Mathematical Statistics
Stockholm University

Some results in the field of epidemic modeling and analysis of a smallpox outbreak

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control

Research Report 2006:5
Licentiate thesis

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2006:5**,
http://www.math.su.se/matstat

Some results in the field of epidemic modeling and analysis of a smallpox outbreak

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control*

May 2006

Abstract

This licentiate thesis consists of two research reports:

1. "Effects of random distributions on infectious time periods in epidemic modeling, Research Report 2005:13"
2. "Lessons learned from a smallpox outbreak in Stockholm 1963, Research Report 2005:10"

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91Stockholm, Sweden. E-mail: tommi.asikainen@smi.ki.se. Financial support from the Swedish Institute for Infectious Disease Control, Karolinska Institute, DG Sanco project MODELREL and Martin Rinds foundation is gratefully acknowledged.

1 Acknowledgments

I am very grateful for the support from my two supervisors. Åke Svensson from Stockholm University and Swedish Institute for Infectious Disease Control (SIIDC) for helping finding suitable solutions to my problems concerning epidemic modeling. Johan Giesecke from Karolinska Institute and European Centre for Disease Prevention and Control for recruiting me and giving "up to date" information on important problems in the field of epidemic modeling.

Thank you also for the epidemic modeling PhD students at Stockholm University, Elinore Bengtsson and Patricia Geli for help and pushing me to finish the licentiate report. Also thanks to all the teachers I have had at Stockholm University.

The department of epidemiology at SIIDC has supported me all the time. Financially, giving new ideas and giving the possibility to come in contact with the people working in field concerning infectious diseases. Without these people my work would probably suffered a lot in the sense of not knowing the real problems.

2 Very, very, short summary

The first article investigates two issues. First how many will totally be infected if no major outbreaks are possible. This has arisen due to small outbreaks of measles which occur annually. We investigate how different random distributions on the time an individual is infectious affects the number of totally infected.

The second part deals with how long time it will elapse until a major outbreak can take place. This can be applied to SARS and the infection which is in "fashion" at the moment, pandemic influenza. Is it efficient to screen people before inter continental flights or does the effect of putting restrictions on traveling within a country stop the arrival of an epidemic? In both cases it can be shown not to be very efficient. The time gained before the epidemic comes is very small.

The second article concerns the latest outbreak of smallpox in Stockholm 1963. We estimate some parameters of interest for epidemic modeling.

Mathematical Statistics
Stockholm University

Effects of random distributions on infectious time periods in epidemic modeling

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control

Research Report 2005:13

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2005:13**,
http://www.math.su.se/matstat

Effects of random distributions on infectious time periods in epidemic modeling

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control*

December 2005

Abstract

This paper aims at making sensitivity analysis in epidemic models. We will concentrate on the effect of different random distributions on the infectious time period. One issue is how many persons will be totally infected during an outbreak when no major outbreaks are possible. With a major outbreak we mean the equivalence with branching processes, when a population will explode in size. We find out that the expected value is low, but the variance can be considerably high. This causes problems since often having more than one case makes it to an epidemic among medically trained individuals.

The second problem considered is how long time will it take for a major epidemic to occur. Given that the population is in a state, that major outbreaks can occur.

KEY WORDS: Herd immunity, epidemic modeling, time until outbreak, basic reproductive distribution.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91Stockholm, Sweden. E-mail: tommi.asikainen@smi.ki.se. Financial support from the Swedish Institute for Infectious Disease Control, Karolinska Institute, DG Sanco project MODELREL and Martin Rinds foundation is gratefully acknowledged.

Contents

List of Figures

1. Introduction

The aim is to investigate the sensitivity in epidemic models. The emphasis will be on the effect of different random distributions on infectious time periods.

But first we will have to define the term epidemic (or outbreak). There is no universal definition for an epidemic. (Heymann (2004)) describes an epidemic as "The occurrence, in a defined community or region, of cases of illness (or an outbreak) with a frequency clearly in excess of normal expectancy", this is for a disease which could occur seasonally. In the case of new diseases it is described as "A single case of a communicable disease long absent from a population or the first invasion by a disease not previously recognized in that area requires immediate reporting and full field epidemiological investigation; 2 cases of such a disease associated in time and place are sufficient evidence of transmission to be considered an epidemic". These definitions are useful for medical people and epidemiologists. Statistically an outbreak could be divided into a minor or major outbreak A minor outbreak could be when the size of a population in a branching process approximation will not explode, a major outbreak would mean that the population explodes in size. It is important to be aware of the differences between the medical and statistical view of outbreaks.

First we study the total number of infected in a community where the probability for a major outbreak (epidemic) is zero. Secondly we study the time until a major outbreak (epidemic) appears of an infectious disease in a community who are subject to infection, with a positive probability for major outbreak.

The infection process will be started by imported cases, that is an infectious person enters the population.

One example is the measles is Sweden, for which the vaccination program started 1982. Before that it was in a state that a major outbreak occurred with 3 - 4 year intervals (Ström (1964)). After the introduction of the vaccine the number of resistant individuals increased. These were enough to put the population in a state, where the probability for a major outbreak is zero. Measles is a disease which in many countries with childhood vaccination programs against measles can cause minor outbreaks by introduction of imported cases. The number of infected in Sweden, started by one case have varied between 1 - 80 since 1997. 80 sounds like a major outbreak from a medical viewpoint, but calculating prediction intervals will not make it to a major outbreak from a statistical viewpoint.

Due to some articles claiming that the vaccine could cause autism (Wakefield et al. (1998), Pounder et al. (1995)), the proportion vaccinated decreased.

If this would have continued, the population could again be in a state where major outbreaks would be possible (Asikainen et al. (2003)). This shows that the question of how long time it will take before a major outbreak will occur is also of importance.

These things are interesting for both known (present and past) and new "emerging" diseases. Known diseases could be measles, rubella or smallpox. New ones pandemic influenza caused by avian flu or some other strain.

Depending on how infectious a disease is and how many in the population are immune against this, through vaccination or having encountered the disease earlier in life, there is a a probability for a major outbreak. The quantity describing infectiousness is called basic reproductive number. We will discuss some problems in interpreting this quantity.

If the proportion immune is high enough, the probability for a major outbreak is zero. This is linked with basic reproductive number.

An other example is the time until a major outbreak might occur. This has been an interesting question concerning the recent outbreaks or fear of it, of SARS and avian flu. Both could be capable of making a major outbreak, but have been kept in control by massive counter measures by the authorities.

The most modeling done in the field do not take the stochasticity in the start of the epidemic into account. The interesting part is how long time from breaking the barrier when major outbreaks are possible, it will take until the major outbreak (epidemic) will take place. It is also a believe among medical people that this will take place instantly. We try to show that this should be viewed as a stochastic process.

2. Models for spread of infection

2.1 *Disease and infection*

The development of disease can be described as shown below and in Figure 1 (Giesecke (2002)):

1. Healthy The person has not yet experienced the disease and is not protected against it.
2. Incubation period A susceptible person gets infected by an infectious individual. It takes a certain amount of time until the disease develops in the body and the first signs of disease appear (these can be sub clinical).
3. Clinical disease Signs of the disease are appearing

[!bp]

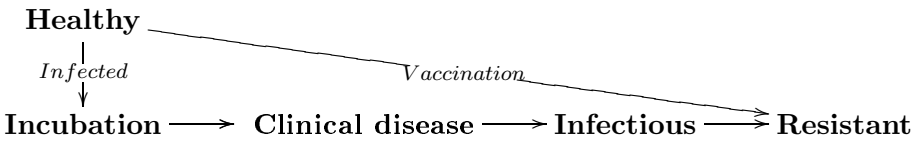

Figure 1. Progress of disease

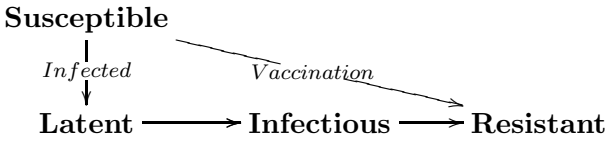

Figure 2. Progress of infection

4. Resistant The person has recovered from the disease, or has been transferred from susceptible through a successful vaccination.

How infection is spread, differs from the above. This is because an individual can be infectious before or even after the stage of clinical disease. This process consists of the following states and is shown in Figure 2 (Giesecke (2002)):

1. Susceptible The person has not yet experienced the disease and is not protected against it.
2. Latent period A susceptible person gets infected by an infectious individual. The virus is fighting with the body's defense system.
3. Infectious period The person is now infectious and can transmit the disease to others.
4. Resistant The person has recovered from the disease, or has been transferred from susceptible through a successful vaccination.

The number of individuals in the different states are usually of interest. Although in many cases, each individual has to be followed separately during the epidemic. These are the two important factors which will decide how the epidemic is progressing:

- The time spent at each state, and its random distribution
- How many susceptible are infected per time unit

When an individual has moved to Resistant state, there will be two possibilities. For individuals who have been infected with measles they will stay there for the rest of their life. For persons vaccinated it is not known how long time the protection holds. The other possibility is with diseases like influenza which come in different forms every year and this is why people can be susceptible even after encountering the disease earlier.

2.2 *Basic assumptions in epidemic models and some terms*

Infectious contacts made by an infectious person in the population are assumed to occur according to a Poisson process $\boldsymbol{K}$ with rate λ. This mean that time periods between contacts are exponentially distributed. The number of contacts during a time period is Poisson distributed.

If a proportion of the population are immune, due to earlier encounter with the disease or through vaccination, the rate λ will reduce in relation to this.

Different kind of mixing is possible in the population, through families, work etc. In this report the population is homogeneously mixing. That is there is no special division of the population into different subgroups.

An index case is the first infective in an outbreak. The others infected by this index case are called secondary cases. Tertiary cases are those infected by the secondary cases. In reality this ordering is very hard to follow due to the uncertainty in who has infected whom. It is also possible to have several index cases, people exposed to a sudden release of an agent for example. The index case will be assumed to be equally infectious as other individuals during the epidemic. Often it might happen that the index case is infectious before entering the population, making him less infectious that the others. This is not investigated in this paper.

In order to start an epidemic the index cases are assumed to appear according to Poisson process $\boldsymbol{K_i}$ with rate λ_i. Index case be a person who gets infected abroad and is still infectious or becomes infectious when returning home.

2.3 *Time spent in different states*

An infectious individual will remain infectious for a stochastic period of time $\boldsymbol{I}$, according to a random distribution.

When trying to model the transitions between different states the time spent in different states can have different distributions. The two most common distributions are delta distribution (constant time period) and exponential distribution.

Constant time periods are useful in certain situations. The case with exponential distribution gives some stochasticity. But very few infections, if any,

follows an exponential distribution in the different states.

In our paper we try a number of different distributions and do comparisons between these. We will use the gamma distribution, since it can be viewed as a sum of exponential distributions. This way we can keep the nice Markovian properties and add more reality to our model.

We will also investigate the effect of time, when the critical level for having major outbreaks is achieved. The time for a major outbreak, can be depending on the assumptions concerning the distribution of time being infectious. The disease might also have different states, where individuals are transmitting different amounts of infection.

2.4 *Basic reproductive distribution*

Assume we have knowledge of the number of infectious contacts per time unit, Chapter 2.2. The random distribution of the infectious time period $\boldsymbol{I}$ is also known, Chapter 2.3.

The number of new cases made by an infectious person will be a stochastic process. We will propose the term "Basic reproductive distribution", $\boldsymbol{RD}$, for the number of new cases. Depending on the random distributions on the input parameters, $\boldsymbol{RD}$ will follow a random distribution. In some situations it will be a very simple expression.

Using $\boldsymbol{RD}$ as a tool, there will be a lot of flexibility. Assume that the infectious time period is constant and that the rate of making infectious contacts is exponentially distributed. This is equivalent for having exponentially distributed infectious time period and a constant rate of making infectious contacts.

2.5 *Basic reproductive number*

An often used parameter in epidemic modeling, is the so called "basic reproductive number", denoted $\boldsymbol{R_0}$. In its easiest form it can be expressed as the expected number of new infections generated by one infectious individual, during the time he is infectious in a population where everyone is susceptible. It can be seen as the mean of the basic reproductive distribution.

This is although a very "misused" parameter since it is often estimated from previous outbreaks. Usually how the population structure is set can give major deviations in the estimation. (Asikainen (2005), Gani and Leach (2001))

If $\boldsymbol{R_0} < 1$ no major outbreaks (epidemics) will occur, only small outbreaks are possible. Although a small can be of size even exceeding 100. If $\boldsymbol{R_0} > 1$ there will be a positive probability for a major outbreak, meaning that a considerable proportion of the population will get infected.

Basic reproductive number is related to the term, herd immunity, which means that $\boldsymbol{R_0} < 1$, giving only small outbreaks. It means that there are so few susceptible in the population that even if an infectious individual enters the population, he will not start an epidemic. This term should be used very carefully. For example 1997 a person infected with measles entered a discotheque in a Stockholm, which due to mass vaccination is in a state of herd immunity. There he infected a number of others and the total epidemic consisted of about 80 cases before it was stopped.

If the population possesses some resistance to the disease, not all infectious contacts will cause new cases. In this case $\boldsymbol{R_0}$ will be scaled to suit the case, often denoted $\boldsymbol{R_a}$, $\boldsymbol{R_*}$ or $\boldsymbol{R_e}$. We will write $\boldsymbol{R_0}$ even in this case.

3. Epidemic models

In Chapter 2 we mentioned the basic assumptions and some common terms. In this chapter we mention a number of models used.

In describing the models all will have the same value on $\boldsymbol{R_0}$, namely λ/γ.

The notation for the number of susceptible, latent, infectious and removed at time point t is $\mathcal{S}(t)$, $\mathcal{E}(t)$, $\mathcal{I}(t)$ and $\mathcal{R}(t)$.

Assume that a person spends in average $1/\beta$ in the latent phase.

3.1 *Constant value on the infectious time period*

Also known as the Reed-Frost epidemic model (Abbey (1952)), assumes that the infectious time period $\boldsymbol{I}$ is constant, with a value of $1/\gamma$.

$$\begin{aligned} \boldsymbol{I} &\sim \delta\left(\frac{1}{\gamma}\right) \\ \mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\ \boldsymbol{K}(\boldsymbol{I}) &\sim \boldsymbol{RD} \\ \boldsymbol{R_0} &= \mathrm{E}[\boldsymbol{K}(\boldsymbol{I})] = \frac{\lambda}{\gamma} \end{aligned}$$

A model can be made for both discrete and continuous time.

Assume that infections occur at discrete time (and that the time between generations is of length 1 for simplicity). The number of infections at different

time points will be binomially distributed. Let N denote the total size of the population.

$$\begin{aligned}
&\text{P(A certain individual not infected by 1 infective)} = e^{-\frac{\lambda}{N\gamma}} \\
&\text{P(A certain individual not infected by } \mathcal{I}(t) \text{ infective)} = e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}} \\
&\text{P(A certain individual infected by } \mathcal{I}(t) \text{ infective)} = 1 - e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}}
\end{aligned}$$

This gives a way to calculate the number of infected at different generations:

$$\begin{aligned}
\mathcal{I}(t+1) &\sim \text{Bin}\Big(\mathcal{S}(t), 1 - e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}}\Big) \\
\text{E}[\mathcal{I}(t+1)] &= \mathcal{S}(t)\Big(1 - e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}}\Big) \approx \frac{\lambda}{\gamma n}\mathcal{S}(t)\mathcal{I}(t)
\end{aligned}$$

In continuous time, the infection is a Poisson process with rate $\frac{\lambda}{\gamma}$. The time between new infections will be exponentially distributed with rate λ, with expected value $\frac{1}{\lambda}$. Let $\boldsymbol{\xi}$ = number of new cases made by $\mathcal{I}(t)$ infective. $\boldsymbol{\xi} \sim \boldsymbol{RD}$.

$$\begin{aligned}
\boldsymbol{\xi} &\sim \text{Po}\Big(\frac{\lambda}{\gamma}\Big) \\
\text{Proportion susceptible} &= \frac{\mathcal{S}}{N} \\
&\Downarrow \\
\boldsymbol{\xi} &\sim \text{Po}\left(\frac{\lambda \mathcal{S}}{\gamma N}\mathcal{I}\right) \\
\text{E}[\boldsymbol{\xi}] &= \frac{\lambda \mathcal{S}}{\gamma N}\mathcal{I}
\end{aligned}$$

The expected value of the processes is the same. The difference is that in continuous time new infections do not have to happen at certain generations.

We would like to point out that in the latter calculations, not everyone in the population is assumed to be susceptible. This is why $\frac{\mathcal{S}}{N}$, proportion of susceptible has been used.

The Basic reproductive distribution, $\boldsymbol{RD}$ with constant value on the infectious time period will be a Poisson distribution.

3.2 *Exponentially distributed infectious time period*

Also called the general epidemic model (Kermack and McKendrick (1927)), the infectious time $\boldsymbol{I}$ is assumed to be exponentially distributed with rate γ.

$$\begin{aligned} \boldsymbol{I} &\sim \text{Exp}(\gamma) \\ \text{E}[\boldsymbol{I}] &= 1/\gamma \\ \text{Var}(\boldsymbol{I}) &= \frac{1}{\gamma^2} \\ \boldsymbol{\xi} &\sim \boldsymbol{RD} \\ \boldsymbol{R_0} &= \text{E}[\boldsymbol{\xi}] = \frac{\lambda}{\gamma} \end{aligned}$$

This model is widely used, specially for simulating epidemics. Since the times are exponentially distributed there can be a deterministic solution by solving a number of differential equations. N denotes the population size, $1/\beta$ the average time spent in the latent phase.

$$\begin{aligned} \frac{d\mathcal{S}}{dt} &= -\frac{\lambda}{N}\mathcal{S}\mathcal{I} \\ \frac{d\mathcal{E}}{dt} &= \frac{\lambda}{N}\mathcal{S} - \beta\mathcal{E} \\ \frac{d\mathcal{I}}{dt} &= \beta\mathcal{E} - \gamma\mathcal{I} \\ \frac{d\mathcal{R}}{dt} &= \gamma\mathcal{I} \end{aligned}$$

The downside is that with this setup, we are modeling what happens in average. It is crucial to decide if this is the quantity of interest before applying this approach.

We would like to point out that in the latter calculations, not everyone in the population is assumed to be susceptible. This is why the differential equations are scaled with $\frac{\mathcal{S}}{N}$.

The Basic reproductive distribution, $\boldsymbol{RD}$ with exponential distribution on the infectious time period will be a geometric distribution.

3.3 *Gamma distributed infectious time period*

The general epidemic model is modified so that the expected infectious time period of length $1/\gamma$ is divided into n independent blocks each having an exponentially distributed length of $1/n\gamma$.

$$
\begin{aligned}
\boldsymbol{I}_j &\sim \text{Exp}(\text{n}\gamma) \\
\boldsymbol{I} &= \sum_{j=1}^{a} \boldsymbol{I}_j \sim \Gamma(n, n\gamma) \\
\text{E}[\boldsymbol{I}_j] &= \frac{1}{n\gamma} \quad \text{Var}(\boldsymbol{I}_j) = \frac{1}{n^2\gamma^2} \\
\text{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \quad \text{Var}(\boldsymbol{I}) = \frac{1}{n\gamma^2} \\
\boldsymbol{\xi} &\sim \boldsymbol{RD} \\
\boldsymbol{R_0} &= \text{E}[\boldsymbol{\xi}] = \frac{\lambda}{\gamma}
\end{aligned}
$$

The main advantage is that the variation compared with the general epidemic model is reduced.

As with an exponential distribution, this can be modeled using differential equations, the difference being that the infectious time period $\mathcal{I}$ will be split into n different blocks.

$$
\begin{aligned}
\frac{d\mathcal{S}}{dt} &= -\frac{\lambda}{N}\mathcal{S}\sum_{j=1}^{n}\mathcal{I}_j \\
\frac{d\mathcal{E}}{dt} &= \frac{\lambda}{N}\mathcal{S}\sum_{j=1}^{n}\mathcal{I}_j - \beta\mathcal{E} \\
\frac{d\mathcal{I}_1}{dt} &= \beta\mathcal{E} - n\gamma\mathcal{I}_1 \\
\frac{d\mathcal{I}_2}{dt} &= n\gamma\mathcal{I}_1 - n\gamma\mathcal{I}_2 \\
&\vdots \\
\frac{d\mathcal{I}_n}{dt} &= n\gamma\mathcal{I}_{n-1} - n\gamma\mathcal{I}_n \\
\frac{d\mathcal{R}}{dt} &= n\gamma\mathcal{I}_n
\end{aligned}
$$

We would like to point out that in the latter calculations, not everyone in the population is assumed to be susceptible. This is why the differential equations are scaled with $\frac{\mathcal{S}}{N}$. The latent period is in the differential equation above assumed to be exponentially distributed.

The Basic reproductive distribution, $\boldsymbol{RD}$ with gamma distribution on the infectious time period will be a negative binomial distribution.

3.4 *Changes for the remaining of the article*

This far the latent period has been taken into account in the models. In the continuation of our report the latent period is abandoned from the model. The focus will be on the probability for a major outbreak and the time point at which that might be initiated.

4. Total number of infected during herd immunity

As stated in Chapter 2.2 the number of infected can become rather big. This even if the population is "protected" by herd immunity ($\boldsymbol{R_0} < 1$). The total number of infected is of interest. As shown in the Introduction, if this is a major outbreak or not, depends in which definition of outbreak to use.

Let $\boldsymbol{X}$ denote the total number of cases during the outbreak, including the index case(s).

Assume that the infectious contacts made by an infectious person is a Poisson process $\boldsymbol{K}$ with rate λ. The distribution of the infectious time period $\boldsymbol{I}$ is assumed to be known.

The progression of disease will be modeled by using techniques from the field of branching processes. The approximation is done by assuming that as infectious person possibly generates new cases, all of these will start new independent processes.

4.1 *Calculation of moments*

Let $\boldsymbol{\xi}_i$ denote the number of new cases made by an infectious individual, $\boldsymbol{\xi}_i \sim \boldsymbol{RD}$ i. The infectious contacts made by him is a Poisson process with rate λ. The probability generating function for $\boldsymbol{\xi} = g_{\boldsymbol{\xi}}(t)$.

$$\begin{aligned} g_{\boldsymbol{\xi}}(t) &= \psi_{\boldsymbol{I}}(\lambda(t-1)) \\ \psi_{\boldsymbol{I}}(t) &= \text{Moment generating function for } \boldsymbol{I} \\ S_x &= \sum_{j=1}^{x} \boldsymbol{\xi}_j \\ g_{S_x}(t) &= \psi_{\boldsymbol{I}}^x(\lambda(t-1)) \\ \boldsymbol{X} &= S_x \end{aligned}$$

As proved in (Sewastjanow (1975)):

$$\begin{aligned} P(\boldsymbol{X}=x) &= \frac{1}{x}P(S_x=x-1) \\ P(S_x=x-1) &= \frac{1}{(x-1)!}\frac{d^{x-1}}{dt^{x-1}}g_{S_x}(t)\Big|_{t=0} \\ P(\boldsymbol{X}=x) &= \frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}g_{S_x}(t)\Big|_{t=0} = \frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}\psi_{\boldsymbol{I}}^{x}(\lambda(t-1))\Big|_{t=0} \end{aligned}$$

$g_{\boldsymbol{X}}(t)$ =Probability generating function for $\boldsymbol{X}$.

$$\begin{aligned} g_{\boldsymbol{X}}(t) &= \mathrm{E}[t^{\boldsymbol{X}}] = \sum_{j=0}^{\infty} t^j \frac{1}{j!}\frac{d^{x-1}}{dt^{x-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \\ \psi_{\boldsymbol{X}}(t) &= \mathrm{E}[e^{t\boldsymbol{X}}] = \sum_{j=0}^{\infty} e^{tj} \frac{1}{j!}\frac{d^{j-1}}{dt^{j-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \\ \mathrm{E}[\boldsymbol{X}] &= \psi'_{\boldsymbol{X}}(0) = \sum_{j=0}^{\infty} \frac{1}{(j-1)!}\frac{d^{j-1}}{dt^{j-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \\ \mathrm{E}[\boldsymbol{X}^{\mathrm{v}}] &= \psi^{v}_{\boldsymbol{X}}(0) = \sum_{j=0}^{\infty} \frac{j^{v-1}}{(j-1)!}\frac{d^{j-1}}{dt^{j-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \end{aligned}$$

This enables to calculate moments of acquired value.

4.2 *Expected value of total number infected*

The expected value of total number infected can be calculated by letting $\boldsymbol{\xi}$ denote the number of infected by an infectious person, $\boldsymbol{\xi} \sim \boldsymbol{RD}$. Let $\boldsymbol{Y}$ denote the total number of infected caused by the index case. Use methods from branching processes. The index case will generate $\boldsymbol{\xi}$ new cases. Each of these $\boldsymbol{\xi}$ will start new independent processes of size $\boldsymbol{Y}$.

$$
\begin{aligned}
\boldsymbol{X} &\stackrel{d}{=} \boldsymbol{Y}+1 \\
\boldsymbol{Y}|\boldsymbol{\xi} &= \boldsymbol{\xi}+\sum_{j=1}^{\boldsymbol{\xi}} \boldsymbol{Y}_j \\
\mathrm{E}[\boldsymbol{Y}|\boldsymbol{\xi}] &= \mathrm{E}[\boldsymbol{\xi}]+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{Y}_{\mathrm{j}}] \\
\text{Since } \boldsymbol{Y}_j \stackrel{d}{=} \boldsymbol{Y} &\quad \text{and} \quad \mathrm{E}[\mathrm{E}[\boldsymbol{Y}|\boldsymbol{\xi}]] = \mathrm{E}[\boldsymbol{Y}] \\
\mathrm{E}[\boldsymbol{Y}] &= \mathrm{E}[\boldsymbol{\xi}]+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{Y}] \\
\mathrm{E}[\boldsymbol{Y}] &= \frac{\mathrm{E}[\boldsymbol{\xi}]}{1-\mathrm{E}[\boldsymbol{\xi}]} \\
\mathrm{E}[\boldsymbol{X}] &= 1+\mathrm{E}[\boldsymbol{Y}] = \frac{1}{1-\mathrm{E}[\boldsymbol{\xi}]}
\end{aligned}
$$

Alternative solution can be obtained by noting that $\boldsymbol{X}$ has the same distribution as the index case plus the total number infected by the offsprings of the index case.

$$
\begin{aligned}
\boldsymbol{X} &\stackrel{d}{=} 1+\sum_{j=1}^{\boldsymbol{\xi}} \boldsymbol{X}_j \\
\mathrm{E}[\boldsymbol{X}] &= 1+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{X}_{\mathrm{j}}] = 1+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{X}] \\
\mathrm{E}[\boldsymbol{X}] &= \frac{1}{1-\mathrm{E}[\boldsymbol{\xi}]} \qquad (1)
\end{aligned}
$$

The expected value of the total number of cases during the epidemic, will be independent of the distribution on the infectious time $\boldsymbol{I}$. As long as it has the same mean $1/\gamma$.

4.3 *Variance of total number infected*

The variance of total number infected can be calculated by letting $\boldsymbol{\xi}$ denote the number of infected by an infectious person, $\boldsymbol{\xi} \sim \boldsymbol{RD}$. Let $\boldsymbol{X}$ denote the total number of infected including the index case. This index case generates $\boldsymbol{\xi}$ new cases. Each of these will start a new independent branching process,

with the same distribution as $\boldsymbol{X}$.

$$\begin{aligned}
\boldsymbol{X} &\stackrel{d}{=} 1 + \sum_{j=1}^{\boldsymbol{\xi}} \boldsymbol{X}_j \\
\mathrm{Var}(\boldsymbol{X}) &= \mathrm{E}[\boldsymbol{\xi}]\mathrm{Var}(\boldsymbol{X}_\mathrm{j}) + (\mathrm{E}[\boldsymbol{X}_\mathrm{j}])^2\mathrm{Var}(\boldsymbol{\xi}) \\
\mathrm{Var}(\boldsymbol{X}) &= \frac{\mathrm{Var}(\boldsymbol{\xi})}{(1-\mathrm{E}[\boldsymbol{\xi}])^3} \qquad (2)
\end{aligned}$$

In contrast to the expected value, which is same for all distributions with the same mean value. The variance will depend on the variance of $\boldsymbol{\xi}$. In our case $\boldsymbol{\xi}$ will be a compound Poisson distribution.

4.4 *Constant value on the infectious time period*

The total number of cases $\boldsymbol{X}$ in a Reed-Frost process (constant infectious time period) can be shown to follow a Borel-Tanner distribution (Height and Breuer (1960), Tanner (1953), Borel (1942)). This distribution is derived from queuing theory, in the case for a constant service period and arrival of new individuals according to a Poisson process with rate λ.

$$\begin{aligned}
\boldsymbol{I} &\sim \delta(\frac{1}{\gamma}) \\
\psi_{\boldsymbol{I}}(t) &= e^{\frac{t}{\gamma}} \\
\psi_{\boldsymbol{I}}(\lambda(t-1)) &= e^{\frac{\lambda}{\gamma}(t-1)} = e^{\boldsymbol{R_0}(t-1)} \\
\psi^x_{\boldsymbol{I}}(\lambda(t-1)) &= e^{x\boldsymbol{R_0}(t-1)} \\
\frac{d^x}{dt^x} &= (x\boldsymbol{R_0})^x e^{x\boldsymbol{R_0}(t-1)} \\
P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{1}{x!}(x\boldsymbol{R_0})^{x-1}e^{x\boldsymbol{R_0}(t-1)}\Big|_{t=0} \\
P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{(x\boldsymbol{R_0})^{x-1}e^{-x\boldsymbol{R_0}}}{x!} \\
\boldsymbol{X} &\sim \text{Borel-Tanner}(\boldsymbol{R_0}) \quad ; \quad \boldsymbol{R_0} < 1 \\
\mathrm{E}[\boldsymbol{X}] &= \frac{1}{1-\boldsymbol{R_0}} \\
\mathrm{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}}{(1-\boldsymbol{R_0})^3} \\
&\Longleftrightarrow
\end{aligned}$$

$$\begin{array}{rcl} & \Longleftrightarrow & \\ \lim_{\boldsymbol{R_0}\to 1-} \mathrm{E}[\boldsymbol{X}] & = & \infty \\ \lim_{\boldsymbol{R_0}\to 1-} \mathrm{Var}(\boldsymbol{X}) & = & \infty \end{array}$$

Alternative solution by using equation (2) and (1), side 19 and 18:

$$\begin{aligned} \boldsymbol{\xi} &\sim Po(\lambda/\gamma) = Po(\boldsymbol{R_0}) \\ \mathrm{E}[\boldsymbol{\xi}] &= \boldsymbol{R_0} \\ \mathrm{E}[\boldsymbol{X}] &= \frac{1}{(1-\boldsymbol{R_0})} \\ \mathrm{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}}{(1-\boldsymbol{R_0})^3} \end{aligned}$$

See Figure 3 for the progress and the increase in the number of infected as $\boldsymbol{R_0} \to 1-$.

4.5 *Exponential distribution on infectious times*

Since the infectious time is exponentially distributed, the offspring distribution in a branching process approach becomes geometric. This given that infectious contacts occur according to a Poisson process with rate λ. The total number of infected $\boldsymbol{X}$ including index cases has been investigated by (Farrington et al. (2003)), by using the Lagrangian generalized negative binomial distribution (or Lagrangian negative binomial distribution).

$$\begin{aligned} \boldsymbol{I} &\sim \mathrm{Exp}(\gamma) \\ \psi_{\boldsymbol{I}}^x(t) &= (1-\frac{t}{\gamma})^{-x} \\ \frac{d^l}{dt^l}\psi_{\boldsymbol{I}}^x(t) &= \frac{(x+x-1)!}{(x-1)!}\frac{1}{\gamma^l}(1-\frac{t}{\gamma})^{-(x+l)} \\ \frac{d^l}{dt^l}\psi_{\boldsymbol{I}}^x(\lambda(t-1)) &= \frac{(x+l-1)!}{(x-1)!}(\frac{\lambda}{\gamma})^l(1-\frac{\lambda(t-1)}{\gamma})^{-(x+l)} \\ P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{1}{x!}\frac{(x+(x-1)-1)!}{(x-1)!}(\frac{\lambda}{\gamma})^{x-1}(1-\frac{\lambda(t-1)}{\gamma})^{-(x+(x-1))}\Big|_{t=0} \\ P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{(2x-2)!}{x!(x-1)!}\frac{{\boldsymbol{R_0}}^{x-1}}{(1+\boldsymbol{R_0})^{2x-1}} \end{aligned}$$

Rewriting the expression as $\boldsymbol{Z} = \boldsymbol{X} - 1 \sim$ Lagrangian binomial distribution (Kotz and Johnson (1983)), also known as Lagrangian negative binomial distribution (Jain and Consul (1971)).

$$
\begin{aligned}
\boldsymbol{Z} &= \boldsymbol{X} - 1 \sim \text{ Lagrangian generalized negative binomial distribution} \\
\mathrm{P}(\boldsymbol{Z} = z) &= \frac{n}{n+\alpha t}\binom{n+\alpha z}{z} p^z q^{n+\alpha z - z} \quad , n > 0 \quad , p < \alpha p < 1 \quad , z = 0, 1, 2, \ldots \\
\mathrm{E}[\boldsymbol{Z}] &= \frac{np}{1-\alpha p} \\
\mathrm{Var}(\boldsymbol{Z}) &= \frac{npq}{(1-\alpha p)^3} \\
&\vdots \\
\mathrm{E}[\boldsymbol{Z}] &= \frac{\frac{\boldsymbol{R_0}}{1+\boldsymbol{R_0}}}{1-\frac{2\boldsymbol{R_0}}{1+\boldsymbol{R_0}}} = \frac{\boldsymbol{R_0}}{1+\boldsymbol{R_0}-2\boldsymbol{R_0}} = \frac{\boldsymbol{R_0}}{1-\boldsymbol{R_0}} \\
\mathrm{Var}(\boldsymbol{Z}) &= \frac{\frac{\boldsymbol{R_0}}{(1+\boldsymbol{R_0})^2}}{(1-\frac{2\boldsymbol{R_0}}{1+\boldsymbol{R_0}})^3} = \frac{\boldsymbol{R_0}(1+\boldsymbol{R_0})}{(1-\boldsymbol{R_0})^3} \\
\mathrm{E}[\boldsymbol{X}] &= \mathrm{E}[\boldsymbol{Z}] + 1 = \frac{\boldsymbol{R_0}}{1-\boldsymbol{R_0}} + 1 = \frac{1}{1-\boldsymbol{R_0}} \\
\mathrm{Var}(\boldsymbol{X}) &= \mathrm{Var}(\boldsymbol{Z}) = \frac{\boldsymbol{R_0}(1+\boldsymbol{R_0})}{(1-\boldsymbol{R_0})^3}
\end{aligned}
$$

Since the offspring is geometrically distributed, the variance of $\boldsymbol{X}$ can also be calculated by equation (2) and (1), side 19 and 18:

$$
\begin{aligned}
\boldsymbol{I} &\sim \mathrm{Exp}(\gamma) \\
\boldsymbol{\xi} \sim \boldsymbol{RD} &\sim \mathrm{Po}(\lambda \boldsymbol{I}) \\
\boldsymbol{\xi} &\sim \mathrm{Ge}(p) \\
\mathrm{Var}(\boldsymbol{\xi}) &= \frac{q}{p^2} = \boldsymbol{R_0}(1+\boldsymbol{R_0}) \\
\mathrm{E}[\boldsymbol{X}] &= \frac{1}{(1-\boldsymbol{R_0})} \\
\mathrm{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}(1+\boldsymbol{R_0})}{(1-\boldsymbol{R_0})^3}
\end{aligned}
$$

See Figure 3 for the progress and the increase in the number of infected as $\boldsymbol{R_0} \to 1-$.

4.6 *Gamma distributed infectious times*

Dividing the expected time $1/\gamma$ into n separate parts. Each of these are exponentially distributed with rate $n\gamma$.

$$\begin{aligned}
\boldsymbol{I} &\sim \Gamma(n, \frac{1}{n\gamma}) \\
\psi_{\boldsymbol{I}}^x(t) &= (1 - \frac{t}{n\gamma})^{-nx} \\
\frac{d^l}{dt^l}\psi_{\boldsymbol{I}}^x(t) &= \frac{(nx+l-1)!}{(nx-1)!}(\frac{1}{n\gamma})^l(1 - \frac{t}{n\gamma})^{-(nx+l)} \\
\frac{d^l}{ds^l}\psi_{\boldsymbol{I}}^x(\lambda(t-1)) &= \frac{(nx+l-1)!}{(nx-1)!}(\frac{\lambda}{n\gamma})^l(1 - \frac{\lambda(t-1)}{n\gamma})^{-(nx+l)} \\
P(\boldsymbol{X} = x|\boldsymbol{R_0}) &= \frac{1}{x!}\frac{(nx+(x-1)-1)!}{(nx-1)!}(\frac{\lambda}{n\gamma})^{x-1}(1 - \frac{\lambda(t-1)}{n\gamma})^{-(nx+(x-1))}\Big|_{t=0} \\
P(\boldsymbol{X} = x|\boldsymbol{R_0}) &= \frac{1}{x!}\frac{(x(n+1)-2)!}{(nx-1)!}(\frac{\boldsymbol{R_0}}{n})^{x-1}\frac{1}{(1+\frac{\boldsymbol{R_0}}{n})^{x(n+1)-1}}
\end{aligned}$$

When having a compound Poisson distribution with a gamma distribution, the resulting distribution will be a negative binomial.

$$\begin{aligned}
\boldsymbol{I} &\sim \Gamma(n, \frac{1}{n\gamma}) \\
\boldsymbol{Y} &\sim \boldsymbol{K}|\boldsymbol{I} = i \sim \mathrm{Po}(\lambda i) \\
\mathrm{P}(\boldsymbol{\xi} = \rho) &= \int_0^\infty \mathrm{P}(\boldsymbol{\xi} = \rho|\boldsymbol{I} = i)\mathrm{P}(\boldsymbol{I} = i)di \\
&= \int_0^\infty e^{-\lambda i}\frac{(\lambda i)^\rho}{\rho!}\frac{1}{\Gamma(n)}i^{n-1}(n\gamma)^n e^{-in\gamma}di = \ldots = \\
&= \frac{(n\gamma)^n\gamma^\rho}{\rho!\Gamma(n)}(\rho+n-1)!(\frac{1}{\lambda+n\gamma})^n(\frac{1}{\lambda+n\gamma})^\rho = \\
&= \frac{(\rho+n-1)!}{\rho!(n-1)!}(\frac{\lambda}{\lambda+n\gamma})^n(\frac{n\gamma}{\lambda+n\gamma})^\rho
\end{aligned}$$

Set $p = \frac{n\gamma}{\lambda+n\gamma}$. Using equation (2) and (1), side 19 and 18:

$$
\begin{aligned}
\boldsymbol{\xi} &\sim \text{NegBin(n, p)} \\
\text{E}[\boldsymbol{\xi}] &= n\frac{q}{p} = n\frac{\lambda}{n\gamma} \\
\text{Var}(\boldsymbol{\xi}) &= n\frac{q}{p^2} = n\frac{\frac{\lambda}{(\lambda+n\gamma)^2}}{\frac{n\gamma}{\lambda+n\gamma}} = \frac{\lambda(\lambda+n\gamma)}{n\gamma^2} = \boldsymbol{R_0}(\frac{\boldsymbol{R_0}}{n}+1) \\
\text{E}[\boldsymbol{X}] &= \frac{1}{(1-\boldsymbol{R_0})} \\
\text{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}(\frac{\boldsymbol{R_0}}{n}+1)}{(1-\boldsymbol{R_0})^3}
\end{aligned}
$$

If $n = 1$ we will get the variance for the exponential distribution on the infectious time period. If $n = \infty$ we will get the variance for the constant infectious time period.

See Figure 3 for the progress and the increase in the number of infected as $\boldsymbol{R_0} \to 1-$.

4.7 *Comparison between different epidemic models*

All models show that the uncertainty in the number of infected rapidly increases when approaching the herd immunity threshold limit ($\boldsymbol{R_0} = \text{E}[\boldsymbol{\xi}] = 1$). In all cases the mean number will not differ. The variance is shifting between the models, see Figure 3. Assuming exponential distribution, shows the largest distribution in our case. This can be explained due to that some individuals have very long infectious periods.

[!ht]

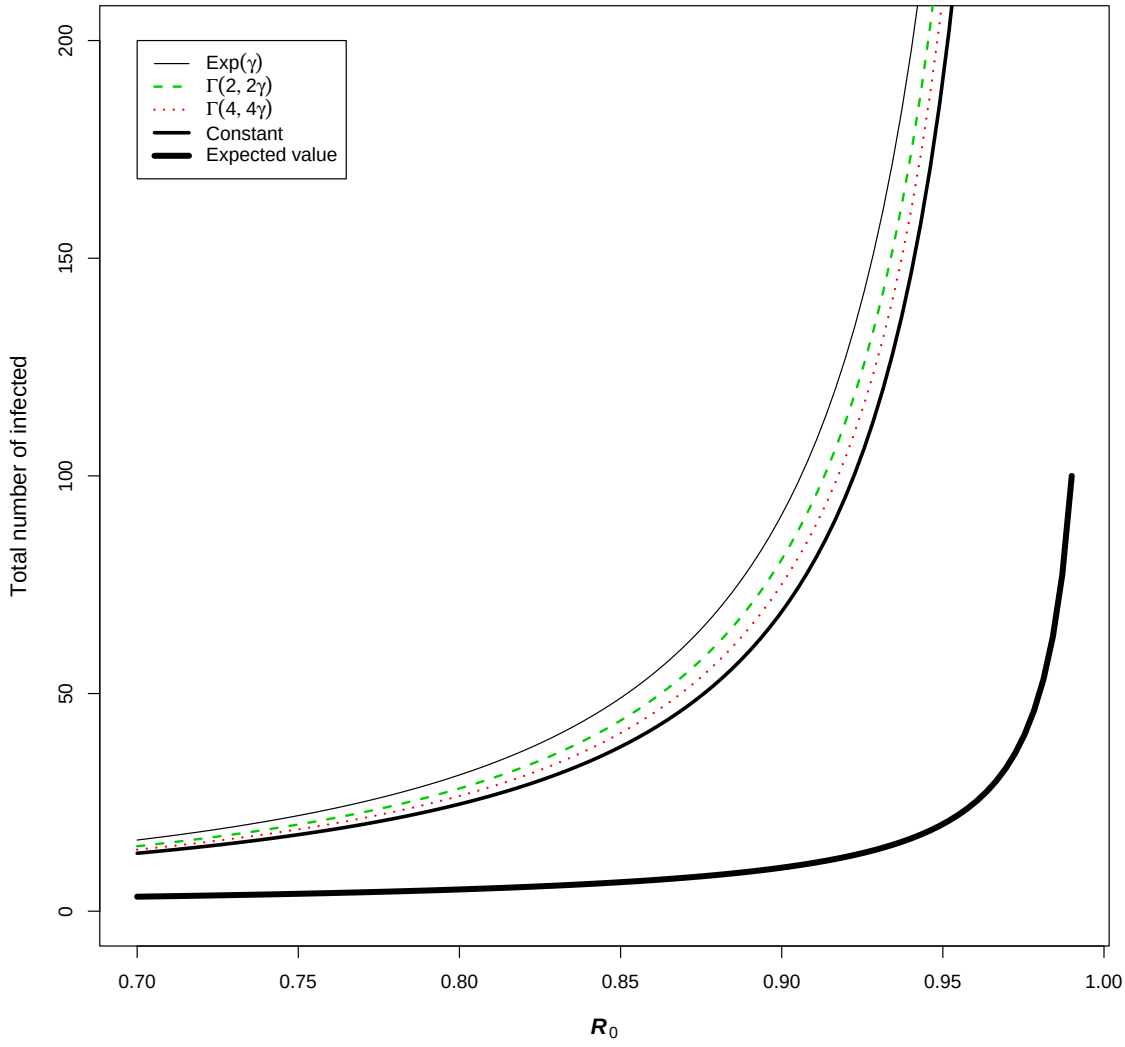

Figure 3. Number of infected as a function of $\boldsymbol{R_0}$, the mean value of the infectious time period is $1/\gamma$. Figure showing the expected value and 95% upper prediction interval.

5. Probability for a major outbreak when no herd immunity

5.1 *Background*

A common problem with different infectious diseases is that the population might not be protected by herd immunity. This can be caused by decreasing vaccination coverage. Or in the case of whooping cough, the vaccine is not efficient enough. Even if everyone would be vaccinated it would not be possible to obtain herd immunity.

When the herd immunity threshold has been reached, there will be a positive probability for a major outbreak. This can happen every time a new index case enters the population. The probability for a major outbreak can be calculated from branching process theory (Sewastjanow (1975)).

$$\begin{aligned} \boldsymbol{\xi} &= \text{The number of new infectious contacts} \\ \boldsymbol{\xi} &\sim \text{Po}(\lambda) \\ g_{\boldsymbol{\xi}}(t) &= \text{probability generating function for } \boldsymbol{\xi} \\ t &= g_{\boldsymbol{\xi}}(t) \end{aligned}$$

The smallest solution t is the probability for not having a major outbreak (epidemic).

Knowing the distribution of the infectious time period $\boldsymbol{I}$ gives a way to calculate the probability for not having a major outbreak.

We will calculate this in different epidemic models. The assumption is that the probability is for one index case. The probability generating function for $\boldsymbol{\xi}$ can be written as:

$$g_{\boldsymbol{\xi}}(t) = \mathrm{E}[t^{\boldsymbol{\xi}}] = \mathrm{E}[\mathrm{E}[t^{\boldsymbol{\xi}}|\boldsymbol{I}]] = \mathrm{E}[e^{-\lambda \boldsymbol{I}(1-t)}] = \mathrm{E}[e^{\lambda \boldsymbol{I}(t-1)}] = \Psi_{\boldsymbol{I}}(\lambda(t-1)) \quad (3)$$

Where $\Psi_{\boldsymbol{I}}(t)$ is the moment generating function for $\boldsymbol{I}$. In equation (3) $\boldsymbol{\xi} \sim \boldsymbol{RD}$, which is the basic reproductive distribution. See Figure 4 for different random distributions on $\boldsymbol{I}$.

[!ht]

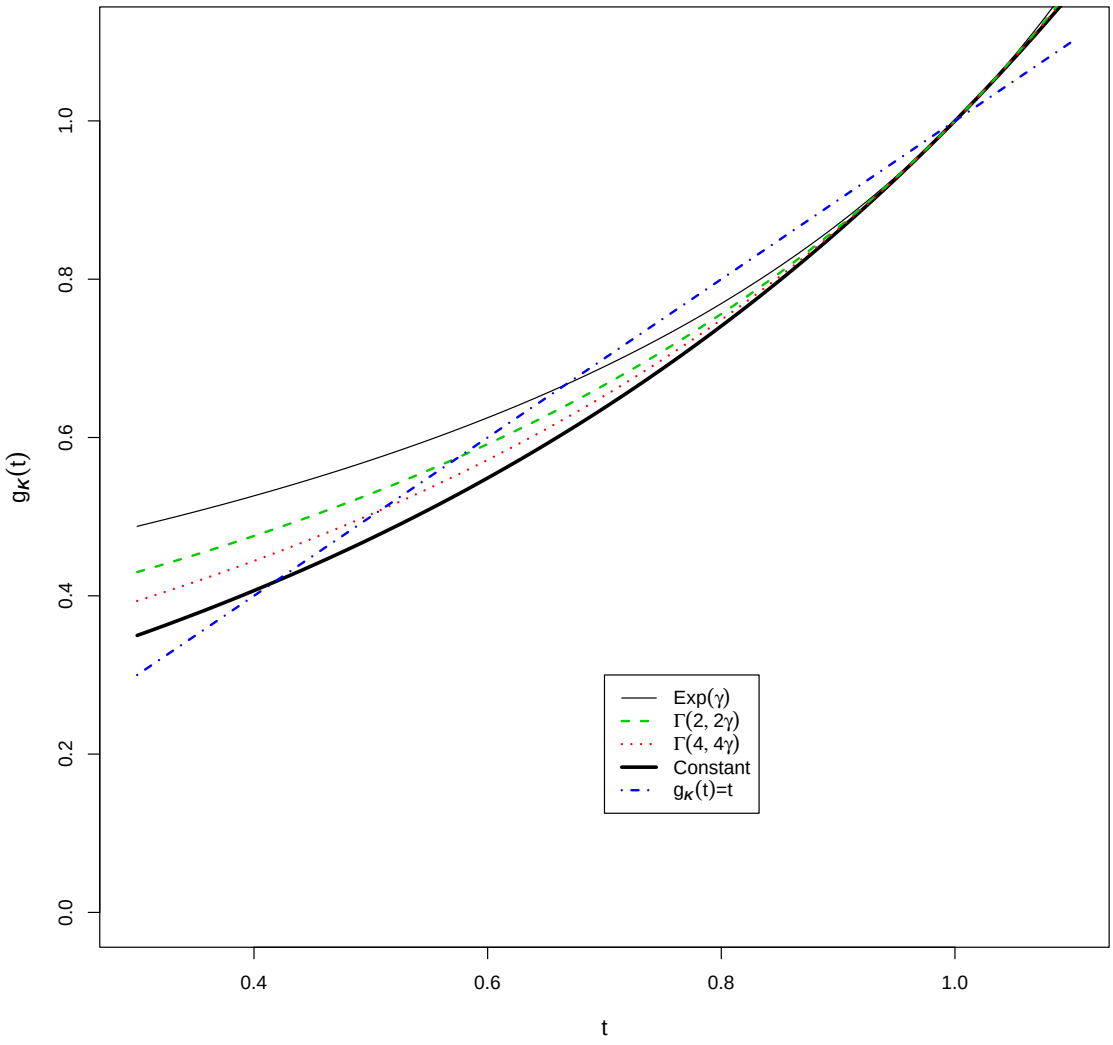

Figure 4. Probability for an outbreak depending on the random distribution of infectious times, $\boldsymbol{R_0} = 1.5$

5.2 *Constant value on the infectious time period*

Also known as the Reed-Frost model.

$$\begin{aligned} \mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\ \Psi_{\boldsymbol{I}}(t) &= e^{\frac{\lambda(t-1)}{\gamma}} \\ t &= e^{\frac{\lambda(t-1)}{\gamma}} \\ s &= 1-t \\ 1-s &= e^{\frac{-\lambda s}{\gamma}} = e^{-\boldsymbol{R_0}s} \end{aligned}$$

Here s will denote the probability of having a major outbreak. It can be solved numerically for different values of $\boldsymbol{R_0}$.

5.3 *Exponentially distributed infectious time period*

Also known as the general epidemic model. The probability for a major outbreak will be less than for the case with a constant infectious time period.

$$\begin{aligned} \boldsymbol{I} &\sim \mathrm{Exp}(\gamma) \\ \mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\ \Psi_{\boldsymbol{I}}(t) &= \frac{1}{1-\frac{t}{\gamma}} = \frac{\gamma}{\gamma - t} \\ t &= \Psi_{\boldsymbol{I}}(\lambda(t-1)) \\ t &= \frac{\gamma}{\gamma - \lambda t + t} \\ t &= \frac{\gamma}{\lambda} \\ s &= 1-t \\ s &= \frac{\lambda-\gamma}{\lambda} = \frac{\boldsymbol{R_0}-1}{\boldsymbol{R_0}} = 1 - \frac{1}{\boldsymbol{R_0}} \end{aligned}$$

Here s will denote the probability of having a major outbreak. The general epidemic model gives an easy way to calculate the probability for a major outbreak.

5.4 *Gamma distributed infectious time period*

$$
\begin{aligned}
\boldsymbol{I} &\sim \Gamma(n, n\gamma) \\
\mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\
\mathrm{Var}[\boldsymbol{I}] &= \frac{1}{n\gamma^2} \\
\Psi_{\boldsymbol{I}}(t) &= \frac{1}{(1-\frac{t}{n\gamma})^n} \\
t &= \Psi_{\boldsymbol{I}}(\lambda(t-1)) \\
&\vdots \\
1 &= t^{1/n}\Big(1 - \frac{\lambda(t-1)}{n\gamma}\Big) \\
1 &= t^{1/n}\Big(1 - \frac{\boldsymbol{R_0}(t-1)}{n}\Big) \\
s &= 1 - t \\
1 &= (1-s)^{1/n}\Big(1 + \frac{\boldsymbol{R_0}s}{n}\Big)
\end{aligned}
$$

Here s will denote the probability of having a major outbreak. It can be solved numerically for different values of $\boldsymbol{R_0}$.

5.5 *Comparison between the different epidemic models*

The probability for a major outbreak given by the Reed-Frost, gamma distributed infectious time period and the general epidemic model are shown in Figure 5. For low values in $\boldsymbol{R_0}$ the difference can be remarkably large. For example with $\boldsymbol{R_0} = \lambda \mathrm{E}[\boldsymbol{I}] = 2$, the general epidemic has a probability of 0.50 while the Reed-Frost has 0.80. The ordering seems to be that constant distribution > gamma distribution > exponential distribution, when it comes to the probability for having a major outbreak.

5.6 *Order of probabilities for major outbreaks with different random distributions on the infectious time periods*

Assume that $\boldsymbol{R_0}$ has the same value in all models.

[!ht]

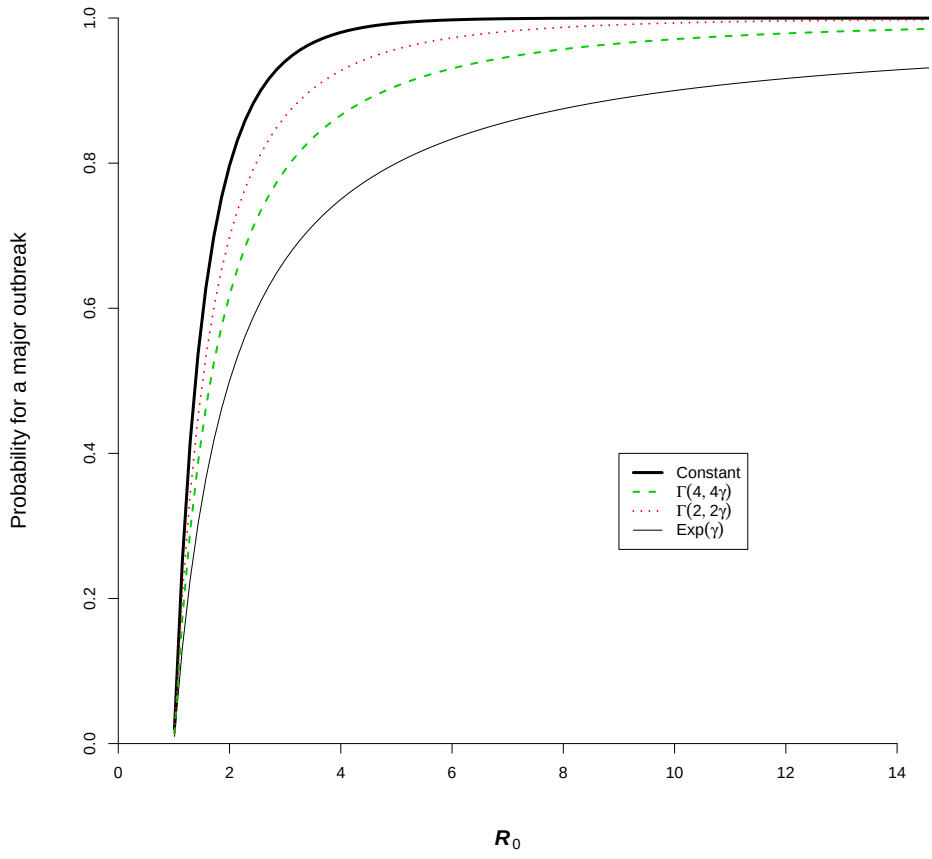

Figure 5. Probability for a major outbreak as a function of $\boldsymbol{R_0}$

Let f be a convex function and $\boldsymbol{U}$ be a stochastic variable.

$$\begin{aligned}
\text{Jensen's inequality} &\Rightarrow f(\mathrm{E}[\boldsymbol{U}]) \leq \mathrm{E}[f(\boldsymbol{U})] \\
\text{Let } f(t) &= e^{-\lambda t} \\
e^{-\lambda \mathrm{E}[\boldsymbol{I}]} &\leq \mathrm{E}[e^{-\lambda \boldsymbol{I}}] \\
\mathrm{P}(\text{Infected during time period } \boldsymbol{I}) &= 1 - e^{-\lambda \boldsymbol{I}} \\
1 - \mathrm{E}[e^{-\lambda \boldsymbol{I}}] &\leq 1 - e^{-\lambda \mathrm{E}[\boldsymbol{I}]}
\end{aligned}$$

In the constant case there is an equality while for the exponential and gamma distributions on the infectious time period there can be an inequality. This gives that the constant distribution has the highest probability for a major outbreak.

Next we show that the probability for a major outbreak with the general epidemic model, is less or equal to the case with a gamma distribution with an integer as the form parameter. This can be shown by methods from (Daley (1990)). $\phi_{\boldsymbol{I}}(t)$ is the Laplace transform for $\boldsymbol{I}$, $\phi_{\boldsymbol{I}}(t) = \mathrm{E}[e^{-t\boldsymbol{I}}]$. Let p_{mo} = probability for a major outbreak.

$$\begin{aligned}
\boldsymbol{I}_1 \sim \mathrm{Exp}(\gamma) \qquad & \qquad \boldsymbol{I}_2 \sim \Gamma(n, n\gamma) \\
\phi_{\boldsymbol{I}_1}(t) = \frac{1}{1+\frac{t}{\gamma}} \qquad & \qquad \phi_{\boldsymbol{I}_2}(t) = \frac{1}{(1+\frac{t}{n\gamma})^n} \\
\phi_{\boldsymbol{I}_1}(t) \quad &\geq \quad \phi_{\boldsymbol{I}_2}(t) \text{ , since } \frac{\phi(\boldsymbol{I}_1)}{\phi(\boldsymbol{I}_2)} = \frac{(1+\frac{t}{n\gamma})^n}{1+\frac{t}{\gamma}} \\
&\Longrightarrow \\
\boldsymbol{I}_1 \quad &\leq_L \quad \boldsymbol{I}_2 \\
&\Longrightarrow \\
p_{mo}(\boldsymbol{I}_1) \quad &\leq \quad p_{mo}(\boldsymbol{I}_2)
\end{aligned}$$

For $n = 1$ there is equality, since it gives the exponential distribution. While for $n > 1$ there is an inequality. Showing that gamma distribution has greater or equal probability for an outbreak than the exponential distribution.

Letting $n \to \infty$ gives $\frac{\phi(\boldsymbol{I}_1)}{\phi(\boldsymbol{I}_2)} = \frac{e^{\frac{t}{\gamma}}}{1+\frac{t}{\gamma}}$.

6. Time until a major outbreak

In Chapter 5 we calculated the probability for having a major outbreak. In real life the probability is not always the most wanted concept. People are more interested in what time span outbreaks might happen. This giving aid

in the planning procedures, such as producing the necessary amounts of anti viral or vaccine doses.

Assume that there is a flow of imported new infectious individuals (index cases) $\boldsymbol{K_i}$ who arrive according to a Poisson process with rate λ_i. If the population is protected by herd immunity, ($\boldsymbol{R_0} < 1$), this contributes with a probability 0 a.s. for a major outbreak. If $\boldsymbol{R_0} > 1$, there is a probability $k(\boldsymbol{R_0}(t))$ for a major outbreak. This at a certain time point t, where index cases enters the population.

The probability function $k(\boldsymbol{R_0}(t))$ has been calculated in Chapter 5. Let $\boldsymbol{T}$ denote the time point for a major outbreak.

$$
\begin{aligned}
\boldsymbol{R_0}(t) < 1 &\Leftrightarrow k(\boldsymbol{R_0}(t)) = 0. \\
\mathrm{P}(\mathrm{t} \leq \boldsymbol{T} \leq \mathrm{t}+\mathrm{h}) &= \lambda_i h \mathrm{P}(\boldsymbol{T} > \mathrm{t}) \mathrm{k}(\boldsymbol{R_0}(\mathrm{t})) \\
\frac{\mathrm{P}(\boldsymbol{T} \leq \mathrm{t}+\mathrm{h}) - \mathrm{P}(\boldsymbol{T} \leq \mathrm{t})}{h} &= \lambda_i \mathrm{P}(\boldsymbol{T} > \mathrm{t}) \mathrm{k}(\boldsymbol{R_0}(\mathrm{t})) \\
\lim_{h \to 0} \frac{\mathrm{P}(\boldsymbol{T} \leq \mathrm{t}+\mathrm{h}) - \mathrm{P}(\boldsymbol{T} \leq \mathrm{t})}{h} &\to \mathrm{P}'(\mathrm{t}) \\
\mathrm{P}(\boldsymbol{T} > \mathrm{t}) &= 1 - \mathrm{P}(\mathrm{t}) \\
\frac{\mathrm{P}'(\mathrm{t})}{1 - \mathrm{P}(\mathrm{t})} &= \lambda_i k(\boldsymbol{R_0}(t)) \\
\ln(1 - \mathrm{P}(\mathrm{t}))' &= -\lambda_i k(\boldsymbol{R_0}(t)) \\
\ln(1 - \mathrm{P}(\mathrm{t})) &= -\lambda_i \int_0^t k(\boldsymbol{R_0}(s)) \mathrm{d}s \\
\mathrm{P}(\mathrm{t}) &= 1 - e^{-\lambda_i \int_0^t k(\boldsymbol{R_0}(s)) ds} \\
\mathrm{P}(\mathrm{t}) &= \mathrm{F}_{\boldsymbol{T}}(t) \\
f_{\boldsymbol{T}}(t) &= \lambda_i k(\boldsymbol{R_0}(t)) e^{-\lambda_i \int_0^t k(\boldsymbol{R_0}(s)) ds} \qquad (4)
\end{aligned}
$$

Looking at the expression for P(t), shows that the probability for a major outbreak increases when $k(\boldsymbol{R_0})$ increases. This is also what to expect. Letting $k(\boldsymbol{R_0})$ increase is similar to that $\boldsymbol{R_0}$ is increasing. Knowing the function $k(\boldsymbol{R_0}(t))$, gives now a way to calculate the probability for a major outbreak at any given time point.

6.1 *Time for an outbreak when $\boldsymbol{R_0}$ is constant*

Assume that $\boldsymbol{R_0}$ will be constant for all time points and $\boldsymbol{R_0} > 1$. It is possible to give an analytic expression the expected time $\boldsymbol{T}$ for a major outbreak and its variance. Equation (4) shows this is the density function for an

exponential distribution with rate $\lambda_i k(\boldsymbol{R_0}(t))$, where $k(\boldsymbol{R_0}(t))$ is constant.

$$\begin{aligned} \boldsymbol{T} &\sim \text{Exp}(\lambda_i k(\boldsymbol{R_0}(t))) \\ \text{E}[\boldsymbol{T}] &= \frac{1}{\lambda_i k(\boldsymbol{R_0}(t))} \\ \text{Var}[\boldsymbol{T}] &= \frac{1}{(\lambda_i k(\boldsymbol{R_0}(t)))^2} \end{aligned}$$

Depending on which assumptions are on the infectious period of time $\boldsymbol{I}$, we can make comparisons between different models.

Figure 6 shows how the distribution of time depends on the value of $\boldsymbol{R_0}$. Figure 7 shows the expected time and the 95% prediction interval for a major outbreak with $\boldsymbol{R_0} = 1.1, 2, 4, 10$. For small values on $\boldsymbol{R_0}$ there is a certain difference.

6.2 *Time for an outbreak when $\boldsymbol{R_0}$ is not constant*

In this situation there rarely exists an exact solution, instead the calculations will be done numerically.

Two typical situations arise with this setting. First situation is diseases like measles which in many countries is controlled by herd immunity. But decreased vaccination coverage can make $\boldsymbol{R_0} > 1$. The second situation is when a new unknown disease enters the population. For example SARS or pandemic influenza. With diseases like these $\boldsymbol{R_0}$ might be far from 1 already at the beginning.

To answer these questions of when after crossing the herd immunity threshold level a major outbreak can occur, two different scenarios are calculated with different assumptions on the development of $\boldsymbol{R_0}$. Assume that one index case enters the population each year. In the first scenario $\boldsymbol{R_0}$ increases with 0.01 units each year, see Figure 8. The second scenario is when $\boldsymbol{R_0}$ first increases, followed by a decrease, see Figure 9.

7. Results and Discussion

When using different epidemic models it is important to consider how the infectious time period is distributed. Both for the case of emerging new infections and for future outbreaks of diseases known which are known today.

The number of infected when the population is protected by herd immunity can also be considerably high. This shows the danger of using deterministic models which will predict the expected value, and this will be very low.

[!ht]

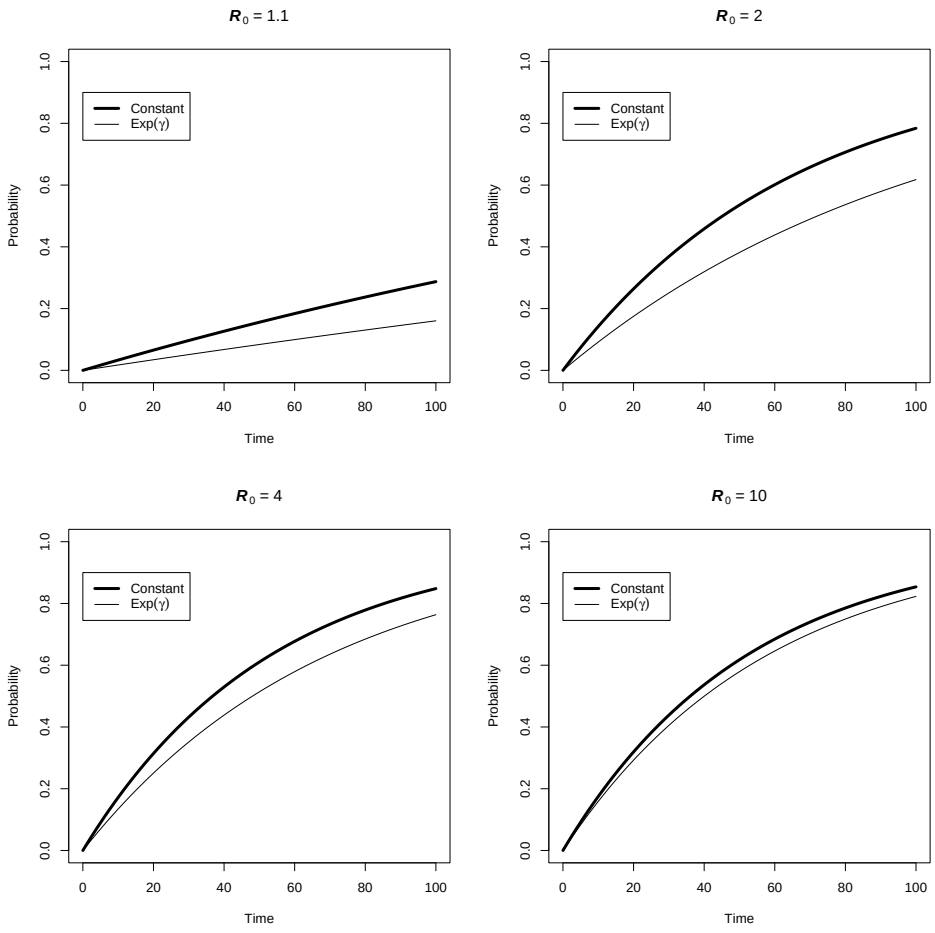

Figure 6. Probability for a major outbreak as a function of $\boldsymbol{R_0}$. The index cases are arriving with a rate of 1/52. This is equal to one new case arriving each year if the time scale is in weeks.

[!ht]

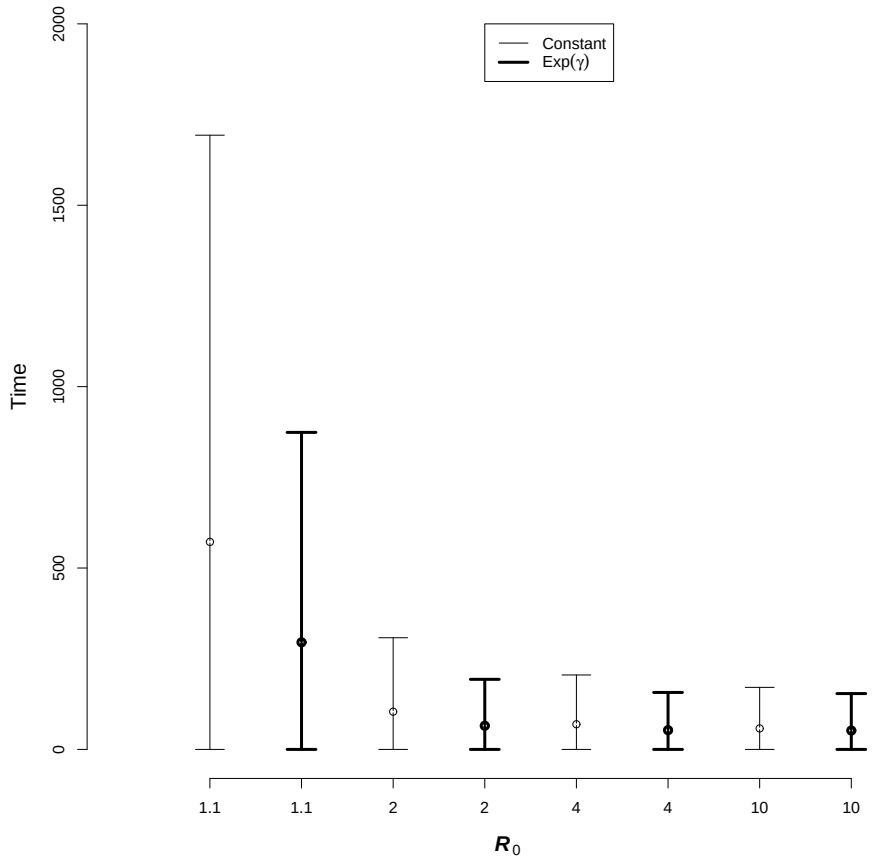

Figure 7. Time until a major outbreak as a function of $\boldsymbol{R_0}$, when $\boldsymbol{R_0}$ is constant. Index cases are arriving at a rate of $1/52$ each week.

[!ht]

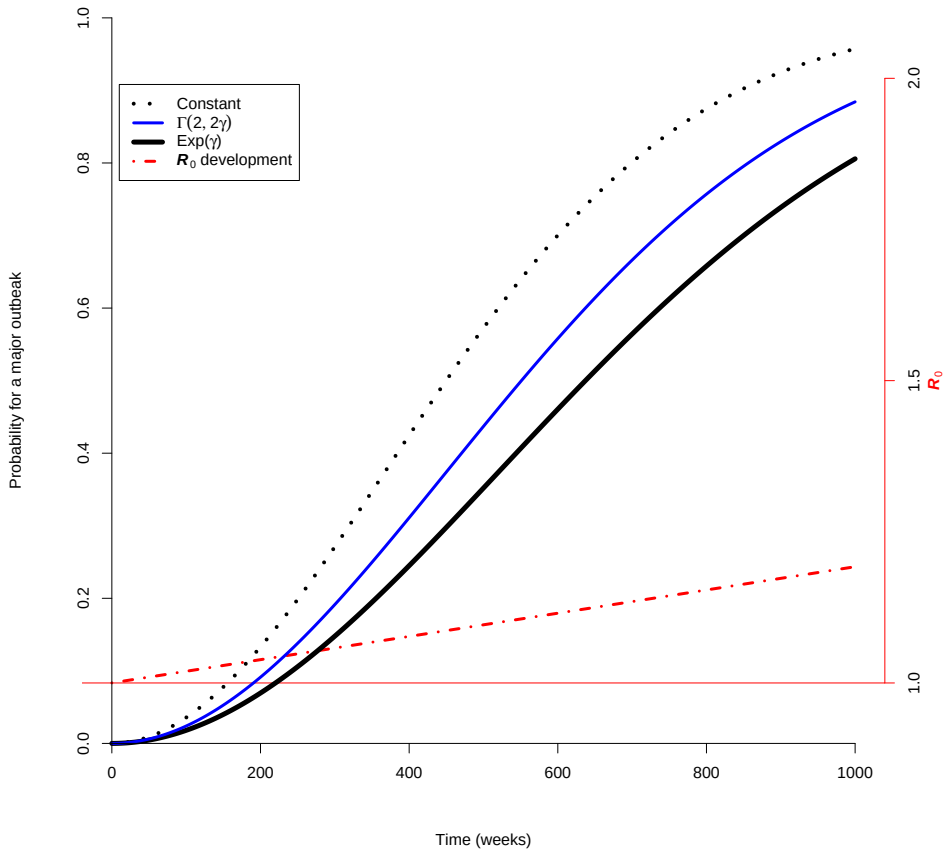

Figure 8. Time until a major outbreak as $\boldsymbol{R_0}$ increases with 0.01 units each year.Index cases are arriving at a rate of $1/52$ each week.

[!ht]

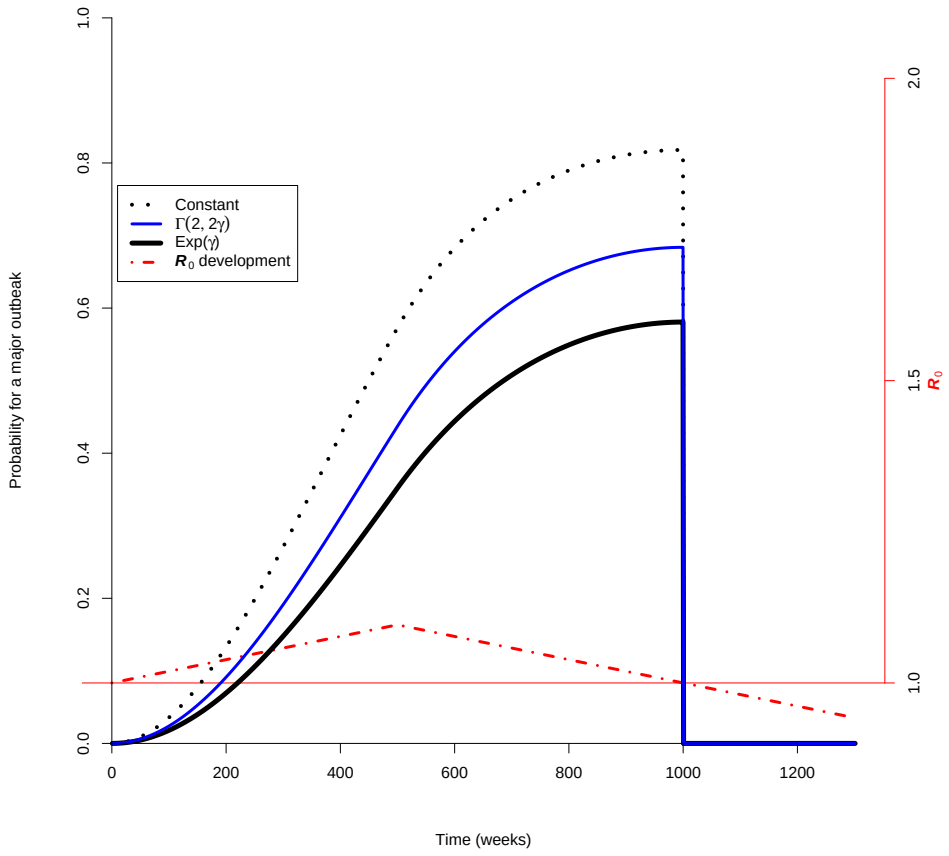

Figure 9. Time until a major outbreak when $\boldsymbol{R_0}$ increases followed by a decrease. Index cases are arriving at a rate of $1/52$ each week.

The time for major outbreaks when crossing the herd immunity level is very sensitive for the assumption on the distribution of the infectious time period.

There are limitations in this report. One is that the population is assumed to be homogeneously mixing. The aim is to concentrate on the random distribution on the infectious time period. Constant, exponentially distributed and gamma distributed time periods have been used. Although the methods can be used for other distributions.

In the future other distributions will be considered. Also the case with different heterogeneity in the population will be considered.

8. Acknowledgments

This work has been financially supported by DG Sanco project MODELREL, Swedish Institute for Infectious Disease Control, Karolinska Institute and Martin Rinds foundation. Thanks also to the epidemic modeling PhD students Elinore Bengtsson and Patricia Geli at Stockholm University for the support to get this report ready. Thanks also to all the staff at Department of epidemiology at the Swedish Institute for Infectious Disease Control for support and help. And making me aware that the problems addressed in this report are based on real problems. Thanks also to my two supervisors, Åke Svensson at Stockholm University and the Swedish Institute for Infectious Disease Control and Johan Giesecke from the European Centre for Disease Prevention and Control.

References

Abbey, H. (1952). An examination of the Reed-Frost theory of epidemics. *Human Biology* **24**, 201–33.

Asikainen, T. (2005). Lessons learned from a smallpox outbreak in stockholm 1963. Technical Report 2005:10, Department of mathematics.

Asikainen, T., Giesecke, J. and Svensson, Å. (2003). Reduced vaccination will result in epidemics, about 25 000 cases are expected to occur annualy within a 25 year period. *Läkartidningen* **40**, 3126–30.

Borel, E. (1942). Sur l'emploi du theoreme de bernoulli pour faciliter le calcul d'un infinite de coefficients. application au probleme de l'attente a un guichet. *C. R. Acad. Sci., Paris* **214**, 452–56.

Daley, D. (1990). The size of epidemics with variable infectious periods. Technical Report SMS-012-90, Statistics Research Section, School of Mathematical Sciences, Australian National University.

Farrington, C., Kanaan, M. and Gay, N. (2003). Brancing process mod-

els for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* **4**, 279–95.

Gani, R. and Leach, S. (2001). Transmission potential of smallpox in contemporary populations. *Nature* **414**, 748–751.

Giesecke, J. (2002). *Modern Infectious Disease Epidemiology*. Arnold, 338 Euston Road, London NW1 3BH, second edition.

Height, F. and Breuer, M. (1960). The Borel-Tanner distribution. *Biometrika* **47**, 143–50.

Heymann, D. L., editor (2004). *Control of communicable diseases manual*. APHA, 18 edition.

Jain, G. C. and Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics* **21**, 501–513.

Kermack, W. and McKendrick, A. (1927). Contributions to the mathematical theory of epidemics, part i. *Proc. Roy. Soc. Edin.* **115**, 700–21.

Kotz, S. and Johnson, N. L. (1983). *Encyclopedia of statistical sciences*, volume 4. Wiley.

Pounder, R., Wakefield, A., Thompson, N. and et al, S. M. (1995). Is measles vaccination a risk factor for inflammatory bowel disease. *Lancet* **34**, 1071–4.

Sewastjanow, B. (1975). *Verzweigungsprozessse*. Akademie-Verlag, 108 Berlin, Leipziger Str. 3-4, second edition.

Ström, J., editor (1964). *Akuta infektionssjukdomar*. Bröderna Ekstrands Tryckeri AB.

Tanner, J. (1953). A problem of inference between two queues. *Biometrika* **40**, 58–69.

Wakefield, A., Munch, S., Anthony, A., Linnell, J., Casson, D. and et al, M. B. (1998). Heallumphoid-nodular, non-specific colitis and pervasive developmental disorder in children. *Lancet* **251**, 637–41.

Mathematical Statistics
Stockholm University

Lessons learned from a smallpox outbreak in Stockholm 1963

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control

Research Report 2005:10

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2005:10**,
http://www.math.su.se/matstat

Lessons learned from a smallpox outbreak in Stockholm 1963

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control*

December 2005

Abstract

Detailed individual data are used to estimate the number of infectious contacts in the population due to an outbreak of smallpox. The data is known due to a contact tracing conducted during the epidemic. This gives individual data on date of probable infection and different stages in the disease.
We have fitted different distributions to the observed times spent in different states of the infection process. The states considered are incubation and period with fever until rash.
We have also considered if there is any difference in infectiousness during the prodromal compared with the rash period.
Estimations are made in respect to community, household and hospital based spread of infection.
The major finding is that many estimates are very dependent of population structure, making comparisons with other studies cumbersome.

KEY WORDS: Smallpox, Variola, vaccination, epidemic model.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91Stockholm, Sweden. E-mail: tommi@math.su.se. Financial support from DG Sanco project MODELREL, Swedish Institute for Infectious Disease Control and the Swedish emergency management agency is gratefully acknowledged.

Contents

List of Tables

List of Figures

Abstract

Detailed individual data are used to estimate the number of infectious contacts in the population due to an outbreak of smallpox. The data is known due to a contact tracing conducted during the epidemic. This gives individual data on date of probable infection and different stages in the disease.
We have fitted different distributions to the observed times spent in different states of the infection process. The states considered are incubation and period with fever until rash.
We have also considered if there is any difference in infectiousness during the prodromal compared with the rash period.
Estimations are made in respect to community, household and hospital based spread of infection.
The major finding is that many estimates are very dependent of population structure, making comparisons with other studies cumbersome.

1 Introduction

This paper describes the analysis of data from the last outbreak of smallpox in Sweden, which occurred during the summer and autumn of 1963. Models for epidemic spread are developed which take households and hospitals into account since most of the spread during the outbreak was concentrated here. We have aimed specifically at two questions:

- How big are the number of infectious contacts per time unit at different locations, i.e. within community, households and hospitals?
- How do the different assumptions of the length of the infectious period fit the data?

Even for a "well known" disease as smallpox there exist many open questions. How infectious are individuals before the onset of rash, is there variability in infectiousness during the rash period? For the Stockholm outbreak we possess individual data from the outbreak concerning probable dates of infection, onset of fever, onset of rash etc. Apart from this it is also known how many were not infected in the cases families and hospital wards if they were hospitalized.

These data are applied into an epidemic model taking hospital and household setting into account. The hypothesis of interest is whether there are differences in infectiousness between the prodromal and the rash period. Also to

calculate the number of infectious contacts at different settings of infection.

We have taken the approach of calculating the infection pressure on individuals in the population. The model takes individual dates for infection into account and calculates the total force of infection experienced by the non infected in the population.

Our population structure shows a great difference in R_0 if we assume a homogeneous population or a heterogeneous mixing assumption. This also means that R_0 is a very unclear measure without a definite meaning. This makes comparisons with other studies very difficult.

The second part concerns time spent in different disease states, which is analyzed separate to the first epidemic model. We use available data from cases and compare these under different distributions, showing that assuming an exponential distribution does not fit the data very well, neither does a period of constant length.

2 Materials and methods

2.1 General description of the epidemic

The outbreak was initiated by a Swedish seaman who flew home from Australia via New Zealand, Singapore, India, Pakistan, Iran, Switzerland, Germany and Denmark. At that time there were smallpox outbreaks in India and Pakistan, and it is probable that this individual was infected in one of these two countries. This is the first documented case of transmission of smallpox during an airport transfer.

The seaman had previously been vaccinated against smallpox. Two weeks after arrival in Stockholm, he developed fever followed by reddish rashes. Since he thought it was acne, he did not seek a physician and stayed at home instead. There he was taken care of by his 80 year old grandmother who two weeks later became ill with similar symptoms and was suspected to have chickenpox. A third family member later also became ill. Later on the grandmother was transfered to a hospital where she infected other people.

A community nurse visited the grandmother on a daily basis became infected and later also infected her family.

It took one month before the first diagnosis of smallpox was made at Danderyds sjukhus.

The major part of the outbreak was concentrated in three families and three different hospitals. The spread within hospitals was due to incorrect diagnosis. The probably infected persons were not isolated but were under observation in large dormitories where nurses went around and observed the patients. After smallpox was diagnosed the patients were isolated.

A total of 27 persons were infected, and four of these died. Three of the 27 had not been vaccinated prior to the epidemic, two of these died. With vaccination we mean that the individual has been vaccinated as a child, in military service and / or before travel to a country with smallpox. Only persons vaccinated before the epidemic are considered as vaccinated, not those receiving vaccine after a probable infection of smallpox.

During the epidemic a massive vaccination campaign was launched. Involving both contact tracing of possible infectious contacts as well as a mass vaccination of more than 112 000 persons. It took approximately four months until the last infected person recovered and no new cases were found.

Some unexpected infectious routes were identified. An embalmer (a person involved in the cleaning of corpses) was infected by contact with a corpse. Another individual was infected through transporting laundry from one of the hospitals, when he came in contact with clothes from one of the isolation rooms. He also delivered food to a nearby hospital causing secondary cases at that location.

For a general overview of how cases infected one another please see Figure 4 on page 23 and Figure 5 on page 24.

2.2 Data

The data from this outbreak was obtained from (Ström [1966]), a TV documentary (Wallén [1964]) and medical journals at the Swedish Institute for Infectious Disease Control. Since contact tracing of possible contacts was undertaken, information is available regarding both cases as well as persons who were exposed but not infected (i.e non cases). The following data are available:

For the cases:

- Whom they were probably infected by
- Which phase of disease the infectious person was in at probable contact
- Possible dates of infection
- Date of onset of fever
- Date of onset of rash
- Date of isolation
- Date discharged from the hospital / death
- Vaccination status for smallpox, in some cases also the time of vaccination
- Which family they belonged to and number of family members
- Where they were hospitalized and in which ward

For the non cases:

- Probable infectious contacts
- Possible dates of infection
- If they belonged to a family
- Where they were hospitalized and in which ward

The date for probable infectious contact is not used in the calculations because not all possible contacts are known. Information regarding vaccination status of non cases was unfortunately not available. We only know that they were given boosters after contact tracing. We assume no asymptomatic cases (i.e infected without symptoms) existed, since this is very unlikely for smallpox.

Short summary of data concerning cases is shown in Table 1.

Variable	Number of observations		
Vaccination status	27	24 vaccinated	
Family size	27	$(1-4)$	
Hospital ward size	27	$(27-51)$	
Time period		Mean	Standard deviation
Infection to onset of fever	26	11.73	4.47
Onset of fever to rash	23	3.17	1.61

Table 1: Short summary of data for cases

2.3 Smallpox in general

Smallpox also known as Variola Major or Variola Minor, is a disease which is eradicated worldwide since 1979. It was a much feared infectious disease and the first vaccines were developed in the 19th century.

The virus transmits through air and body fluids and often through close contacts with infectious individuals.

The typical presentation of disease is that first symptoms appear two weeks after infection in the form of high fever. This lasts for about three days after which a rash appears on the body, which is similar to chickenpox or acne in the beginning. These rashes last 2-3 weeks and the person is bedridden during this stage.

Variola Major can be of four different types: ordinary (more than 90% of cases), modified (mild, common for previously vaccinated persons), flat and hemorrhagic.

The case fatality rate (CFR) for Variola Major of ordinary type is approximately 30%. For flat and hemarrhagic it is almost 100%, modified is rarely lethal. For Variola Minor CFR is approximately 1%.

Protection against smallpox can be acquired in two ways, immunity following infection or through vaccination. The vaccine is called Variola and gives protection for a certain period, the protective effect decreasing with time. One advantage is that it decreases the severity of infection, i.e individuals will develop Variola Major modified as opposed to Variola Major ordinary. This gives a much higher probability of survival given infection.

In Sweden vaccination was a part of the childhood vaccination program until 1976, with an indicator of successful vaccination being a circular scar. After the virus was eradicated from the world, vaccination ceased. The last known case was diagnosed in Somalia october 1977. Global eradication was certified by World Health Organisation in 1979. A laboratory associated smallpox death occurred at University of Birminghan, England 1978.

The long term protection of the vaccine is questionable, with some studies having shown protection after 20 years (Eichner [2003]).

2.4 Progress of infection

We will give a crude sketch of both of how the infectious disease and the infection spread develops in theory (Giesecke [2002]).

As an infectious disease it can be viewed in the following stages, see also Figure 1:

1. Healthy: The person has not yet been infected and is not protected against it.
2. Incubation period: A susceptible person gets infected by an infectious individual. Incubation period is the period of time between initial infection and the onset of clinical disease (these manifestations may not always be clinically visible).
3. Clinical disease: Appearance of visible symptoms
4. Resistant: The person has recovered from the disease, or has been successfully immunized through vaccination.

How infection is spread differs from the above because an individual can be infectious before the stage of clinical disease, see also Figure 2:

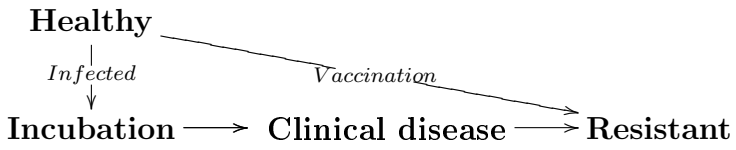

Figure 1: Progress of infectious disease

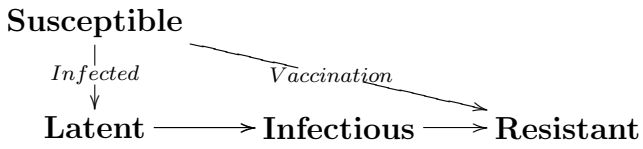

Figure 2: Progress of infection

1. Susceptible: The person has not yet been infected and is not protected against it.
2. Latent period: When a person becomes infected by an infectious individual.
3. Infectious period: The person is now infectious and can transmit the disease to others
4. Resistant: The person has recovered from the disease, or has been successfully immunized through vaccination.

This is modeled by assuming that there is a certain amount of transition between the different states. The transition in our case is in one direction:

For infectious agents like influenza the recovered person can be susceptible again, depending on which type of influenza is circulating the next influenza season. But in our case everyone who is resistant will remain resistant.

Infected persons are assumed to meet other individuals in the population according to a Poisson process. This implies that the number of contacts in any time period is a Poisson distributed random variable. If a proportion of the population is immune the number of contacts that will result in a new infection will reduce in correspondence to this.

2.4.1 Fitting distributions to observed time observations

The data used will be the available data for the periods. These transition periods investigated are the following: infected to onset of fever and fever to onset of rash. The time from onset of rash until discharged has not been analyzed since it does not measure how long time a person is infectious. We want to emphasize that we are not estimating these parameters through our epidemic model. Each phase in infection will be analyzed separately.

2.4.2 Time spent in different states

When trying to model the transitions between different states the time spent in different states, many times it is assumed to be exponentially distributed. This allows the possibility to use Markovian properties to calculate, for example, the size of the epidemic. Many studies have used or are using this distribution. The limitation is that no diseases have shown to follow an exponential distribution in the different states, making this approach mostly of academic interest.

Another approach is to assume that the time spent in the different states is constant or to find a suitable balance between the exponentially and constantly distributed transition times.

We will be using an approach by assuming the occupation times to follow a gamma distribution.

2.4.3 Basic reproductive number R_0

A parameter of interest in this case is the so called "basic reproductive number", denoted R_0. This is a crude measure of how many infected persons are generated by one infectious individual entering a totally susceptible population. If $R_0 < 1$ no large outbreaks will occur, only a small outbreak is possible. If $R_0 > 1$ there will be a positive probability for a large outbreak, meaning that a considerable proportion of the population could get infected.

2.5 The epidemic model used

As in (Eichner and Dietz [2003]) we are using a $SEIR$ model to describe the progress of the disease.

- $\boldsymbol{S}$ =Susceptible
- $\boldsymbol{E}$ =Exposed (latent period)
- $\boldsymbol{I}$ =Infectious period, which is divided into two parts:
  - $\boldsymbol{I}_{Fever}$ = Fever to onset of rash
  - $\boldsymbol{I}_{Rash}$ = Rash until recovered / death
- $\boldsymbol{R}$ = Resistant / death

Given infection the person will first be in the exposed (latent) period for about two weeks. This is followed by the fever phase which last about three days. Then a rash starts to appear and recovering from this stage takes an additional 1-3 weeks. There are two possible outcomes, namely death or recovery. Schematically this model is showed in Figure 3.

When establishing the model we are also interested in any differences in "infectiousness" in the states $\boldsymbol{I}_{Fever}$ and $\boldsymbol{I}_{Rash}$, this ratio being denoted by b.

2.6 Notations

The following notation is used:

- $t^i_{\mathbf{Infected}}$ = Time point for infection, when several days exist we have taken their mean value.
- $t^i_{\boldsymbol{F}}$ = Time point for onset of fever
- $t^i_{\boldsymbol{Rash}}$ = Time point for onset of rash
- $t^i_{\boldsymbol{Isolated}}$ = Time point for isolation / recovery (recovery for cases not isolated)
- N =population size
- n = number of cases

2.7 Calculation of infectiousness

We have adopted the approach to calculate the infection pressure on individuals in the population. This is done by calculating the number of individuals who are in the fever or rash period at a certain point in time. The effect of different infectiousness in the fever and rash period is also taken into account

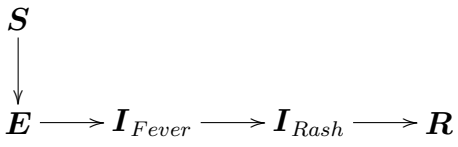

Figure 3: Epidemic model used

through the variable b.

Let $h_j(t)$ measure the infectious status of person j at time j. These are $\boldsymbol{I}_{Fever}$, $\boldsymbol{I}_{Rash}$ and not infectious. Let $\mathbf{J}$ be an indicator function.

$$h_j(t) = b\mathbf{J}_{\{t^j_{\boldsymbol{F}} \leq t < t^j_{\boldsymbol{Rash}}\}} + \mathbf{J}_{\{t^j_{\boldsymbol{Rash}} \leq t \leq t^j_{\boldsymbol{Isolated}}\}}$$

To calculate the total "infection pressure" individual i is exposed to, $\lambda_i(t)$ at a certain time point t. The $h_j(t)$ for all cases j are summarized and multiplied by the number of infectious contacts per day.

$$\lambda_i(t) = \sum_{j=1}^{27} h_j(t) * \text{number of infectious contacts per day}$$

2.7.1 Population structure

A number of different structures for the population is possible. We have assumed two different models for dynamics in our population. One assumes homogeneous mixing, that is no family or hospital effects are taken into account. The second approach takes into account the effect of family and hospital on infectious contacts.

In the homogeneous mixing model a number of contacts per day κ is assumed and this is spread uniformly within the rest of the population. If the population size is N, one infectious individual is in mean transmitting $\frac{\kappa}{N-1}$ infectious contacts to any other individual.

In the heterogeneous mixing we have attempted to set up a population structure with households and hospitals taken into account. If a persons belongs to multi households he will be exposed to two rates, κ and κ_F. Here κ_F is an addition per family members. If he belongs to a single household his exposure is the rate κ. If a person belong to a hospital he is exposed to a rate κ_H. We assume only infectious contacts within hospitals.

- $\widehat{\kappa}$ = Number of infectious contacts per day with any person in a population with hospitals and households

- κ = Number of infectious contacts per day with any person in a population with hospitals and households
- κ_F = Number of infectious contacts per day with a certain family member
- κ_H = Number of infectious contacts per day with any person in a hospital
- N = Population size
- n_H^i = Number of people in the hospital ward where person i is situated

Letting $h_j(t)$ denote in which infectious state person j is at a certain time point we can describe the population structure. Assuming an homogeneously mixing population the force of infection $\lambda_i(t)$ follows as:

$$\lambda_i(t) = \sum_{j=1}^{n} h_j(t)\left(\frac{\kappa}{N-1}\right)$$

If a heterogeneously mixing population is used:

$$\begin{aligned}\lambda_i(t) &= \sum_{j=1}^{27} h_j(t)\mathbf{J}(hosp_i(t)=0)\left(\mathbf{J}(f_j=f_i)\left(\frac{\kappa}{N-1}+\kappa_F\right)+\mathbf{J}(f_j\neq f_i)\frac{\kappa}{N-1}\right)+ \\ &+ \sum_{j=1}^{27} h_j(t)\mathbf{J}(hosp_i(t)=1)\left(\frac{\kappa_H}{n_H^i-1}\right)\end{aligned}$$

Here f_i is the family identity for person i, $hosp_i(t)$ is an indicator if person i is in hospital at time t. The procedure is to first check if person i is in hospital. If yes ($hosp_i(t)=1$), he is exposed to certain number of infectious contacts. If no ($hosp_i(t)=0$), he is exposed to some other infectious contacts on which family he belongs to. **J** is an indicator function.

2.7.2 Calculation of R_0

Assuming that we know the parameters of interest, this allows us to calculate R_0, the expected number of infectious contacts during the infectious time period. We define it as a product of number of infectious contacts per day multiplied with the total length of the two infectious periods. Using the same notations as before, the parameters of interest are the scale factor b, number of infectious contacts per day κ, κ_F, κ_H and the length of the infectious time

periods (prodromal + rash).

Concerning the length of the infectious periods, the prodromal period (fever to rash) has the mean length based on our data. The rash period is probably overestimated in this data, and this is the reason for including other time periods as well. We are assuming that the two time periods are distributed as a gamma distribution.

$$\begin{aligned} \boldsymbol{I}_{Fever} &\sim \Gamma(\alpha_f, \beta_f) \\ \mathrm{E}[\boldsymbol{I}_{Fever}] &= \frac{\alpha_f}{\beta_f} = \mu_{Fever} \\ \boldsymbol{I}_{Rash} &\sim \Gamma(\alpha_r, \beta_r) \\ \mathrm{E}[\boldsymbol{I}_{Rash}] &= \frac{\alpha_r}{\beta_r} = \mu_{Rash} \end{aligned}$$

Now R_0 can be estimated in the situation with homogeneous or heterogeneous mixing assumptions:

Homogeneously mixing population:

$$R_0 = (N-1)\frac{\widehat{\kappa}}{N-1}\Big(b\mu_{Fever} + \mu_{Rash}\Big) = \widehat{\kappa}\Big(b\mu_{Fever} + \mu_{Rash}\Big)$$

Heterogeneous mixing population with a single household:

$$R_0 = (N-1)\frac{\kappa}{N-1}\Big(b\mu_{Fever} + \mu_{Rash}\Big) = \kappa\Big(b\mu_{Fever} + \mu_{Rash}\Big)$$

Heterogeneous mixing population with a multi person household:
Let f_i and f_j be family indicators for person i and j. Calculate the number of infectious contacts caused by person i. $\mathbf{J}$ is an indicator function. Let n_F denote the number of additional family members in a household.

$$\begin{aligned} R_0^F &= \sum_{j \neq i} \Big(\mathbf{J}(f_i = f_j)\frac{\kappa}{N-1} + \mathbf{J}(f_i = f_j)\kappa_F\Big)\Big(b\mu_{Fever} + \mu_{Rash}\Big) = \\ &= \Big(\frac{n_F}{N-1}\kappa + n_F\kappa_F\Big)\Big(b\mu_{Fever} + \mu_{Rash}\Big) \end{aligned}$$

Heterogeneous mixing population with within hospital infection:
Let $hosp(i)$ and $hosp(j)$ be hospital ward indicators for person i and j. Calculate the number of infectious contacts caused by person i. $\mathbf{J}$ is an indicator function.

$$\begin{aligned} R_0^H &= \left(\frac{\kappa_H}{n_H^i}\sum_{j\neq i}\mathbf{J}(hosp(i)=hosp(j))\right)\Big(b\mu_{Fever}+\mu_{Rash}\Big) = \\ &= \left(\frac{\kappa_H}{n_H^i}n_H^i\right)\Big(b\mu_{Fever}+\mu_{Rash}\Big) = \kappa_H\Big(b\mu_{Fever}+\mu_{Rash}\Big) \end{aligned}$$

We note that in the case of heterogeneous mixing the calculation is made as an expected value, not as most often the largest eigenvalue of a mixing matrix.

2.7.3 Likelihood calculations

The infectious contacts are assumed to follow a Poisson process with intensity $\lambda_i(t)$ where $\lambda_i(t)$ is calculated in Section 2.7. The probability to become infected at time s, given no infection before is $\lambda_i(s)e^{-\int_0^s \lambda_i(\tau)d\tau}$. If we know the time of infection $t^i_{\boldsymbol{Infected}}$ and the vaccine efficacy v we can calculate the likelihood for the cases and non cases as follows:

- v = vaccine efficacy.
- δ_i = Vaccination status for case number i, 1 if vaccinated, 0 otherwise
- $t^i_{\mathbf{Infected}}$ = Day of infection for case number i.
- $\lambda_i(t)$ = Force of infection on individual i, day number t.
- $\mathcal{N}_c$ = set consisting of all cases
- $\mathcal{N}_p^v$ = set consisting of vaccinated non cases cases not belonging to a household or hospital during the epidemic
- $\mathcal{N}_p$ = set consisting of non vaccinated non cases cases not belonging to a household or hospital during the epidemic
- $\mathcal{N}_F$ = set consisting of non cases belonging to a household but not belonging to a hospital during the epidemic

- $\mathcal{N}_H^v$ = set consisting of vaccinated non cases not belonging to a household but belonging to a hospital during the epidemic
- $\mathcal{N}_H$ = set consisting of non vaccinated non cases not belonging to a household but belonging to a hospital during the epidemic

The likelihood functions can be calculated as follows (Eichner and Dietz [2003]):

$$\begin{aligned}
P_i(\text{case}) &= (1-v)^{\delta_i}\lambda_i(t^i_{\boldsymbol{Infected}})e^{-\int_0^{t^i_{\boldsymbol{Infected}}}\lambda_i(\tau)d\tau} \\
P_i(\text{non case}) &= (1-v)^{\delta_i}e^{-\int_0^{\infty}\lambda_i(\tau)d\tau} \\
L &= \prod_{i\in\{\text{cases}\}} P_i(\text{case}) \prod_{\text{i}\in\{\text{non case}\}} \text{P}_\text{i}(\text{non case}) = \\
&= \prod_{i\in\mathcal{N}_c} P_i(\text{case}) \prod_{\text{i}\in\mathcal{N}_\text{p}^\text{v}} \text{P}_\text{i}(\text{non case}) \prod_{\text{i}\in\mathcal{N}_\text{p}} \text{P}_\text{i}(\text{non case}) \\
&\quad \prod_{i\in\mathcal{N}_F} P_i(\text{non case}) \prod_{i\in\mathcal{N}_H^v} P_i(\text{non case}) \prod_{i\in\mathcal{N}_H} P_i(\text{non case})
\end{aligned}$$

$$\begin{aligned}
\log L &= \sum_{i=1}^{n}\delta_i \log(1-v) + \sum_{i=1}^{n}\log\lambda_i(t) - \sum_{i=1}^{n}\int_0^{t^i_{\boldsymbol{Infected}}}\lambda_i(\tau)d\tau + \\
&+ \sum_{i\in\mathcal{N}_p^v}\log\left(v+(1-v)e^{-\int_0^{\infty}\lambda_i(\tau)d\tau}\right) + \\
&+ \sum_{i\in\mathcal{N}_F}\log\left(v+(1-v)e^{-\int_0^{\infty}\lambda_i(\tau)d\tau}\right) + \\
&+ \sum_{i\in\mathcal{N}_H^v}\log\left(v+(1-v)e^{-\int_0^{\infty}\lambda_i(\tau)d\tau}\right) - \\
&- \sum_{i\in\mathcal{N}_p}\int_0^{\infty}\lambda_i(\tau)d\tau - \sum_{i\in\mathcal{N}_H}\int_0^{\infty}\lambda_i(\tau)d\tau
\end{aligned}$$

The likelihood for the cases is calculated by taking the family- and hospital place at different times into account. This is included in the calculation of $\lambda(t)$. For non cases there is a division made depending if they are in a multi person household or hospital. In a household due to the high vaccination proportion in the population all are considered to be vaccinated. In a hospital the number of vaccinated and unvaccinated are rounded off to the closest

Time period	α	β	Expected value	Std	n
infected to fever	7.58	0.64	11.73	4.26	26
fever to rash	4.05	1.28	3.17	1.58	23

Table 2: Parameters for duration in different states

integer.

The likelihood for non cases is depending on the proportion vaccinated prior the epidemic. We have assumed this to be the same as the proportion for cases. This is of course questionable, sensitivity analysis could be made or an implication of the the EM-algorithm to estimate the "expected proportion".

To estimate the parameters of interest profile likelihood methods have been used. This is done by fixing one parameter and varying the others. The maximum of these is taken as the likelihood for the fixed parameter. Then the fixed parameter is given a new value and it is maximized again over the other ones. The maximum likelihood estimation for the fixed parameter will be the point at which the maximum value for the likelihood is obtained.

3 Results

3.1 Times spent at different states

Assuming that the amount of time spent in different states follows a $\Gamma(\alpha, \beta)$-distribution with expected value $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$. The parameters α (shape) and β (scale) are estimated through maximum likelihood. The yielded estimates are shown in Table 2.

The times for rash to recovered are overestimated since the recorded recovery date is probably over-exaggerated in order to ensure that the patient is no longer infective.

The infected to fever period and the prodromal period (2 weeks and 3 days respectively) are similar to those found in other studies.

When comparing this with an exponential distribution we know that the

variable	value	95% confidence interval
scale	0.51	(0.1 - 1.5)
vaccine efficacy	0.98	(0.95 - 0.99)
κ	0.47	(0.25 - 0.85)
κ_H	0.54	(0.31 - 0.92)
κ_F	0.057	(0.015 - 0.12)

Table 3: Estimates for parameters of interest, heterogeneous mixing

shape parameter α should be around 1 in the Γ-distribution. Following our estimations this value is far from one, indicating that an exponential distribution is not suitable in the case of smallpox. This is enough only if we are interested in how big the outbreak will become.

3.2 Estimation of parameters of interest

Estimating our parameters of interest are shown in Table 3.

The wide interval on the scale parameter b indicates that there is uncertainty concerning a difference in infectiousness at different infectious states.

The vaccine efficacy v seems to be high. We are unsure of the number of vaccinated individuals in the total population and how the protective effect diminishes over time.

The number of infectious contacts are approximately 0.5 per day in community and hospitals. We want to emphasize that by varying the population structure these values can change considerably.

4 Discussion

The major finding is that although very detailed data, it is very hard to find "good" estimates for even the most basic measures of interest. In our case these are the number of infectious contacts per day and the length of the infectious period.

These two parameters are crucial to estimate R_0 and are very sensitive to the structure given to the population. We have also compared with a homogeneously mixing population and the results differ dramatically from the values obtained in our heterogeneous population. Also when changing the infectiousness within families, assuming some infectious contacts with all population and extra infectiousness spread among family members, altered the results. This means that interpreting and comparing R_0 values obtained in different studies is very difficult since usually the population structures are different. When simulating possible smallpox outbreaks the results should vary. This also for the case with same value on R_0 but different population structures.

Testing infectiousness in the prodromal period with the rash period gives no significant difference. The estimate is about 0.5. If we assume a homogeneous mixing population the infectiousness in the prodromal period is about the same as with heterogeneous mixing. Still this is far from the assumption used in (Halloran et al. [2002]) where a 10 times higher infectiousness is assumed in the prodromal period.

Our calculations are also perturbated by the fact that a number of infectious contacts occur in unexpected places. Some scenarios are hard to put into the model. For example in the beginning of the outbreak a home visiting nurse is infected. In the model she belongs to the first affected family in the beginning but must at one time point be transfered to her own family where she infects her relatives. Also the person transporting clothes from the hospital and the embalmer are unusual occupations to include in a realistic way. In our case these infectious contacts are taking place at a hospital.

We have also assumed that no more infectious contacts will take place while a person is isolated in the infectious disease hospital. The embalmer is an example that this is not a completely true assumption.

The largest error is probably the fact that in calculating "R_0" it is assumed that the populations remain the same. But in the event of an epidemic situation people are probably much more attentive to their contacts thus the structure could change dramatically. In our case the extensive vaccination campaign of over 112 000 individuals vaccinate should also matter concerning the proportion resistant in the population at different time points. Also the fact that large populations surrounding the infected households were mass vaccinated must affect the result. These events would decrease the number of infectious contacts per day, suggesting that at the beginning the the number of infectious contacts should be higher than at the end of the epidemic.

Our results show that on using values of R_0 in simulations of possible small-

pox outbreaks, great care should be taken when interpreting the meaning of outputs. Are the assumptions similar to our model, if not how will they change the outcome?

As noted in the beginning of discussion, only a small number of parameters are possible to estimate reasonably. One can only raise the question in many simulations where maybe more than 15 parameters are used. How reliable are these? Probably only two are known and the rest are "estimated" through some medical expertise. The known parameters may be the length of the infectious period and the number of infectious contacts which are already very unsure. Leading to the conclusion that we probably know nothing with certain for these parameters. In such case it is probably wiser to use much simpler models for more realistic situations, than using many values for the unknown parameters as in (Kretzschmar et al. [2004] and E.H.Kaplan et al. [2002]). An extreme example is (Epstein et al. [2004]) where 30 parameters are used.

The inportant feature with the Stockholm outbreak was that a large part of it occurred within households. It is a striking contrast with many other study findings where hospitals have been the coins for infection transmission.

5 Acknowledgments

This project started as a small project initiated by Fredrik Liljeros. It has involved all epidemic modelling and many medical experts in Stockholm. Thanks to Anders-Martin Löf and epidemic modelling PhD students Elinore Bengtsson and Patricia Geli at Stockholm University for theoretical and practical help. Thanks also to Sven Britton, Anders Ternhag, Mikael Stenhem and Anders Tegnell with medical expertise. The EPIET-fellows Lara Payne and Benn Sartorius at Swedish Institute for Infectious Disease Control (SIIDC) for help with corrections. Special thanks to Pasi Penttinen at SIIDC with many interesting medical views on how epidemic modelling should be done. And finally the biggest thanks to Åke Svensson at SIIDC and Stockholm University for interesting views on how to adapt models to reality and not the opposite which is almost always done in this field. The project has been funded through DG Sanco project MODELREL, Swedish Institute for Infectious Disease Control and the Swedish Emergency Management Agency.

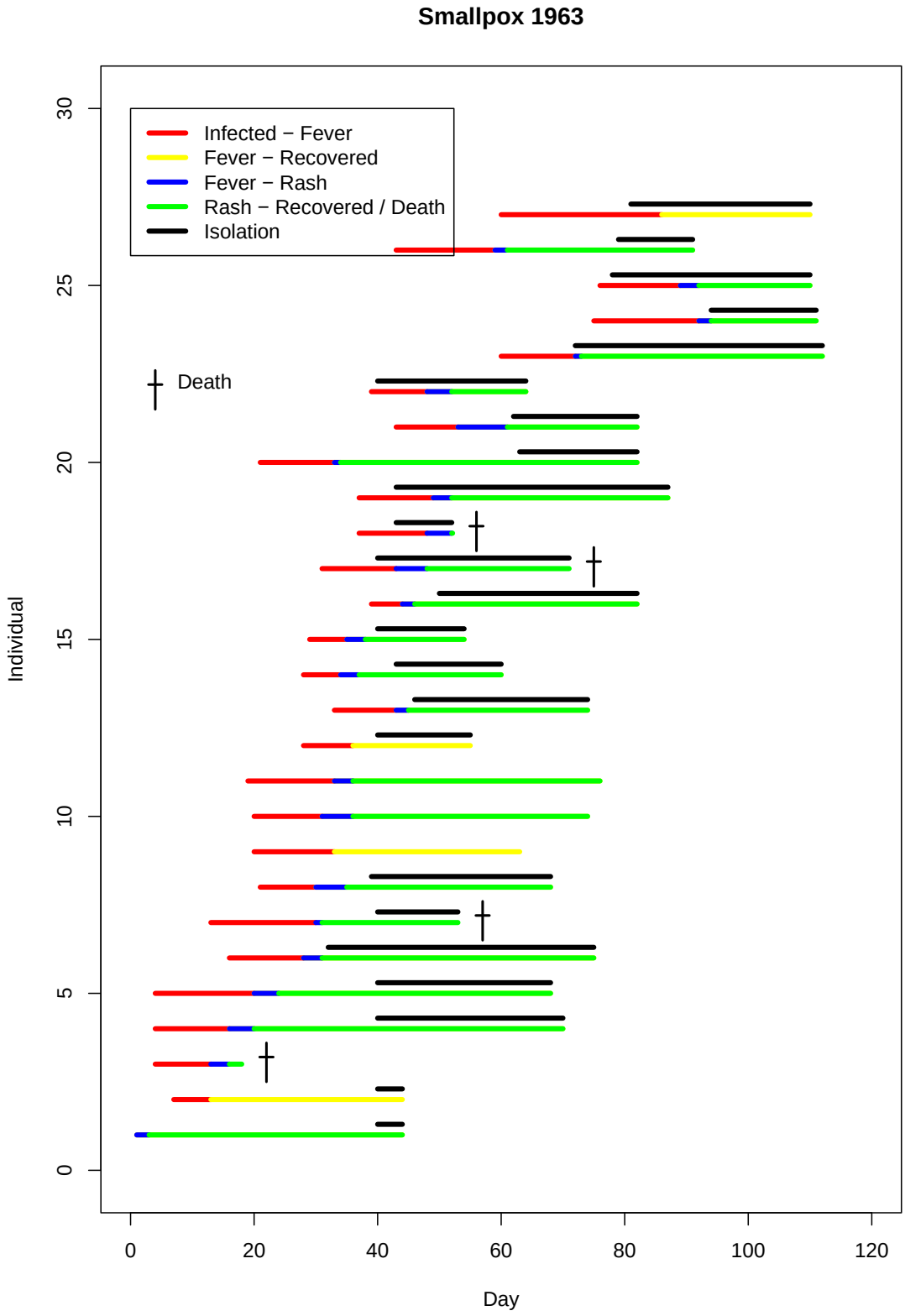

Figure 4: Available data on infected persons after the onset of fever of the first person. Scale on x-axis = Number of days.

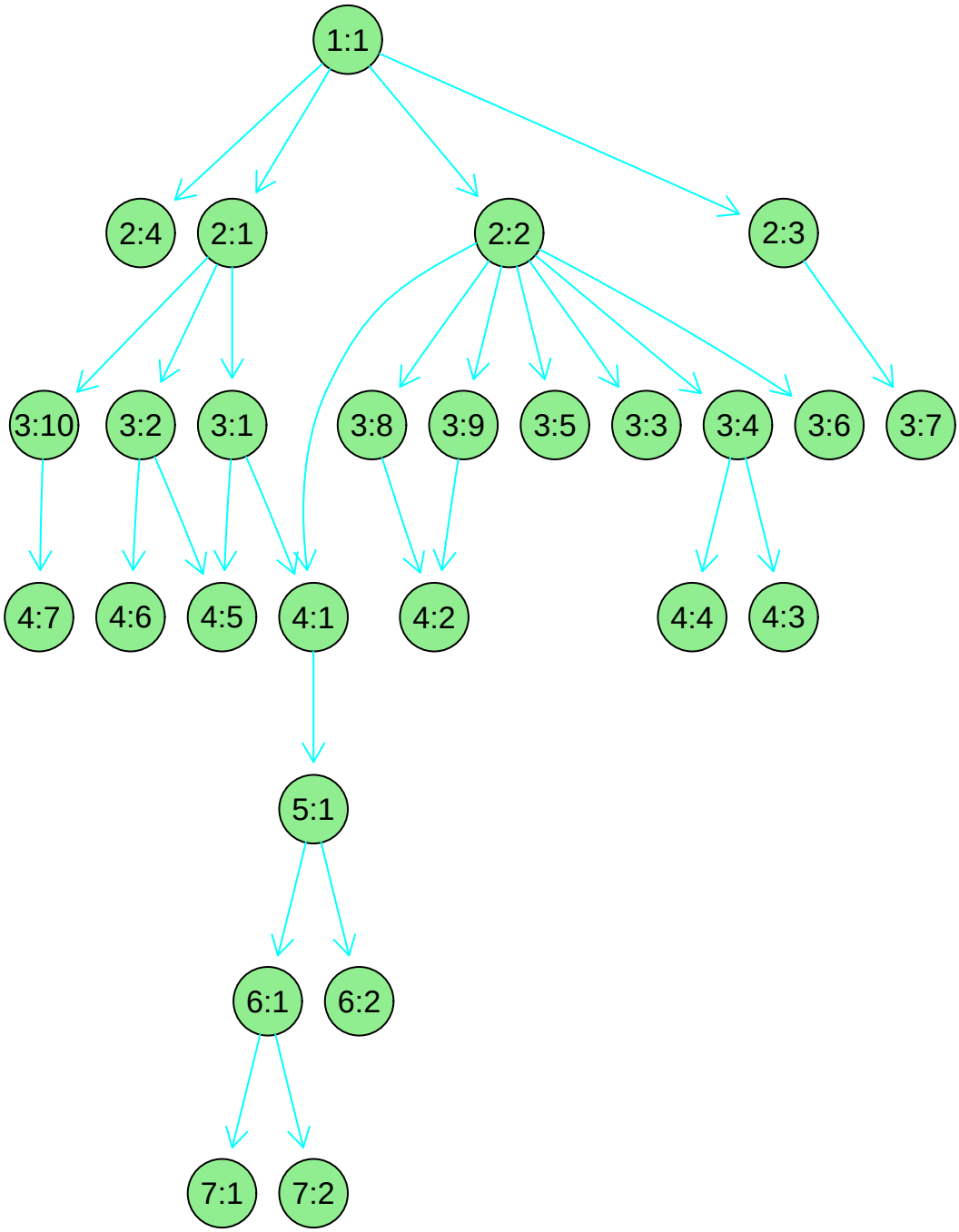

Figure 5: Schematic overview of the epidemic, how cases infected each other.

References

E.H.Kaplan, D.L.Craft, and L.M.Wein. Emergency response to a smallpox attack: The case for mass vaccination. *Proc Natl Acad Sci U S A*, 99: 10935–40, 2002.

M. Eichner. Analysis of historical data suggests long-lasting protective effects of smallpox vaccination. *American Journal of Epidemiology*, 158:717–23, 2003.

M. Eichner and K. Dietz. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *American Journal of Epidemiology*, 158:110–17, 2003.

J. M. Epstein, D. A. Cummings, S. Chakravarty, R. M. Singa, and D. S. Burke. *Toward a Containment Strategy for Smallpox Bioterror An Individual-Based Computational Approach.* Brooking Institution Press, Brookings Institution Press, Attn: Book Order Department, 1775 Massachusetts Avenue, NW, Washington DC 20036, first edition, 2004.

J. Giesecke. *Modern Infectious Disease Epidemiology.* 2002.

M. E. Halloran, I. M. L. Jr, A. Nizam, and Y. Yang. Containing bioterrorist smallpox. *Science*, 298:1428–32, 2002.

M. Kretzschmar, S. van den Hof, J. Wallinga, and J. van Wijngaarden. Ring vaccination and smallpox control. *Emerging Infectious Diseases*, 10:832–41, 2004.

J. Ström, editor. *Smallpox outbreak and vaccination problems in Stockholm, Sweden 1963.* Kungl. boktryckeriet P.A. Norstedt & Söner, 1966.

L. Wallén. *NMAPQ KWAPJ BADBO - om smittkoppsepidemin i Stockholm 1963.* Telefilm, Prod. nr 5331/64, 1964.