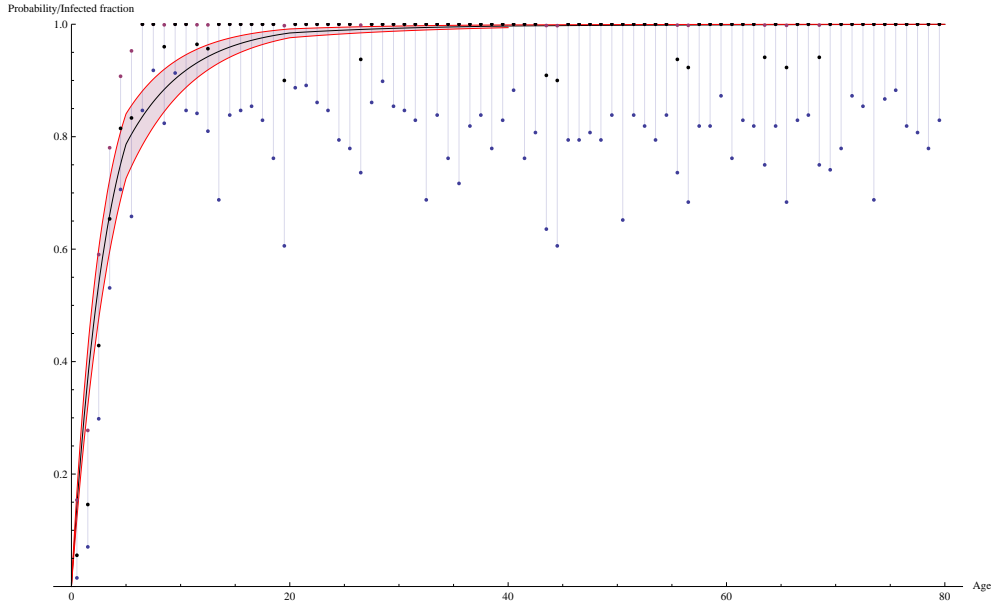
Figure 8: Estimated probability of infection (black line) with associated 95% confidence region (transparent purple area) for the first maternal immunity scenario, together with the data (cf Figure 3) adapted for the first year.

Table 5: Maximum likelihood estimate of the epidemiological parameters β_1, β_2, and ρ, with associated 95% confidence intervals. Also shown are derived parameters such as the forces of infection in the five age groups at the maximum likelihood estimate, the reproduction number of varicella R_1, and the mean age at infection with varicella $\bar{a}$. Here the data without individuals up to 6 months of age is used and implemented into a discretised version of equation (2.12).

Parameter	MLE	95%CI	Accuracy	Range
β_1	4.89	$(4.45, 5.60)$	$\frac{1}{100}$	3 - 6
β_2	0	$[0, \infty)$	$\frac{1}{10}$	0 - 4

ρ	0	$[0,\infty)$		$\frac{1}{3000}$	$\frac{9}{3000}$ - $\frac{15}{3000}$
$\lambda_{(0)}$	0				
$\lambda_{(1)}^*$	0.342	$[0.284,0.410)$	by β_1, β_2 and ρ		
$\lambda_{(2)}$	0.194	$(0.180,0.214)$	by β_1, β_2 and ρ		
$\lambda_{(3)}$	0.088	$(0.075,0.104)$	by β_1, β_2 and ρ		
$\lambda_{(4)}$	0.055	$(0.049,0.070)$	by β_1, β_2 and ρ		
$\lambda_{(5)}$	0.076	$(0.065,0.090)$	by β_1, β_2 and ρ		
R_1	9.0	$(8.1,10.4)$	by β_1		
$\bar{a}$	4.0	$[3.3,4.8]$	by the $\lambda_{(i)}$		

At first sight everything seems higher compared to section 4.2.1 and 4.2.3, but we should take care in interpreting these results. Because the first half year every newborn is immune there is less time in a lifespan to become infected, but our data has not changed compared to section 4.2.3. This means that we have the same amount of infected individuals, in a shorter timespan. Naturally $\lambda_{(1)}$ and β_1 rise. More surprising is that the rest of the force of infection also rises compared to sections 4.2.1 and 4.2.3. This should be explained by the fact that the model hangs on equation (2.12). When β_1 we expect the force of infection to rise along.

Figure 9 gives us the data of the fraction of seropositive individuals with CI and the probability to be seropositive as a function of age according to the estimates of the model. Note that the seropositive individuals of the first year are actually only of the second half year. Therefore we put the fraction of seropositives not halfway the first year, but at 3 quarters. We see that with this approach the second year data reaches the probability with the upper bound CI. Indicating that taking maternal immunity like this improves the model. For the 8th year this way of modeling does not change much.

It seems that this way of modeling with maternal immunity is an improvement over section 4.2.3, because our probability to become infected is closer to the data.
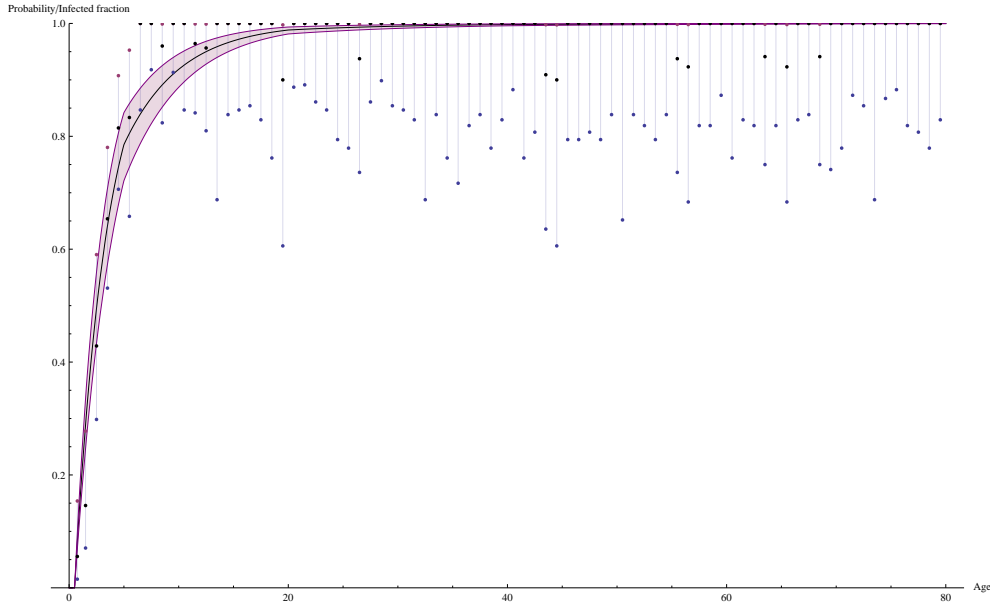
Figure 9: Estimated probability of infection (black line) with associated 95% confidence region (transparent lila area) for the second maternal immunity scenario using equation (2.12), together with the data (cf Figure 3) adapted for the first year.

5 Numerical method

For the model we described in section 2.2 we start with a description of a partial differential equations (PDEs) in section 2.1.5. But when the endemic equilibrium is introduced, the PDEs are let go. Because of this reason we are able to ignore the time spend in the infectious states for such a model and simply add to the force of infection. In a PDE this would mean that the recovery parameter would be ∞, but then infection could not happen. However the model is still closely related to PDE models. The point made is that the discretised PDE model of this chapter, that is meant to run simulations with, will coincide very closely with the model of section 2.2.

We use Iannelli's numerical method [2] to integrate the PDEs describing the dynamics of a reactivating pathogen in an age structured population. Our method and exposition closely follows [2]. In [2] the method is developed for the SIR and SIS type epidemic model. Here we extend the method for the SIR model in [2] to incorporate reactivation using the SI_1DI_2R model. The symbols he uses will be different than ours, because we want the symbols to be consistent with the rest of the thesis. However because the numerical model by [2] uses a lot of subscripts we use different notation for the SI_1DI_2R model. We use small letters for the states of the actual model and capital letters (but not exactly S, I_1, D, I_2 and R)

for the states of the numerical algorithm.

An overview of the PDEs is given on page 50. An explanation is given in the following sections.

5.1 Iannelli's method

5.1.1 Demography

First we set up the demography. We assume that the disease we model is non-lethal, nor does it induce less newborn individuals, hence the disease has no effect on the demography and we can model the demography independent of the disease. We have $n(a, t)$ as our true population function and N in the algorithm, which will be explained in the next section. The a and t are age and time respectively and both are measured in years. Throughout this section we will use small letters for the true functions of the states and capital letters for the numerical approximation. The initial age distribution of the population, $n(a, 0)$ is given by $n_0(a)$. The function $p(t)$ is the total population at time t and $b(a, p)$ is used for the birth rate function. Actually it is $b(a, p(t))$ but we will omit this double variable notation. The death rate function $\mu(a, p)$ is divided up in $m(a)$, dependent on age only and $M(a, p)$ dependent on both age and population size: $\mu(a, p) = m(a) + M(a, p)$. The maximum age of an individual is denoted a_d. In the simulations we will do later we use type one mortality, so the individuals all die at a chosen age. So the simulations will not use any death rate functions, they merely stop at the chosen age. This leads to the following system for the demography

$$
\begin{aligned}
\frac{\partial n}{\partial t}(a, t) + \frac{\partial n}{\partial a}(a, t) + \mu(a, p(t))n(a, t) &= 0 \\
n(a, 0) &= n_0(a) \\
n(0, t) &= \int_0^{a_d} b(a, p(t))n(a, t)\, da \\
p(t) &= \int_0^{a_d} n(a, t)\, da.
\end{aligned}
\tag{5.1}
$$

5.1.2 Numerical approximation of the demography

Let $T_f > 0$ be the final time and let K be the number of steps needed to arrive at T_f. So $u = \frac{T_f}{K}$ is the step size. Choose A_d such that $A_d u < a_d \leq (A_d + 1)u$ and set the age steps $a_j = ju$ for $0 \leq j \leq A_d$. We use a superscript for the time steps, this is not a power: $t^k = ku$, $0 \leq k \leq K$. For a function $f(a, t)$ we set $f_j^k = f(a_j, t^k)$ unless specified otherwise. To sum up all the functions we have seen up till now $n(a, t)$ is discretised by $N_j^k = n(a_j, t^k)$, $p(a)$ by P^k, $b(a, p)$ by b_j^k,

$m(a)$ by m_j and $M(a, p)$ by M_j^k. The latter two cover $\mu a, p$. Note that for b, and M we have no direct reference in the notation to the population variable. The way to compute those is $b_j^k = b(a_j, P^k)$ and $M_j^k = M(a_j, P^k)$. Then the algorithm wil compute the approximations $N_j^k = n(a_j, t^k)$ and $P^k = p(t^k)$. Initialize by:

$$N_j^0 = n_0(a_j)$$
$$P^0 = \sum_{j=0}^{A_d} N_j^0 u$$

The Riemann variant of the splitting algorithm of Iannelli *et al.* [2] approximates system (5.1) by:

$$\frac{N_j^{k-\frac{1}{2}} - N_{j-1}^{k-1}}{u} + m_j N_j^{k-\frac{1}{2}} = 0$$

$$\frac{N_j^k - N_j^{k-\frac{1}{2}}}{u} + M_j^{k-1} N_j^k = 0$$

$$N_0^k = \sum_{j=1}^{A_d} b_j^{k-1} N_j^k u \qquad (5.2)$$

$$P^k = \sum_{j=0}^{A_d} N_j^k u. \qquad (5.3)$$

The $N_j^{k-\frac{1}{2}}$ are variables to compute the algorithm, they have no meaning of their own. We eliminate $N_j^{k-\frac{1}{2}}$ from the equations and obtain the relation

$$N_j^k = \frac{1}{(1 + M_j^{k-1} u)(1 + m_j h)} N_{j-1}^{k-1}. \qquad (5.4)$$

The order of procedure is first (5.4) for ($1 \le j \le A_d$), then (5.2) to get the newborns and then (5.3) to initiate $k + 1$.

See Iannelli *et al.* [2] for convergence theorems and proofs.

5.1.3   Disease dynamics

Now we incorporate the SIR approximation scheme. We use $s(a, t)$ and $i(a, t)$ for the true functions of the susceptibles and infectious individuals, so we can use $S_j^k = s(a_j, t^k)$ and $I_j^k = i(a_j, t^k)$ in the algorithm. $s_0(a)$ and $i_0(a)$ are the

distributions at time t. We do not explicitly model the recovered state where individuals are immune, as $n(a,t) - s(a,t) - i(a,t)$ represents the density of recovered individuals of age a at time t. The constant q gives the fraction of newborns of infectious parents that are born into state i and, while the other fraction $1 - q$ will be born into s. The constant g gives the fraction of recovered individuals that are born into the recovered state. (Both q and g will be zero throughout the simulations later on and the author wonders whether a positive g is ever realistic, for it is not the same as temporary maternal immunity, but gives lifelong immunity.). For the disease dynamics we have $\lambda(a, i(\cdot, t))$ as the force of infection, dependent on age a in years and the total infectious population at time t. It is given by $\lambda(a, i) = \lambda(a, i(\cdot, t)) = \frac{\int_0^{a_d} h(a,a^*) i(a^*, t)\, da^*}{p(t)}$ with $h(a, a^*)$ the general infection rate depending on both the age of the recipient a and the age of the infectious individual a^*. [2] also introduces a γ, but we will not use it here. Finally we have $\alpha(a)$ as the recovery rate dependent on age a of the infectious individual.

$$
\begin{aligned}
\frac{\partial s}{\partial a}(a,t) + \frac{\partial s}{\partial t}(a,t) + \mu(a, p(t))s(a,t) &= -\lambda(a,i)s(a,t) \\
s(a,0) &= s_0(a) \\
s(0,t) &= \int_0^{a_d} b(a, p(t))(s(a,t) + (1-q)i(a,t) + \\
& \qquad (1-g)(n(a,t) - s(a,t) - i(a,t)))\, da \\
\frac{\partial i}{\partial a}(a,t) + \frac{\partial i}{\partial t}(a,t) + \mu(a, p(t))i &= \lambda(a,i)s(a,t) - \alpha(a)i(a,t) \\
i(a,0) &= i_0(a) \\
i(0,t) &= q \int_0^{a_d} b(a, p(t))i(a,t)\, da
\end{aligned}
$$

5.1.4   Numerical approximation of the SIR model

To simulate we discretise again. $s(a,t)$ and $i(a,t)$ are approximated by S_j^k and I_j^k with starting values $s_0(a_j)$ and $i_0(a_j)$. Furthermore $\lambda(a,i)$ is approximated by Λ_j^k, and $\alpha(a)$ by $\alpha_j = \alpha(a_j)$. $h(a, a^*)$ is discretised $h_{j,l} = h(a_j, a_l)$ We then have the following algorithm

46

$$I_j^0 = i_0(a_j)$$

$$S_j^0 = s_0(a_j)$$

$$\frac{S_j^{k-\frac{1}{2}} - S_{j-1}^{k-1}}{u} + m_j S_j^{k-\frac{1}{2}} = 0 \tag{5.5}$$

$$\frac{S_j^k - S_j^{k-\frac{1}{2}}}{u} + M_j^{k-1} S_j^k = -\Lambda_{j-1}^{k-1} S_j^k \tag{5.6}$$

$$S_0^k = \sum_{j=1}^{A_d} b_j^{k-1}(S_j^k + (1-q)I_j^k +$$

$$(1-g)(N_j^k - S_j^k - I_j^k))u \tag{5.7}$$

$$\frac{I_j^{k-\frac{1}{2}} - I_{j-1}^{k-1}}{u} + m_j I_j^{k-\frac{1}{2}} = 0 \tag{5.8}$$

$$\frac{I_j^k - I_j^{k-\frac{1}{2}}}{u} + (M_j^{k-1} + \alpha_j)I_j^k = \Lambda_{j-1}^{k-1} S_j^k \tag{5.9}$$

$$I_0^k = q\sum_{j=1}^{A_d} b_j^{k-1} I_j^k u$$

$$\Lambda_j^k = \frac{\sum_{l=0}^{A_d} h_{j,l} I_l^k u}{uP^k}. \tag{5.10}$$

The $I_j^{k-\frac{1}{2}}$ and the $S_j^{k-\frac{1}{2}}$ are variables for the computation, they have no meaning on their own. These variables can be eliminated.

$$I_j^k = \frac{1}{(1 + u(M_j^{k-1} + \alpha_j))(1 + um_j)}(I_{j-1}^{k-1} + u(1 + um_j)\Lambda_{j-1}^{k-1} S_j^k) \tag{5.11}$$

$$S_j^k = \frac{1}{(1 + u(M_j^{k-1} + \Lambda_{j-1}^{k-1}))(1 + um_j)} S_{j-1}^{k-1} \tag{5.12}$$

So (5.5) and (5.6) are replaced by (5.11) and (5.8) and (5.9) by (5.12). Then the algorithm can be run in this order (so with replacements in place), where we take one step of k at the time doing all the j of that step. This means that for (5.11), (5.12) and (5.10) we run a full cycle of j for each k. There is one exception to the order described and that is that one cycle of j of Λ_j^0 must be run immediately after the two starting equations. We need it for $k = 1$ in the other equations. Of course the two starting equations are run only once for a full cycle of j.

This completes the Riemann version of the SIR model described in Iannelli *et al.* [2]. Next we will setup a natural extension to reactivation.

5.1.5 Extension to incorporate reactivation

In this section we set up the SI_1DI_2R model. The notation for the states is as follows, the small letters are for the true functions of the model: $s(a,t) = S(a,t)$, $i(a,t) = I_1(a,t)$, $d(a,t) = D(a,t)$ and $j(a,t) = I_2(a,t)$.[16] The recovered state where individuals are immune is $n - s - i - d - j$. First we need a new parameter function for our next step. Later on we introduce more parameters. $\bar{h}(a, a^*)$ is the general transmission rate function depending on the age of the recipient a and the age of the reactivated individual a^*. The discretisation is $\bar{h}_{j,l} = \bar{h}(a_j, a_l)$ for $\bar{h}(a, a^*)$. We have no parameters that bring a fraction of the newborns of states d and j to their own state upon birth. It is not so clear in which state newborns should get when having seropositive parents. One could think that a reactivated individual gives birth to a first time infected individual or a general seropositive individual could give birth do a dormant or immune individual. We also adopt the convention that $q = g = 0$ and with that all newborns are born susceptible. We will use this in the simulations and it is consistent with the new states. A convenient side effect is that the formulas for newborns are cleaner than with new parameters. With these additions little of the original algorithm will change, they do of course effect most equations and hence states, but not visibly in the formula. $\lambda(a, i)$ will be replaced by $\lambda(a, (i, j))$ because of the new state j, but the discrete approximation will still be denoted by Λ_j^k. From the equations we already had, 3 of them change. Namely the equations for $s(0, t)$, $i(0, t)$ and for $\lambda(a, i)$, which changes in notation as said.

$$
\begin{aligned}
s(0, t) &= \int_0^{a_d} b(a, p(t)) n(a, t)\, da \\
i(0, t) &= 0 \\
\lambda(a, (i, j)) &= \lambda(a, (i(\cdot, t), j(\cdot, t))) \\
&= \frac{\int_0^{a_d} h(a, a^*) i(a^*, t) + \bar{h}(a, a^*) j(a^*, t)\, da^*}{p(t)}
\end{aligned}
$$

Their discretised counterparts are:

$$
\begin{aligned}
S_0^k &= \sum_{j=1}^{A_d} b_j^{k-1} N_j^k u \\
I_0^k &= 0 \\
\Lambda_j^k &= \frac{\sum_{l=0}^{A_d} (h_{j,l} I_l^k + \bar{h}_{j,l} J_l^k) u}{u P^k}.
\end{aligned}
$$

---

[16]the function j is not to be confused with the index $_j$ and the function d is not to be confused with the integration d.

Now the additional parameters will be introduced. $\rho(a)$ is the reactivation rate function dependent on the age a in years of the dormant individual. $\bar{\alpha}(a)$ is the recovery function dependent on age a reactivated individuals. The new states will have the following equations.

$$
\begin{aligned}
\frac{\partial d}{\partial t}(a,t) + \frac{\partial d}{\partial a}(a,t) + \mu(a,p(t))d(a,t) &= \alpha(a)i(a,t) - \rho(a)d(a,t) \\
d(a,0) &= d_0(a) \\
d(0,t) &= 0 \\
\frac{\partial j}{\partial t}(a,t) + \frac{\partial j}{\partial a}(a,t) + \mu(a,p(t))j(a,t) &= \rho(a)d(a,t) - \bar{\alpha}(a)j(a,t) \\
j(a,0) &= j_0(a) \\
j(0,t) &= 0
\end{aligned}
$$

The complete list of the PDEs is on page 50

5.1.6 Discretisation of the reactivation

The functions $\bar{h}(a,a^*)$ and $\lambda(a,(i,j))$ are discretised by $h_{j,l}$ and Λ_j^k. The others are $\rho_j = \rho(a_j)$ and $\bar{\alpha}_j = \bar{\alpha}(a_j)$. When we implement them into the algorithm we get the discretised states of the individuals with a dormant infection and the reactivated individuals

$$
\begin{aligned}
D_j^0 &= d_0(a_j) \\
J_j^0 &= j_0(a_j) \\
\frac{D_j^{k-\frac{1}{2}} - D_{j-1}^{k-1}}{u} + m_j D_j^{k-\frac{1}{2}} &= 0 \\
\frac{D_j^k - D_j^{k-\frac{1}{2}}}{u} + (M_j^{k-1} + \rho_j)D_j^k &= \alpha_j I_j^k \\
D_0^k &= 0 \\
\frac{J_j^{k-\frac{1}{2}} - J_{j-1}^{k-1}}{u} + m_j J_j^{k-\frac{1}{2}} &= 0 \\
\frac{J_j^k - J_j^{k-\frac{1}{2}}}{u} + (M_j^{k-1} + \bar{\alpha}_j)J_j^k &= \rho_j D_j^k \\
J_0^k &= 0.
\end{aligned}
$$

Like before with the other states the $D_j^{k-\frac{1}{2}}$ and $J_j^{k-\frac{1}{2}}$ are only to help with the calculations and can be eliminated.

$$D_j^k = \frac{1}{(1 + u(M_j^{k-1} + \rho_j))(1 + um_j)}(D_{j-1}^{k-1} + u(1 + um_j)\alpha_{j-1}I_j^k)$$

$$J_j^k = \frac{1}{(1 + u(M_j^{k-1} + \bar{\alpha}_j))(1 + um_j)}(J_{j-1}^{k-1} + u(1 + um_j)\rho_{j-1}D_j^k)$$

The order of algorithm should be run in the order of the states, for the rest all will go similar as discussed before in section 5.1.

5.1.7 Overview of the PDEs

The PDEs together are given below without the initial conditions and the boundary conditions which equal 0.

$$\frac{\partial n}{\partial t}(a, t) + \frac{\partial n}{\partial a}(a, t) + \mu(a, p(t))n(a, t) = 0$$

$$n(0, t) = \int_0^{a_d} b(a, p(t))n(a, t)\,da$$

$$p(t) = \int_0^{a_d} n(a, t)\,da$$

$$\lambda(a, (i, j)) = \frac{\int_0^{a_d} h(a, a^*)i(a^*, t)\,da^*}{p(t)} +$$
$$\frac{\int_0^{a_d} \bar{h}(a, a^*)i(a^*, t)\,da^*}{p(t)}$$

$$\frac{\partial s}{\partial a}(a, t) + \frac{\partial s}{\partial t}(a, t) + \mu(a, p(t))s(a, t) = -\lambda(a, (i, j))s(a, t)$$

$$s(0, t) = \int_0^{a_d} b(a, p(t))n(a, t)\,da$$

$$\frac{\partial i}{\partial a}(a, t) + \frac{\partial i}{\partial t}(a, t) + \mu(a, p(t))i(a, t) = \lambda(a, (i, j))s(a, t) - \alpha(a)i(a, t)$$

$$\frac{\partial d}{\partial t}(a, t) + \frac{\partial d}{\partial a}(a, t) + \mu(a, p(t))d(a, t) = \alpha(a)i(a, t) - \rho(a)d(a, t)$$

$$\frac{\partial j}{\partial t}(a, t) + \frac{\partial j}{\partial a}(a, t) + \mu(a, p(t))j(a, t) = \rho(a)d(a, t) - \bar{\alpha}(a)j(a, t)$$

5.1.8 Simplifications

For the simulations our model will have a number of simplifications. First we have type one mortality. The functions μ, m and M will no longer be necessary and will be set to zero and to model death we simply use A_d as age of death. The birth

rate function $b(a, p)$ will be set constant, effectively making S_0^k constant. This means that after a generation the population is automatically constant and every age is equally represented. In our simulations we will set the initial population at this constant. So $N_j^k = c$ for a constant c, depending on the total population size and step size u, for all j and k. Furthermore the parameter functions $\alpha(a)$, $\bar{\alpha}(a)$ and $\rho(a)$ will be set constant.

5.2   Simulations

5.2.1   Matching the force of infection model with the Iannelli model

The first deterministic simulations of the Iannelli method, Figures 10 through 15, are to check the correctness of our model. These are based on the maximum estimate. So $\beta_1 = 4.49$, $\beta_2 = 0$ and $\rho = 0$ from section 4.2.1 for most of them. Additional assumption is that the infectious period has a mean of 5 days. This comes back in $\beta_1 = hT$, where we fill in $T = 5$ days and determine h with that. The method converges to an equilibrium. It is the equilibrium we compare to the results of the model. We choose a population of 16 million individuals, but the fractions are independent of population size so we will not see any effect of that particular choice.

In figures 10 end 11 we have done simulations with steps of 1 resp. $\frac{1}{4}$ year where we filled in the β_1 of the first maximum likelihood estimate, so $\beta_1 = 4.49$ and $\beta_2 = \rho = 0$. In figure 10 we see that the simulation approximately follows the given seropositive probability, in figure 11 we see that the simulation is off. This raises the question why the sharper simulation is more off from the model of section 4.1.3 and before. An explanation can be found in the fact that age steps of 1 year is much closer to the age steps described in section 4.1.3 we used in equation (2.17) to find our β_1, than age steps of $\frac{1}{4}$ year is. We should take into account that the relation of the $\lambda_{(i)}$ and β_1, β_2 and ρ is dependent on the chosen age steps in section 4.1.3.

To see whether the model and the simulation method coincide when the age steps are taken the same we have used age steps of $\frac{1}{4}$ year in both in Figure 12. We have taken $\beta_1 = 4.49$ and $\beta_2 = \rho = 0$, the first maximum likelihood estimate. The model needs roughly a factor 1000 more time to calculate the force of infection for this particular point, making it take 2 minutes instead of $\frac{1}{8}$ second. So this is a checkup and will not become routine. The simulation and the model coincide closely indicating that the difference noted before is indeed due to the difference in age steps. The strongest difference is found in the neighbourhood of the age of 5. This is possibly because an individual whom gets his 5th birthday during his infectious period does not change his contact pattern in the model whereas an individual in the simulation does. This check indicates parameters should
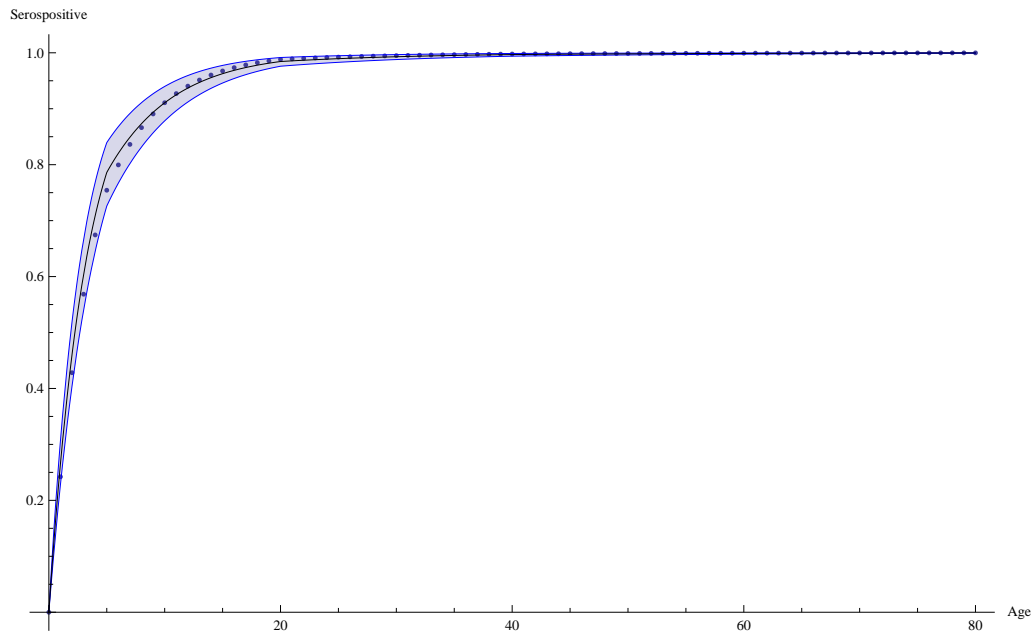
Figure 10: The probability of being seropositive 95% CI together with a simulation (dots) with steps of 1 year against age.
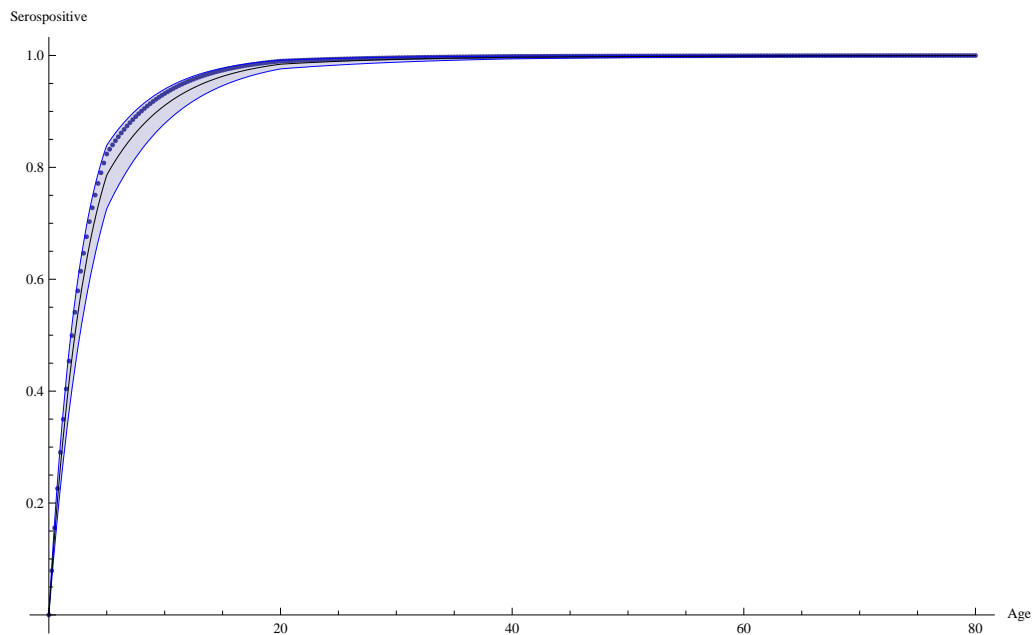

Figure 11: The probability of being seropositive 95% CI together with a simulation (dots) with steps of $\frac{1}{4}$ year against age.
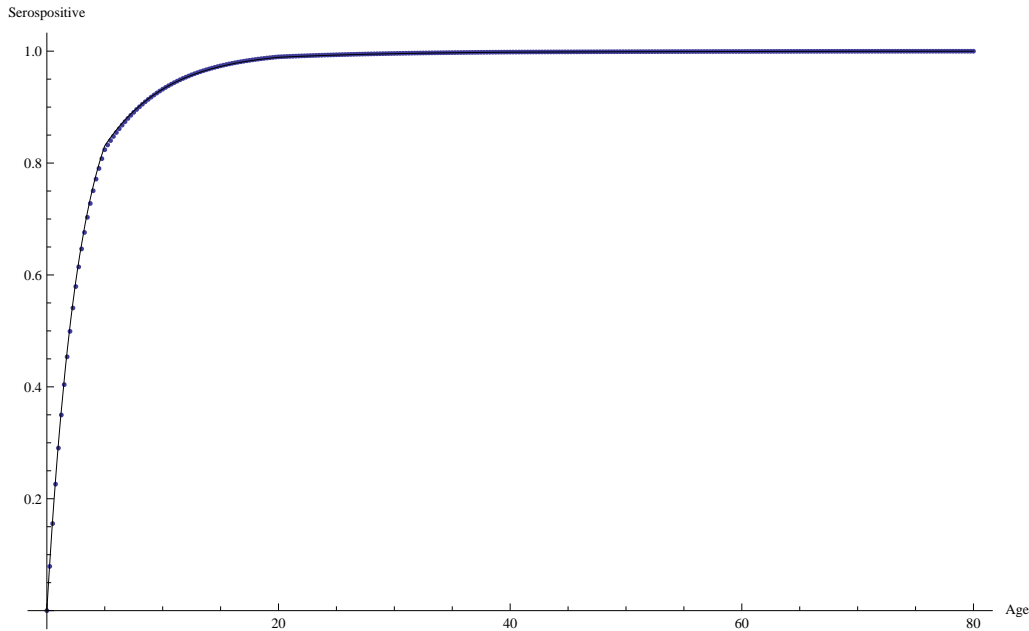
Figure 12: Probability of being seropositive (line) together with a simulation (dots), both in age steps of $\frac{1}{4}$ against age.

always be taken into account with respect to equation (2.17). The parameters are estimated through the force of infection. So the force of infection should stay the same for different age steps, but the parameters should vary.

A second check with age steps of $\frac{1}{8}$ is seen in Figure 13. It gives a slightly better accuracy indicating that the change of contact pattern during infectious period may not have such an impact after all. Still the greatest difference (which is small) between the model and the simulation is around the age of 5. The duration for the computation for the model increased by approximately a factor 8 and the simulation cannot handle much smaller steps at the computer we used.

In Figure 14 we have a check with reactivation, we take $\beta_2 = 4$ and $\rho = 0.005$. We have chosen the strongest variant of reactivation with respect to our grid to make reactivation have an impact in the disease dynamics. This check also coincides closely indicating that the reactivation part of both the model and the simlation are correct too.

In Figure 15 we have the same check, now with age steps of $\frac{1}{8}$. The result is better again with smaller age steps.

All checks with take age steps for both models identical indicate that the models coïnside closely and we assume they do it always based on these examples.
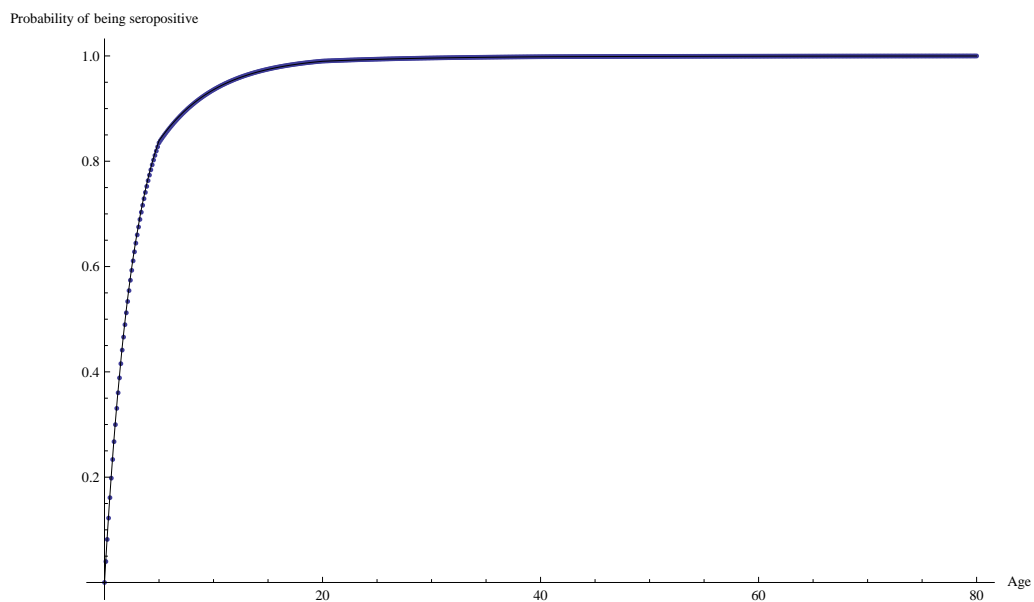
Figure 13: Probability of being seropositive (line) together with a simulation (dots), both in age steps of $\frac{1}{8}$ against age.
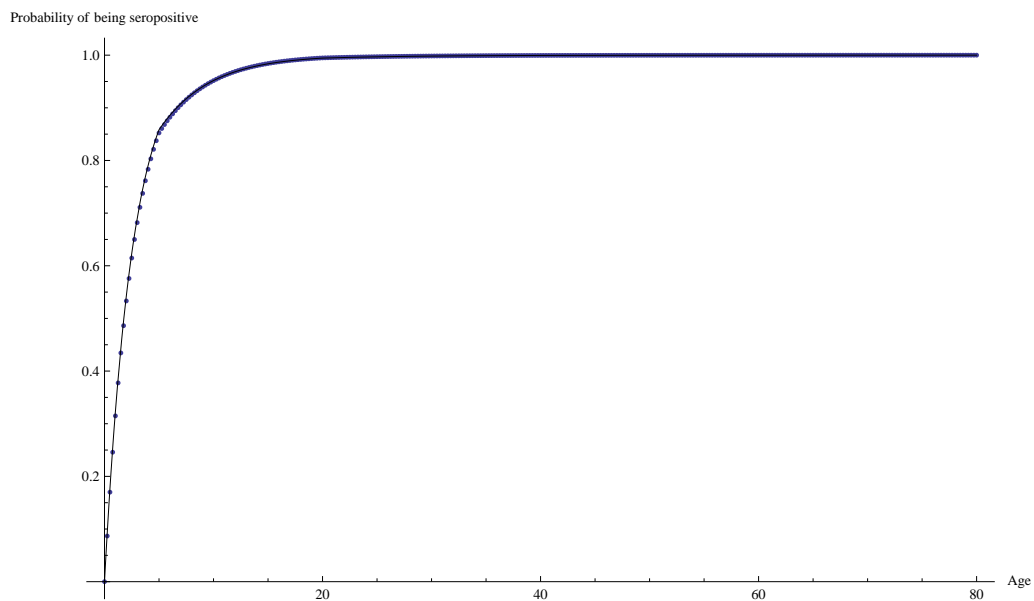

Figure 14: Probability of being seropositive (line) together with a simulation (dots), both in age steps of $\frac{1}{4}$ and reactivation against age.
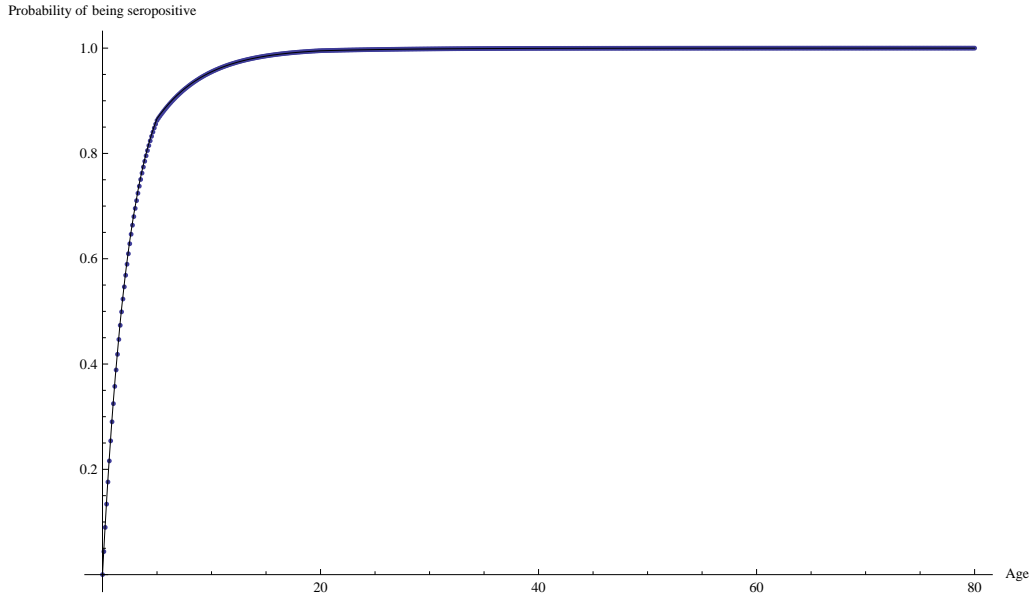
54

Figure 15: Probability of being seropositive (line) together with a simulation (dots), both in age steps of $\frac{1}{8}$ and reactivation against age.

The force of infection is directly fitted into the likelihood function. So even if we use a larger age step size the force of infection remains roughly the same. The Iannelli approximation closely follows the model when the same step size is used. This indicates that with a large step size still a reliable simulation may be made.

5.2.2 Simulations with vaccination

Here we have a simulation with the Iannelli method. We used the results from section 4.2.2. So we have time/age steps of $\frac{1}{2}$ year and $\beta_1 = 4.17$. We choose $\rho = 0.004$ which a realistic option in the sense that 23% to 30% of the individuals should reactivate. The choice $\beta_2 = 1$ is more random (and probably to large if we look at the low amount of known infections due to herpes zoster) but still lies within the 95% confidence region if we take it together with $\rho = 0.004$. To look at changes in time we have taken the time dynamics of individuals of 3, 10, 30, 50 and 70 year old (approximately the middle of an age contact group) and let it run over 270 years. We have taken up a 90% vaccination coverage[17] for newborns from the 50th year of our simulation till the 150th year and we assume a perfect vaccine. Note that vaccination in this simulation means that individuals are simply born into the in section 5.1.5 mentioned recovered state where they are immune. So we see a drop of susceptibles for the young ages because only 10% of the individuals are born as susceptibles during the vaccination period. The

---

[17]So 90% of the newborns are vaccinated.

simulations are in Figures 16, 17 and 18, containing a total of 20 graphs. The 4 explicit states of section 5.1.5 where the Iannelli method is described are included, so the susceptible, the first time infectious, the dormant infected and the reactivated infectious individuals. The situation at time $t = 0$ is an approximate equilibrium state. We actually started 50 years earlier ($270 + 50 = 320$ years is 4 full generations) with a single infectious individual. To prevent any confusion and to show the important bits better we do not show these 50 years. When studying the graphs, pay close attention to the scaling of the vertical axis, it is different almost every time.

The graphs clearly show that a 90% vaccine coverage is sufficient to obtain herd immunity, which was to be expected, since $R_1 \approx 7.5$ (with CI < 8.7, see section 4.2.2) and reactivation should not have a to big impact on R_0, since only 23-30% reactivates and $\beta_2 \approx \frac{\beta_1}{4}$, furthermore reactivation occurs not mainly to young individuals, like infection, but is more spread out and young individuals have a higher contact rate in general, see section 3.2. The vaccination has much impact on young individuals as the graphs display, seeing an almost immediate drop in all 4 graphs for 3 year olds. The effect to 10 year olds seems less, but considering the dormant and second time infectious individuals results are still quite significant. Also there is a natural delay of 10 years after vaccination before effects become visible. This coincides with the time that the first vaccinated individuals become 10 years old, hence the term 'natural delay'. The 30 year olds actually have a much larger fraction of susceptibles some years after vaccination and we see a slight peak of 30 years old infectious individuals after vaccination. This is an important age because this is the fertile age for women amongst the ages we chose. Chickenpox during pregnancy increases the risk of complications. To have a better look we have a magnified graph of infectious individuals of 30 year old in Figure 19. There we see that the fraction of infectious individuals increases above original level until the first vaccinated individuals turn 30. Then the fraction of infected individuals decreases below original level and keeps decreasing until vaccination stops. The vaccination stop induces a relatively large outbreak amongst 30 year old individuals in a few years, oscillating back to the original level in 50 years. The dropping below the original level is good, but the increase over the first 30 years is troublesome. A known strategy is to vaccinate women of fertile age during the first years until the newborn vaccination takes over. The magnified graphs of infectious individuals of 50 resp 70 year old individuals (not shown here) show a similar behavior, but the peaks are higher, especially for the 50 year olds. The 70 year olds do not drop below original of infectious level at all. The fractions may have lost a part of their meaning at such low levels though. (i.e. it should be modeled stochastically, because the fractions may decrease below 1 over the population size, meaning we are talking about not even 1 individual, but it is still an average over time). Also the individuals of ages 50 and 70 get more susceptible individuals due to vaccination and the drop of dormant and

Figure 16: The time dynamics of the 3 year olds and the 10 year olds. Started form an equilibrium state without vaccination. Perfect vaccine coverage 90% from year 50 till 150. The input of parameters is given in page 55
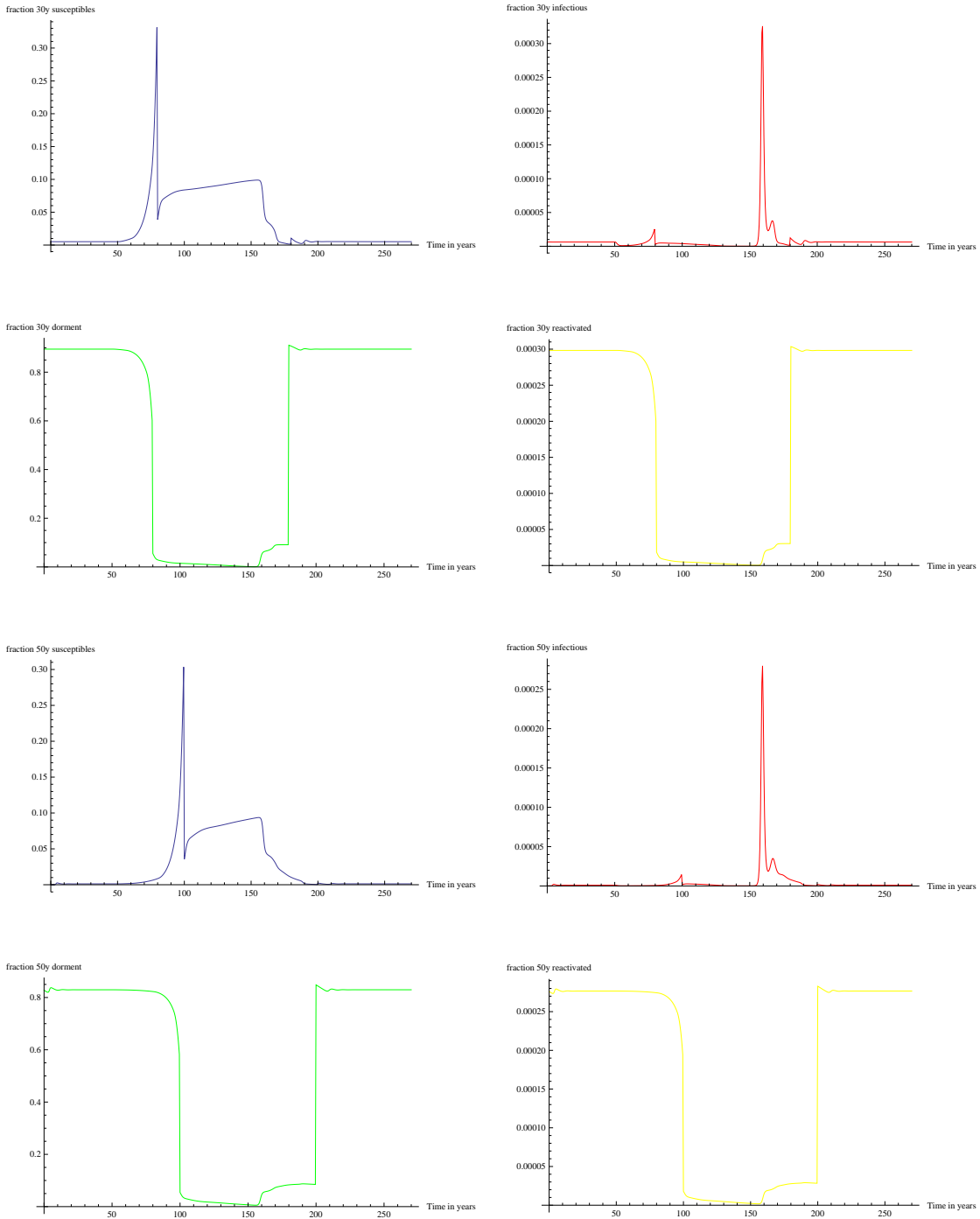
Figure 17: The time dynamics of the 30 year olds and the 50 year olds. Started form an equilibrium state without vaccination. Perfect vaccine coverage 90% from year 50 till 150. The input of parameters is given in page 55
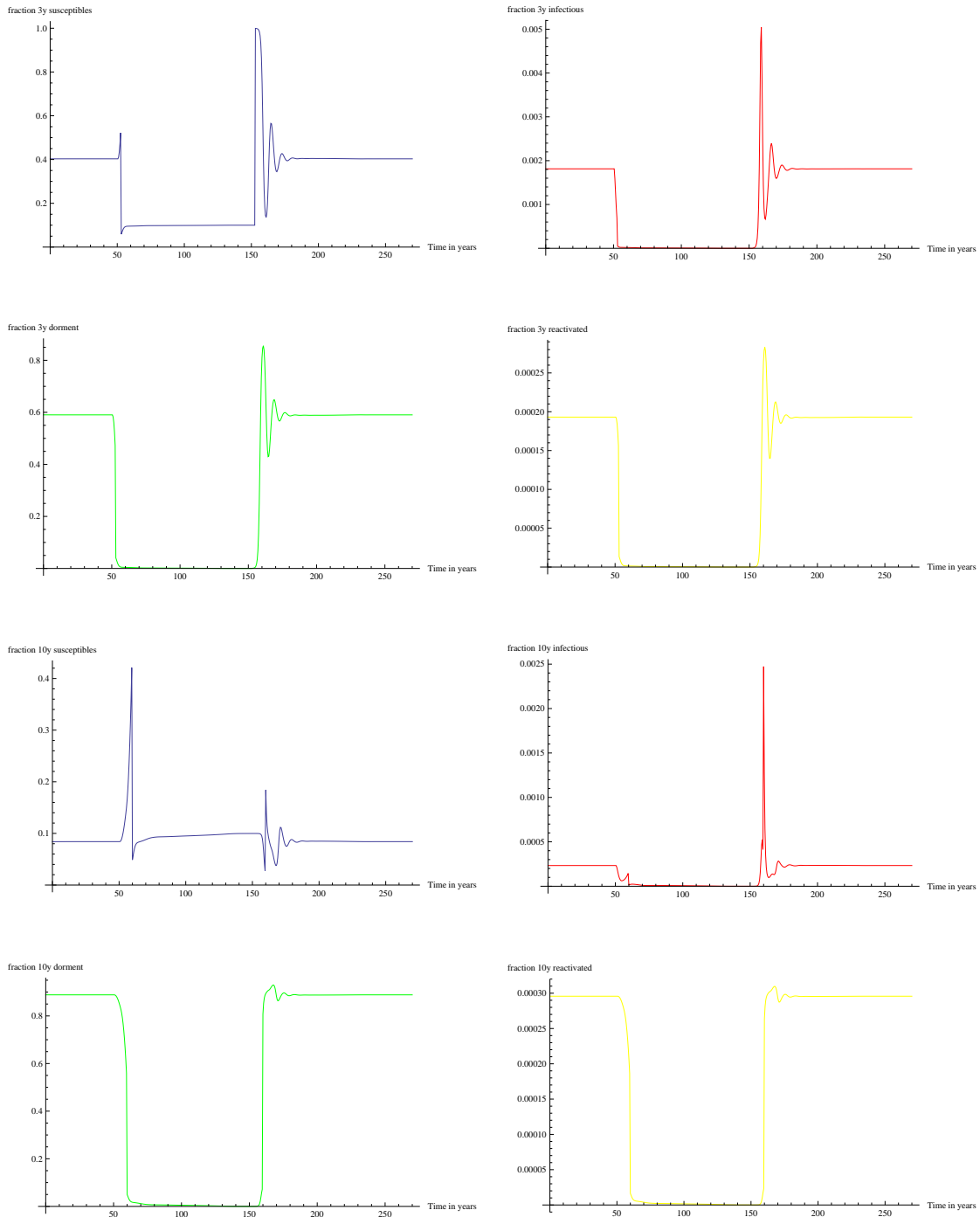
58

Figure 18: The time dynamics of the 70 year olds. Started form an equilibrium state without vaccination. Perfect vaccine coverage 90% from year 50 till 150. The input of parameters is given in page 55

59

Figure 19: Fraction of infectious individuals of 30 years old, magnified around the lower levels. Started form an equilibrium state without vaccination. Perfect vaccine coverage 90% from year 50 till 150. The input of parameters is given in page 55
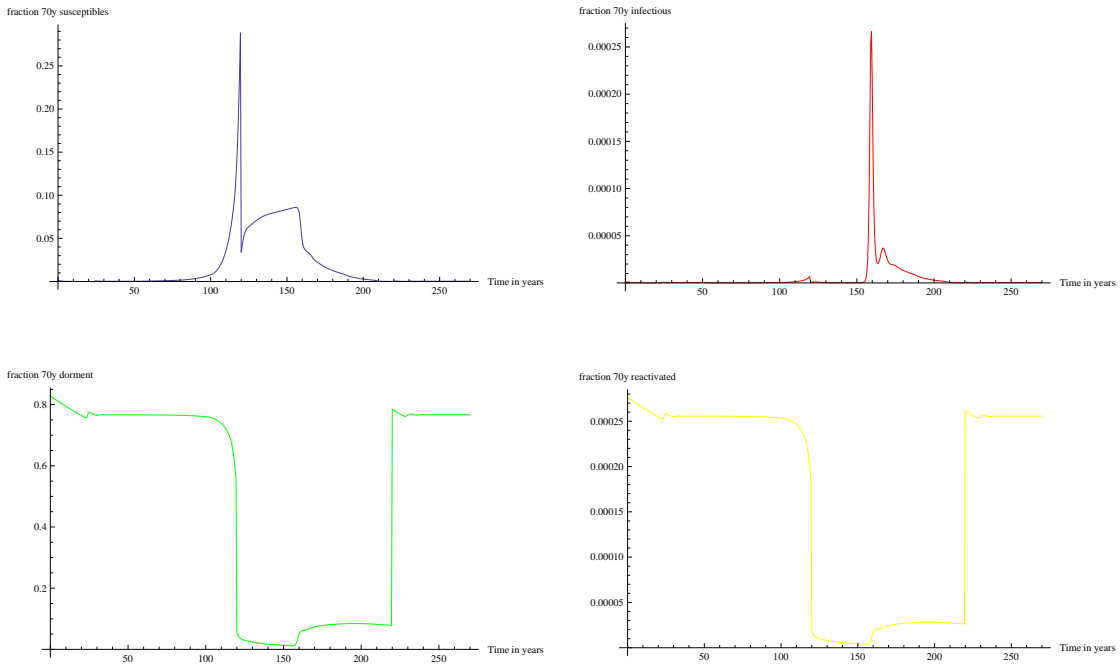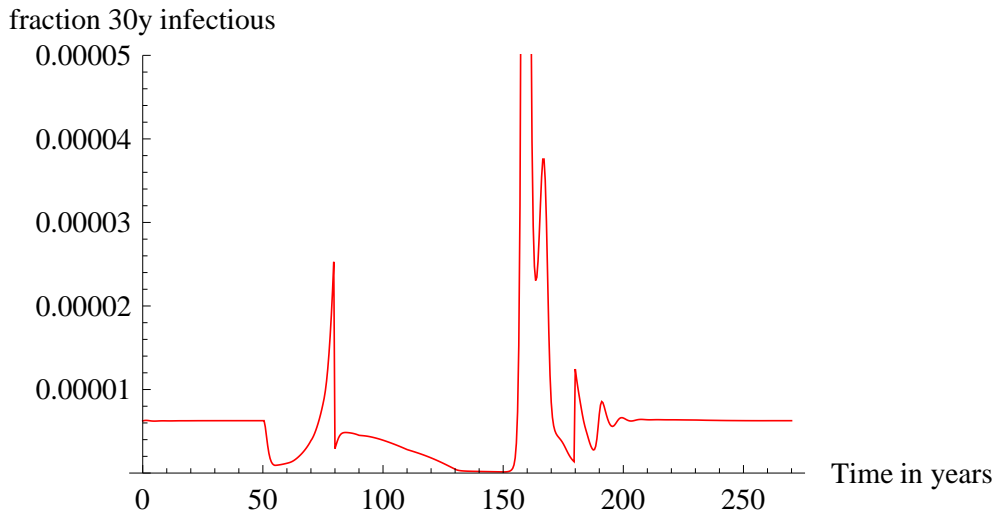
reactivated individuals drop 50 resp 70 years after the first vaccination round. It would be interesting to make reactivation dependent on the force of infection to see whether this has a great impact on reactivation (see section 6.4 for ideas how to do it.). Now we see no change in the reactivation of individuals until the vaccinated individuals reach the ages we are looking at. There have been serological studies on this particular subject (even wikipedia:herpes zoster mentions some).

If we take a 80% coverage perfect vaccine we should have to little. We use it until the 100th year, then starting using it again from year 200. To illustrate that it is not enough we have in Figure 20 the 30 year old susceptible population as a function of time. The fraction of susceptibles is approximate 0.1 where it should approximate 0.2 if the vaccine would have covered a greater fraction of the newborns than $(1 - \frac{1}{R_0})$. In Figure 21 we have an approximate steady state of seropositives (excluded vaccinees) with an 80% vaccination program. Here we also see that a significant portion of the remaining (after vaccination) susceptible population still gets infected and vaccination is not sufficient to eradicate the virus. Furthermore many individuals become infected at older age and that is a worst case scenario for the host. Chickenpox is more severe and painful at later ages. Here we can clearly see that when vaccination is applied, we have to make sure that the coverage is sufficient. If the vaccination does not reach the critical threshold, money is spend on making things worse.
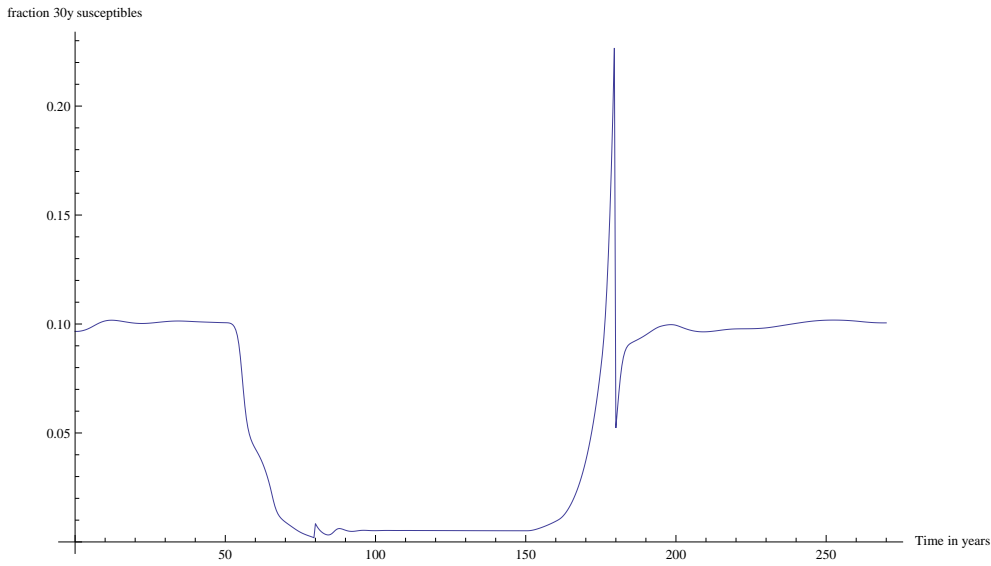
Figure 20: 30 year old susceptible with 80% perfect vaccination till 100 years in time and from 200 years onward as a function of time. $\beta_1 = 4.17$, $\beta_2 = 1$ and $\rho = 0.004$
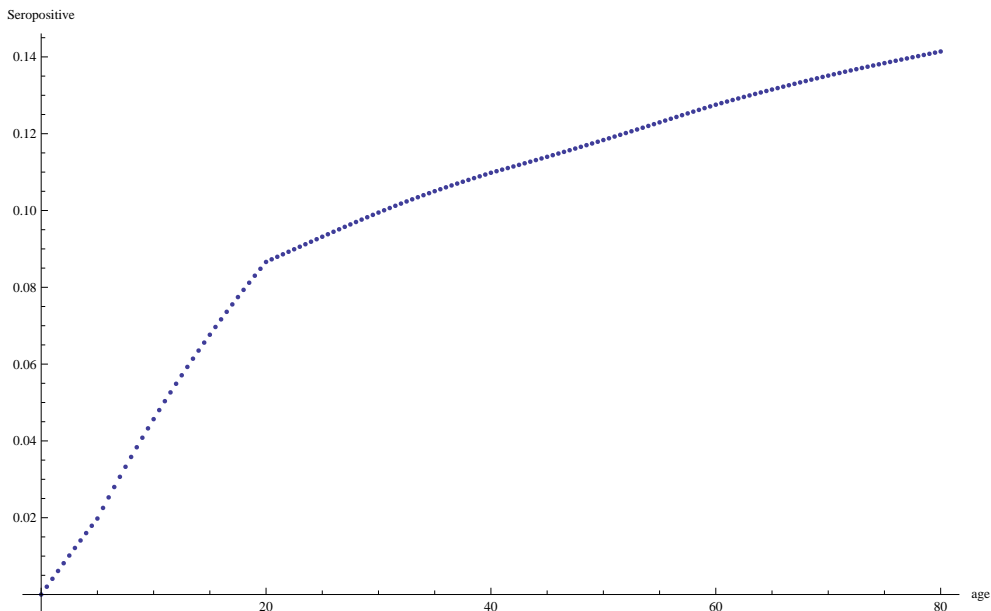

Figure 21: Steady state of the seropositive fraction of the population when a 80% coverage perfect vaccine is applied. Vaccinated individuals not counted. $\beta_1 = 4.17$, $\beta_2 = 1$ and $\rho = 0.004$

6 Discussion

6.1 Transmission dynamics of varicella zoster virus.

The main aim of this research project is to obtain quantitative insight in the transmission dynamics of VZV in the Netherlands, and to quantitatively investigate the likely impact of vaccination strategies. To this end, we have formulated a model in section 2.2 centered around the equilibrium force of infection for the age-specific incidence of infection with VZV and incidence of zoster. We have introduced age-specific data about the infection status of individuals obtained by a survey and age-dependent contact patterns. We have made estimates on a model-based likelihood of the parameters, given the data.

The results of the estimation procedure indicate that VZV is highly transmissible. Our estimates of the reproduction number R_1 vary from 7.7 (6.8,8.6 CI) in the smaller age step scenario 4.2.2 to $R_1 = 9.0$ (8.1,10.4 CI) in the second maternal immunity scenario 4.2.4. The base scenario returns $R_1 = 8.3$ (7.3,9.3 CI). The implication of this result is that already 90% of the 10 year old individuals are seropositive. The smaller age steps indicate that R_1 is overestimated, but the force of infection λ remains approximately equal for a different time step size and the first maternal immunity scenario. The second maternal immunity scenario, where individuals up to 6 months old are completely taken out of consideration (maternal immune or no) estimates a higher force of infection, because the same amount of infections must take place in a shorter time span.

Our results also show that the reactivation rate and transmissibility of zoster cannot be estimated with precision separately. If fact the MLE gives either no reactivation or no transmissibility of herpes zoster, leaving the confidence intervals unchecked. When we take them together the contribution of herpes zoster to new infections has an upper bound of 7.3%. This implies that R_0 is not much higher than R_1 and the transmission of herpes zoster is in itself not powerful enough to sustain VZV in the Dutch population. Note that this is also because of the chosen range of β_2 and ρ considered reasonably a priory, since the corresponding β_2 is a boundary point.

6.2 Vaccination strategies

A large scale program against varicella zoster virus in children can serve several goals, not necessarily mutually exclusive. First, the goal may be to reduce infection and the burden of disease in children. Second, the goal may be to reduce the burden of disease of herpes zoster.

Our simulations based on the estimated parameters show that a large-scale vac-

cination program with a sufficiently effective vaccine will be able to reduce the circulation of varicella, and therewithin reduce the overall probability of infection, which leads to an overall reduction of varicella and herpes zoster cases.

However, the analyses also show that the mean age at infection will increase with decreasing circulation of varicella. If infections in older persons are associated with more serious disease, then vaccination may actually increase the burden of disease, rather than decreasing it. It depends on whether the absolute number of infections in older individuals increases. An example of worsening of the situation by vaccination is found in Greece, where a vaccination program for rubella was introduced. The vaccination increased the number of infections amongst individuals of fertile age resulting in serious pregnancy and birth complications and damage to the unborn life. Such complications of large-scale vaccination programs may be especially severe if the vaccination effort is not sufficient to achieve elimination of the pathogen, as illustrated in the simulations.

The simulations also indicate that the groups that are just too old to be included in the vaccination program experience most of the negative effects of the vaccination campaign. They may benefit from a catch up vaccination program to be vaccinated before fertile age.

Furthermore, even with a sufficient coverage vaccination program reducing the incidence of herpes zoster is a long term goal. When vaccination is introduced the majority of the population still carries the varicella zoster virus dormantly. Reactivation and dormant individuals will steadily decrease during approximately a generation.

Terminating the vaccination program would probably result in a substantial fraction of infectious individuals for years, and is therefore not recommended.

6.3 Assumptions and limitations

The model we use in this thesis is relatively simple. Not all necessarily limiting, the assumptions we use are

- Individuals are born at a constant rate and die at age of 80. Therefore every age is equally represented. This can certainly be improved for the Dutch population, where there are more elderly individuals than young ones.

- The infections status is in equilibrium, while local outbreaks usually go with waves. This mainly effects the young individuals because older individuals have usually already experienced multiple waves.

- Every individual is equally susceptible and contagious. This is of course

not true on individual level, but over a large population this is a convenient average.

- We use a fixed contact pattern. It is a given estimate, but we may be able to estimate it ourselves, for instance by applying Goeyvaerts contraints to the contacts.

- The contact pattern is only dependent on age.

- The parameter ρ for reactivation is constant for all ages and independent of the force of infection. In reality ρ may change depending on age and be dependent on other influences like the force of infection. In section 6.4 several possible extensions are presented.

- The cut off for determination whether an individual in the data is seropositive is fixed, there are no false positive or false negative individuals.

- We have assumed the data sample out of the population to be random. Actually the sample was from a low vaccinee area in the Netherlands.

- The discretisation of the model approximates the true situation.

6.4   Avenues for future research

For the assumption that the rate of reactivation is given by a fixed parameter ρ there exist several extensions. A first step would be to take parameter ρ and make it dependent on age, so it becomes a parameter function $\rho(a)$, like in the general description of the model. A few further extensions are mentioned below.

There are surveys (wikipedia: herpes zoster, with source references) that indicate that the number of cases of herpes zoster increased after vaccination for varicella. This could be explained by the decrease of the force of infection the vaccine induces, if we assume that a high force of infection boosts the immunity against varicella and hence reduces reactivation. For a nation that wonders whether to vaccinate against varicella this is an important issue. Therefore a model which incorporates this boosting of immunity is a possible next step in this research. A first idea would be to make the reactivation rate dependent on the force of infection. I do not know about the realistic effect of the force of infection, so I choose a mathematical convenient form for the reactivation rate

$$\tilde{\rho}(a) = \frac{\rho(a)^2}{c\lambda(a) + \rho(a)}.$$

c is a to be chosen constant and recall that $\lambda(a)$ is the age dependent force of infection. Another option is

$$\tilde{\rho}(a) = e^{-c\lambda(a)}\rho(a).$$

The most convenient properties are that for $\lambda(a) > 0$ we always have that $\tilde{\rho}(a) < \rho(a)$ and that if $\lambda(a) = 0$ then $\tilde{\rho}(a) = \rho(a)$.[18] This automatically makes sure that $\lambda(a) \to 0$ does not generate an asymptote in the sense that $\tilde{\rho}(a)$ grows beyond all limits. Also for $\lambda(a) \to \infty$ we have that $\tilde{\rho}(a) \to 0$.[19]

Another option would be to look at time since infection. In our equilibrium model, we would have to change the incidence of reactivation. More specifically the function $\rho(a)$ would become $\rho(a,\xi)$ with ξ the age the individual became infected. Inserting this into the incidence of reactivation we obtain

$$f_{I_2}(a) = N(a) \int_0^a \lambda(\xi) e^{-\int_0^\xi \lambda(\eta)\,d\eta} \rho(a,\xi) e^{-\int_\xi^a \rho(\eta,\xi)\,d\eta}\,d\xi$$

We may use a combination of the above suggestions. Then we have also the possibility of using the cumulative force of infection since infection, $\int_\xi^a \lambda(\eta)\,d\eta$, instead of the current force of infection for the boosting of immunity. This may be a more realistic way of modeling, since then short periods with low and high force of infection have a lesser impact. The cumulative force of infection is a non-decreasing function (because it is cumulative), so we may need to choose a different form for $\rho(a)$ than when we use the current force of infection. If we drop the convention that an individual may reactivate only once, we need to consider the possibility that reactivation also boosts immunity.

In the PDEs we could also add 'time since last reïnfection' for an individual who is in state D, the state of individuals carrying a dormant infection. $D(a,t)$ becomes $D(a,b,t)$, where b could either be age at last infection, or time since last infection. This model would give us the most flexible framework, in which we can study the impact of reïnfection on the incidence of herpes zoster.

We saw that for the estimation of the parameters of herpes zoster, β_2 and ρ, the 95% confidence interval is wide. To narrow down the confidence region specific incidence data on herpes zoster may be used. When it is available mainly the likelihood function of section 4.1 needs to be extended compared to our current model, the rest of the procedure will mostly remain the same.

---

[18] of course $\rho(a) > 0$, otherwise there is no reactivation.
[19] This is for a fixed. We do not let a run to reach 0 or ∞.

A Stochastic timings of the outbreak, extinction certain after an outbreak.

A.1 Intro

With diseases with reactivation the rates of infection and recovery are typically mùch larger than the rate of reactivation. This gives rise to the idea of a model where there are waves of epidemics, introduced by reactivation. A nice model for this was published by A.A. de Koeijer *et al.* [5], where the probability and final size of an outbreak depend on the number of susceptibles and infectives at that time. She uses a time scale where 1 unit is an entire expected lifespan.[20] Therefore an epidemic is chosen to happen over an instant, so we get the final size immediately and birth and death during the epidemic can be neglected. This requires that the infection we are looking at will die out in a reasonable time per epidemic wave. We include the use of the parameters from section 2 and new ones will be explained. We get a bit different R_0 and R_1. $R_1 = \frac{\beta}{\alpha}$ like in the simplest model and $R_0 = R_1 + \frac{\rho}{\rho+\mu}R_0$, giving us $R_0 = \frac{\beta(\rho+\mu)}{\alpha\mu}$. Here I present a concise overview of the paper by De Koeijer *et al.* [5].

A.2 Overview

The idea of the model is that there is a number of susceptible individuals that increases in time by birth and death and a number of seropositive individuals that decrease in time. The seropositive individuals can reactivate, possibly triggering an outbreak. When that happens the final size of the outbreak is calculated and every infected susceptible has become seropositive, then the process starts anew until the fraction of susceptibles reaches 1 before an outbreak happens. We let $f(s)$ be the function that returns the fraction of susceptibles that have escaped infection,[21] depending on the fraction of susceptibles $s(t)$. So if the fraction of susceptibles is s when an outbreak occurs, the remaining fraction of susceptibles after the outbreak will be $f(s)$. f is implicitly defined by.

$$\log(f(s)) - f(s)R_1 = \log(s) - sR_1$$

where the log is the standard e-log. (Note that $f(s) = s$ would solve this equation, but that is not what we are after for all s.) The fraction of susceptibles determines whether an outbreak can occur. For $\frac{1}{s} \geq R_1$ we have $f(s) = s$ and for $\frac{1}{s} < R_1$ we get $f(s) = \frac{\text{productlog}(-e^{sR_1}sR_1)}{R_1}$, where $\text{productlog}(z)$ is the solution w of $z = we^w$. Note that productlog is still an implicit function. We have

---

[20]This has influence on the given formulas, one should consider the effects of changing the timescale if that is desirable.

[21]This is not the final size from section 2.1.1.3. There it was the fraction of the population that eventually did get infected. Here it is the population that does not get infected.

$s_0 = \frac{1}{R_1}$ the critical point. As seen from f, when $s \leq s_0$ a reactivation does not trigger an outbreak, since then fraction of susceptibles before the 'outbreak' s is the same as the fraction of susceptibles after the 'outbreak' $f(s)$, but for $s > s_0$ a reactivation does trigger an outbreak. Here it is good to point out that actually $R_1 \leq 1$ already means that outbreaks cannot occur, because $\frac{1}{s} \geq 1$ for every s in $(0, 1]$. (In the case of $s = 1$ there are no seropositive individuals anymore and hence reactivation cannot happen.) This is different from the rest of this thesis where we need $R_0 < 1$ before an infection can no longer prevail. Therefore this model from De Koeijer *et al.* [5] cannot deal with heavily reactivating viruses, since then reactivation does not only trigger an outbreak, but also influences the course of the epidemic.

The change of the fraction of susceptibles in time is defined, after some modeling, by $\frac{ds}{dt} = 1 - s$. Solving this ODE with an unspecified starting value gives us $s(t) = 1 - (1 - s(0))e^{-t}$. We define X to be the stochast to tell us when an epidemic strikes due to reactivation and Y to be time of a single round with one outbreak. A single round with one outbreak is defined as the time between two successive passes of the critical point $s_0 = \frac{1}{R_1}$. De Koeijer *et al.* [5] explains how to get to $\mathbb{E}(Y)$ and we need the following ingredients. $r(s) = \log \frac{1-f(s)}{1-s}$ and the meaning of r lies in the fact that $Y = r(X)$ and so is the relation between the fraction of susceptibles at time of the outbreak the duration of that particular round. $\bar{\rho}$ is the probability of reactivation in a fully seropositive population. The probability of an outbreak when the fraction of susceptibles is s, is $u(s) = 1 - \frac{1}{sR_1}$ for $s \geq \frac{1}{R_1}$ and 0 otherwise. $G(s) = e^{-\bar{\rho}(s - \frac{1+\log(sR_1)}{R_1})}$ is the survival function describing the probability for a population to go from s_0 to s without a new epidemic phase. The probability that after an outbreak a population becomes totally virus free is hence $G(1)$.

$$\mathbb{E}(Y) = \int_{s_0}^{1} \frac{r(s)\bar{\rho}u(s)G(s)}{1 - G(1)} \, ds$$

Next we define $\bar{T}$ to be the stochastic time till eradication, that is the time from the start until s_0 is passed for the last time. At that time there is actually proportion of the population still seropositive, but if we take the condition that no outbreak occurs anymore we can say that the virus is exterminated in that sense. Again the derivation is in De Koeijer *et al.* [5].

$$\mathbb{E}(\bar{T}) = \int_{s_0}^{1} \frac{r(s)\bar{\rho}u(s)G(s)}{G(1)} \, ds = \bar{\rho} \int_{s_0}^{1} \log(\frac{1-f(s)}{1-s})(1 - \frac{1}{sR_1})e^{\bar{\rho}(1-s+\frac{\log s}{R_1})} \, ds$$

Finally we also have a formula for the variance of $\bar{T}$ and Y. We start with defining $\phi(y) = s$ as the inverse of r, so the function that tells us when an epidemic has started. Then with $\Lambda(y) = \frac{G'(\phi(y))\phi(y)}{1-G(1)}$ and $\kappa_i = \int_0^{\infty} x^i \Lambda(x) \, dx$ we get

$$\begin{aligned}
\mathrm{Var}(Y) &= \kappa_2 - \kappa_1^2 \\
\mathrm{Var}(\bar{T}) &= \frac{1 - G(1)}{G(1)}\kappa_2 + (\frac{1 - G(1)}{G(1)}\kappa_1)^2.
\end{aligned}$$

B Tables of notation

The following table gives an explanation of the most symbols uses. Symbols that nowhere require reading earlier sections in this thesis for their interpretation are not included.

Table 6: Variables, parameters and functions in this thesis with a short description and their appearance. Per year* means that, when a contact pattern is defined for the specific model, the multiplication of a rate and a contact pattern together is actually per year.

Notation	Meaning	Quantity	Sections
a, a^*	Age variables	years	All
$\bar{a}$	Mean age at infection	years	2.2.3, 4.2, 4.1.3
$\mathbf{a}$	Vector of all ages of individuals in the data	years	4.1, 4.2.1
a_d, A_d	Age of death in case of type one mortality. a_d is for the normal model and A_d for the discretised version.	years	2.2, 5
a_j	Length of the jth age class	years	2.2.5, 4.1, 4.2.1
a_j	$a_j = ju$, $0 \leq j \leq A_d$ is the discretisation of age in the Iannelli algorithm	years	5
a_s	Starting age for the force of infection in case of maternal immunity	years	2.2.3
A	Total number of age classes		2.2.5
$A(\iota)$	Function that gives the remaining age of individual ι after the maximum number of age classes has been substracted	years	4.1, 4.2.1
α	Rate of recovery	per year	2, A
$\alpha(a), \alpha_j$ $\bar{\alpha}(a), \bar{\alpha}_j$	$\alpha(a)$ is the recovery parameter function of the first infection, $\bar{\alpha}(a)$ for the reactivated infection of an individual of age a. $\alpha_j = \alpha(a_j)$ and $\bar{\alpha}_j = \bar{\alpha}(a_j)$ are their discetisations for the numerical version	per year	5

Notation	Meaning	Quantity	Sections
α_1, α_2	Rate of first resp. reactivated infection recovery	per year	2.1.4
$\alpha(a)$ or $\alpha_1(a)$, $\alpha_2(a)$	Rate of first resp. reactivated infection recovery, depending on age	per year	2.1.5, 2.1.7
$b(a)$	Birth rate function dependent on age a	per year	2.1.5
$b(a, p), b_j^k$	$b(a, p)$ is the birth rate function dependent on age a and total population p. $b_j^k = b(a_j, p(t^k))$ is for the discretised numerical version	per year	5
β	Infection transmission parameter	per year*	2, 2.2, A
β_1, β_2	Infection transmission parameter for the first resp reactivated infectious period	per year*	2.1.4, 2.2, 4.1 4.2, 5.2, 6.1 6.4
$\beta(a)$	Infection transmission function dependent on age a	per year*	2.1.5, 2.1.6
$c(a, a^*)$	Rate of contact of an individual of age a with an individual of age a^*	per year*	2.1.5, 2.1.6, 2.1.7 2.2
$C = \{c_{ij}\}_{(0 \le i, j \le A)}$	Contact matrix corresponding to the choice of dividing up the age classes. It is used as the discretisation $c(a, a^*)$	per year*	2.2.5, 2.2.6, 3.2 4.1.3, 4.2.1
$D(a, t)$	Function in time and age of the number of individuals with dormant infection		2.1.6, 2.1.7
$D(t)$	Function in time of the number of individuals with dormant infection		2.1.4
D^∞	Equilibrium state of the number of individuals with a dormant infection		2.1.4
$d(a, t), D_j^k$	$d(a, t)$ is the density of individuals with dormant infection of age a at time t. $D_j^k = d(a_j, t^k)$ is the discretised version for the numerical algorithm		5
$f_I(a)$ or $f_{I_1}(a)$, $f_{I_2}(a)$	Incidence of infection resp reactivation at age a	per year	2.2, 6.4
$F_d(a)$	Probability for an individual to be		2.1.5, 2.1.6, 2.2

Notation	Meaning	Quantity	Sections
	still alive at age a		
$F_D(a)$	Probability for an individual to have been infected, but not yet reactivated as of age a		2.2.3
$F_S(a)$	Probability for an individual to have escaped infection up till age a		2.2
$h(a,a^*)$, $\bar{h}(a,a^*)$ $h_{j,l}, \bar{h}_{j,l}$	$h(a,a^*)$ and $\bar{h}(a,a^*)$ are the transmission rates upon contact of an first resp. reactivated infectious individual of age a^* with a susceptible of age a. $h_{j,l} = h(a_j, a_l)$ and $\bar{h}_{j,l} = \bar{h}(a_j, a_l)$ are the discretised versions for the numerical method	per year*	5
$h(\tau,a)$	Transmission rate upon contact, τ time after the age a of the infectious individual got infected	per year*	2.1.6, 2.2
$i(a,t), I_j^k$	$i(a,t)$ is the infectious population of age a at time t in years. $I_j^k = i(a_j, t^k)$ is the discretised version for the numerical method		5
I^∞	Equilibrium state of the number of infectious individuals		2.1.2, 2.1.3
$I(a,t)$	Function in time and age of the number of infectious individuals		2.1.5
$I(t)$	Function in time of the number of infectious individuals		2
I_1^∞, I_2^∞	Equilibrium state of the number of initially resp reactivated infectious individuals		2.1.4
$I_1(t)$, $I_2(t)$	Function in time t of the number of initially resp reactivated infectious individuals		2.1.4, 2.1.7
IU/ml	Quantity in which titre taken from. blood samples is measured.	IU/ml	3.1

Notation	Meaning	Quantity	Sections
ι	Summation parameter that indicates which individual in the data is used		4.1
$\hat{\iota}$	The function that takes the indicator value of the age class of individual ι		4.1
$j(a,t)$, J_j^k	$j(a,t)$ is the number of infectious individuals due to reactivation of age a at time t in years. $J_j^k = j(a_j, t^k)$ is the discretised versions for the numerical method		5
$l(\beta_1, \beta_2, \rho \| \mathbf{a}, \mathbf{u})$, $L(\beta_1, \beta_2, \rho \| \mathbf{a}, \mathbf{u})$	log-likelihood resp normal likelihood function with parameters to be estimated β_1, β_2 and ρ using data $\mathbf{a}$ and $\mathbf{u}$	(log-) likelihood	4
$\lambda(a)$	The force of infection experienced by an individual of age a in years	per year	2.2, 6.4
$\lambda(a,i)$, $\lambda(a,(i,j))$, Λ_j^k	The force of infection dependent on the infectious individuals of state i resp state i and j of all ages, experienced by an individual of age a. Λ_j^k is the discritised version (for both) for the numerical method	per year	5
λ_i	The force of infection experienced by an individual in the ith age class	per year	2.2.5, 4.1, 4.2.1
$\lambda_{(i)}$	The force of infection experienced by an individual in the ith contact age group	per year	4.1.3, 4.2, 5.2.1
$m(a), m_j$	Age dependent only part of $\mu(a,p)$. $m_j = m(a_j)$ is the discretised version for the numerical method	per year	5
$M(a,p), M_j^k$	Age and population dependent part of $\mu(a,p)$. M_j^k is the discretised version for the numerical method	per year	5
μ	Death and birth rate parameter	per year	2, A
$\mu(a)$	Death and birth rate function dependent on age a	per year	2.1.5, 2.1.6, 2.1.7

Notation	Meaning	Quantity	Sections
$\mu(a,p)$	$\mu(a,p) = m(a) + M(a,p)$ the death rate function dependent on age a in years and total population size p	per year	5
$n(a,t), N_j^k$	$n(a,t)$ is the total population function with respect to age a in and time t. $N_j^k = n(a_j, t^k)$ is the discretised version for the numerical method		5
N	Total population counted per individual		2
$N(a)$	normalised age a distribution of the total population		2.1.5, 2.1.6, 2.2 6.4
$N(a,t)$	Total population of age a at time t	absolute	2.1.5, 2.1.7
N_j	Fraction of the total population with respect to the jth age class		2.2.5, 4.2.1
$p(t), P^k$	$4p(t)$ is the total population at time t. $P^k = p(t^k)$ is the discretised version for the numerical method		5
$\mathbf{Q}$	Basic contact matrix	per year*	2.2.6, 3.2, 4.1.3
$r(\beta_1, \beta_2, \rho)$ $r_i(\beta_1, \beta_2, \rho)$	$r(\beta_1, \beta_2, \rho)$ is the fractional contribution of herpes zoster to new infections in the point (β_1, β_2, ρ). $r_i(\beta_1, \beta_2, \rho)$ is the contribution in age class i		2.2.5, 4.2.1
R_0	Basic reproduction number		All (2.1.1.2)
R_1	Basic reproduction number of the first infectious period		2.1.3, 2.1.4, 2.2.6 4.2, 5.2.2, 6.1
R_2	Basic reproduction number of the reactivated infectious period(s)		2.1.4
R^∞	Equilibrium state of the number of recovered individuals		2.1.2, 2.1.3, 2.1.4
$R(a,t)$	Function of age a and time t of the number of recovered individuals		2.1.5, 2.1.7
$R(t)$	Function in time of the number of recovered individuals		2
ρ	Reactivation rate parameter	per year	2.1.3, 2.1.4, 4.1

Notation	Meaning	Quantity	Sections
			4.2, 5.2, 6.3 6.1, 6.4, A
$\rho(a)$	$\rho(a)$ is the age a dependent rate of reactivation.	per year	2.1.7, 2.2.3, 2.2.4 2.2.5, 2.2.7, 5 6.4
ρ_i	The rate of reactivation of an individual in age class i	per year	2.2.5, 4.2.1
ρ_j	discretised version of $\rho(a)$. Reactivation rate of an individual in age step j	per year	5.1.6
$s(\infty)$	$1-$The final size of an outbreak		2.1.1.3
$s(a,t), S_j^k$	$s(a,t)$ is the total of the susceptible population of age a and at time t in years. $S_j^k = s(a_j, t^k)$ is the discretised version for the numerical method		5
$s(t)$	Function in time of the proportion of susceptibles		2.1.1.3, A
S^∞	Equilibrium state of the number of susceptibles		2.1.2, 2.1.3, 2.1.4
$S(a)$	Fraction of the total susceptible population of age a with respect to the total population		2.2
$S(a,t)$	Function in time and age of the number of susceptibles		2.1.5
$S(t)$	Function in time of the number of susceptibles		2
t, τ	Time variables	years	All
t^k	$t^k = ku$, $0 \leq k \leq K$ is time step number k in the numerical Iannelli algorithm	years	5
u	$u = \frac{T_f}{K}$ step size of time in the Iannelli algorithm	years	5
u_ι	titre corresponding to individual ι	IU/ml	4.1
$\mathbf{u}$	Vector of all titres of individuals in the data.	IU/ml	4.1
χ_d^2	The Chi-squared distribution with		4.1.2, 4.1.3

Notation	Meaning	Quantity	Sections
	d degrees of freedom. $\chi^2_{d,95\%}$ is the Value at 95% for χ^2_d		

Table 7: Variables, parameters and functions of the De Koeijer *et al.* [5] in appendix A.

Notation	Meaning	Quantity	Sections
α	Rate of recovery	per year	2, A
β	Infection transmission parameter	per year	2, 2.2, A
$f(s)$	Fractional final size function dependent on the fraction s of susceptibles before the outbreak		A
$G(s)$	Survival function describing the probability of the population of going from state s_0 to s without a new epidemic phase		A
κ_i	ith moment of Y		A
$\Lambda(y)$	Density function of Y		A
μ	Death and birth rate parameter		2, A
$r(s)$	Relation function between the fraction of susceptibles s at time of the outbreak and the final size		A
ρ	Reactivation rate parameter	per year	2.1.3, 2.1.4, 4.1 4.2, 5.2, 6.3 6.1, 6.4, A
s_0	Critical fraction of susceptibles from which onward an outbreak can occur		A
$s(t)$	Function in time of the proportion of susceptibles	fraction	2.1.1.3, A
$\tilde{T}$	Stochast determining the time in generations of the host population until the virus is eradicated.	generations	A
$u(s)$	Probability of a major outbreak depending on the susceptible fraction s		A
X	Stochast to tell when an epidemic strikes due to reactivation depending on the fraction of		A

Notation	Meaning	Quantity	Sections
	the susceptible host population		
Y	Stochast determining the time in generations of the host population of a single round with one outbreak		A

References

[1] O. Diekmann and J.A.P. Heesterbeek. *Mathematical Epidemilogy of Infectious Diseases.* John Wiley & Sons Ltd, 2000.

[2] M. Iannelli and M.Y. Kim. Splitting methods for the numerical approximation of some models of age-structured population dynamics and epidemiology. *Elsevier Science Inc.*, 1, 1997.

[3] R. Johnson, J. McElhaney, B. Pedalino, and M. Levin. Prevention of herpes zoster and its painful and debilitating complications. *Elsevier Inc.*, 2007.

[4] F.R. van der Klis, L Mollema, G.A. Berbers, H.E. de Melker, and R.A. Coutinho. Second national serum bank for population-based seroprevalence studies in the netherlands. *Netherlands Journal of Medicine*, 67:301–308, 2009.

[5] A.A. de Koeijer, O. Diekmann, and M.C.M. de Jong. Calculating the time to extinction of a reactivating virus, in particular bovine herpes virus. *Elsevier Inc.*, 1, 2007.

[6] P. McCullagh and J.A. Nelder. *Generalized Linear Models.* Chapman & Hall/CRC, second edition, 1989.

[7] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W.J. Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. March 2008.

[8] J. Prins, D. McCormack, D. Michelson, and K. Horrell. *NIST/SEMATECH e-Handbook of Statistical Methods chapter 7.2.4.1.* http://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm.

[9] J. Wallinga, P. Teunis, and H. Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 167:936–944, 2006.