Mathematical Statistics
Stockholm University

Stochastic epidemic models for endemic diseases: the effect of population heterogeneities

Mathias Lindholm

Research Report 2007:10
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 **2007:10**,
Licentiate thesis
http://www.math.su.se/matstat

Stochastic epidemic models for endemic diseases: the effect of population heterogeneities

Mathias Lindholm*

May 2007

Abstract

This thesis consists of two papers

1. Lindholm M., Britton T. (2007): Endemic persistence or disease extinction: the effect of separation into sub-communities, (to appear in Theoretical Population Biology).
2. Lindholm M. (2007): On the time to extinction for a two-type version of Bartlett's epidemic model, Stockholm University Research Reports in Mathematical Statistics 2007:9 (submitted).

Both papers deal with stochastic epidemic models for endemic diseases, and in particular how population heterogeneities affects such a disease's ability to persist in a population over a long period of time. As a measure of persistence we use the time to extinction, for which we describe approximations. These approximations make it possible to draw some conclusions about the behaviour of the underlying epidemic that are also supported by simulations.

KEY WORDS: Stochastic epidemic models, Time to extinction, Markov processes, Quasi-stationary distribution, Diffusion approximation.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Stockholm, Sweden. E-mail: lindholm@math.su.se. The author is grateful to Swedish Foundation for Strategic Research (SSF).

Acknowledgements

Tom Britton, thank you for your endless patience and many inspiring and enlightening discussions. Without your supervision I would probably be driving a bus instead of doing mathematical statistics. Another source of inspiring discussions is my co-supervisor Anders Martin-Löf with his wide knowledge and experience of mathematical statistics in general.

I would also like to thank all of you at the Department of Mathematical statistics at Stockholm University for providing such a friendly and stimulating atmosphere. I would especially like to thank my friends and colleagues Andreas Nordvall Lagerås, Elinore Gustafsson and Patricia Geli for all the very much needed coffee breaks, informal discussions about all kinds of (un)important intriguing topics, and for just being good friends.

Apart from all the people at work, the ones who really have to endure the consequences of my efforts, good and bad, is my family, which always succeed in putting my perspective on the really important issues in every day life. For this I am enormously grateful. The biggest part of this job is by all means carried out by you Emma and I am very glad that you still manage to put up with me.

1 A brief non-technical introduction to epidemic models

On p. 295 in The Concise Oxford Dictionary (1976) we find the following definition of the word "disease":

> **dĭsea´se** (-zē´z) n. Unhealthy condition of body, mind, plant, or some part thereof, illness, sickness; particular kind of this with special symptoms or location.

In light of this, diseases, and in particular infectious diseases, are something that we want to be able to control, and understand the behaviour of. By constructing mathematical models describing how a disease can be transmitted between individuals in a population, often with a simple mechanism, we gain insight to the dynamics of the disease. Early work of this kind is from the beginning of the 20th century, and perhaps the first really influential mathematical model is a deterministic model due to Kermack and McKendrick from 1927, see p. 7 in Andersson and May (1991). The probably most important stochastic model from that era is the chain-binomial model introduced by Reed and Frost in 1928, see p. 8 in Andersson and Britton (2000). Since then the area of epidemic modelling has grown rapidly, and a good overview of other important work can be found in e.g. Bailey (1975), Anderson and May (1991) and Andersson and Britton (2000).

In the present thesis we are interested in the dynamics of endemic diseases. A disease is called endemic if it is able to persist in a population for a long time, without the need of introducing new infectious individuals from some external population, see p. 73 in Andersson and Britton (2000). Common diseases which can become endemic are for example childhood diseases such as measles. If we use measles as an example, we have that once an individual becomes infected this individual will stay so for a random period of time, the so-called infectious period, and when this individual eventually recovers from the disease she/he becomes immune. This tells us that disease transmission occurs according to Susceptible $\rightarrow$ Infectious $\rightarrow$ Recovered and immune, if we assume that all individuals are born susceptible. This type of disease transmission is often abbreviated as being of SIR-type, and epidemic models with this mechanism are hence called SIR models. Our main interest is to model endemic diseases, and especially diseases which have the SIR-type of disease transmission. Note that we under these conditions all individuals that become infected eventually will recover and become immune. Thus we need to have some demographic mechanism so that there is an influx of susceptible individuals in order for the disease to become endemic, since we need

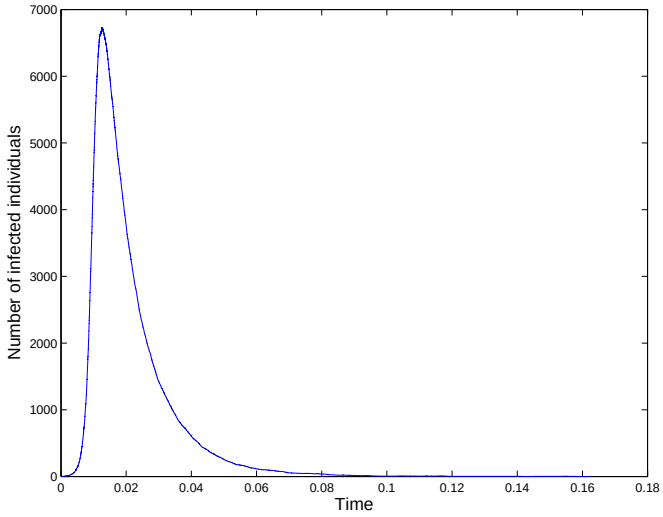

Figure 1: One simulation of the SIR model with demography when one infected individual was introduced into a susceptible population of size 10,000 and $R_0 = 10$. Here the disease does not become endemic.

to have a sufficiently large susceptible population at all time points. Apart from this, we also need that the disease is "infectious enough". This is quantified with the so-called basic reproduction number, denoted R_0, defined as the expected number of individuals that a single infectious individual infects in an otherwise susceptible population, see p. 6 in Andersson and Britton (2000). As it turns out, R_0 works as a threshold such that if $R_0 \leq 1$, the disease can not become endemic, whereas if $R_0 > 1$ the disease has a positive probability of becoming endemic, see p. 75 in Andersson and Britton (2000). If we are in a situation when $R_0 > 1$ and introduce a single infected individual, this individual will start an epidemic, that a) infects a large number of individuals, but the disease goes extinct rather quickly, or b) infects a large number of individuals in the first "wave of infection" and the following waves stabilise around a certain level of number of infectious individuals, known as the endemic level, and the disease is present in the population over a long period of time. For an illustration of these two scenarios, see Figs. 1 and 2. Up to this point we have primarily discussed the dynamics of the disease, and not mentioned much about the underlying population. When building mathematical models one of the simplifying assumptions, which is often used, is to assume that the population lacks social structure, i.e. everyone knows everyone equally much.

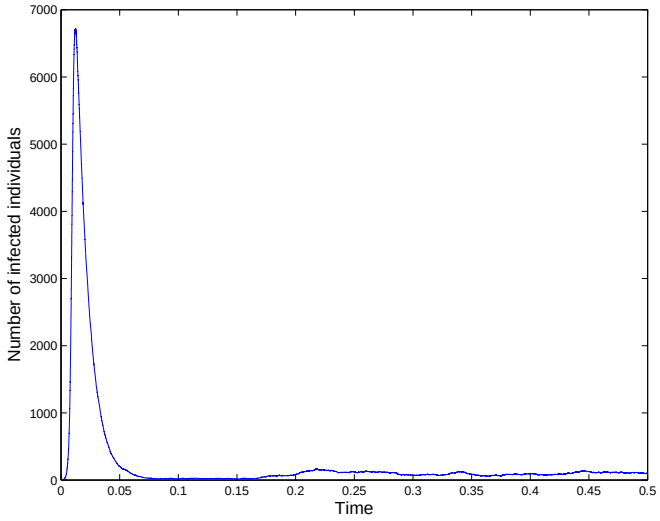

Figure 2: One simulation of the SIR model with demography when one infected individual was introduced into a susceptible population of size 10,000 and $R_0 = 10$. Here the disease becomes endemic.

This heuristically described simple model could be used to model an endemic disease. By imposing some additional assumptions we can formalise this model a bit more mathematically. A reasonable assumption is to assume that infectious contacts are rare, which makes it reasonable to model this with a Poisson process. If we also consider the inflow of susceptible individuals as a rather rare event, we can again model this with a Poisson process. A less realistic, but mathematically tractable, assumption is to assume that the infectious periods are exponentially distributed, so that the epidemic process becomes Markovian. In many aspects this heuristic epidemic model resembles the models that we have analysed in the two papers that is the main part of this thesis, and we will now give a brief summary of these two papers.

In the first paper, Lindholm and Britton (2006), we analyse a version of the SIR model with demography, see Nåsell (1999), where the population is divided into a small number of (equally) large sub-communities, following the definition of this model as it is presented in Hagenaars et al. (2004). That is, we are interested in how the time to extinction, which is used as a measure of persistence, is affected by varying the degree of interaction between sub-communities, and we present two approximations of the time to extinction. Both approximations are based on the idea that the qualitative

dynamics of the epidemic can be captured by regarding sub-communities as meta-individuals, and use theory from Markov jump processes to describe the epidemic from a sub-community view. Simulations indicate that the time to extinction is increasing in the degree of interaction between sub-communities, which also is seen from the approximations in relevant parameter regions.

In the second paper, Lindholm (2007), we analyse a version of Bartlett's epidemic model, a model with a simple demographic mechanism, see Bartlett (1956). In this model there is only a single community and all individuals are classified as one of two types, e.g. female or male, young or adult, etc., and the types give the individuals different features in terms of susceptibility and infectivity. This paper is also concerned with finding approximations of the time to extinction, and especially, to understand how it is affected by differences in susceptibility and infectivity between the two types. We present an approximation of the time to extinction which is based on diffusion approximation techniques. From this approximation we get that if we increase the difference in infectivity between the two types, the expected time to extinction decreases, whereas the situation is more complicated when varying the difference in susceptibility between the two types, and even non monotonic behaviour can be seen in certain parameter regions. These results are supported by simulations.

References

Anderson R.M., May R.M. (1991): Infectious Diseases of Humans - Dynamics and Control, Oxford University Press.

Andersson H., Britton T. (2000): Stochastic Epidemic Models and Their Statistical Analysis, Lecture Notes in Statistics 151, Springer-Verlag, New York.

Bailey N.T.J. (1975): The mathematical theory of infectious diseases and its applications, Griffin, London.

Bartlett M.S. (1956): Deterministic and stochastic models for recurrent epidemics, in: J. Neyman, (Ed.), Proceedings of the third Berkeley symposium on mathematical statistics and probability vol IV, University of California Press, pp. 81-109.

The Concise Oxford Dictionary, New Edition, Sixth Edition (1976), (Ed.) Sykes J.B., Oxford at the Clarendon Press.

Hagenaars T.J., Donnelly C.A., Ferguson N.M. (2004): Spatial heterogeneity and the persistence of infectious diseases, Journal of Theoretical Biology, 229, 349-359.

Lindholm M., Britton T. (2007): Endemic persistence or disease extinction: the effect of separation into sub-communities (to appear in Theoretical Population Biology).

Lindholm M. (2007): On the time to extinction for a two-type version of Bartlett's epidemic model, Stockholm University Research Reports in Mathematical Statistics 2007:9 (submitted).

Nåsell I. (1999): On the time to extinction in recurrent epidemics, Journal of the Royal Statistical Society Series B 61 Part 2, 309-330.

Paper I

Endemic persistence or disease extinction: the effect of separation into sub-communities

Mathias Lindholm[†,*]
Tom Britton[‡,*]

[†]Phone: +46 8 16 45 61, E-mail: lindholm@math.su.se
[‡]Phone: +46 8 16 45 34, E-mail: tomb@math.su.se
[*]Postal Address: Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden.

Abstract

Consider an infectious disease which is endemic in a population divided into several large sub-communities that interact. Our aim is to understand how the time to extinction is affected by the level of interaction between communities.

We present two approximations of the expected time to extinction in a population consisting of a small number of large sub-communities. These approximations are described for an SIR epidemic model, with focus on diseases with short infectious period in relation to life length, such as childhood diseases. Both approximations are based on Markov jump processes.

Simulations indicate that the time to extinction is increasing in the degree of interaction between communities. This behaviour can also be seen in our approximations in relevant regions of the parameter space.

KEY WORDS: Endemic diseases, SIR-epidemic, Population heterogeneity, Expected time to extinction, Quasi-stationary distribution

1 Introduction

When modelling infectious diseases a simplifying assumption often made is that the social interaction within the population is homogeneous. This assumption becomes less credible as the population size increases, i.e. there is a need for including some kind of population heterogeneity. A way to include population heterogeneity is to divide the population into sub-communities. By doing so we can, in the easiest setting, allow two different levels of social interaction, one level within and one between sub-communities. Under these conditions it is natural to let the social interaction within sub-communities be homogeneous. In the present paper we study how this new level of social interaction affects the epidemic behaviour as an infectious disease is introduced into the population. This we do for the situation when there are k sub-communities each of size n, where typical values of k is $2, \ldots, 5$ and n is 50,000 or larger. Throughout this report we will focus on infectious diseases which have a short infectious period in relation to life length and give rise to life long immunity, e.g. childhood diseases. When an outbreak of such a disease occurs in a community, we have three possible scenarios. The first being that only a few become infected and the time to extinction is short. The second one being that many become infected but the time to extinction is short. We are interested in diseases that behave as in the third scenario, namely when many individuals become infected and the time to extinction is long. When a disease behaves in this way it is called endemic. During the progression of an endemic disease there is only a rather small fraction of infectious individuals present in the population at each time point, but since the time until disease extinction is long, the accumulated number of infected individuals may still be large. Usually the fraction of infected individuals at each time point fluctuates around some specific level, the endemic level, until disease extinction.

Whether a disease becomes endemic or not depends on a number of factors, such as population size, the basic reproduction number, length of the latency and infectious period, seasonal effects etc., see pp. 128-143 in Anderson and May (1991). In the present paper we are primarily interested in the effect of a community being divided into sub-communities. For this reason we neglect most other hetereogeneities, and the most important factor that remains is a sufficiently large population. The basic reproduction number, R_0, is defined as the expected number of individuals that a single infectious individual infects in a large susceptible population during its infectious period. One can show that the basic reproduction number works as a threshold which determines the dynamics of the disease and that it is dimensionless, see Anderson and May (1991) pp. 13-19. If $R_0 \leq 1$, the disease will go extinct

rather quickly. On the other hand, if $R_0 > 1$, the disease has a positive probability to persist in the population over a long time period. Henceforth R_0 is assumed to be greater than one. The notion of 'sufficiently large population' which we have used above is not a trivial question, and this is something we discuss in Section 5 and Section 6.

When the population is divided into sub-communities rather than being homogeneously mixing, the dynamics of the spread of disease becomes more intricate. Now, some sub-communities may be disease-free, while others contain infected individuals, and infectious contacts between individuals from different sub-communities may re-infect disease-free sub-communities. It seems reasonable to expect that the mean time to extinction of an endemic disease depends on the social activity between the different communities when keeping everything else fixed. This has been shown to be true by Hagenaars et al. (2004).

Endemic diseases can be modelled stochastically in different ways. Depending on the model, different aspects of the qualitative behaviour of the dynamics of the disease can be studied. Our aim is to study the expected time to extinction of an endemic disease in the situation with a small number of large sub-communities, when each sub-community starts at the so-called endemic level. The model used is an SIR model for a population divided into sub-communities. This model will henceforth be denoted SIR-SC. For a homogeneously mixing population, from here on denoted SIR-HM, much work has been done, see for example van Herwaarden and Grasman (1995), Nåsell (1999, 2005), and Andersson and Britton (2000b). For more on epidemic models in general, see Anderson and May (1991), Andersson and Britton (2000a) and Diekmann and Heesterbeek (2000).

A short heuristic description of the SIR-HM model is that all individuals in the community are equally likely to meet, and that each individual may switch between being Susceptible, Infectious and Recovered (and immune). Thus, switches occur according to $\mathrm{S} \to \mathrm{I} \to \mathrm{R}$. Another important property is that susceptible individuals are born into the community and that individuals eventually die, i.e. demographic aspects. This will also give us a non constant community size. The version of this model which we use is from Nåsell (1999), and does not allow for birth of infectious individuals. Important results for the SIR-HM model that we will use are from Nåsell (1999, 2005). In those papers he derives approximations for the expected time to extinction when starting at the quasi-stationary level of infection. Hagenaars et al. (2004) study the same expected time to extinction as Nåsell but for the case with a small number of sub-communities. They obtain an approximation of this expected time, but the approximation is derived under the assumption of low mixing between communities and that the infectious period is fairly long in

relation to life length. Human childhood diseases, having infectious periods of 1-2 weeks out of life lengths of 70-80 years, fall outside of this domain. For more information on infectious periods of infectious diseases see Table 3.1 on pp. 31 in Anderson and May (1991).

In the present paper we study the SIR model for a population divided into sub-communities previously studied by Lloyd and May (1996) (who treat a more general model) and Hagenaars et al. (2004). Similar models can be found in the metapopulation literature (sometimes also referred to as patch models), see e.g. Keeling (2000a, 2000b), Etienne and Heesterbeek (2000), van den Driessche and Watmough (2002), where they address related topics to those which we treat here. In a paper by Wonham et al. (2004) they also treat the situation when there are seasonal effects. Other versions of SIR models with heterogeneities are household models and models with several levels of mixing, where the population is divided into many small groups, see e.g. Ball et al. (1997), Ball and Lyne (2001) and Ball and Neal (2002).

In the present paper we have adopted ideas from both Nåsell (1999) and Hagenaars et al. (2004), trying to find better approximations for $\tau = E(T_Q)$, the expected time to extinction for a population divided into sub-communities given that all sub-communities are started at the endemic level, for diseases with short infectious period with respect to life length. We present two approximations of τ, the first one is based on similar arguments as in Hagenaars et al. (2004) and the second one is based on more heuristic arguments motivating the use of an exponential form. Further, we show that T_Q is exponentially distributed, and by approximating τ, we get that $T_Q \sim \text{Exp}(1/\tau)$. Simulations indicate that our approximations are more suitable for situations when there is low mixing between sub-communities and that the second, more heuristically motivated approximation, performs somewhat better.

In Section 2 we define the sub-community epidemic model and present results needed later on. In Sections 3 and 4 we describe our two approximations. Section 5 is devoted to a small simulation study and numerical evaluations of these approximations. A closing discussion and summary of our results is given in Section 6.

2 The SIR model for a population divided into sub-communities (SIR-SC)

We start with a brief look at the SIR model with homogeneous mixing (SIR-HM) defined in Nåsell (1999). We have a population which lacks social

structures and where individuals may switch between the states Susceptible, Infectious and Recovered (and immune) according to S $\rightarrow$ I $\rightarrow$ R. Once an individual becomes infected, this individual will stay so for an exponentially distributed time with mean $1/\nu$. During this time period the infected individual makes infectious contacts with a given susceptible individual according to a Poisson process with rate β/n. All infectious contacts are assumed to result in instantaneous infection. The demographic aspects of the model are as follows: Susceptible individuals are 'born' according to a Poisson process with constant rate μn, and all individuals live for an exponentially distributed time with mean $1/\mu$. A consequence of the constant birth rate is that the population size will fluctuate around n, which is thought of as being large. This is deliberate to avoid that the dynamics of the disease depend on extensive population fluctuations. Once the population becomes disease-free, it will remain so forever on, since there is no birth or immigration of infectious individuals. Hence, the disease-free states are absorbing, and all other states are transient.

When the population is divided into sub-communities with higher mixing within, the dynamics of the disease becomes more involved. The simplest case is to let all sub-communities be equally large, having size n, and to let all individuals have the same within sub-community contact rates and the same between sub-community contact rates. We are interested in the situation when the number of sub-communities, k, is fixed and small in relation to n. With this model the population structure is symmetric and we only need to add one parameter, ε, which is the proportion of an individuals contacts that are with other sub-communities. This parameter, ε, is defined such that $\varepsilon = 0$ corresponds to having k isolated sub-communities, and $\varepsilon = 1$ corresponds to the case where all k sub-communities act as a single large community of size kn. One can also think of ε as an inverse distance, where $\varepsilon = 0$ corresponds to that all sub-communities lie infinitely far apart and $\varepsilon = 1$ corresponds to the case when they coincide, and then ε works as a measure of spatial heterogeneity or spatial coupling, see e.g. Keeling (2000a, 2000b).

A natural way to model the situation with sub-communities is to do so such that the overall infectious pressure in the entire population is kept constant regardless of the value of ε. This also has the advantage that we get the same basic reproduction number, R_0, as for the SIR-HM model and hence the two models become easier to compare. The basic reproduction number for the SIR-HM model is defined as the average number of individuals which a single infectious individual infects in an otherwise susceptible population during its infectious period. That is, a single infectious individual makes infectious contacts at rate β/n with any given individual of the surrounding n

susceptible individuals, during an infectious period with mean $1/(\mu+\nu)$, since death can occur before recovery. Thus R_0 for the SIR-HM model becomes

$$R_0 = \frac{\beta}{\mu+\nu} = \frac{\beta}{\mu\alpha}, \tag{1}$$

where $\alpha = (\mu+\nu)/\mu$. For the case with sub-communities, an infected individual makes contacts with any given individual within its own sub-community at rate β'/n, and at rate $\varepsilon\beta'/n$ with a given individual in any of the $k-1$ surrounding sub-communities. This gives us that the probability that a contact is within the sub-community is

$$\frac{n\beta'/n}{n\beta'/n + (k-1)n\varepsilon\beta'/n} = \frac{1}{1+\varepsilon(k-1)}.$$

If we have a single infected individual in an otherwise susceptible population, this individual will infect a given individual within its own sub-community at rate β'/n during an exponentially distributed infectious period with mean $1/(\mu+\nu)$, and infect a given individual in any of the $k-1$ neighbouring sub-communities at rate $\varepsilon\beta'/n$, hence the basic reproduction number becomes

$$R_0 = \frac{1}{\mu+\nu}(n\beta'/n + (k-1)n\varepsilon\beta'/n) = \frac{\beta'}{\mu+\nu}(1+\varepsilon(k-1)).$$

Thus, if we let $\beta' = \beta/(1+\varepsilon(k-1))$ we see that we have found the proper scale in order to keep R_0 independent of ε. For this β' we have

$$R_0 = \frac{\beta}{\mu\alpha}.$$

The possible transitions and their rates for the SIR-SC model are specified in Table 1, which are the same as in Hagenaars et al. (2004).

We now derive the endemic level. In the stochastic model this corresponds to the mean in the quasi-stationary distribution. Let $(\mathbf{X}(t), \mathbf{Y}(t)), t \geq 0$, denote a $2k$ dimensional Markov jump process, where $X_j(t) = s_j$ and $Y_j(t) = i_j$ denote the number of susceptible and infectious individuals in sub-community j at time t, with random transition rates defined in Table 1. If we now look at the process of proportions $(\mathbf{X}(t)/n, \mathbf{Y}(t)/n)$, when n is large, this process can be approximated by the solution of a deterministic system of differential equations corresponding to the transition rates defined in Table 1. This system is given by

$$\begin{aligned} \tfrac{dx_i}{dt} &= \mu - \tfrac{\beta}{(1+\varepsilon(k-1))} x_i \left(y_i + \varepsilon \textstyle\sum_{j\neq i} y_j \right) - \mu x_i \\ \tfrac{dy_i}{dt} &= \tfrac{\beta}{(1+\varepsilon(k-1))} x_i \left(y_i + \varepsilon \textstyle\sum_{j\neq i} y_j \right) - (\mu+\nu) y_i \end{aligned} \tag{2}$$

Setting these equations equal to zero for $i = 1, \ldots, k$ gives us the stationary points, which turn out to be (1, 0), the disease-free state, and

$$(\hat{x}_i, \hat{y}_i) = (\hat{x}, \hat{y}) = \left(\frac{1}{R_0}, \frac{1}{\alpha} \left(1 - \frac{1}{R_0} \right) \right), \tag{3}$$

which corresponds to the endemic level, and which only exists if $R_0 > 1$. Equating the differential equations from (2) to 0 when $k = 1$, gives us the same endemic level as for the SIR-HM model.

A quasi-stationary distribution is defined as the distribution after a long time conditioned on that the process has not been absorbed. The endemic level can be thought of as the mean of this distribution, which the process fluctuates around. The quasi-stationary distribution is important when modelling endemic diseases, since we are interested in the behaviour of the epidemic until it goes extinct. But, quasi-stationary distributions give rise to many difficulties such as questions of uniqueness and existence, see Pollett and Roberts (1990), and the occurrence of quasi-cycles Bartlett (1957) and Dushoff (2004). A longer treatment of quasi-stationarity concerning birth and death process is given by van Doorn (1991).

Let $Q = \{q_{\mathbf{x},\mathbf{y}}\}$ denote the quasi-stationary distribution, where $q_{\mathbf{x},\mathbf{y}}$ is the probability that the process $(\mathbf{X}(t), \mathbf{Y}(t)) = (\mathbf{x}, \mathbf{y})$ as $t \to \infty$, conditioned on that the process has not been absorbed. Recalling that the lack of memory property implies an exponential distribution, we have

$$\begin{aligned} &P(T_Q > t + s \mid T_Q > t,\ (\mathbf{X}(0), \mathbf{Y}(0)) \sim Q) \\ &\quad = P(T_Q > t + s \mid T_Q > t,\ (\mathbf{X}(t), \mathbf{Y}(t)) \sim Q) \\ &\quad = P(T_Q > s \mid (\mathbf{X}(0), \mathbf{Y}(0)) \sim Q), \end{aligned}$$

which establishes that T_Q is exponentially distributed. The rate parameter for this exponential distribution is the intensity with which the process leaves the set of transient states. For the case with sub-communities the set of states from which the process can be absorbed is $(\mathbf{X}(t), \mathbf{Y}(t)) = \{(\mathbf{x}, \mathbf{y}); \mathbf{y} = \mathbf{e}_i, i = 1, \ldots, k\}$. We state the conclusions from above in the following Proposition:

Proposition 1 *The time to extinction given that the process is started in the quasi-stationary distribution, T_Q, is exponentially distributed with mean*

$$\tau = \frac{1}{\mu \alpha q_{\bullet,1}}, \tag{4}$$

where

$$q_{\bullet,1} = \sum_{\mathbf{x}} \sum_{i=1}^{k} q_{\mathbf{x},\mathbf{e}_i}, \tag{5}$$

and where $\mathbf{e}_i$ *is the i'th unit vector.*

The reasoning here is the same as in the proof of Proposition 4.1 in Andersson and Britton (2000b), but this result was first derived for the homogeneous case by Nåsell in Nåsell (1999), and if we set $k = 1$ in Proposition 1 we obtain the result for a homogeneously mixing population. Another way of obtaining (4) and (5) from Proposition 1 is via the Kolmogorov forward equations for the process $(\mathbf{X}(t), \mathbf{Y}(t))$ conditioned on that it has not gone extinct by time t, and then use the identity $P(T_Q \leq s) = P(\mathbf{Y}(s) = 0)$.

Table 1: SIR model for a population divided into sub-communities

Rates for sub-community j		
From	To	Rate
(s_j, i_j)	$(s_j + 1, i_j)$	μn
(s_j, i_j)	$(s_j - 1, i_j)$	μs_j
(s_j, i_j)	$(s_j - 1, i_j + 1)$	$\frac{\beta}{n}\frac{1}{(1+\varepsilon(k-1))} s_j \left(i_j + \varepsilon \sum_{u \neq j} i_u\right)$
(s_j, i_j)	$(s_j, i_j - 1)$	$(\mu + \nu) i_j$

To completely determine the distribution of T_Q, it remains to derive $q_{\bullet,1}$. We will now give a short description of methods used to derive approximations for $q_{\bullet,1}$ for the SIR-HM model, but as it turns out these methods do not work for the SIR-SC model. These results are also needed later on when we describe approximations for the SIR-SC model. One way to obtain an approximation for $q_{\bullet,1}$ is to use use a diffusion approximation. Let $(\widetilde{X}_n(t), \widetilde{Y}_n(t))$ be the process defined by

$$(\widetilde{X}_n(t), \widetilde{Y}_n(t)) = \sqrt{n}\left(\frac{X(t)}{n} - \hat{x}, \frac{Y(t)}{n} - \hat{y}\right). \tag{6}$$

One can show that this process converges to an Ornstein-Uhlenbeck process, $(\widetilde{X}(t), \widetilde{Y}(t))$, as n tends to infinity, see e.g. Chapter 11 in Ethier and Kurtz (1986). From the theory of diffusion processes it is known that this process has a Gaussian *stationary* distribution with mean zero and computable covariance matrix, see e.g. pp. 357 in Karatzas and Shreve (1991). This together with (6) gives us that $Y(t) \approx \sqrt{n}\widetilde{Y}(t) + n\hat{y} = \sqrt{n}\widetilde{Y}(t) + \mu_Y$ for large n, so that $Y(t)$ is approximately $N(\mu_Y, \sigma_Y^2)$ when t is large. But, now the approximate marginal distribution of $Y(t)$ for large t is defined on $\mathbb{R}$, whereas the original process $Y(t)$ is integer valued and always greater or equal to zero (we can not have a negative number of individuals). Thus, if we truncate the approximate marginal distribution for the number of infectious individuals at

zero and use continuity correction we get an approximation of the underlying quasi-stationary distribution according to

$$q_{\bullet,1} = \frac{1}{\sigma_Y} \frac{\varphi((\mu_Y - 1)/\sigma_Y)}{\Phi((\mu_Y - 0.5)/\sigma_Y)}, \tag{7}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal distribution function and density function respectively, and

$$\begin{aligned} \sigma_Y &= Sd(\sqrt{n}\widetilde{Y} + n\hat{y}) = \tfrac{\sqrt{n}}{R_0}\sqrt{R_0 - 1 + R_0^2/\alpha} \\ \mu_Y &= E(\sqrt{n}\widetilde{Y} + n\hat{y}) = n\tfrac{R_0 - 1}{\alpha R_0} \end{aligned} \tag{8}$$

Here σ_Y and μ_Y are the expectation and standard deviation of the marginal process of number of infected individuals, see Eq. (2.10) in Nåsell (1999), which are obtained with methods from Karatzas and Shreve (1991). Using (7) together with Proposition 1 gives us that T_Q is exponentially distributed with approximate mean

$$\tau_n \approx \frac{\sigma_Y}{\mu\alpha} \frac{\Phi((\mu_Y - 0.5)/\sigma_Y)}{\varphi((\mu_Y - 1)/\sigma_Y)} \tag{9}$$

with σ_Y and μ_Y from Eq. (8), (c.f. Eq. (2.13) in Nåsell (1999)). From here on τ_n refers to the case with a homogeneously mixing population of size n, and all other types of references to τ are for the case with sub-communities unless otherwise stated. Note that usually $\mu_Y \gg 1$ which gives us that $q_{\bullet,1}$ from (7) simplifies to

$$q_{\bullet,1} \approx \frac{1}{\sigma_Y} \frac{\varphi(\mu_Y/\sigma_Y)}{\Phi(\mu_Y/\sigma_Y)} \approx \frac{1}{\sigma_Y}\varphi(\mu_Y/\sigma_Y) \tag{10}$$

where the last approximation is good when $\mu_Y/\sigma_Y > 3$, since then $\Phi(\mu_Y/\sigma_Y) \approx 1$ holds. When $\alpha \gg R_0$, which is the case for childhood diseases, then $\sigma_Y \approx \sqrt{n(R_0 - 1)}/R_0$, and $q_{\bullet,1}$ from (10) becomes

$$q_{\bullet,1} \approx \frac{1}{\sigma_Y}\varphi\left(\sqrt{n(R_0 - 1)}/\alpha\right) = \frac{R_0}{\sqrt{2\pi n(R_0 - 1)}} \exp\left(-n\frac{R_0 - 1}{2\alpha^2}\right) \tag{11}$$

and τ_n from (9) simplifies to

$$\tau_n \approx \frac{\sqrt{2\pi n(R_0 - 1)}}{\mu\alpha R_0} \exp\left(n\frac{R_0 - 1}{2\alpha^2}\right). \tag{12}$$

When the average life length is long in relation to the average infectious period, Nåsell (2005) shows that (9) is a too crude approximation when n

is only moderately large, e.g. $n \leq 2,000,000$ for a specific set of parameter values corresponding to measles, see Section 5 below. In Nåsell (2005) he instead proposes that the quasi-stationary distribution of the number of infected individuals could be approximated with a geometric distribution with $p = 1/\mu_Y$ where μ_Y is from (8). If $Y \sim Geo(p)$ then $E(Y) = 1/p = \mu_Y$ which together with Proposition 1 with $k = 1$ yields the following: When the quasi-stationary distribution of Y is approximated with a $Geo(1/\mu_Y)$ distribution with mean μ_Y from (8), then T_Q is exponentially distributed with approximate mean

$$\tau_n \approx n \frac{R_0 - 1}{\mu \alpha^2 R_0}, \tag{13}$$

c.f. Eqs. (8.3) and (9.2) in Nåsell (2005).

Returning to the case with sub-communities again, we would like to use the techniques described above, but the resulting Ornstein-Uhlenbeck diffusion process approximated at the endemic level is independent of ε, and hence not of much help. The second approach will also give us an approximation of the quasi-stationary distribution which is independent of ε, since it was a geometric distribution with parameter $p = 1/\mu_y$, where μ_y is as (8), but with n replaced with kn. Despite of this, we can still say something about the expected time to extinction for the two extreme cases, $\varepsilon = 0$ and $\varepsilon = 1$, using results from the SIR-HM model. Let $\tau(\varepsilon)$ be the expected time to extinction when all k sub-communities, each of size n, are started at the endemic level, when there is a proportion ε of contacts between sub-communities. (Note that $\tau(\varepsilon) = \tau(\varepsilon, n, k, \mu, \alpha, R_0)$.) When $\varepsilon = 0$ all k sub-communities are isolated and independent, and all k sub-communities start at the endemic level of infection, the expected time until one of the k infected sub-communities recovers is τ_n/k, due to independence and that the expected duration of an epidemic within a sub-community is exponentially distributed with mean τ_n, where τ_n is from one of Eqs. (9) or (13). Due to the Markov property and that a disease-free community never can be re-infected when $\varepsilon = 0$, the expected time until one of the $k - 1$ remaining communities recovers is $\tau_n/(k-1)$. Repeating this argument gives us

$$\tau(0) = \tau_n \sum_{i=1}^{k} \frac{1}{i}, \tag{14}$$

where τ_n can be approximated using either of Eqs. (9) or (13). On the other hand, when $\varepsilon = 1$, all k communities behave as one large community of size kn, and we can again make use of (9) with n replaced by kn, i.e.

$$\tau(1) = \tau_{kn}. \tag{15}$$

If n is small we suggest to approximate $\tau(0)$ and $\tau(1)$ by using the geometric approximation of τ_n from Eq. (13), which gives us

$$\frac{\tau(0)}{\tau(1)} = \frac{\sum_{i=1}^{k} \frac{1}{i}}{k} \tag{16}$$

which is smaller than one for $k > 1$, i.e.

$$\tau(0) < \tau(1). \tag{17}$$

If n is large we recommend to use the truncated normal approximation of τ_n from (9) instead. For n such that $\mu_Y/\sigma_Y > 3$ and when we are in the parameter region corresponding to childhood diseases we can approximate τ_n with (12). Inserting this into $\tau(0)$ and $\tau(1)$ from Eqs. (14) and (15) gives us that

$$\frac{\tau(0)}{\tau(1)} \approx \frac{\sum_{i=1}^{k} \frac{1}{i}}{\sqrt{k}} \exp\left(-n\frac{R_0 - 1}{2\alpha^2}(k-1)\right) \tag{18}$$

which is smaller than 1 for sufficiently large n, i.e. $\tau(0) < \tau(1)$.

3 Approximation using a recovered (and immune) state

As we have seen, it is hard to find approximations of the quasi-stationary distribution which depend on ε. But, if we rely on Proposition 1, that T_Q is exponentially distributed, together with results for the SIR-HM model, we can approximate $\tau = E(T_Q)$ directly, instead of going via approximations of the quasi-stationary distribution.

In Hagenaars et al. (2004) they look at the case when $0 < \varepsilon \ll 1$ and α is thought of as small, such as $\alpha = 2$ or 160. An example of a disease with small α is scrapie among sheep, see Hagenaars et al. (2004). For scrapie the average incubation period is a few years which is of the same order of magnitude as the average life length of sheep. Hence, for diseases with small α one can assume that when an individual recovers from infection, this individual will likely be removed due to death within a relatively short time period. This motivates that we can look at the system from a sub-community view, classifying each sub-community as either endemic or susceptible. That a sub-community is endemic here means that the sub-community on average has a fraction of infected individuals corresponding to the endemic level $\hat{y}$. A sub-community that is susceptible only contains susceptible individuals. Further,

switches between these two states are assumed to occur instantaneously. This is a reasonable approximation, since the time it takes from that a single individual becomes infected until the endemic level of infection is reached is short in relation to the time it takes for an endemic sub-community to become disease-free.

When defining the rate with which susceptible sub-communities becomes endemic, it is natural to think that this rate depends on the infectious pressure generated by the endemic sub-communities. But, we are only interested in those infectious contacts between sub-communities that result in a disease invasion and not those that fade out by chance, so we must take this fact into account. Thus, we need to derive the probability of this event. Suppose a sub-community with a fraction x susceptible and $1-x$ recovered (and immune) individuals has just been re-infected, i.e. a single susceptible individual has become infected. In the early stages of an epidemic it behaves approximately as a branching process. When there is only one infected individual in a population with a fraction x susceptible individuals, the effective reproduction number becomes xR_0. Since the infectious period is exponentially distributed, the number of children of this one infected individual, D, will be $Geo(1/(1+xR_0))$, and we get that the probability that the epidemic started by this single infected individual will not fade out by chance, p, is the solution to the following equation:

$$1-p = E((1-p)^D), \tag{19}$$

see Andersson and Britton (2000a) pp. 22-25. Solving this gives us the solution

$$p = 1 - \frac{1}{xR_0}. \tag{20}$$

From this we get that the probability that the introduced disease will not fade out by chance in a fully susceptible population is $1-1/R_0$. The more general result from (20) is needed later on.

One individual contacts a given individual from one of the surrounding sub-communities at rate $\varepsilon\beta'/n$, and hence contacts a fully susceptible sub-community at rate $\varepsilon\beta'$. Consequently, a sub-community at the endemic level, having $\hat{y}n$ infected individuals, infects a given susceptible sub-community at rate $\varepsilon\beta'\hat{y}n = \varepsilon\beta\hat{y}n/(1+\varepsilon(k-1))$, which is the same as the infectious pressure each endemic sub-community expose each susceptible sub-community to. This together with that each infectious contact has the probability $1-1/R_0$ that the introduced disease will become endemic, gives us the rate with which susceptible sub-communities become endemic.

If we again look at the rate with which sub-communities becomes disease-free, this is thought of occurring independently of everything else, i.e. the time to disease extinction in a sub-community is exponentially distributed with mean parameter τ. From this we can define a birth and death process of number of endemic sub-communities, with transition rates

$$\begin{cases} \zeta_j &= (k-j)j\varepsilon n \frac{\mu R_0}{1+\varepsilon(k-1)} \left(1-\frac{1}{R_0}\right)^2 \\ \eta_j &= j/\tau_n \end{cases} \tag{21}$$

where ζ_j is the rate for a transition from j to $j+1$ endemic sub-communities, and η_j is the rate for a transition from j to $j-1$ endemic sub-communities.

Since ε is small, the probability of re-infection will also be small. Based on this fact Hagenaars et al. (2004) assume that the probability of more than one re-infection during the epidemic is negligible. Their approximation can be described as the expected time to absorption of a birth and death process for the number of endemic sub-communities, with rates as in (21), which only allow one birth, or more formally:

Approximation (Hagenaars et al. (2004)) *The expected time to extinction given that the process is started at the endemic level can be approximated by*

$$\tau_{\mathrm{SI}}(\varepsilon) = \tau(0) + \varepsilon \left((k+1) \sum_{j=1}^{k} \frac{1}{j} - 2k \right) \tau_n^2 \mu R_0 \left(1-\frac{1}{R_0}\right)^2 + O(\varepsilon^2) \tag{22}$$

where $\tau(0)$ is from (14) and τ_n is any approximation for a homogeneously mixing population of size n, e.g. (9) or (13).

This corresponds to Eq. (6) in Hagenaars et al. (2004). Here SI in τ_{SI} is used to emphasise that they only use the two sub-community states, Susceptible and Infected, in their approximation.

When α is large the approach to approximate the expected time to extinction described above is not completely feasible. Because in this situation, the approximation that an endemic sub-community that becomes disease-free instantaneously becomes susceptible is not reasonable. One way to avoid this problem is to add a recovered (and immune) state to our approximating sub-community Markov process. A sub-community is defined as being recovered (and immune) when it is disease-free but not possible to infect. The difference between this state and the susceptible state is that, when a sub-community is recovered (and immune) there is on average a fraction $\hat{x}$ susceptible and $1-\hat{x}$ immune individuals, as opposed to the susceptible state which only contain susceptible individuals.

By introducing this type of transitions for the sub-communities we have a communication between the states of sub-communities that can be described as S $\to$ I $\to$ R $\to$ S, so what we need to define is the rate, ξ, with which a community makes a transition from R to S, since (21) can be used for the other transitions. A transition from R $\to$ S is similar to a I $\to$ S transition, so that one way of defining this rate is to assume that a sub-community stays immune for an exponentially distributed time with mean τ_{R}. We will return to the definition of τ_{R} later.

Let s be the number of susceptible sub-communities and i be the number of endemic sub-communities out of a total of k sub-communities, so that $k-(s+i)$ are recovered (and immune), then the transition rates become

$$\begin{cases} \zeta_{s,i} &= si\varepsilon \frac{\mu n R_0}{1+\varepsilon(k-1)} \left(1-\frac{1}{R_0}\right)^2 \\ \eta_{s,i} &= \frac{s}{\tau_n} \\ \xi_{s,i} &= \frac{k-(s+i)}{\tau_{\mathrm{R}}}. \end{cases} \tag{23}$$

There are $k(k+1)/2+k$ possible states, and k of them are disease-free and hence an absorbing class of states. For a schematic graph of the dynamics of this process, see Fig. 1.

Based on the rates (23) we are able to set up a difference equation for $\tilde{t}_{s,i}$, the expected time to extinction when starting with i endemic and s susceptible sub-communities out of k possible, by conditioning on the first transition and using the process' lack of memory. We get the following relation

$$\begin{aligned} \tilde{t}_{s,i} = E(\{&\text{time to extinction from } (s,i)\}) = E(\{\text{time spent in } (s,i)\}) \\ &+ P((s,i) \to (s-1,i+1))E(\{\text{time to extinction from } (s-1,i+1)\}) \\ &+ P((s,i) \to (s,i-1))E(\{\text{time to extinction from } (s,i-1)\}) \\ &+ P((s,i) \to (s+1,i))E(\{\text{time to extinction from } (s+1,i)\}) \end{aligned}$$

which gives us that

$$\begin{aligned} \tilde{t}_{s,i} = &\frac{1}{\zeta_{s,i}+\eta_{s,i}+\xi_{s,i}} + \frac{\zeta_{s,i}}{\zeta_{s,i}+\eta_{s,i}+\xi_{s,i}}\tilde{t}_{s-1,i+1} \\ &+ \frac{\eta_{s,i}}{\zeta_{s,i}+\eta_{s,i}+\xi_{s,i}}\tilde{t}_{s,i-1} + \frac{\xi_{s,i}}{\zeta_{s,i}+\eta_{s,i}+\xi_{s,i}}\tilde{t}_{s+1,i}, \end{aligned} \tag{24}$$

see pp. 148-150 in Karlin and Taylor (1975). In general this system has no closed form solution. Even so, by looking at the transition rates (23) and the relation (24), we see that we can write this as an equation system of the form

$$\tilde{\mathbf{t}} = \mathbf{v} + \mathbf{A}\tilde{\mathbf{t}}, \tag{25}$$

where $\tilde{\mathbf{t}} = (\tilde{t}_{0,k}, \tilde{t}_{0,k-1}, \ldots)'$, $\mathbf{v} = (v_{0,k}, v_{0,k-1}, \ldots), v_{s,i} = 1/(\zeta_{s,i} + \eta_{s,i} + \xi_{s,i})$, and $\mathbf{A}$ is the matrix with transition probabilities defined by the rates (23). A general solution to (25) is given by

$$\tilde{\mathbf{t}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{v}, \tag{26}$$

which can be solved numerically, where $\mathbf{I}$ is the identity matrix of the same dimension as $\mathbf{A}$. As before, we are mainly interested in the expected time to extinction when all k sub-communities are initially endemic, i.e. $\tilde{t}_{0,k}$. This defines our first approximation of the expected time to extinction.

Approximation 1 *The expected time to extinction given that the process is started at the endemic level can be approximated by* $\tau_{SIR}(\varepsilon) = \tilde{t}_{0,k}$*, where* $\tilde{\mathbf{t}}$ *solves (25).*

If we look at τ_{SIR} for the case $k = 2$, using the rates from (23), we get the following explicit expression

$$\tau_{\text{SIR}}(\varepsilon) = \tau(0) + \frac{\varepsilon\mu n R_0(1 - 1/R_0)^2\tau_{\text{R}}\tau_n^3}{2(\varepsilon\mu n R_0(1 - 1/R_0)^2\tau_{\text{R}}\tau_n + \tau_{\text{R}} + \tau_n)}, \tag{27}$$

where $\tau(0)$ is from (14). From (27) one sees that τ_{SIR} is increasing in ε, and that if $\varepsilon = 0$ then $\tau_{\text{SIR}} = \tau_{\text{SI}}$. For larger values of k the calculations becomes more tedious, since the number of unknown equations increases rapidly.

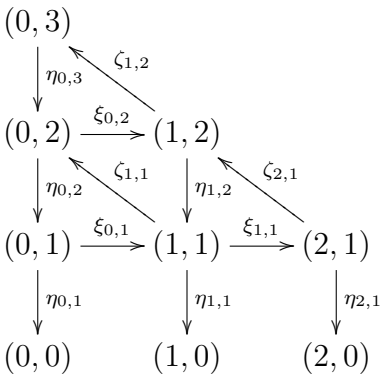

Figure 1: Schematic graph of the dynamics in our approximating SIR Markov jump process for $k = 3$ sub-communities, where each node is (no. of susceptible sub-communities, no. of endemic sub-communities) and the rates are from (23).

We now return to the derivation of τ_{R}, the expected time which a sub-community stays recovered (and immune). When a sub-community becomes

recovered (and immune), there is approximately a fraction $\hat{x} = 1/R_0$ susceptible individuals and a fraction $1 - \hat{x}$ immune individuals. The problem now is that the probability that an introduced disease will be able to persist depends on the fraction of susceptible individuals in the sub-community.

If we look at the probability that the introduced disease will become endemic, (20), we see that this probability is zero when we have a proportion of susceptible corresponding to the endemic level, and we know that this probability is $1 - 1/R_0$ when a sub-community is fully susceptible. Thus, we can define the expected time which a sub-community stays recovered (and immune) in terms of the average time it takes until a fraction $\tilde{x} > \hat{x}$ becomes susceptible in a sub-community such that the introduced disease will persist in the population with a pre-specified probability. A natural, but somewhat arbitrary, choice of this probability is $(1 - 1/R_0)/2$, i.e. half way between 0 and $1 - 1/R_0$. This gives us that the fraction of susceptible $\tilde{x}$ is the solution to

$$1 - \frac{1}{\tilde{x}R_0} = \frac{1}{2}\left(1 - \frac{1}{R_0}\right),$$

which is $\tilde{x} = 2/(R_0 + 1) \approx 2\hat{x}$ when R_0 is fairly large.

While a sub-community is treated as recovered (and immune), no infectious contacts may occur, and the expected fraction of susceptible $x(t)$ at a certain time point t after becoming disease-free is given by the solution to the differential equation

$$\begin{cases} \frac{dx}{dt} & = \mu(1 - x) \\ x(0) & = \hat{x} \end{cases} \tag{28}$$

Solving this equation gives us the relation

$$x(t) = 1 - (1 - \hat{x})\exp(-\mu t). \tag{29}$$

If we set $x(t) = \tilde{x} = 2/(R_0 + 1)$ and solve (29) in terms of t, we get

$$t = \tau_{\mathrm{R}} = \frac{1}{\mu}\log\left(\frac{R_0 + 1}{R_0}\right). \tag{30}$$

Note that the longer we treat a sub-community as recovered (and immune), the harder it gets for the infection to persist in the rest of the population. If τ_{R} is close to zero, we loose the effect of the recovered (and immune) state and the approximation resembles that of Hagenaars et al. (2004), and if τ_{R} tend to infinity it is the same as removing a sub-community which becomes disease-free. Our suggestion of an approximation of τ_{R}, (30), will give relatively small values. But, as said before, it is hard to find a natural definition of this quantity.

4 Approximation using an exponential form

When we introduced the SIR model for a population divided into sub-communities, we derived the expected time to extinction both for the case when all communities are isolated and the case when they are mixing as one large homogeneous community, corresponding to $\varepsilon = 0$ and $\varepsilon = 1$ respectively. We have also mentioned that these two approximations can not be improved along the present lines without improving the approximations for the SIR-HM model, Eqs. (9) and (13).

For $0 < \varepsilon < 1$ we now introduce a new approximation, $\tau_{\text{Exp}}(\varepsilon)$, by simply fitting an exponential curve having $\tau(0)$ as starting point and approximately $\tau(1)$ as end point such that $\tau'_{\text{Exp}}(0) = \tau'_{\text{SIR}}(0)$, i.e. we make use of the behaviour of τ_{SIR} where we expect it to work satisfactory. These imposed restrictions on the exponential curve determines it completely, and we propose the following approximation:

Approximation 2 *The expected time to extinction given that the process is started in quasi-stationarity can be approximated by*

$$\tau_{Exp}(\varepsilon) = \tau(1) - (\tau(1) - \tau(0)) \exp\left(-\frac{\tau'_{SIR}(0)}{\tau(1) - \tau(0)}\varepsilon\right), \tag{31}$$

where $\tau'_{SIR}(\cdot)$ is the first derivative of (26) with respect to ε, and $\tau(0)$ and $\tau(1)$ are from Eqs. (14) and (15) respectively.

One can easily verify that $\tau_{\text{Exp}}(0) = \tau_{\text{SIR}}(0)$, $\tau'_{\text{Exp}}(0) = \tau'_{\text{SIR}}(0)$, and we see that when $\tau'_{\text{SIR}}(0) \gg \tau(1) - \tau(0)$ then $\tau_{\text{Exp}}(1) \approx \tau(1)$, as desired. To see that this is reasonable, look at the exponent of (31), $-\tau'_{\text{SIR}}(0)\varepsilon/(\tau(1)-\tau(0))$, when $k = 2$ and use τ_{SIR} from (27). We then get that $\tau'_{\text{SIR}}(0) = \mu n(R_0 - 1)^2\tau_{\text{R}}\tau_n^3/(2R_0(\tau_{\text{R}} + \tau_n))$ and a first order expansion of τ_{R} around 1 gives us that $\tau_{\text{R}} \approx 1/(\mu R_0)$ which together with the geometric approximation of τ_n from (13) yields

$$-\frac{\tau'_{\text{SIR}}(0)}{\tau(1) - \tau(0)}\varepsilon \approx -n^3\frac{(R_0 - 1)^4}{R_0^3\mu\alpha^2(\alpha^2 + n(R_0 - 1))}\varepsilon, \tag{32}$$

which is a very small number for reasonable parameter values and choices of n. We illustrate this with a numerical example: Suppose that we have a population which is separated into two equally large sub-communities of size $n = 50,000$. Suppose further that the average infectious period is one week and a typical individual lives for ca. 70 years, i.e. $\alpha \approx 3500$. This together with $R_0 = 14$ and $\varepsilon = 1$ gives us that the exponent (32) is approximately -475, and $\exp(-475) \approx 0$, thus $\tau_{\text{SIR}}(1) \approx \tau(1)$.

5 Examples and simulations

We now compare our two approximations with simulations for some different parameter values and number of sub-communities. For childhood diseases the average infectious period is typically one to two weeks, see pp. 81-86 in Anderson and May (1991). This together with the assumption that the average life length among individuals in the population is 70 years, gives us α-values between 1,800 and 3,500. Usually, these kind of diseases have values of R_0 around 10 or higher, see Table 4.1 on pp. 70 in Anderson and May (1991). We have chosen to set R_0 to 14 in compliance with Nåsell (2005). These are the parameter values which we will use. As for the number of sub-communities we have chosen to concentrate on $k = 3$ and 5.

All simulations have been done using Monte Carlo simulation and the routines are written in the C-programming language. For graphical presentation `MATLAB` has been used. The expected time to extinction when starting in quasi-stationarity is estimated from the simulations as follows: Initially, 500 epidemics were started at the endemic level, which is the mean in the limiting quasi-stationary distribution. Then the epidemics were simulated long enough for 100 of them to go extinct, and at this time point, the clock for the remaining 400 simulations was started. These starting points will be approximately from the quasi-stationary distribution since the epidemics have been started at the endemic level together with the fact that they had not gone extinct for some time.

Depending on the parameter values we have alternated between using the truncated normal and geometric approximation of the quasi-stationary distribution. For more on the appropriate choice of approximation in different parts of the parameter region see Nåsell (2005), Fig. 3. Roughly one can say that for the present parameter values, the geometric approximation, Eq. (13), is to prefer if n is smaller than ca. 5-600,000 and when n is greater than ca. 2,000,000 one should use the truncated normal approximation, Eq. (9). For values of n in between, neither of the approximations work well. In Fig. 2.a we have a total population size of $n_{tot} = 150,000$ and $k = 3$, i.e. $n = 50,000$, and in this parameter region we use the geometric approximation of the quasi-stationary distribution. By studying the graphs we see that both approximations work well for small values of ε, but that τ_{Exp} also works satisfactory for intermediate values of ε. Notice, however, that the expected time to extinction is too short to say that the disease is endemic, since the average time to extinction is ≈ 0.01 life lengths, i.e. less than a year. Still, the simulations indicate that the expected time to extinction is increasing in the degree of social interaction between sub-communities, and that already for small values of ε the expected time to extinction is close to the case when

the population is mixing homogeneously. In Fig. 2.b we have used the same parameter values as in Fig. 2.a, but now with $k = 5$. Note that since we keep n_{tot} fixed the sub-community size is smaller. We again see that τ_{Exp} performs better than τ_{SIR}.

In Figs. 2.c and 2.d we have increased the total population size, n_{tot}, to 900,000. Now we are in a situation where neither of the approximations of the quasi-stationary distribution, Eqs. (9) and (13), work satisfactory, but we have chosen to use the geometric approximation, Eq. (13). In Figs. 2.c and 2.d it is seen that neither of τ_{SIR} and τ_{Exp} work that well unless ε is very small, i.e. $\varepsilon \sim 10^{-5} - 10^{-4}$, and both $\tau_{\text{Exp}}(0)$ and $\tau_{\text{Exp}}(1)$ are quite far away from the corresponding simulated values. Values of $\varepsilon \sim 10^{-5} - 10^{-4}$ are, however, probably not very realistic. A rough more realistic 'small' value of ε is for example if individuals visit other sub-communities one day per year, then an estimate of ε is $1/365$. Instead of using the analytical approximations of $\tau_n, \tau(0)$ and $\tau(1)$, we can use the corresponding 'true' values obtained from the simulations. If we do so our approximations performs better and this is illustrated in Fig. 2.e for the same parameter values as before when $n_{tot} = 900,000$ and $k = 3$. The behaviour of τ_{SIR} and τ_{Exp} is very similar to that in Figs. 2.a and 2.b. That is, τ_{Exp} seems to perform better and the functional form of τ_{Exp} gives a rather good description of the expected time to extinction. The improvement in the behaviour of τ_{SIR} and τ_{Exp} between Figs. 2.c and 2.d, and Fig. 2.e is an indication of that both our approximations are sensitive to the initial approximations of τ_n, $\tau(0)$ and $\tau(1)$.

In Fig. 2.f we have five sub-communities and a total population size of 2,500,000, and in this situation we use the truncated normal approximation of the quasi-stationary distribution of $\tau(1)$, and the geometric approximation for τ_n and $\tau(0)$. Once again the behaviour of our approximations is similar to what we have seen before, but τ_{Exp} does not work particularly well for intermediate values of ε.

6 Discussion

In the present paper we have been concerned with approximations of τ, the expected time to extinction for an SIR model for a population divided into sub-communities, when each sub-community is started at the endemic level of infection. Our aim has been to understand the effect of the level of population subdivision on the time to extinction. We have mainly focused on endemic diseases which have a short average infectious period in relation to average life length, such as childhood diseases.

Our first approximation, τ_{SIR} (Approximation 1), extends a method pre-

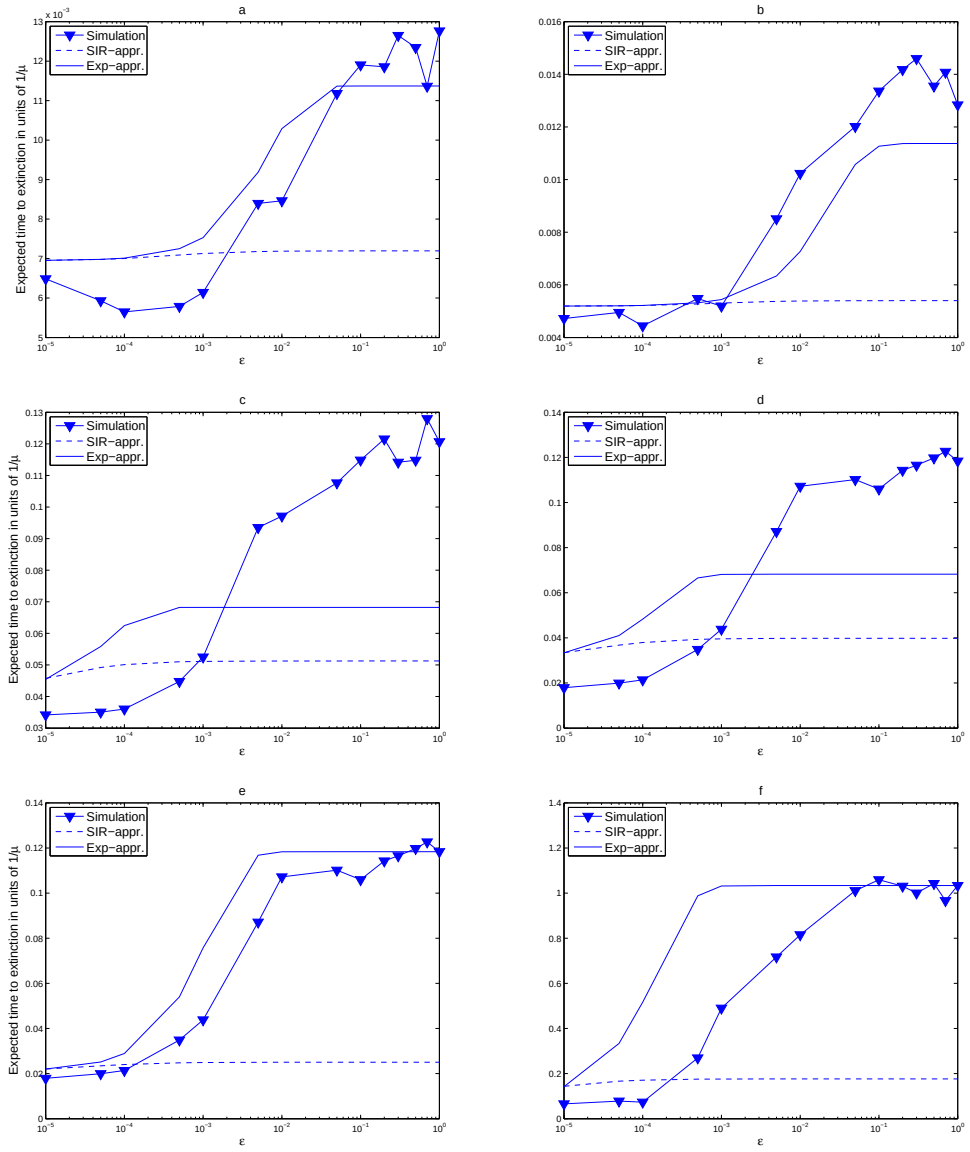

Figure 2: In all Figs. a-f we have used $\alpha = 3,500$ and $R_0 = 14$, and time is measured in units of life lengths. In Figs. a-b we have a total population size of $n_{tot} = 150,000$, but in Fig. a there are $k = 3$ sub-communities where as in b $k = 5$. In Figs. c and d $n_{tot} = 900,000$ and $k = 3$ and $k = 5$ respectively. Further, Fig. e is the same as Fig. d, but here $\tau(0)$ and $\tau(1)$ are approximated with the values corresponding to $\varepsilon = 0$ and $\varepsilon = 1$ from the simulation, and τ_n can be obtained from $\tau(0)$. In Fig. f $n_{tot} = 2,500,000$ and $k = 5$.

sented in a paper by Hagenaars et al. (2004). In Hagenaars et al. (2004) they are mainly interested in diseases with long infectious period (small α), such as scrapie among sheep, in situations where the social activity between sub-communities is low, i.e. $0 < \varepsilon \ll 1$. Under these circumstances they argue that the underlying SIR model can be analysed from a sub-community view, where each sub-community is classified as either fully susceptible or endemic, and they approximate the dynamics in the population with a birth and death process for the number of endemic sub-communities, which only allow for one re-infection. The expected time to extinction for this process, τ_{SI} from Eq. (22), is then a reasonable approximation of τ. Here we are mainly interested in childhood diseases. Based on similar arguments as those made in Hagenaars et al. (2004), we argue that it is necessary to introduce a recovered (and immune) state when classifying sub-communities in order to avoid over-estimation of τ. For this extended model we approximate the underlying SIR model with a Markov jump process for the number of endemic sub-communities, see Fig. 1, and estimate τ with the corresponding expected time to extinction τ_{SIR}. We present a general solution form for an arbitrary number of k sub-communities in Eq. (26), which can be solved numerically, and we present an explicit expression for the case when $k = 2$ in (27).

Simulations indicate that the expected time to extinction is increasing in the degree of social interaction between sub-communities, which also can be seen in τ_{SIR}. Further, τ_{SIR} is more suitable to use when the degree of social activity between sub-communities is very low. One crucial part with this approximation is that it is difficult to find a natural way of defining the time which we let a sub-community stay recovered (and immune), and perhaps the choice for this quantity which we suggest in (30) is chosen to restrictively. If one could improve this part of the approximation, it is possible that τ_{SIR} could work better when the degree of social activity between sub-communities is somewhat higher. It is also hard to say something general about at which point τ_{SIR} becomes independent of ε. Furthermore τ_{SIR} is rather sensitive to the initial approximations of τ_n, $\tau(0)$ and $\tau(1)$. Depending on the parameter values we use either the truncated normal or geometric distribution respectively.

Our second approximation, τ_{Exp} (Approximation 2), is motivated in a slightly different way. At the end of Section 2 the time to extinction for the two extreme cases where derived, i.e. $\tau(0)$ and $\tau(1)$. These two approximations only rely on results for the SIR-HM model, and can not be improved without improving results for that model. The idea behind τ_{Exp} is to approximate the expected time to extinction with an exponential curve starting at $\tau(0)$ and which has $\tau(1)$ as approximate end point and having $\tau'_{\text{Exp}}(0) = \tau'_{\text{SIR}}(0)$, since τ_{SIR} is reasonable to use when the social activity be-

tween sub-communities is very low. When comparing τ_{Exp} with simulations the same behaviour as for τ_{SIR} is seen. It works satisfactory when the degree of social activity is low, and it is increasing in the degree of social activity.

To conclude, τ_{SIR} is only suitable to use for small values of ε and τ_{Exp} is suitable to use for intermediate as well as small values of ε. Thus, τ_{Exp} is the best approximation of the two. Note, however, that the use of these approximations is not recommended unless we are in the parts of the parameter region where either of the SIR-HM approximations of the quasi-stationary distribution is good. To our knowledge these approximations are the only ones at hand which deal with the expected time to extinction when α is large.

In the present paper it was also shown that T_Q, the time to extinction given that the epidemic process is started in the quasi-stationary distribution, is exponentially distributed, see Proposition 1. This result is important when talking about other quantities of interest such as critical community size.

Some possible improvements of the present results has been commented above, but as always it would be tractable to leave the assumption of exponential infectious periods which in most situations is not realistic. This would be interesting to do for both the SIR-HM and the SIR-SC model. By following the framework provided in Andersson and Britton (2000b) the infectious periods could be generalised to gamma distributions, and it would be possible to use the methods discussed in the present paper. Another, perhaps easier, generalisation of the model would be to allow for differently sized sub-communities, and we believe that the methods described in the present paper could be used.

7 Acknowledgements

The authors are grateful to Swedish Foundation for Strategic Research (SSF). The authors would also like to thank two anonymous reviewers for helpful suggestions and clarifying remarks.

References

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans - Dynamics and Control. Oxford University Press, pp. 13-19, 31, 70, 81-86, 128-143.

Andersson, H., Britton, T., 2000a. Stochastic Epidemic Models and Their Statistical Analysis. Lecture Notes in Statistics 151, Springer-Verlag, New York, pp. 22-25.

Andersson, H., Britton, T., 2000b. Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. J. Math. Biol. 41, 559-580.

Ball, F., Mollison, D., Scalia-Tomba, G., 1997. Epidemics with Two Levels of Mixing. Ann. Appl. Probab. 7, 46-89.

Ball, F., Owen, D.L., 2001. Stochastic multi-type SIR epidemics among a population partitioned into households. Adv. in Appl. Probab. 33, 99-123.

Ball, F., Neal, P., 2002. A general model for stochastic SIR epidemics with two levels of mixing. Math. Biosci. 180, 73-102.

Bartlett, M.S., 1957. Measles periodicity and community size. J. R. Statist. Soc. A 120, 48-70.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Chichester: Wiley, cop.

van Doorn, E.A., 1991. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. Adv. Appl. Prob. 23, 683-700.

van den Driessche, P., Watmough, J., 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 180, 29-48.

Dushoff, J., Plotkin, J.B., Levin, S.A., Earn, D.J.D., 2004. Dynamical resonance can account for seasonality of influenza epidemics. PNAS, vol. 101, no. 48, 16915-16916.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes: Characterization and Convergence. Wiley, New York.

Etienne, R.S., Heesterbeek, J.A.P., 2000. On Optimal Size and Number of Reserves for Metapopulation Persistence. J. Theor. Biol. 203, 33-50.

Hagenaars, T.J., Donnelly, C.A., Ferguson, N.M., 2004. Spatial heterogeneity and the persistence of infectious diseases. J. Theo. Biol. 229, 349-359.

van Herwaarden, O.A., Grasman, J., 1995. Stochastic epidemics: major outbreaks and the duration of the endemic period. J. Math. Biol. 33, 581-601.

Karlin, S., Taylor, H.M., 1975. A first course in stochastic processes (Second edition). Academic Press, pp. 148-150.

Karatzas, I., Shreve, S.E., 1991. Brownian motion and stochastic calculus (Second edition). Graduate Texts in Mathematics 113, Springer-Verlag, New York, pp. 357.

Keeling, M.J., 2000a. Metapopulation moments: coupling, stochasticity and persistence. J. Animal Ecol. 69, 725-736.

Keeling, M.J., 2000b. Multiplicative Moments and Measures of Persistence in Ecology. J. Theor. Biol. 205, 269-281.

Lloyd, A.L., May, R.M., 1996. Spatial Heterogeneity in Epidemic Models. J. Theor. Biol. 179, 1-11.

Nåsell, I., 1999. On the time to extinction in recurrent epidemics. J. R. Statist. Soc. B 61, 309-330.

Nåsell, I., 2005. A new look at the critical community size for childhood infections. Theo. Pop. Biol. 67, 203-216.

Pollett, P.K., Roberts, A.J., 1990. A Description of the long-term behaviour of absorbing continuous-time markov chains using a centre manifold. Adv. Appl. Prob. 22, 111-128.

Wonham, M.J., de-Camino-Beck, T., Lewis, M.A., 2004. An epidemiological model for West Nile virus: invasion analysis and control applications. Proc. Roy. Soc. Lond B. 271, 501-507.

Paper II

Mathematical Statistics
Stockholm University
Research Report **2007:9**,
http://www.math.su.se/matstat

On the time to extinction for a two-type version of Bartlett's epidemic model

Mathias Lindholm*

April 2007

Abstract

We are interested in how the addition of type heterogeneities affects the long time behaviour of endemic diseases. We do this by analysing a two-type version of a model introduced by Bartlett under the restriction of proportionate mixing. This model is used to describe diseases for which individuals switch states according to susceptible $\rightarrow$ infectious $\rightarrow$ recovered and immune, where the immunity is lifelong. We describe an approximation of the distribution of the time to extinction given that the process is started in the quasi-stationary distribution, and we analyse how the variance and the coefficient of variation of the number of infectious individuals depends on the degree of heterogeneity between the two types of individuals. These are then used to derive an approximation of the time to extinction. From this approximation we get that if we increase the difference in infectivity between the two types the expected time to extinction decreases, and if we instead increase the difference in susceptibility the behaviour of the expected time to extinction depends on which part of the parameter space we are in, and we can also obtain non monotonic behaviour. These results are supported by simulations.

KEY WORDS: Stochastic SIR epidemic model, Quasi-stationary distribution, Diffusion approximation, Endemic diseases

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Stockholm, Sweden. E-mail: lindholm@math.su.se.

1 Introduction

Diseases that are able to persist in a population for a long time without the need of introducing new infectious individuals from an external population are called endemic, see e.g. pp. 73 in [1]. A question which has received a lot of attention in the literature is the behaviour of an endemic disease after a long time, see e.g. [2] and the references therein, [3], [4], [5]. In the present paper we look at the situation when all individuals in a homogeneously mixing population is classified as one of two types, e.g. children or adults, female or male etc. These types give the individuals different characteristics in terms of their ability to become infected while susceptible, and their ability to infect while infectious. We are interested in how this addition of heterogeneity affects the persistence of an infectious disease. As a measure of persistence we use the additional time to extinction conditioned on that the disease has not gone extinct for a long time. We have used a stochastic epidemic model introduced by Bartlett in [6], and modified it to take types into account. Bartlett proposed a very similar deterministic two-type model in [6], and the model treated in the present paper is a stochastic analogue to this model when Bartlett's deterministic model is suitably parametrized. The definition of Bartlett's original stochastic one-type model which we will use is from ch. VII in [7]. This model is used to describe diseases where individuals switch between the states susceptible, S, infectious, I, and recovered (and immune), R, according to $\mathrm{S} \to \mathrm{I} \to \mathrm{R}$. Henceforth we make no distinction between recovered and immune, and only refer to this state as being recovered. Immunity is here regarded as life-long, hence the state R is absorbing up until death. The big difference between Bartlett's model and the standard SIR epidemic, see e.g. [1] pp. 11, is that it has a simple demographic mechanism. This mechanism is necessary in order for the disease to become endemic. That is, since the model has a recovered state, we need to have some inflow of susceptible individuals in order for this behaviour to appear. Bartlett's model has been widely studied over the years, e.g. [2] and the references therein, and different versions of this model has emerged, see e.g. [8]. A central issue in many of these papers is the behaviour of the epidemic after a long time.

We analyse an extension of this model when two types of individuals are present under the restriction of so-called proportionate mixing, that is when the contact parameter λ_{ij} for a contact between an infected i-individual and a susceptible j-individual, has the form $\lambda_{ij} = \alpha_i\beta_j$ (this is sometimes called separable), where $\{\alpha_i\}$ and $\{\beta_j\}$ are called infectivities and susceptibilities respectively. Under this extra restriction we describe an approximation of the distribution of the time to extinction conditioned on that the disease

has not gone extinct for a long time, i.e. an approximation of the quasi-stationary distribution. We also analyse the variance and the coefficient of variation of the number of infectious individuals under quasi-stationarity. When comparing our analytical approximations with simulations, it is seen that the approximations are not perfect, but that they capture the qualitative behaviour of the epidemic in relevant regions of the parameter space. That is, if we increase the difference in infectivity between the two types the expected time to extinction decreases, and if we instead increase the difference in susceptibility the behaviour of the expected time to extinction depends on which parameter region we are in, and we can in fact observe non monotonic behaviour.

For the disposition of this paper, we define the two-type version of Bartlett's epidemic model and describe our main results in Section 2 and in Section 3 we make some numerical comparisons with simulations. Section 4 is devoted to a closing discussion.

2 Bartlett's epidemic model with two types of individuals

We have an open population without social structures where each individual lives forever and is classified as either a type one or a type two individual, e.g. female or male, young or adult etc., and individuals are not allowed to change type. An alternative interpretation of the model is that individuals never die while susceptible or infectious. This interpretation is realistic when we are interested in childhood diseases, and this is the interpretation we will use henceforth. The population is open in that new susceptible individuals are brought into the population via birth/immigration, or more formally: susceptible individuals of type j are born according to a pure birth process with constant rate, $\mu n_j, j = 1, 2$. Here $1/\mu$ is considered to be of the same order of magnitude as the average life-length of a real-life human, e.g. $1/\mu \approx 70$ years, and where n_j is more or less the average number of type j individuals which are either susceptible or infectious, or if we suppose that type j individuals in R dies at rate μ then $n_j \approx \mathrm{S}_j + \mathrm{I}_j + \mathrm{R}_j$. By introducing new susceptible individuals into the population in this way, we will on average have a proportion $\pi_j = n_j/n, n = n_1 + n_2$, type j individuals in the population which are either susceptible or infectious. Turning to the spread of disease, an infectious i individual stays so for an exponentially distributed infectious period with mean $1/\nu_i = 1/\nu$, where $1/\nu$ is considered small, e.g. $1/\nu$ is about one week for measles (if $1/\mu \approx 70$ then $1/\nu \approx 1/52$), see e.g. pp. 31 in [9]. During this time period, the infectious i individual makes contacts with a given j

individual according to a Poisson process with rate λ_{ij}/n. If a contacted j individual is susceptible at the time of contact then this individual becomes infectious. After the infectious period is over the previously infectious individual recovers and becomes immune, that is, individuals switch between being susceptible, infectious and recovered according to S $\to$ I $\to$ R. Since recovered individuals never again will contribute to the infectious pressure in the population, we only need to keep track of the number of susceptible and infectious individuals at each time point to be able to describe the dynamics of the epidemic. Due to this, we can always interpret the model as if death occurs at rate μ in the recovered state, which is reasonable when talking about childhood diseases. Note, that since susceptible individuals are born at a constant rate there will always be a large number of susceptible individuals present in the population at all time points, hence the epidemic will not go extinct due to lack of susceptible individuals to infect.

In the present paper we will use this model under the extra restriction of proportionate mixing, which is the situation when λ_{ij} can be written as $\lambda_{ij} = \alpha_i\beta_j$ where $\{\alpha_i\}$ and $\{\beta_j\}$ are called infectivities and susceptibilities respectively. Let $(\mathbf{X}(t), \mathbf{Y}(t)) = (X_1(t), X_2(t), Y_1(t), Y_2(t))$ denote the process of number of susceptible and infectious individuals of the two types at t. Rates for all possible transitions are defined in Table 1. From these rates we also see that the disease-free set of states is absorbing. Denote this set by $\mathcal{S}_{\text{abs}}$, where

$$\mathcal{S}_{\text{abs}} = \{(\mathbf{x}, \mathbf{y}); \mathbf{x} \in \mathbb{N} \times \mathbb{N}, \mathbf{y} = \mathbf{0}\} \tag{1}$$

where $\mathbb{N} = \{0, 1, 2, \ldots\}$. If we instead look at the process of proportions, $(\mathbf{X}(t)/n, \mathbf{Y}(t)/n)$, where $n = n_1 + n_2$, then its deterministic counterpart is described by the following system of differential equations:

$$\begin{cases} x_j'(t) &= \mu\pi_j - \beta_j x_j(t)\left(\alpha_1 y_1(t) + \alpha_2 y_2(t)\right) \\ y_j'(t) &= \beta_j x_j(t)\left(\alpha_1 y_1(t) + \alpha_2 y_2(t)\right) - \nu y_j(t) \end{cases} \tag{2}$$

where $\pi_j = n_j/n,\ j = 1, 2$. An important epidemiological quantity is R_0, the so-called basic reproduction number. For general multi-type epidemic models R_0 is defined as the largest eigenvalue of the mean offspring matrix $\{\lambda_{ij}\pi_j/\nu\}$, see pp. 51-61 in [1]. For Bartlett's model with two types of individuals, under the restriction of proportionate mixing, the mean offspring matrix becomes $\{\alpha_i\beta_j\pi_j/\nu\}$, and we get that

$$R_0 = \frac{\alpha_1\beta_1\pi_1 + \alpha_2\beta_2\pi_2}{\nu}. \tag{3}$$

In general, for a disease to be able to become endemic R_0 must be strictly above one, see e.g. ch. 8 in [1]. This is however *not* the case for Bartlett's

model, which can become endemic for *any* $R_0 > 0$. That is, regardless of the value of $R_0 > 0$, the epidemic process has always a positive probability of stabilising around some equilibrium.

That a disease becomes endemic is the same as to say that it has been able to persist in the population for a long but finite time. In the present paper we use the additional time to extinction conditioned on that the process has not gone extinct for a long but finite time, as a measure of persistence. If the process has not gone extinct after a long time it is likely to have stabilised, making small fluctuations, around the endemic level. The endemic level is obtained by finding the stationary points to the system of differential equations defined in (2). For this particular model we get one stationary point

$$(\hat{x}_j, \hat{y}_j) = \left(\frac{\pi_j \mu \gamma}{\alpha_\pi \beta_j}, \frac{\pi_j}{\gamma} \right), \ j = 1, 2 \tag{4}$$

where $\alpha_\pi = \alpha_1\pi_1 + \alpha_2\pi_2$, $\beta_j > 0, j = 1, 2$, and $\gamma = \nu/\mu$. Here γ denotes the ratio of average lifetime and average infectious period. Note that as discussed above, the stationary point from (4) will always exist as soon as $R_0 > 0$. Under these circumstances it is natural to look at the quasi-stationary distribution of the epidemic process denoted Q, i.e. the distribution conditioned on that the epidemic has not died out by time t when we let t tend to infinity. Hence, we are interested in T_Q, the time to extinction given that the process is started in the quasi-stationary distribution. Regardless of whether or not we *know* the quasi-stationary distribution, we can still say something about T_Q:

$$\begin{aligned} P(T_Q > t + s \mid T_Q > t, \ (\mathbf{X}(0), \mathbf{Y}(0)) \sim Q) &= \\ &= P(T_Q > t + s \mid T_Q > t, \ (\mathbf{X}(t), \mathbf{Y}(t)) \sim Q) \\ &= P(T_Q > s \mid (\mathbf{X}(0), \mathbf{Y}(0)) \sim Q) \end{aligned}$$

i.e. T_Q has the lack of memory property which implies that T_Q is exponentially distributed. This follows the same reasoning as in [3]. To determine this exponential distribution completely we need to know its mean. Following the lines of [2] for Bartlett's original model, we know that $P(T_Q \leq t) = P(\mathbf{Y}(t) = \mathbf{0})$, and by defining the Kolmogorov forward equations for $P(\mathbf{Y}(t) = \mathbf{i})$ one can show that $E(T_Q) = \tau_Q = 1/\mu\gamma q_{\bullet,1}$. Here $q_{\bullet,1} = \sum_{\mathbf{x}}(q_{\mathbf{x},\mathbf{e}_1} + q_{\mathbf{x},\mathbf{e}_2})$, where $\mathbf{e}_j$ is the j'th unit vector. Thus, the distribution of T_Q will depend on the rate with which $(\mathbf{X}(t), \mathbf{Y}(t))$ enters the set of absorbing states, $\mathcal{S}_{\text{abs}}$ from (1). We state this more precisely in the following proposition:

Table 1: Bartlett's model with two types of individuals, possible transitions for individuals of type $j = 1, 2$. Here s_j and i_j denote the number of susceptible and infectious type j individuals at a certain time point. Note that if a change of state occurs for the type 1 individuals, the state of the type 2 individuals remains unchanged, and vice versa.

From	To	Rate
(s_j, i_j)	$(s_j + 1, i_j)$	μn_j
(s_j, i_j)	$(s_j - 1, i_j + 1)$	$\frac{\beta_j}{n} s_j(\alpha_1 i_1 + \alpha_2 i_2)$
(s_j, i_j)	$(s_j, i_j - 1)$	νi_j

Proposition 1 *The time to extinction given that the process is started in the quasi-stationary distribution, T_Q, is exponentially distributed with mean*

$$\tau_Q = \frac{1}{\mu\gamma q_{\bullet,1}} \tag{5}$$

where

$$q_{\bullet,1} = \sum_{\mathbf{x} \geq \mathbf{0}} (q_{\mathbf{x},\mathbf{e}_1} + q_{\mathbf{x},\mathbf{e}_2}) \tag{6}$$

and where $\mathbf{e}_j$ is the j'th unit vector.

A way of approximating the quasi-stationary distribution, $q_{\bullet,k}$, is via a diffusion approximation. Introduce the scaled and centred process

$$(\widetilde{\mathbf{X}}_n(t), \widetilde{\mathbf{Y}}_n(t)) = \sqrt{n}\left(\frac{\mathbf{X}(t)}{n} - \hat{\mathbf{x}}, \frac{\mathbf{Y}(t)}{n} - \hat{\mathbf{y}}\right) \tag{7}$$

where $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ corresponds to the endemic level of infection. From the theory of diffusion processes it is known that this process converges weakly to an Ornstein-Uhlenbeck process, $(\widetilde{\mathbf{X}}(t), \widetilde{\mathbf{Y}}(t))$, as n tends to infinity, see e.g. ch. 11 in [10]. Since the limiting process is of Ornstein-Uhlenbeck type, it has a *stationary* Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\hat{\mathbf{\Sigma}} = \{\hat{\sigma}_{ij}\}$, which is the solution to the following equation

$$\hat{\mathbf{B}}\hat{\mathbf{\Sigma}} + \hat{\mathbf{\Sigma}}\hat{\mathbf{B}}^{\mathrm{T}} = -\hat{\mathbf{S}}, \tag{8}$$

see e.g. pp. 357 in [11]. Here $\mathbf{B}$ and $\mathbf{S}$ are the local drift and covariance matrices of $(\widetilde{\mathbf{X}}(t), \widetilde{\mathbf{Y}}(t))$. The local drift matrix is the Jacobian of the first

order infinitesimal moment of $(\widetilde{\mathbf{X}}_n, \widetilde{\mathbf{Y}}_n)$ and the local covariance matrix is the infinitesimal covariance matrix of $(\widetilde{\mathbf{X}}_n, \widetilde{\mathbf{Y}}_n)$. We are interested in the behaviour of the epidemic process close to the endemic level, and we therefore approximate $\mathbf{B}$ and $\mathbf{S}$ at the endemic level, denoted $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{S}}$. We can now conclude the following: after a long but finite time the process is likely to have stabilised around the stationary point (4), and then $\mathbf{Y}(t) \approx \sqrt{n}\widetilde{\mathbf{Y}}(t) + n\hat{\mathbf{y}}$ implying that $\mathbf{Y}(t) \sim apprN(n\hat{\mathbf{y}}, nCov(\widetilde{\mathbf{Y}}))$. In particular we have that $Y_1(t) + Y_2(t) \sim apprN(\mu_Y, \sigma_Y^2)$, where

$$\begin{aligned} \mu_Y &= E(Y_1 + Y_1) = n(\hat{y}_1 + \hat{y}_2) \\ \sigma_Y^2 &= Var(Y_1 + Y_2) = n(\hat{\sigma}_{11}^2 + \hat{\sigma}_{22}^2 + 2\hat{\sigma}_{12}) \end{aligned} . \tag{9}$$

This, however, contradicts our original definition of the process $\mathbf{Y}(t)$ which is non-negative and integer valued, since we can not have a negative number of individuals, whereas the approximate distribution of the total number of infected individuals after a long time is defined on $\mathbb{R}$. But, if we truncate this distribution at 0 (or at 0.5 using continuity correction) we get an approximation of the the probability that $Y_1(t) + Y_2(t) = k$, i.e. $q_{\bullet,k}$:

$$q_{\bullet,k} \approx \frac{1}{\sigma_Y} \frac{\varphi((k - \mu_Y)/\sigma_Y)}{\Phi((\mu_Y - 0.5)/\sigma_Y)}, \ k \geq 0 \tag{10}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal distribution and density functions, and μ_Y and σ_Y are from (9). By using $q_{\bullet,1}$ from (10) together with Proposition 1 the distribution of T_Q is determined. This is analogous to what Nåsell did for Bartlett's original model in [2].

In Appendix A we derive $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{S}}$ for Bartlett's model with two types, with which we can find a solution to equation (8). This amounts to solving a ten dimensional equation system, and for this task we have used the symbolic software `MAPLE`. Unfortunately the closed expression for $\widehat{\Sigma}$ is lengthy and not easy to grasp, and is hence omitted. We can, however, calculate μ_Y and σ_Y^2 from (9). Calculating μ_Y gives us

$$\mu_Y = \frac{n}{\gamma}. \tag{11}$$

A general expression for σ_Y^2 turns out to be long and not illuminating, but if we either set $\alpha_1 = \alpha_2 = \alpha$ and vary β_1 and β_2, or set $\beta_1 = \beta_2 = \beta$ and vary α_1 and α_2, we can simplify σ_Y^2 quite a lot.

If we set $\alpha_1 = \alpha_2 = \alpha$ we get that

$$\sigma_{Y,\beta}^2 = \frac{n}{\gamma} + n\frac{\mu\gamma}{\alpha} \frac{\beta_1^2\pi_2 + \beta_2^2\pi_1 + \beta_1\beta_2 + (\beta_1\pi_1 + \beta_2\pi_2)\frac{\mu\gamma^2}{\alpha}}{\beta_1^2\left(\beta_2 + \frac{\mu\gamma^2}{\alpha}\pi_1\right) + \beta_2^2\left(\beta_1 + \frac{\mu\gamma^2}{\alpha}\pi_2\right)} \tag{12}$$

and if we instead set $\beta_1 = \beta_2 = \beta$ we get

$$\sigma^2_{Y,\alpha} = \frac{n}{\gamma} + n\frac{\mu\gamma}{\alpha_\pi}\frac{2\beta + ((\alpha_1+\alpha_2)\alpha_\pi - \alpha_1\alpha_2)\,\frac{\mu\gamma^2}{\alpha_\pi^3}}{2\beta + \frac{\mu\gamma^2}{\alpha_\pi}} \tag{13}$$

where $\alpha_\pi = \alpha_1\pi_1 + \alpha_2\pi_2$. Our particular interest is to analyse the effect of the heterogeneity caused by including types into the model, and this we would do by looking at the α's and β's separately, i.e. $\lambda_{ij} = \alpha\beta_j$ or $\lambda_{ij} = \alpha_i\beta$, so from this point of view we have not limited ourselves. A measure used to get an idea of how far the process is from extinction, is the coefficient of variation, CV_Y, defined as

$$CV_Y = \frac{\sigma_Y}{\mu_Y} \tag{14}$$

but, since μ_Y from (11) is independent of all α's and β's, we can analyse σ^2_Y as a function of either the α's or β's instead. That it is enough to analyse σ^2_Y, and no higher order moments, is due to that we approximate the quasi-stationary distribution with a truncated normal distribution, i.e. the approximating distribution lacks skewness. Intuitively, as σ^2_Y increases, we are more likely to make larger fluctuations around the endemic level, and are hence more likely to hit $\mathcal{S}_{\text{abs}}$, the disease-free set of states. Thus, increasing the variance ought to shorten the expected time to extinction, and vice versa.

2.1 Analysing $\sigma^2_{Y,\beta}(\delta)$: the effect of difference in susceptibility between the two types

To be able to compare our results with those for Bartlett's original one-type model, we parametrize β_1 and β_2 according to

$$\begin{cases}\beta_1 = \beta(1-\delta) \\ \beta_2 = \beta(1+\frac{\pi_1}{\pi_2}\delta)\end{cases}, \ \delta \in [0,1] \tag{15}$$

where δ correspond to the degree of heterogeneity between the two types. By using this parametrization we get that $R_0 = \alpha\beta/\gamma$, hence independent of δ and thus compatible with R_0 for the original Bartlett model. Note that the limits for δ are chosen so that both β_1 and β_2 will remain positive, and due to symmetry we only look at $\delta \in [0,1]$, since we do not gain any extra information by including $\delta \in [-\pi_2/\pi_1, 0)$. Further, since R_0 is independent of δ and we are interested in the heterogeneity caused by differences in the susceptibilities, we can set $\alpha_1 = \alpha_2 = 1$, because we can always scale the β's

such that the α's can be set to unity. When parametrizing the β's in this way and setting $\pi_1 = 1-\pi$ and $\pi_2 = \pi$ then $\sigma^2_{Y,\beta}$ from (12) can be written as

$$\sigma^2_{Y,\beta}(\delta) = \frac{n}{\gamma} + n\frac{f(\delta)}{\mu\gamma R_0 f(\delta) + g(\delta)}\mu\gamma \tag{16}$$

where

$$\begin{aligned} f(\delta) &= \beta(2\beta + \mu\gamma^2)\pi^2 + 3\beta^2(1-2\pi)\pi\delta + \beta^2(1-2\pi)^2\delta^2 = \\ &= R_0(\mu\gamma)^2(2R_0+\gamma)\pi^2 + 3(R_0\mu\gamma)^2(1-2\pi)\pi\delta + (R_0\mu\gamma)^2(1-2\pi)^2\delta^2 \\ &= a_1 + a_2\delta + a_3\delta^2 \qquad (17) \\ g(\delta) &= \beta^2(\mu\gamma^2 - 2\beta)\pi(1-\pi)\delta^2 - \beta^3(1-\pi)(1-2\pi)\delta^3 \\ &= R_0^3(\mu\gamma)^4(\gamma - 2R_0)\pi(1-\pi)\delta^2 - (R_0\mu\gamma)^3(1-\pi)(1-2\pi)\delta^3 \\ &= a_4\delta^2 + a_5\delta^3 \qquad (18) \end{aligned}$$

Analysing $\sigma^2_{Y,\beta}(\delta)$ at its end points gives us

$$\sigma^2_{Y,\beta}(\delta) = \begin{cases} \frac{n}{\gamma} + \frac{n}{R_0} & \text{if } \delta = 0 \\ \frac{n}{\gamma\pi} + n\frac{\pi}{R_0} & \text{if } \delta = 1 \end{cases} \tag{19}$$

so that if $\pi = R_0/\gamma$ then $\sigma^2_{Y,\beta}(0) = \sigma^2_{Y,\beta}(1)$. Setting $\pi = R_0/\gamma$, then straightforward calculations gives us that $\sigma^2_{Y,\beta}(\delta)$ is *not* independent of δ, hence $\sigma^2_{Y,\beta}(\delta)$ is *not* monotone for all choices of parameters. We also see that when $\pi \geq R_0/\gamma$ then $\sigma^2_{Y,\beta}(0) \geq \sigma^2_{Y,\beta}(1)$, and that when $\pi < R_0/\gamma$ then $\sigma^2_{Y,\beta}(0) < \sigma^2_{Y,\beta}(1)$. Note that if $\delta = 1$ then $\beta_1 = 0$, thus, susceptible type one individuals can never become infectious and $X_1(t)$ is a strictly *growing* process. This gives us that $X_1(t)$ does not have an endemic level and from (2) we see that the only stable point for $Y_1(t)$ is 0. Hence, $\sigma^2_{Y,\beta}(1)$ only describes the variation among infectious type two individuals. From (16) together with (17) and (18) we see that $\sigma^2_{Y,\beta}(\delta)$ depends on δ in a non-trivial way, that is, if we change δ in $0 \leq \delta \leq 1$ we can not tell whether $\sigma^2_{Y,\beta}(\delta)$ increases or decreases for arbitrary choices of R_0, γ, μ and π. In order to find the extreme points of $\sigma^2_{Y,\beta}(\delta)$ we equate the first derivative of $\sigma^2_{Y,\beta}(\delta)$ to 0, which immediately gives us that $\delta_0 = 0$ is a root, and if we calculate the second derivative of $\sigma^2_{Y,\beta}(\delta)$ in the point 0 we get

$$\frac{d^2}{d\delta^2}\sigma^2_{Y,\beta}(0) = n\frac{2}{(\mu\gamma)^2 R_0^3}\frac{2R_0 - \gamma}{2R_0 + \gamma}\frac{1-\pi}{\pi} \tag{20}$$

so that $\sigma^2_{Y,\beta}(0)$ is a local maximum if $0 < R_0 < \gamma/2$ and a local minimum if $R_0 > \gamma/2$. Note that for all practical purposes 0 will be a local maximum to

$\sigma^2_{Y,\beta}$, since typical values of R_0 and γ for diseases like measles are $R_0 \approx 10-15$ and $\gamma \approx 1800-3500$, see e.g. pp. 31 and 70 in [9]. The remaining three roots can be solved explicitly by using Cardano's formula, see e.g. pp. 65 in [12], and this is done in Appendix B. Using these roots together with the values of $\sigma^2_{Y,\beta}(\delta)$ in the end points of the interval $\delta \in [0,1]$ we can determine the functional form of $\sigma^2_{Y,\beta}(\delta)$.

2.2 Analysing $\sigma^2_{Y,\alpha}(\delta)$: the effect of difference in infectivity between the two types

If we instead set $\beta_1 = \beta_2 = 1$ and parametrize α_1 and α_2 analogously to (15), then $\sigma^2_{Y,\alpha}$ from (13) simplifies to

$$\sigma^2_{Y,\alpha}(\delta) = \frac{n}{\gamma} + \frac{n}{R_0}\frac{2R_0 + \gamma + \gamma\frac{1-\pi}{\pi}\delta^2}{2R_0+\gamma} \tag{21}$$

which increases monotonically as δ increases, such that

$$\sigma^2_{Y,\alpha}(\delta) = \begin{cases} \frac{n}{\gamma} + \frac{n}{R_0} & \text{if } \delta = 0 \\ \frac{n}{\gamma} + \frac{n}{R_0}\frac{2R_0+\frac{\gamma}{\pi}}{2R_0+\gamma} & \text{if } \delta = 1 \end{cases} \tag{22}$$

Worth noticing is that if we set $\delta = 1$, we have moved the *entire* infectious pressure to the type two individuals, and the only way that a susceptible individual may become infected is via an infectious type two individual. To see that this only corresponds to a shift in the infectious pressure, we can look at the endemic level from (4) which is unchanged.

2.3 The effect of type heterogeneities on τ_Q

Using the approximations $\sigma^2_{Y,\beta}(\delta)$ and $\sigma^2_{Y,\alpha}(\delta)$ from Eqs. (16) and (21) respectively, together with Proposition 1 gives us approximations $\tau_{Q,\beta}(\delta)$ and $\tau_{Q,\alpha}(\delta)$, the expected time to extinction when the epidemic process is started at quasi-stationarity as a function of the degree of heterogeneity in terms of susceptibility or infectivity. We will in the remainder of this exposition sometimes use the notation $\tau_{Q,\cdot}$ and $\sigma^2_{Y,\cdot}$ when we do not want to stress the effect of neither varying susceptibility nor infectivity.

From the definitions of μ_Y and $\sigma_{Y,\cdot}(\delta)$ we know that for large enough n the relation $\mu_Y > \sigma_{Y,\cdot}(\delta)$ holds for all δ, i.e. $\mu_Y/\sigma_{Y,\cdot}(\delta) > 1$ for all δ. If we

see to $q_{\bullet,1}$ from (10) we can write it as

$$q_{\bullet,1}(\delta) \approx \frac{1}{\sigma_{Y,\cdot}(\delta)} \frac{\varphi\left(\frac{1-\mu_Y}{\sigma_{Y,\cdot}(\delta)}\right)}{\Phi\left(\frac{\mu_Y-0.5}{\sigma_{Y,\cdot}(\delta)}\right)} \approx \frac{1}{\sigma_{Y,\cdot}(\delta)} \frac{\varphi\left(\frac{\mu_Y}{\sigma_{Y,\cdot}(\delta)}\right)}{\Phi\left(\frac{\mu_Y}{\sigma_{Y,\cdot}(\delta)}\right)} = \frac{1}{\mu_Y} \frac{\mu_Y}{\sigma_{Y,\cdot}(\delta)} \frac{\varphi\left(\frac{\mu_Y}{\sigma_{Y,\cdot}(\delta)}\right)}{\Phi\left(\frac{\mu_Y}{\sigma_{Y,\cdot}(\delta)}\right)} \tag{23}$$

$$= \frac{1}{\mu_Y} \frac{u\varphi(u)}{\Phi(u)} = \frac{1}{\mu_Y} g(u) > 0. \tag{24}$$

Differentiating $g(u)$ w.r.t. u and using that $\Phi'(u) = \varphi(u)$ and $\varphi'(u) = -u\varphi(u)$ we get that

$$g'(u) = \frac{g(u)}{u}(1 - u^2 - g(u)). \tag{25}$$

This together with that $g(u)/u > 0$ for all $u > 0$ gives us that $g'(u) < 0$ for $u \geq 1$, i.e. $g(u)$ decreases monotonically when $u \geq 1$. Thus for $u \geq 1$ all non monotonic behaviour of $q_{\bullet,1}$ is a result of the non monotonic behaviour of $\sigma_{Y,\cdot}(\delta)$. We will henceforth only consider this situation. Using the relation that $\mu_Y/\sigma_{Y,\cdot}(\delta) \geq 1$ we can get bounds on n for this to hold, and these bounds on n are needed in the next section when we compare our analytical results with simulations. If we see to $\sigma_{Y,\beta}(\delta)$ we know 1) that when $\pi \geq R_0/\gamma$ then $\sigma^2_{Y,\beta}(0) \geq \sigma^2_{Y,\beta}(1)$ and 2) that when $\pi < R_0/\gamma$ then $\sigma^2_{Y,\beta}(0) < \sigma^2_{Y,\beta}(1)$. If we start with 1) and assume that $\sigma_{Y,\beta}(0)$ is the largest value of $\sigma_{Y,\beta}(\delta)$ for $\delta \in [0, 1]$, we get that

$$\frac{\mu_Y}{\sigma_{Y,\beta}(\delta)} \geq \frac{\mu_Y}{\sigma_{Y,\beta}(0)} \geq 1 \tag{26}$$

$$\Rightarrow \frac{n}{\gamma} = \mu_Y \geq \sigma_{Y,\beta}(0) = \sqrt{\frac{n}{\gamma} + \frac{n}{R_0}} \tag{27}$$

which gives us the following lower bound on n

$$n_\beta \geq \gamma^2 \left(\frac{1}{\gamma} + \frac{1}{R_0}\right). \tag{28}$$

Likewise for 2) we get that

$$\frac{\mu_Y}{\sigma_{Y,\beta}(\delta)} \geq \frac{\mu_Y}{\sigma_{Y,\beta}(1)} \geq 1 \tag{29}$$

which gives us

$$\Rightarrow n_\beta \geq \gamma^2 \left(\frac{1}{\gamma\pi} + \frac{\pi}{R_0}\right). \tag{30}$$

Note that for almost all practical situations $\pi \geq R_0/\gamma$ will hold, since typical values of $R_0 \approx 10 - 15$ and $\gamma \approx 1800 - 3500$, see pp. 31 and 70 in [9].

In the same way we get a lower bound on n for $\sigma_{Y,\alpha}(\delta)$, which becomes

$$n_\alpha \geq \gamma^2 \left(\frac{1}{\gamma} + \frac{2R_0 + \frac{\gamma}{\pi}}{2R_0 + \gamma} \right). \tag{31}$$

Note that n_α is very sensitive to the choice of π.

Returning to the effect of type heterogeneities on $\tau_{Q,\cdot}(\delta)$, we have that if we are in the situation when $\mu_Y/\sigma_{Y,\cdot}(\delta) \geq 1$ holds for all $\delta \in [0,1]$, then an increase in $\sigma_{Y,\cdot}(\delta)$ leads to an increase in $q_{\bullet,1}(\delta)$ that, in turn, leads to a decrease in $\tau_{Q,\cdot}(\delta)$. Thus, we have a more formal statement supporting the heuristic arguments that was made when the coefficient of variation was introduced above.

To conclude, an increase of the difference in infectivity between the two types ought to decrease the expected time to extinction, where as it is a more complicated situation when the difference in susceptibility between the two types are changed, and non monotonic behaviour may occur.

3 Examples

In this section we compare the analytical approximations for $\tau_{Q,\cdot}(\delta)$ with simulations. We give some examples where we apart from varying δ, focus on varying π for different values of n when keeping R_0, μ and γ fix. When comparing the analytical approximations with simulations it is seen that they are not perfect, but that they capture the qualitative behaviour of the underlying epidemic.

All simulations have been performed using Monte Carlo simulation, and all the routines are written in the `C` programming language. The graphical presentation has been done using `Matlab`. The quasi-stationary behaviour of the epidemic has been approximated by simulating 1000 epidemics, and when the first 800 had gone extinct, we restarted the clock for the remaining 200 and kept them as our sample from the quasi-stationary distribution.

We have concentrated on the following parameter values: $R_0 = 10, \mu = 1$ (when $\mu = 1$ time is measured in units of life-lengths) and $\nu = 500$. The reason for choosing $\nu = 500$ instead of $\nu \approx 1800 - 3500$, which is typical for childhood diseases and which are the values we have been referring to in the previous sections, is of practical nature. That is, for the concept of quasi-stationarity to have any meaning, the expected time to extinction should be at least 5-10 years, which in turn corresponds to that we roughly have at least 10 infected individuals at the endemic level of infection, but when

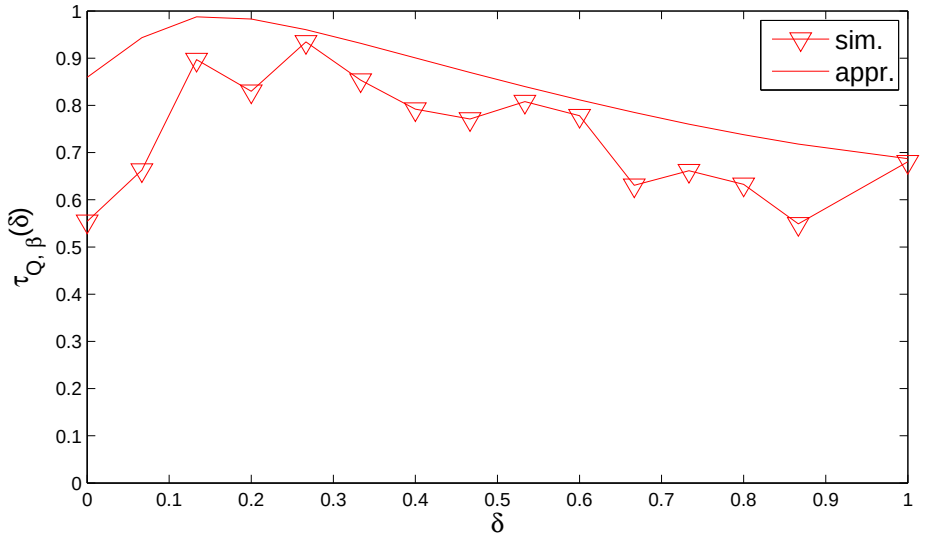

Figure 1: The expected time to extinction as a function of the difference in susceptibility between the two types ($\alpha_1 = \alpha_2 = 1$) when $R_0 = 10, \mu = 1, \nu = 500, n = 50,000$ and $\pi = 1/100$.

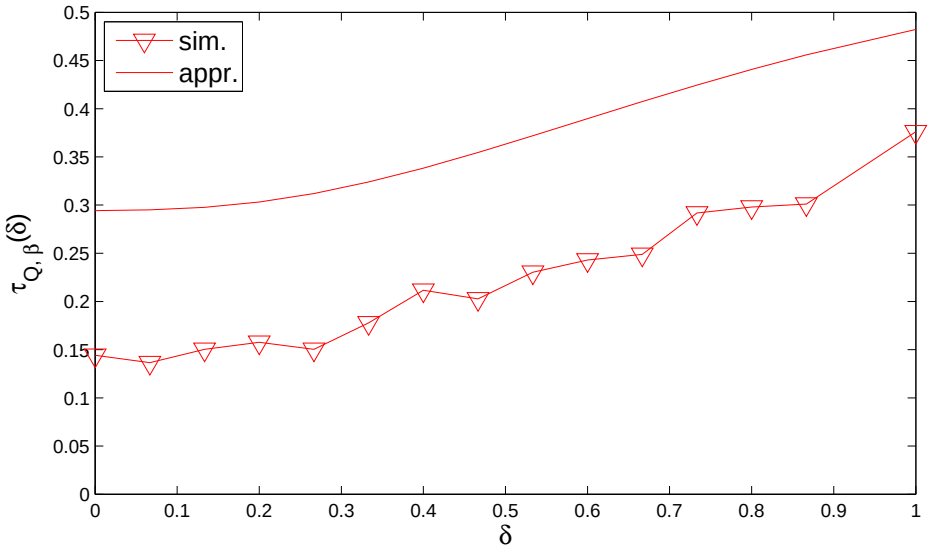

Figure 2: The expected time to extinction as a function of the difference in susceptibility between the two types ($\alpha_1 = \alpha_2 = 1$) when $R_0 = 10, \mu = 1, \nu = 500, n = 22,000$ and $\pi = 1/11$.

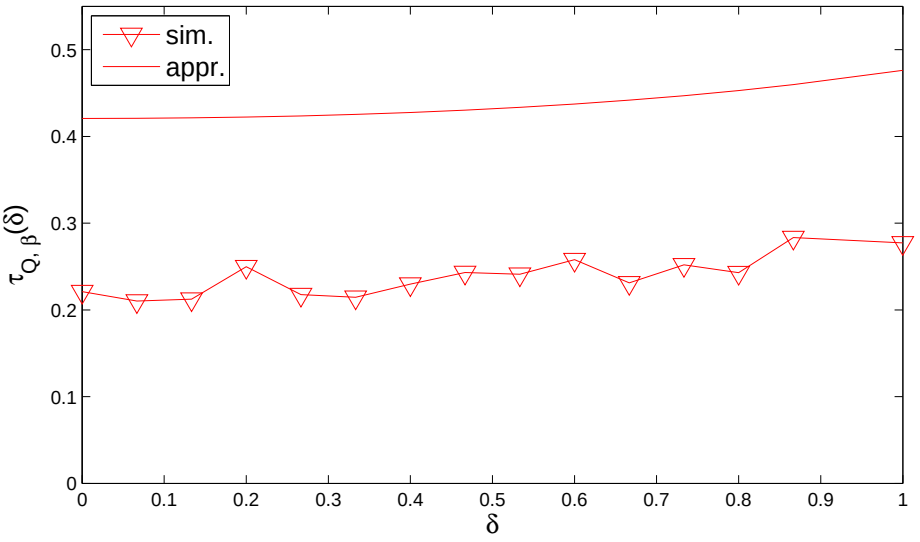

Figure 3: The expected time to extinction as a function of the difference in susceptibility between the two types ($\alpha_1 = \alpha_2 = 1$) when $R_0 = 10, \mu = 1, \nu = 500, n = 30,000$ and $\pi = 2/3$.

$\nu \approx 1800 - 3500$ we need to have a very large population or that $\pi \approx 1/2$, see Eq. (4), which is very time consuming to simulate. Thus, if we decrease ν we can still have a long expected time to extinction (the process is close to quasi-stationary) with a much smaller population, and the theory from the previous sections does not give any support to that $\nu = 500$ should not be regarded as an arbitrary parameter choice.

For the parameters chosen above $\tau_{Q,\beta}(\delta)$ gives a good description of the qualitative behaviour of the underlying epidemic already for small values of n (e.g. $n \approx n_\beta \approx 25,500$ where n_β is from (28)), see Figs. 1-4. From the previous section the analysis of $\sigma^2_{Y,\beta}(\delta)$ showed that it under certain conditions was not increasing/decreasing monotonically in δ which indicated a non monotonic behaviour of $\tau_{Q,\beta}(\delta)$. This behaviour can be seen from simulations, see e.g. Fig. 1, but note that $\pi = 1/100$ so it can not be regarded as a typical value of π. The behaviour of $\tau_{Q,\alpha}(\delta)$ is more sensitive to the choices of n, and it is especially important that n_2 is large. The reason for this is that the infectious pressure gets more and more shifted to the type two individuals as δ tends to α, so it becomes more and more important that the size of n_2 yields a sufficient number of infected type two individuals in order for the process to reach quasi-stationarity. Simulations indicate that if $n \approx n_\alpha$, where n_α is from (30), then $\tau_{Q,\alpha}(\delta)$ captures the qualitative

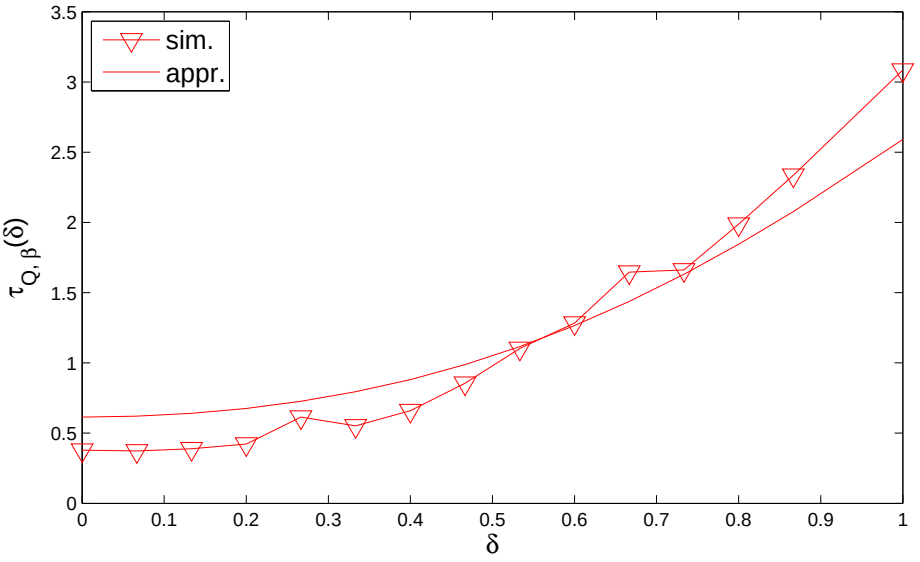

Figure 4: The expected time to extinction as a function of the difference in susceptibility between the two types ($\alpha_1 = \alpha_2 = 1$) when $R_0 = 10, \mu = 1, \nu = 500, n = 40,000$ and $\pi = 1/8$.

behaviour of the underlying population, but note that we need a much larger n, see Figs. 5.

To conclude, as long as $n \approx n_\alpha$, from (28) or (29), or $n \approx n_\beta$, from (30), depending on the situation, our approximations captures the qualitative behaviour of the epidemic.

4 Discussion

Bartlett's two-type epidemic model which has been analysed in the present paper is perhaps not the most realistic model, but we still believe that it captures some of the relevant behaviour which would appear in more complex models. A more realistic model which would be interesting to analyse in a two-type version is the so-called SIR model with demography which in its one-type version has been thoroughly analysed by Nåsell in [2]. A two-type version of this model is however much harder to analyse and obtain explicit expressions for. It would also be of interest to extend Bartlett's model to k types, which could be done using similar methodology as we have used in the present paper.

From a more general point it is always of interest to try and relax the

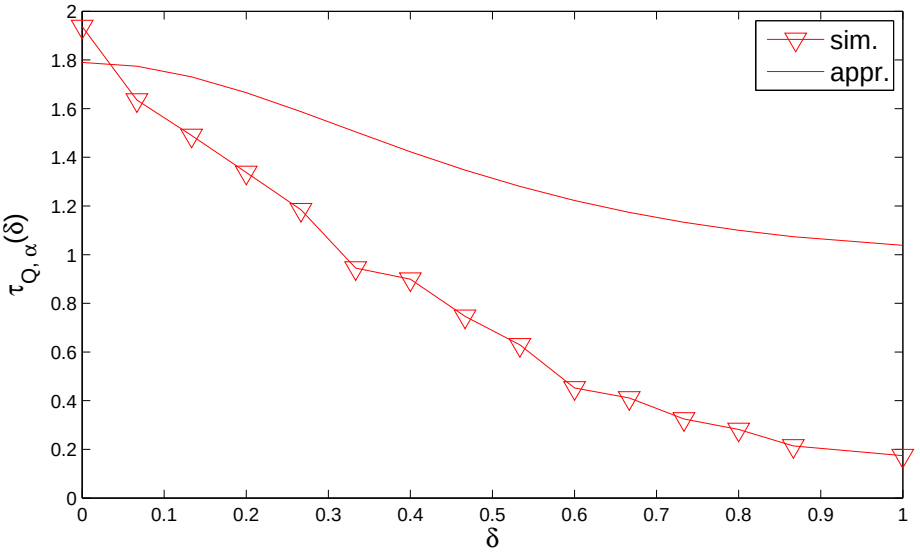

Figure 5: The expected time to extinction as a function of the difference in infectivity between the two types ($\beta_1 = \beta_2 = 1$) when $R_0 = 10, \mu = 1, \nu = 500, n = 75,000$ and $\pi = 1/3$.

assumption of exponentially distributed infectious periods, life-lengths and to include latency periods, see e.g. [3]. Another interesting extension could be to add some structure to the population, see e.g. [13] and [5], and it would also be of interest to analyse seasonal effects in a two-type setting, see e.g. [14] and [15].

5 Acknowledgements

The author would like to thank Tom Britton for good discussions and for comments on the manuscript. The author is grateful to Swedish Foundation for Strategic Research (SSF).

A Appendix

Here we derive the local drift and covariance matrices, $\mathbf{B}$ and $\mathbf{S}$ respectively, for $(\widetilde{\mathbf{X}}(t), \widetilde{\mathbf{X}}(t))$ from (7). To start, set $\widetilde{\mathbf{Z}}(t) = (\widetilde{X}_1(t), \widetilde{Y}_1(t), \widetilde{X}_2(t), \widetilde{Y}_2(t))^{\mathrm{T}}$ to simplify the notation, and let $\mathcal{F}_t$ denote the σ-algebra generated by the process $\widetilde{\mathbf{Z}}(t)$ up to time t. The infinitesimal first moment and covariance is

then defined as:

$$E_t\left[\Delta\widetilde{Z}_j\right] = \frac{1}{h}E\left[\widetilde{Z}_j(t+h) - \widetilde{Z}_j(t) \;\middle|\; \mathcal{F}_t\right] \tag{32}$$

and

$$Cov_t\left(\Delta\widetilde{Z}_i, \Delta\widetilde{Z}_j\right) = E_t\left[\Delta\widetilde{Z}_i\Delta\widetilde{Z}_j\right] - E_t\left[\Delta\widetilde{Z}_i\right]E_t\left[\Delta\widetilde{Z}_j\right] = E_t\left[\Delta\widetilde{Z}_i\Delta\widetilde{Z}_j\right] \tag{33}$$

where $i, j = 1, \ldots, 4$ and h is small. The matrices $\mathbf{B}$ and $\mathbf{S}$ are then defined as

$$\{\mathbf{B}\}_{ij} = \frac{\partial}{\partial\widetilde{Z}_j}E_t\left[\Delta\widetilde{Z}_i\right] \tag{34}$$

and

$$\{\mathbf{S}\}_{ij} = Cov_t\left(\Delta\widetilde{Z}_i, \Delta\widetilde{Z}_j\right) \tag{35}$$

where $i, j = 1, \ldots, 4$. For relief of notation we use the shorthand x_j and y_j for $x_j(t)$ and $y_j(t)$ respectively. For the two-type version of Bartlett's model treated here we get

$$\mathbf{B} = \begin{pmatrix} -\beta_1(\alpha_1 y_1 + \alpha_2 y_2) & -\beta_1\alpha_1 x_1 & 0 & -\beta_1\alpha_2 x_1 \\ \beta_1(\alpha_1 y_1 + \alpha_2 y_2) & \beta_1\alpha_1 x_1 - \mu\gamma & 0 & \beta_1\alpha_2 x_1 \\ 0 & -\beta_2\alpha_1 x_2 & -\beta_2(\alpha_1 y_1 + \alpha_2 y_2) & -\beta_2\alpha_2 x_2 \\ 0 & \beta_2\alpha_1 x_2 & \beta_2(\alpha_1 y_1 + \alpha_2 y_2) & \beta_2\alpha_2 x_2 - \mu\gamma \end{pmatrix} \tag{36}$$

and

$$\mathbf{S}_j = \begin{pmatrix} \mu\pi_j + \beta_j x_j(\alpha_1 y_1 + \alpha_2 y_2) & -\beta_j x_j(\alpha_1 y_1 + \alpha_2 y_2) \\ -\beta_j x_j(\alpha_1 y_1 + \alpha_2 y_2) & \beta_j x_j(\alpha_1 y_1 + \alpha_2 y_2) - \mu\gamma y_j \end{pmatrix},\ j = 1, 2 \tag{37}$$

such that

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{pmatrix} \tag{38}$$

Evaluating these matrices at the endemic level yields

$$\widehat{\mathbf{B}} = \begin{pmatrix} -\frac{\alpha_\pi\beta_1}{\gamma} & -\frac{\alpha_1\pi_1\mu\gamma}{\alpha_\pi} & 0 & -\frac{\alpha_2\pi_1\mu\gamma}{\alpha_\pi} \\ \frac{\alpha_\pi\beta_1}{\gamma} & \mu\gamma\left(\frac{\alpha_1\pi_1}{\alpha_\pi} - 1\right) & 0 & \frac{\alpha_2\pi_1\mu\gamma}{\alpha_\pi} \\ 0 & -\frac{\alpha_1\pi_2\mu\gamma}{\alpha_\pi} & -\frac{\alpha_\pi\beta_2}{\gamma} & -\frac{\alpha_2\pi_2\mu\gamma}{\alpha_\pi} \\ 0 & \frac{\alpha_1\pi_2\mu\gamma}{\alpha_\pi} & \frac{\alpha_\pi\beta_2}{\gamma} & \mu\gamma\left(\frac{\alpha_2\pi_2}{\alpha_\pi} - 1\right) \end{pmatrix} \tag{39}$$

and

$$\widehat{\mathbf{S}} = \mu \begin{pmatrix} 2\pi_1 & -\pi_1 & 0 & 0 \\ -\pi_1 & 2\pi_1 & 0 & 0 \\ 0 & 0 & 2\pi_2 & -\pi_2 \\ 0 & 0 & -\pi_2 & 2\pi_2 \end{pmatrix} \tag{40}$$

with which we can find a solution $\widehat{\mathbf{\Sigma}} = \{\hat{\sigma}_{ij}\},\ i, j = 1, \ldots, 4$.

B Appendix

Here we derive the local min/max points of $\sigma^2_{Y,\beta}(\delta)$ from Section 2. The local extreme points are found by locating points δ_0 such that $\frac{d}{d\delta}\sigma^2_{Y,\beta}(\delta_0) = 0$. From (16) we have that $\sigma^2_{Y,\beta}(\delta)$ can be written on the following form:

$$\sigma^2_{Y,\beta}(\delta) \sim \frac{f(\delta)}{\mu\gamma R_0 f(\delta) + g(\delta)}, \tag{41}$$

where $f(\cdot)$ and $g(\cdot)$ are two polynomials of order two and three respectively defined as

$$\begin{aligned} f(\delta) &= R_0(\mu\gamma)^2(2R_0 + \gamma)\pi^2 + 3(R_0\mu\gamma)^2(1-2\pi)\pi\delta + (R_0\mu\gamma)^2(1-2\pi)^2\delta^2 && (42) \\ &= a_1 + a_2\delta + a_3\delta^2 && (43) \\ g(\delta) &= R_0^3(\mu\gamma)^4(\gamma - 2R_0)\pi(1-\pi)\delta^2 - (R_0\mu\gamma)^3(1-\pi)(1-2\pi)\delta^3 \\ &= a_4\delta^2 + a_5\delta^3 && (44) \end{aligned}$$

Equating the first derivative w.r.t. δ of $\sigma^2_{Y,\beta}(\delta)$ to 0 gives us

$$\begin{aligned} &\frac{d}{d\delta}\sigma^2_{Y,\beta}(\delta) = f'(\delta)g(\delta) - f(\delta)g'(\delta) = 0 && (45) \\ \Rightarrow &(a_2 + 2a_3\delta)(a_4\delta^2 + a_5\delta^3) - (a_1 + a_2\delta + a_3\delta^2)(2a_4\delta + 3a_5\delta^2) = 0 && (46) \end{aligned}$$

from which it follows that $\delta_0 = 0$ is a root. Continuing, we can simplify further which gives us

$$a_3a_5\delta^3 + 2a_2a_5\delta^2 + 2a_1a_4 = 0 \tag{47}$$

and in order to solve this polynomial of order three we use the substitution $\delta = x - a_2/3a_1 = x - b_0$. After some further simplifications we will get a new polynomial:

$$x^3 + b_1x + b_2 = 0 \tag{48}$$

where

$$b_1 = \frac{1}{a_1}\left(a_3 - \frac{a_2^2}{3a_1}\right) \tag{49}$$

$$b_2 = \frac{1}{a_1}\left(a_4 + \frac{2a_2^3}{27a_1^2} - \frac{a_2 a_3}{3a_1}\right) \tag{50}$$

Introduce D defined as

$$D = \left(\frac{b_1}{3}\right)^3 + \left(\frac{b_2}{2}\right)^2 \tag{51}$$

By using Cardano's formula, see e.g. pp. 65 in [12], we get that if $D > 0$ there exist one real root, if $D = 0$ there exist three real roots where at least two are equal, and if $D < 0$ there exist three distinct real roots, and the roots of (47) are given by

$$\delta_1 = \sqrt[3]{-\frac{b_2}{2} + \sqrt[2]{D}} + \sqrt[3]{-\frac{b_2}{2} - \sqrt[2]{D}} - b_0 \tag{52}$$

$$\delta_{2,3} = -\frac{\sqrt[3]{-\frac{b_2}{2} + \sqrt[2]{D}} + \sqrt[3]{-\frac{b_2}{2} - \sqrt[2]{D}}}{2} - b_0 \pm \frac{\sqrt[3]{-\frac{b_2}{2} + \sqrt[2]{D}} + \sqrt[3]{-\frac{b_2}{2} - \sqrt[2]{D}}}{2} i\sqrt{3} \tag{53}$$

References

[1] H. Andersson, T. Britton, Stochastic Epidemic Models and Their Statistical Analysis, Lecture Notes in Statistics 151, Springer-Verlag, New York, 2000.

[2] I. Nåsell, On the time to extinction in recurrent epidemics, Journal of the Royal Statistical Society Series B 61 Part 2 (1999) 309-330.

[3] H. Andersson, T. Britton, Stochastic epidemics in dynamic populations: quasi-stationarity and extinction, Journal of Mathematical Biology 41 (2000) 559-580.

[4] I. Nåsell, Stochastic models of some endemic infections, Mathematical Biosciences, 179 (2002) 1-19.

[5] T.J. Hagenaars, C.A. Donnelly, N.M. Ferguson, Spatial heterogeneity and the persistence of infectious diseases, Journal of Theoretical Biology, 229 (2004) 349-359.

[6] M.S. Bartlett, Deterministic and stochastic models for recurrent epidemics, in: J. Neyman, (Ed.), Proceedings of the third Berkeley symposium on mathematical statistics and probability vol IV, University of California Press, 1956, pp. 81-109.

[7] M.S. Bartlett, Stochastic Population Models in Ecology & Epidemiology, London, Methuen, 1960.

[8] D. Clancy, P.D. O'Neill, P.K. Pollett, Approximations for the Long-Term Behaviour of an Open-Population Epidemic Model, Methodology and Computing in Applied Probability, 3 (2001) 75-95.

[9] R.M. Anderson, R.M. May, Infectious Diseases of Humans - Dynamics and Control, Oxford University Press, 1991.

[10] S.N. Ethier, T.G. Kurtz, Markov processes - Characterization and Convergence, Wiley, New York, 1986.

[11] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, 2 ed. , Graduate Texts in Mathematics 113, Springer-Verlag, New York, 1991.

[12] L. Råde, B. Westergren, Mathematics handbook for engineers and Science, 4 ed. , Studentlitteratur, Lund, 1998.

[13] J.S. Koopman, S.E. Chick, C.P. Simon, C.S. Riolo, G. Jacquez, Stochastic effects on endemic infection levels of disseminating versus local contacts, Mathematical Biosciences 180 (2002) 49-71.

[14] R. Kuske, L.F. Gordillo, P. Greenwood, Sustained oscillations via coherence resonance in SIR, Journal of Theoretical Biology (2006) doi:10.1016/j.jtbi.2006.10.029

[15] L. Stone, R. Olinky, A. Huppert, Seasonal dynamics of recurrent epidemics, Nature, 446 (2007) doi:10.1038/nature05638