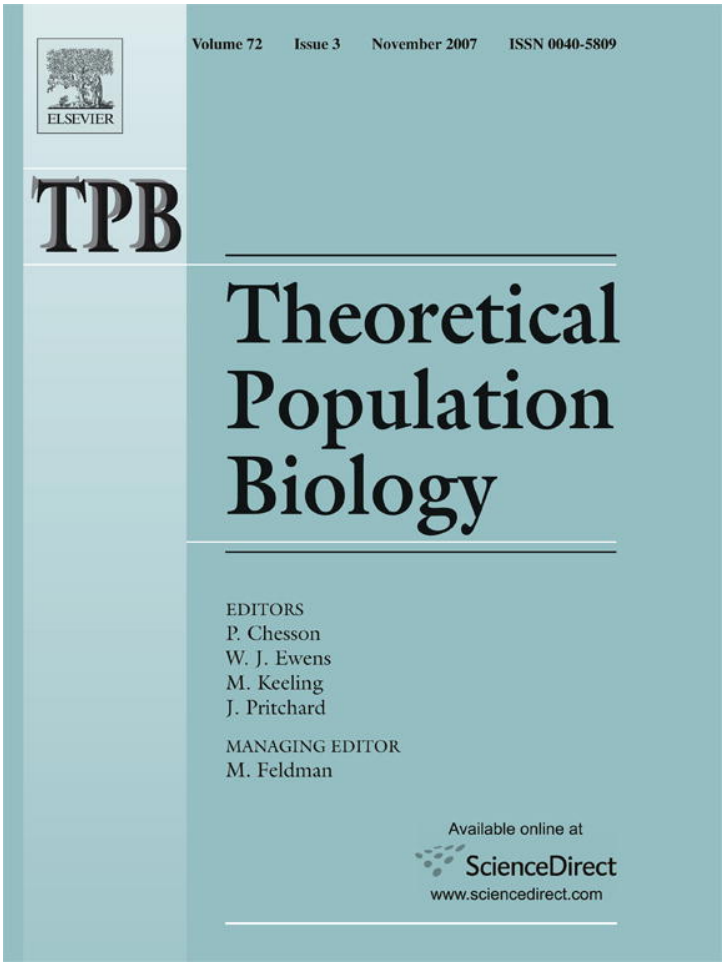

Available online at www.sciencedirect.com
ScienceDirect
Theoretical Population Biology 72 (2007) 389–399
Theoretical Population Biology
www.elsevier.com/locate/tpb

Modelling sexually transmitted infections: The effect of partnership activity and number of partners on R_0

Tom Britton[a,*], Monica K. Nordvik[b], Fredrik Liljeros[b]

[a]Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden
[b]Department of Sociology, Stockholm University, SE-106 91 Stockholm, Sweden

Received 14 September 2006
Available online 10 July 2007

Abstract

We model a sexually transmitted infection in a network population where individuals have different numbers of partners, separated into steady and casual partnerships, where the risk of transmission is higher in steady partnerships. An individual's number of partners of the two types defines its degree, and the degrees in the community specify the degree distribution. For this structured network population a simple model for disease transmission is defined and the basic reproduction number R_0 is derived, R_0 being a size-biased (i.e. biasing individuals with many partners) average number of new infections caused by individuals during the early stages of the epidemic. First a homosexual population is considered and then a heterosexual population. The heterosexual model is fitted to data from a census survey on sexual activity from the Swedish island of Gotland. The main empirical finding is that, for relevant transmission rates, the effect that so-called superspreaders have on R_0 is over-estimated when not admitting for different types of partnerships.

Keywords: Basic reproduction number; Branching process approximation; Network epidemic model; Sexually transmitted infections

1. Introduction

In the present paper we investigate, by means of mathematical modelling, the potential spread of a sexually transmitted infection (STI) in a community of interest, with particular focus on heterogeneities in terms of number of sex partners and sexual activity within partnerships.

Similar questions have been addressed elsewhere in the literature using various approaches. A complete survey is beyond the scope of the present paper, but here are some examples. Inspired by Hethcote and Yorke (1984), May and Anderson (1987) acknowledge variations in sexual activity by dividing the community into subgroups, where the subgroup an individual belongs to is defined by the average number of new sex partners per unit of time s/he has. By approximating the initial stages of the outbreak by a set of differential equations they derive an expression for the basic reproduction number, an expression containing the variance of the number of new sex partners. In Diekmann et al. (1991) the forming and separation of partnerships is modelled dynamically in time—partnerships break up and new partnerships are formed with specified intensities. This leads to a Markov-type model and the number of partners during the infectious period will typically follow the Poisson distribution (or a sum of Poisson variables when there is more than one disease state). The initial phase of the disease outbreak is approximated by a suitable branching process assuming a large population and an expression for the basic reproduction number R_0 is obtained. A similar model, but where variation in sexual activity within partnerships is acknowledged by distinguishing between "steady" and "casual" partnerships, is analysed by Kretzschmar et al. (1994) and R_0 is derived. A somewhat different approach is taken by Diekmann et al. (1998) where each individual has k fixed "acquaintances" chosen uniformly within the community, and contacts between each pair of acquaintances occur

*Corresponding author. Fax: +46 8 6126717.
E-mail address: tom.britton@math.su.se (T. Britton).

0040-5809/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2007.06.006

independently and randomly in time. For such a model they derive expressions for R_0 but also the final size of the epidemic in case a major outbreak occurs. Eames and Keeling (2004) study a model in which each individual has at most one active partnership among a set of fixed potential partnerships but these active partnerships may change over time. The number of sexual partners over a given period can thus be fitted to empirical data. The spread of the disease in the dynamic network of active partnerships is approximated by differential equations relying on a large population and the moment closure method. Newman (2002) studies a general model for infectious diseases where the focus is in modelling social relationships by means of static network models (an approach used also in the present paper). A model for a heterosexually transmitted disease is also presented and briefly analysed. Nordvik and Liljeros (2006) look at an STI-model where the transmission probability depends on the number of sex acts, fits the model to the data reanalysed in the present paper, and derives an expression for the expected number of new infections caused by a randomly selected individual in the community.

In the present paper we study a model which takes into account both individual variation of number of sex partners and variation in sexual activity between partnerships. Empirical evidence shows that the distribution of the number of sexual partners, the degree distribution, is usually heavy-tailed (Colgate et al., 1989; Liljeros et al., 2001). For this reason our model allows an arbitrary degree distribution. Further, in the model we incorporate two types of partnerships, "steady" and "casual" (cf. Kretzschmar et al., 1994), having distinct transmission probabilities. This distinction refers to the number of sex acts over a period corresponding to the length of the infectious period. We distinguish between steady and casual (or high and low transmission risk) partnerships for two reasons. First, the number of sex acts affects the probability of disease transmission in such a way that the more sex acts the higher over-all probability of disease transmission (e.g. Rottingen and Garnett, 2002). Secondly, individuals having many sex partners tend to have fewer sex acts per partner compared to individuals with one or few partners (e.g. Giesecke et al., 1992; Blower and Boe, 1993; Nordvik and Liljeros, 2006).

Assuming a large population we approximate the initial phase of the epidemic by a suitable multitype branching process and derive an expression for the basic reproduction number R_0 which determines whether a major outbreak is possible or not. We then neglect the fact that there are steady and casual partnerships, simply treating all partnerships identically, and derive R_0 under this assumption. By calibrating parameters in the two models we can compare R_0 under the different assumptions. These questions are first addressed for a homosexual community (Section 2) and thereafter for a heterosexual community in which the degree distributions as well as transmission probabilities may differ between sexes (Section 3).

We model the sexual partnerships in the community by a static random network. It would of course be more realistic to allow new partnerships to be formed and old to break up, as is done in some of the papers cited above. However, if in reality new partnerships are formed and old break up in a time-stationary way, then the fixed set of partners of a given individual in the present model can be interpreted as the partners of that individual during a time period corresponding to the typical length of the infectious period in a dynamic network model. With this interpretation the present model can approximate a time-dynamic network model.

In Section 4 the heterosexual model is applied to data on sexual activity collected in the island of Gotland, Sweden (Giesecke et al., 1992, recently reanalysed in Nordvik and Liljeros, 2006). The data are from a representative community sample and contain the number of sex partners of the individuals and also information about the sexual activity in each partnership. R_0 is then computed numerically in terms of transmission parameters for two distinctions between steady and casual partnerships, as well as for the case when not distinguishing between partnerships. A comparison is made between the three partnership definitions. The comparison is "fair" because the transmission probabilities of the different partnership definitions are calibrated by computing them as means, within the specified partnership, of a simple per-sex-act transmission model. The main empirical finding from the analysis is that, for relevant transmission probabilities, R_0 is higher (and hence over-estimated) when neglecting differences in partnerships as compared to the case when partnerships are separated into steady and casual. The paper is concluded with a discussion on limitations of the present work and important future problems.

2. An STI model for a homosexual population

Consider a homosexual community of size n where we assume n to be large.

2.1. Distinguishing between steady and casual partnerships

Assume that there are two types of partnerships, steady and casual partnerships, and that the probability of transmission between an infectious and a susceptible individual in a steady partnership is p_S and p_C in a casual partnership ($p_S \geqslant p_C$). (This model can also be used for non-STI diseases. The important feature is that there is a social network structure having two types of relationships with different transmission probabilities.)

Different individuals have different number of steady and casual partners. Let $\pi_{i,j}$ denote the proportion of individuals having i steady and j casual partners (during a time period corresponding to the typical length of the infectious period), and categorise an individual as an (i,j)-type accordingly. Assume further that steady (casual) partners are chosen randomly among the steady (casual)

partnerships available in the community. This implies that the probability that a steady partnership is with an (i,j)-individual (i.e. having i steady and j casual partners) is

$$\tilde{\pi}_{i,j,S} = \frac{i\pi_{i,j}}{\sum_{k,l} k\pi_{k,l}}.$$

Similarly, a casual partner is an (i,j)-individual with probability

$$\tilde{\pi}_{i,j,C} = \frac{j\pi_{i,j}}{\sum_{k,l} l\pi_{k,l}}.$$

These probabilities are said to be size-biased reflecting the fact that individuals with many steady (casual) partners are more likely to be selected as a steady (casual) partner.

The transmission of infection is modelled by simply assuming that an infectious individual infects his/her susceptible partners independently, and with probability p_S for steady partners and p_C for casual partners. An individual can only become infected (and infectious) once, after which s/he recovers and becomes immune.

For most stochastic epidemic models the early stages of the epidemic outbreak can be approximated by a suitable multitype branching process, and the approximation becomes better the larger n is (e.g. Andersson and Britton, 2000). This should be true also in the present model, although a formal proof requires a more detailed description/construction of the random sexual network (cf. Britton et al., 2006) and extensive model analysis (cf. Bollobás et al., 2007).

A multitype branching process (e.g. Jagers, 1975) studies the number of individuals of different types as time evolves, and the crucial assumption for branching processes is that individuals "give birth" (i.e. infects in our application) independently of each other—and the distribution of the number of individuals of different types an individual infects is type-specific.

During the early stages of an outbreak in the present model, all partners of a newly infected individual, except the partner by whom the individual was infected by, will be susceptible with large probability, and the chance that two individuals will contact the same individual is negligible. Individuals therefore infect approximately independently, justifying the branching process approximation.

The number of individuals an infectious will infect depends not only on the number of steady and casual partners s/he has, but also on whom s/he was infected by. For this reason, the types in the approximating branching process are specified not only by the number of partners the individual has but also by whom the individual was infected by. Consequently, we say that an individual is of type (i,j,S) if s/he has i steady partners, j casual partners and if s/he was infected by a steady partner (of course $i \geqslant 1$ for this to make sense). Similarly, an individual is of type (i,j,C) if s/he has the same partner structure but was infected by a casual partner.

In branching process theory the mean offspring matrix plays an important role. To this end, let $\lambda_{(i,j,S)(k,l,S)}$ denote the expected number of (k,l)-individuals an infectious (i,j)-individual, who him-/herself was infected by a steady partner, infects through its steady partnerships. Let $\lambda_{(i,j,S)(k,l,C)}$ be the same, but here the transmission is through a casual partner. The corresponding $\lambda_{(i,j,C)(k,l,S)}$ and $\lambda_{(i,j,C)(k,l,C)}$ are when the infectious (i,j)-individual was him-/herself infected by casual partner. These quantities can be derived from the model and shown to equal:

$$\begin{aligned}
\lambda_{(i,j,S)(k,l,S)} &= (i-1)\tilde{\pi}_{k,l,S}p_S,\\
\lambda_{(i,j,S)(k,l,C)} &= j\tilde{\pi}_{k,l,C}p_C,\\
\lambda_{(i,j,C)(k,l,S)} &= i\tilde{\pi}_{k,l,S}p_S,\\
\lambda_{(i,j,C)(k,l,C)} &= (j-1)\tilde{\pi}_{k,l,C}p_C.
\end{aligned}$$

For example, in the expression for $\lambda_{(i,j,S)(k,l,S)}$ the (i,j,S)-individual has i steady partners, but s/he was infected by one of them, so there are $i-1$ remaining susceptible steady partners, and each of them is with a (k,l)-individual with probability $\tilde{\pi}_{k,l,S}$, and any such partnership results in infection with probability p_S. The matrix Λ having these λ's as elements for the different $(i,j,X)(k,l,Y)$ combinations present in the community, the first triplet specifying the row and the second the column, defines the mean offspring matrix. The largest eigenvalue ρ of Λ is an important parameter. In particular, depending on whether ρ is smaller than, equal to, or larger than 1, specifies whether the branching process is sub-critical, critical or super-critical, and only super-critical branching processes have positive probability to grow beyond all limits.

When applied to epidemic modelling the notation ρ is often replaced by R_0 (or R) and denoted the basic reproduction number. In epidemic terminology we have the result that a major outbreak (infecting a positive fraction) can occur if and only if $R_0 > 1$, where R_0 is, as previously mentioned, the largest eigenvalue to the matrix Λ having elements $\lambda_{(i,j,X)(k,l,Y)}$ specified above.

For the general case it is hard to say any qualitative results concerning R_0 as a function of model parameters. However, if the community structure, i.e. degree distributions, is kept fixed and p_S and p_C are varied such that $p_S/p_C = \alpha$ is kept fixed, then R_0 increases linearly with p_S and $p_C = p_S/\alpha$ since all components of Λ then are linear p_S.

2.2. No distinction between steady and casual partnerships

If we neglect that there are different types of partnerships having different transmission probabilities, then the corresponding eigenvalue is easier to compute as we now illustrate. To this end, let p denote the probability of transmission in a partnership where one partner is infectious and the other susceptible, and let $\{\pi_k\}$ denote the population distribution of number of partners (within the time horizon corresponding to the length of an infectious period). Further, let K denote a random variable having distribution $\{\pi_k\}$ (i.e. the number of partners of a randomly selected individual), and let $\tilde{K}$ denote a random variable with distribution $\{\tilde{\pi}_k\}$, where $\tilde{\pi}_k = k\pi_k/\sum_i i\pi_i$ (the

size-biased distribution reflecting that individuals with many partners are more likely to be selected). It is worth pointing out that this model does not coincide with the special case of the previous model having $p_S = p_C = p$. The reason for this is that even when the transmission probabilities are identical partners are selected differently in the two models.

This model has been studied previously (Andersson, 1999) where it was shown that the basic reproduction number equals

$$R_0 = pE(\tilde{K}-1) = p\left(E(K) + \frac{V(K)-E(K)}{E(K)}\right) = p\left(\frac{E(K^2)}{E(K)} - 1\right), \quad (1)$$

where $E(\cdot)$ and $V(\cdot)$ denote expected value and variance, respectively. When each individual has exactly k acquaintances (so $K \equiv k$) we have $R_0 = p(k-1)$ which agrees with Diekmann et al. (1998) treating this particular model. In Eq. (1) it is seen that for fixed average number of partners $E(K)$ and fixed transmission probability p, R_0 increases with the variance of the number of partners $V(K)$, and if the variance is infinite, so is R_0 (Pastor-Satorras and Vespignani, 2002). In a finite community the variance can of course not be infinite but empirical evidence show that the number of partners (the degree distribution) is heavy tailed (Colgate et al., 1989; Liljeros et al., 2001). The fact that R_0 increases with the variance of the number of partners in a dynamic partnership model was already discovered by May and Anderson (1987).

In order to compare this new R_0 (neglecting differences in partnerships) with the previous where we acknowledge steady and casual partnerships, the partnership distributions $\{\pi_{ij}\}$ vs. $\{\pi_k\}$, as well as the transmission parameters p_L and p_S vs. p must be calibrated. Clearly, the calibration of partnership distributions should be $\pi_k = \sum_i \pi_{i,k-i}$, so the number of partners equals the number of casual plus the number of steady partners. The most natural calibration for the transmission parameters is to assume that the new p is a weighted average of p_L and p_S, where the relative weights correspond to the population proportions of the two contacts, i.e. to let p satisfy

$$p = \frac{p_S \sum_{i,j} i\pi_{i,j} + p_C \sum_{i,j} j\pi_{i,j}}{\sum_{i,j}(i+j)\pi_{i,j}}.$$

This choice of $\{\pi_k\}$ and p makes the two models as similar as possible thus justifying a comparison of R_0. It seems hard to make any general conclusions when comparing R_0 for the two calibrated models. We have found simple examples for which the model acknowledging steady and casual partnerships gives a larger R_0 (half of the community has 3 steady and 3 casual partners and the other half has 1 steady and 1 casual partners) and simple examples going in the opposite direction (half of the community has 1 steady and 3 casual partners and the other half has 3 steady and 1 casual partner).

3. An STI model for a heterosexual population

Consider now a population consisting of males and females and assume that all sexual contacts are heterosexual.

3.1. Distinguishing between steady and casual partnerships

We still assume that there are two types of partnerships, steady and casual, but the probability of transmission also depends on who is infectious and who is susceptible: $p_S^{(mf)}$ is the probability of transmission in a steady partnership in which the male is infectious and the female is susceptible, and similarly for $p_S^{(fm)}$, $p_C^{(mf)}$ and $p_C^{(fm)}$. Further, let $\pi_{ij}^{(m)}$ denote the community proportion of males having i steady and j casual partners (during a period of equal length as a typical infectious period), and let $\pi_{i,j}^{(f)}$ denote the corresponding proportion for females. In order to the total number of steady and casual partnerships of males and females to be identical (assuming there are equally many males and females) these proportions should satisfy

$$\sum_{i,j} i\pi_{i,j}^{(m)} = \sum_{i,j} i\pi_{i,j}^{(f)} \quad \text{and} \quad \sum_{i,j} j\pi_{i,j}^{(m)} = \sum_{i,j} j\pi_{i,j}^{(f)}.$$

Similar to before, a female has an (i,j)-male as a steady partner with probability $\tilde{\pi}_{i,j,S}^{(m)}$, the corresponding probability for a casual partner equals $\tilde{\pi}_{i,j,C}^{(m)}$, and males have corresponding partnerships with specified female types with probabilities $\tilde{\pi}_{i,j,S}^{(f)}$ and $\tilde{\pi}_{i,j,C}^{(f)}$, respectively, where

$$\tilde{\pi}_{i,j,S}^{(m)} = \frac{i\pi_{i,j}^{(m)}}{\sum_{k,l} k\pi_{k,l}^{(m)}}, \quad \tilde{\pi}_{i,j,C}^{(m)} = \frac{j\pi_{i,j}^{(m)}}{\sum_{k,l} l\pi_{k,l}^{(m)}},$$
$$\tilde{\pi}_{i,j,S}^{(f)} = \frac{i\pi_{i,j}^{(f)}}{\sum_{k,l} k\pi_{k,l}^{(f)}}, \quad \tilde{\pi}_{i,j,C}^{(f)} = \frac{j\pi_{i,j}^{(f)}}{\sum_{k,l} l\pi_{k,l}^{(f)}}.$$

Consider an infectious (i,j)-female who was infected through a steady partnership. Let $\lambda_{(i,j,S)(k,l,S)}^{(f)}$ denote the expected number of males, having (k,l)-partners that she infects through a steady partnership. Similarly, $\lambda_{(i,j,S)(k,l,S)}^{(m)}$ is the corresponding expected number, but for a male infecting females. The suffix hence indicates the sex of the infector, the first three indices show how many partners this type has and through which type of partnership s/he was infected, and the last three indices specifies the type to be infected: his/her number of partners and through which type of partnership s/he gets infected. It follows that

$$\lambda_{(i,j,S)(k,l,S)}^{(f)} = (i-1)\tilde{\pi}_{k,l,S}^{(m)} p_S^{(fm)},$$
$$\lambda_{(i,j,S)(k,l,C)}^{(f)} = j\tilde{\pi}_{k,l,C}^{(m)} p_C^{(fm)},$$
$$\lambda_{(i,j,C)(k,l,S)}^{(f)} = i\tilde{\pi}_{k,l,S}^{(m)} p_S^{(fm)},$$
$$\lambda_{(i,j,C)(k,l,C)}^{(f)} = (j-1)\tilde{\pi}_{k,l,C}^{(m)} p_C^{(fm)},$$
$$\lambda_{(i,j,S)(k,l,S)}^{(m)} = (i-1)\tilde{\pi}_{k,l,S}^{(f)} p_S^{(mf)},$$
$$\lambda_{(i,j,S)(k,l,C)}^{(m)} = j\tilde{\pi}_{k,l,C}^{(f)} p_C^{(mf)},$$

$$\lambda^{(m)}_{(i,j,C)(k,l,S)} = i\tilde{\pi}^{(f)}_{k,l,S}p^{(mf)}_S,$$
$$\lambda^{(m)}_{(i,j,C)(k,l,C)} = (j-1)\tilde{\pi}^{(f)}_{k,l,C}p^{(mf)}_C.$$

In the branching process approximation of the epidemic starting with, say, a man, the next generation will consist of females, the generation thereafter of males, etc. This implies that the offspring distribution between generations typically differ every other generation. We therefore look at the offspring distribution two generations later. Suppose we start with a male of type (i,j,S), which hence had i steady and j casual partners and was infected by a steady partner. Let $\lambda^{(2m)}_{(i,j,S),(k,l,S)}$ denote the expected number of infected males of type (k,l,S) infected two generations later that are caused by our (i,j,S)-male. Then, by conditioning and summing over possible female routes, $\lambda^{(2m)}_{(i,j,S),(k,l,S)}$ satisfies

$$\lambda^{(2m)}_{(i,j,S),(k,l,S)} = \sum_{r,s}(\lambda^{(m)}_{(i,j,S)(r,s,S)}\lambda^{(f)}_{(r,s,S)(k,l,S)} + \lambda^{(m)}_{(i,j,S)(r,s,C)}\lambda^{(f)}_{(r,s,C)(k,l,S)}),$$

and similarly

$$\lambda^{(2m)}_{(i,j,S),(k,l,C)} = \sum_{r,s}(\lambda^{(m)}_{(i,j,S)(r,s,S)}\lambda^{(f)}_{(r,s,S)(k,l,C)} + \lambda^{(m)}_{(i,j,S)(r,s,C)}\lambda^{(f)}_{(r,s,C)(k,l,C)}),$$

$$\lambda^{(2m)}_{(i,j,C),(k,l,S)} = \sum_{r,s}(\lambda^{(m)}_{(i,j,C)(r,s,S)}\lambda^{(f)}_{(r,s,S)(k,l,S)} + \lambda^{(m)}_{(i,j,C)(r,s,C)}\lambda^{(f)}_{(r,s,C)(k,l,S)}),$$

$$\lambda^{(2m)}_{(i,j,C),(k,l,C)} = \sum_{r,s}(\lambda^{(m)}_{(i,j,C)(r,s,S)}\lambda^{(f)}_{(r,s,S)(k,l,C)} + \lambda^{(m)}_{(i,j,C)(r,s,C)}\lambda^{(f)}_{(r,s,C)(k,l,C)}).$$

If the (i,j,X)-indices are ordered in some way and we let $\Lambda^{(2m)}$ denote the matrix containing the elements above, ordered accordingly, then $\Lambda^{(2m)}$ describes the mean offspring matrix after two generations, starting with an infectious male. The largest eigenvalue to this matrix specifies the growth rate after two generations, and depending on whether it is larger or smaller than 1, determines if a major outbreak is possible or not. If we take the square root we get the more natural one-generation correspondence.

The basic reproduction number R_0 is hence the square root to the largest eigenvalue of $\Lambda^{(2m)}$ having elements defined above. If we did the same reasoning above, only starting with an infectious female, and derived the square root to the largest eigenvalue to that matrix, we would get the same value, so one can choose either one. As for the case of a homosexual community it is hard to state any general conclusions about R_0 for the heterosexual community model. In Section 4 the partnership network is fitted to real data and R_0 is calculated for some choices of transmission parameters.

3.2. No distinction between steady and casual partnerships

If we neglect that there are different types of partnerships having different transmission probabilities, then the corresponding eigenvalue is easier to compute.

To this end, let $p^{(mf)}$ denote the probability of transmission in a partnership where the male is infectious and the female is susceptible, and $p^{(fm)}$ is the opposite transmission probability. Further, let $\pi^{(m)}_k$ denote the proportion males having k partners (during a time period of length as a typical infectious period), and let $\pi^{(f)}_k$ be the corresponding for females. Similar to before, the probability that a female has contact with a male having k partners, equals

$$\tilde{\pi}^{(m)}_k = \frac{k\pi^{(m)}_k}{\sum_j j\pi^{(m)}_j}$$

and

$$\tilde{\pi}^{(f)}_k = \frac{k\pi^{(f)}_k}{\sum_j j\pi^{(f)}_j}$$

is the probability that a male contact is with a female having k partners. Let $\tilde{K}^{(f)}$ and $\tilde{K}^{(m)}$, respectively, denote random variables having these distributions. Similar to the homosexual model, the present model does not coincide with the special case of the model in Section 3.1 having $p^{(fm)}_S = p^{(fm)}_C = p^{(fm)}$ and $p^{(mf)}_S = p^{(mf)}_C = p^{(mf)}$ due to the difference in partnership selection between the two models.

As before, the early stages of the epidemic, assuming a large community, is approximated by a multitype branching process. Using similar arguments as in (1) it can be shown that the expected number of males a typical female infects during the early stages of an outbreak equals

$$p^{(fm)}E(\tilde{K}^{(f)}-1) = p^{(fm)}\left(E(K^{(f)}) + \frac{V(K^{(f)})-E(K^{(f)})}{E(K^{(f)})}\right).$$

Similarly, a typical male on average infects

$$p^{(mf)}E(\tilde{K}^{(m)}-1) = p^{(mf)}\left(E(K^{(m)}) + \frac{V(K^{(m)})-E(K^{(m)})}{E(K^{(m)})}\right)$$

females. This implies that the average increase typically varies between odd and even generations. It is therefore natural to define the basic reproduction number as the square root of their product:

$$R_0 = \sqrt{p^{(fm)}\left(E(K^{(f)}) + \frac{V(K^{(f)})-E(K^{(f)})}{E(K^{(f)})}\right)p^{(mf)}\left(E(K^{(m)}) + \frac{V(K^{(m)})-E(K^{(m)})}{E(K^{(m)})}\right)}.$$

In order to be able to compare this R_0 (neglecting different transmission probabilities for steady and casual partnerships) with the one obtained admitting such

differences, the transmission parameters as well as population frequencies of number of partners have to be calibrated. For the number of partners the natural calibration is clearly

$$\pi_k^{(f)} = \sum_{i=0}^{k} \pi_{i,k-i}^{(f)} \quad \text{and} \quad \pi_k^{(m)} = \sum_{i=0}^{k} \pi_{i,k-i}^{(m)}.$$

For the comparison to be fair the transmission parameters should be a weighted average of the steady and casual transmission probabilities, where the weights take into account how frequent the different types of partnerships are. More precisely, they should satisfy

$$p^{(mf)} = \frac{p_S^{(mf)} \sum_{i,j} i\pi_{i,j}^{(f)} + p_C^{(mf)} \sum_{i,j} j\pi_{i,j}^{(f)}}{\sum_{i,j}(i+j)\pi_{i,j}^{(f)}},$$

$$p^{(fm)} = \frac{p_S^{(fm)} \sum_{i,j} i\pi_{i,j}^{(m)} + p_C^{(fm)} \sum_{i,j} j\pi_{i,j}^{(m)}}{\sum_{i,j}(i+j)\pi_{i,j}^{(m)}}.$$

With these definitions, the model ignoring different transmission probabilities for steady and casual partnerships can both underestimate and overestimate R_0. In the next section the two models are fitted to data and computed.

4. Application to a study on sexual partnerships on the island of Gotland, Sweden

The calculations in this study are based on data from a study on sexual behaviour which was conducted in 1988 on Gotland, a Swedish island in the Baltic Sea (Giesecke et al., 1992).

4.1. The data

A random sample of 10% of the individuals aged 16–31 was drawn, and the response rate was 68% (775 individuals, 426 women and 349 men). It was concluded that the material was free from systematic biases (Giesecke et al., 1992).

The respondents were asked for the number of sex partners during the last year and for each such partner, the number of sex acts they had with him/her during the last year (and also when and for how long the partnership continued), thus making the study quite unique in detail. Individuals reporting no sexual activity (22% of the women and 31% of the men) will not contribute to the spread and are left out in what follows. All remaining individuals were used as the data to be analysed. In this data 332(= 58%) were women and 241(= 42%) men. In Fig. 1 the distribution of the number of partners during the last year are shown for men and women, respectively.

The mean number of sexual partners reported were 1.5 for women and 1.7 for men. In Fig. 2 the average *total* number of sex acts during a year is shown, where we have distinguished between men and women, and also according to how many partners the individual had during the year.

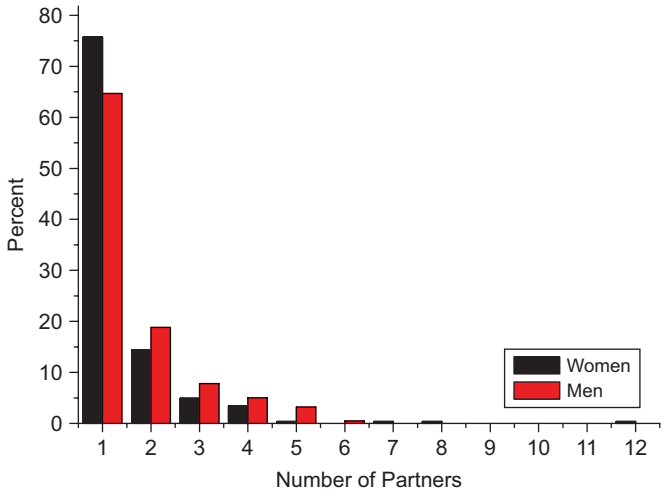

Fig. 1. Distribution of the number of sex partners during one year for men and women in the Gotland data (from Nordvik and Liljeros, 2006).

It is seen that the average total number of sex acts decreases with the number of partners for women, and at least not increases for men. This clearly implies that the average number of sex acts *per partner* decreases with the number of partners (as also observed in Blower and Boe, 1993). As a consequence, individuals having many sex partners can potentially spread the disease to more individuals, but on the other hand, the chance of spreading the disease to a specific partner is lower because of fewer sex acts.

4.2. Estimation of R_0 from the Gotland data

The data presented in the previous section contain no information about any specific STI. Instead we use the data to fit the heterosexual network described in Section 3.1, where the degree of an individual corresponds to the number of sex partners s/he has during one year. As we have no specific STI in mind we use this length for the infectious period. But, if a specific disease was in mind with e.g. shorter infectious period, then the degrees of individuals should reflect the number of partners (and sex acts) for such a period hence resulting in a network having fewer partnerships. Given the resulting network, we estimate R_0 for the community in case an STI, with specific transmission parameters, enters the community.

The "true" probability that an infectious individual infects a susceptible individual in a sexual relationship depends on: the gender of the infector and of the susceptible, the number of sex acts during the infectious period, the infectivity of the disease, the type of sexual activity, individual heterogeneities in terms of susceptibility and infectivity, and other things. To obtain detailed information about all this is, however, hard. In the present data material we have information about the (approximate) number of sex acts in each partnership. We can thus define an infection probability which, beside depending on

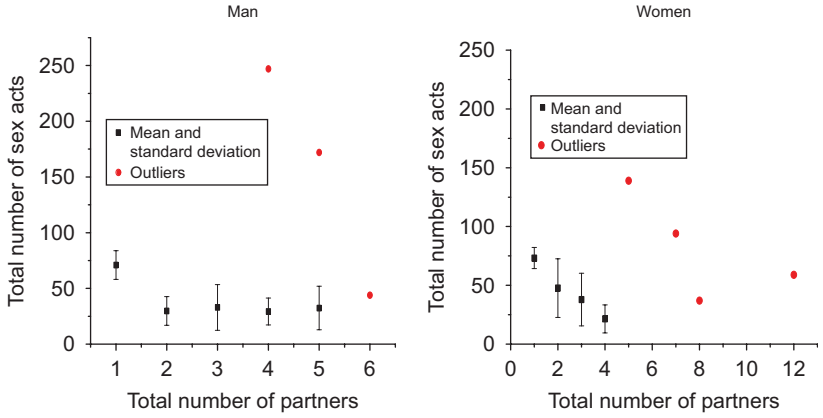

Fig. 2. Average total number of sex acts per year, separated for men and women, and on the number of sex partners (from Nordvik and Liljeros, 2006).

the gender of the infectious individual, varies with the number of sex acts. A natural model is to assume that each sex act, independently and with equal probability, results in transmission to the susceptible individual (see Rottingen and Garnett, 2002, for a further discussion on related models). The probability that a female susceptible escapes infection from an infectious male partner after k sex acts, and with per sex act infection probability p_{mf}, then equals

$$(1-p_{mf})^k,$$

and the probability that she becomes infected is $1-(1-p_{mf})^k$. Similarly, the probability that an infectious female infects her susceptible male partner after k sex acts equals $1-(1-p_{fm})^k$, where p_{fm} is the per-sex-act probability that a female transmits the disease to her male susceptible partner. In the analysis below where R_0 is estimated, we will treat these per-sex-act transmission probabilities as the parameters being varied keeping everything else fixed. We treat two specific scenarios: $p_{fm}=p_{mf}$ and $p_{fm}=0.5p_{mf}$.

In the model in Section 3 we did not allow the transmission probability in a partnership to explicitly depend on the number of sex acts. The reason why not is that such a model would make the number of different types of partnerships enormous and the number of individuals in each type of partnership very small, thus introducing over-parametrisation. Instead we only consider two different types of partnership, steady and casual, where steady and casual refer to the number of sex acts rather than duration of the partnership, the number of sex acts being more relevant for disease transmission. (One could of course distinguish between 3 and 4 types of partnership, but even then the level of complexity for deriving R_0 increases drastically, and the gain in precision is probably minor.) Using our more detailed data we can distinguish between steady and casual partnerships in many ways. In our analysis we have performed two such separations. The first is where a casual partnership is defined by a single sex act and everything more than one sex act is considered a steady partnership. With this definition $p_C^{(mf)}=1-(1-p_{mf})^1=p_{mf}$ (and similarly $p_C^{(fm)}=p_{fm}$) since a casual partnership always has one single sex act. In order to obtain expressions for $p_S^{(mf)}$ as a function of p_{mf} we have, for each partnership in which a male has more than one sex act, computed $1-(1-p_{mf})^{\#\text{sex acts}}$. The transmission probability $p_S^{(mf)}$ is then set to equal the arithmetic mean of these quantities (the geometric mean could have been used instead but the difference was minor). The parameter $p_S^{(fm)}$ is computed similarly, only replacing p_{mf} by p_{fm} and considering averages over females in different partnerships.

We have also treated a different separation in which anything less than 10 sex acts is considered a casual partnership and 10 or more sex acts is defined as a steady partnership. In Fig. 3 we show a histogram of the number of sex acts per partnership (remember that the questionnaire was for one calender year, so partnership durations are limited to 1 year).

It is of course arguable what cut-off to use to distinguish between steady and casual partnerships. A motivation to use 1 is that this is clearly the smallest possible cut-off and also because this is by far the most common number of sex acts in a partnership (see magnification in Fig. 3). The other cut-off, 10, was chosen by the motivation that after 10 the frequency really seemed to drop (see magnified histogram). We could of course have used 5 instead of 10 as our alternative cut-off but we suspect that our conclusions would not be very different.

The estimation of $p_S^{(mf)}$ and $p_C^{(mf)}$ as functions of p_{mf} (and $p_S^{(fm)}$ and $p_C^{(fm)}$ as functions of p_{fm}) is done by taking similar arithmetic means, only now over the new separations of steady and casual partnerships. In Fig. 4 we show plots of the resulting $p_S^{(mf)}$ (left) and $p_C^{(mf)}$ (right) as functions of p_{mf} (plots of $p_S^{(fm)}$ and $p_C^{(fm)}$ look similar but are not shown).

There it is seen that $p_C^{(mf)}=p_{mf}$ (i.e. a straight line) for the situation where casual partnership means one sex act. In the left figure the case where all partnerships are treated as equal is also plotted (with the interpretation that no partnerships are casual). It is seen that the transmission probability $p_S^{(fm)}$ quickly comes close to 1, in

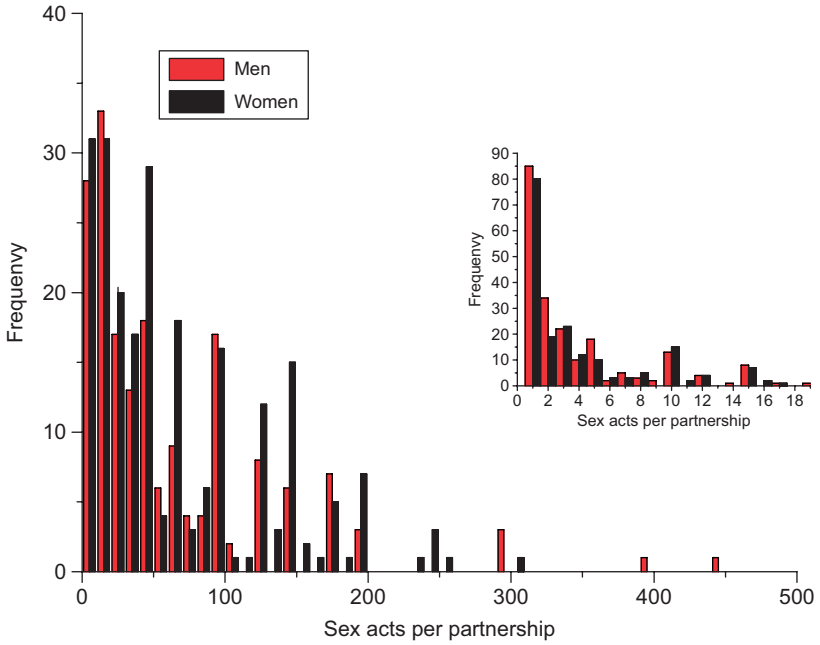

Fig. 3. Histogram of the number of sex acts per partnership (limited to a year duration). The inserted histogram is a magnified histogram for few sex acts.

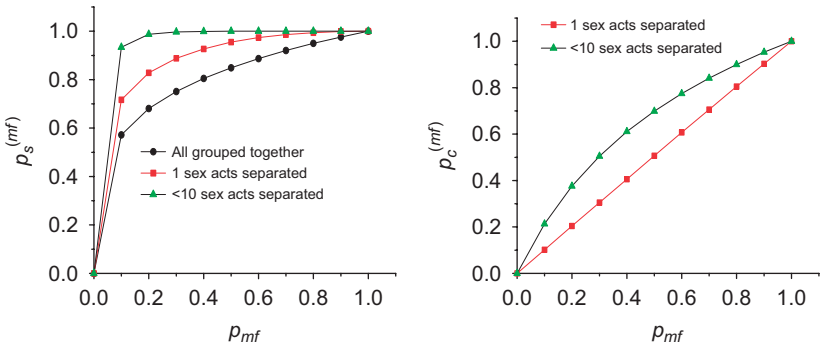

Fig. 4. $P_S^{(mf)}$ and $P_C^{(mf)}$ as a function of p_{mf} for the data and different separations between steady and casual partnerships.

particular, for the case where steady partnership means more than 10 sex acts.

Given these separations between steady and casual partnerships it is possible to compute R_0. In Fig. 5 we have, for a community estimated from the Gotland data, plotted R_0 as a function of the per-sex-act male-to-female transmission probability p_{mf} under the assumption that the disease transmission probability is the same from female to male (i.e. $p_{fm} = p_{mf}$), and that the duration of a typical infectious period is 1 year. The curve with squares is for the separation with one sex act being a casual partnership and the curve with triangles corresponds to the separation with a casual partnership defined as less than 10 sex acts. Also in the figure is a curve with diamonds corresponding to the case where there is no distinction between casual and steady partnerships (corresponding to the model of Section 3.2). For this curve, $p^{(mf)}$ and $p^{(fm)}$ are estimated similarly to earlier, only now the averages are taken over *all* partnerships.

In Fig. 5 it is seen that when the per-sex-act transmission probability is smaller than ≈ 0.4, as is the case for many STI's, the effect of *not* distinguishing between casual and steady partnerships is that R_0 is systematically over-estimated. (For HIV one estimate for p_{mf} is 0.001 (Leynaert et al., 1998), for Chlamydia p_{mf} has been estimated to 0.2 (Lycke et al., 1980) and 0.35 (Katz et al., 1990). Almost always it is believed that $p_{fm} < p_{mf}$.) By "systematically" we mean that the difference would remain in much larger communities and is thus not explained by uncertainty. An intuitive explanation to the observed systematic over-estimation, also taking the contact behaviour in the community into account, is the following. When all partnerships are assumed to have equal transmission probability, this transmission probability is over-estimated for partnerships having few sex acts. In the Gotland data it was seen that partnerships with few sex acts tend to be between individuals having many partners

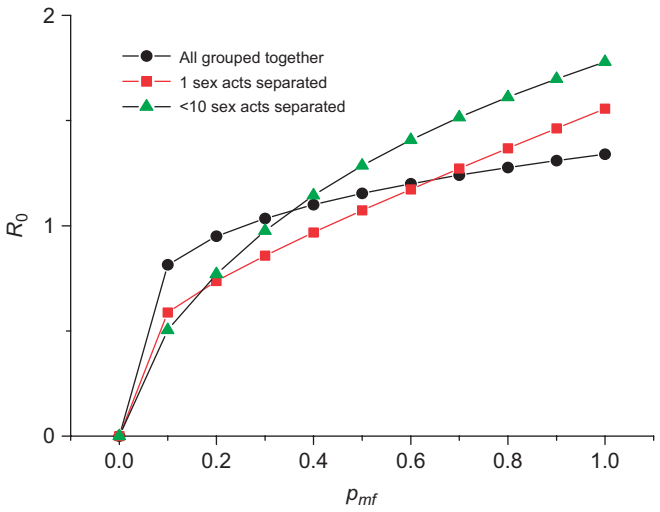

Fig. 5. Plot of R_0 for the Gotland community as a function of the male-to-female per contact probability p_{mf} and under the assumption that $p_{fm} = p_{mf}$. The three curves correspond to the situations where a casual partnership is defined as being one sex act, being 10 or less sex acts, and to the case where steady and casual partnerships are not distinguished.

as would be expected. And, when all partnerships have the same transmission probability, such individuals influence R_0 more than individuals with few partners do because of the size-biasing effect mentioned earlier. As a consequence, the effect of the over-estimation of transmission probabilities for individuals with many partners is not compensated by the corresponding under-estimation for individuals with few partners, with the effect that R_0 is over-estimated when all partnerships are treated equally. We suspect that the difference in R_0 between treating all partnerships as identical compared to when admitting different types of partnerships would be even bigger if more than two types of partnerships would have been considered.

A relevant question is why the observed over-estimation only holds when the per-sex-act transmission is low or moderate (in Fig. 5 it no longer holds when p_{mf} exceeds about 0.5). A possible explanation to this, mentioned in Section 3.2, is that the model with two types of partnerships for the special case where both type of partnerships have equal transmission probability does *not* coincide with the model having one type of partnership. A simple example illustrates this. Suppose that half of the community has 1 steady and 0 casual partnership and the other half of the community has 0 steady and 3 casual partners (males and females are assumed to behave equivalently). Then, all individuals with one partner will have partners also having only one partnership, and individuals with three partners will all have partners with three partnerships (note that this is not the case if only one type of partnership is considered and 50% have one partner and 50% have three partners). If the transmission probability is not very different between steady and casual partnerships (which is necessarily the case when the per-sex-act probability is large!) then this so-called assortative mixing will increase R_0 more than the previously mentioned effect will decrease it (that R_0 increases in this case follows because a core subgroup will dominate the early stages in an outbreak). As an effect, R_0 will be higher with two types of partnerships whenever the per-sex-act transmission probability is large enough. Of course, the example is extreme but since individuals with a steady partnership tend to have fewer casual partnerships than those without steady partners, the explanation should be valid in real communities.

Another observation in Fig. 5 is that there seems to be little difference between the two separations between steady and casual partnerships when the transmission probability is low whereas separating at 10 sex acts gives a higher R_0 estimate when the per-sex-act transmission probability is large. We have no immediate explanation to this but believe it also has to do with the difference in partnership distribution affecting the assortative mixing.

In Fig. 6, the same type of plot is shown, but now under the more realistic assumption that $p_{(fm)} = 0.5p_{(mf)}$, i.e. that the per-sex-act transmission probability from a female to a male is only half that of male to female. This relation between $p_{(fm)}$ and $p_{(mf)}$ is more realistic for most sexually transmitted diseases (for e.g. gonorrhea $p_{(fm)}$ was estimated to $0.55p_{(mf)}$ (Yorke et al., 1978), and for HIV $p_{(fm)}$ was estimated to $0.5p_{(mf)}$ (Leynaert et al., 1998). The same type of qualitative results holds also under this assumption, but now for $p_{(mf)}$ up to about 0.5. The only qualitative difference (compared to the case when $p_{(fm)} = p_{(mf)}$) seems to be that when the per-sex-act transmission probability is large, the models distinguishing between steady and casual partnerships give lower R_0 estimates as compared to both per-sex-act transmission probabilities being equal.

As mentioned earlier the per-sex-act transmission probability is often smaller than 0.5 and sometimes much smaller. So how can an STI become endemic since Fig. 6 indicates an R_0 estimate of about 0.6, which would be even less when the infectious period is shorter than 1 year? Of course, the model does not capture all features of an STI, in particular, not individual heterogeneities in terms of susceptibility, infectivity and sexual behavior, but even if it would, this could be explained by the disease being endemic only in a more restrictively defined core group within which $R_0 > 1$ (the data were a community sample of individuals aged 16–31).

5. Discussion

In the paper a model for the spread of an STI was defined allowing arbitrary partnership distribution and distinguishing between steady and casual partnerships. An approximation using branching processes was derived thus giving an expression for R_0 determining whether a major outbreak is possible or not. When fitted to data the main conclusion was that neglecting differences between partnerships has the effect that R_0 is systematically over-estimated.

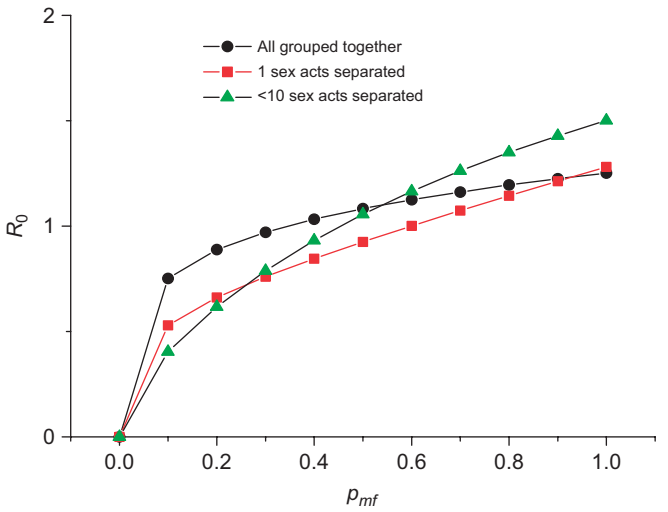

Fig. 6. Plot of R_0 for the Gotland community as a function of the male-to-female per contact probability p_{mf} and under the assumption that $p_{fm} = 0.5p_{mf}$. The three curves correspond to the situations where a casual partnership is defined as being one sex act, being 10 or less sex acts, and to the case where steady and casual partnerships are not distinguished.

The model can be made more realistic in several ways. One underlying assumption is that an individual chooses his/her partners randomly among all partnerships of opposite sex "on the market". As an effect, an individual with few, say 1, partner chooses his/her partner in the same way as someone who has many partners. It would perhaps be more realistic to allow for some assortative mixing meaning that promiscuous people tend to have partners who are also promiscuous. However, to include this into the model makes the analysis harder and to fit the degree of assortative mixing to our data is hard since no such information about partner's sexual activity was available.

Another assumption is that the sexual network is considered fixed over time. In reality, partnerships break up or are interrupted and new are created, with the possibility of dual simultaneous partnerships. Community properties of such dynamics, which are at least partly available in the analysed data, of course affect the potential spread of disease. However, as mentioned previously, an interpretation of this static network is that the number of partners of an individual reflects the number of partners such an individual has during a period corresponding to the infectious period. This should make the interpretation of R_0 the same as in a dynamic network having the same numbers of partnerships over that period of time. The advantage with the present static network model of course being that it can be analysed in more detail.

A different and important model relaxation would be to allow individual heterogeneity in terms of transmission probabilities. Some individuals may have a higher risk of getting infected (in any of his/her partnerships) and others have lower risk, and similarly the per-sex-act probability of infecting others may vary between individuals. Such heterogeneity will have the effect that transmission will tend to take place during the first few sex acts if ever (for partnerships with high transmission and rarely for partnerships with low transmission probability). We believe that this should somewhat reduce the magnitude of over-estimation of R_0, but how much is an open question. To actually fit such a generalised model to data would be the problematic part. Another step towards realism is to allow individuals to be either homosexual, heterosexual or bisexual. Such an extension would be quite straightforward, and estimates of community fractions of the different types can often be obtained although this is not available in the Gotland data set.

One important question not addressed in the present paper is to study different preventive measures affecting the transmission probability and/or the degree distribution. To derive and compare expressions for the reproduction number when different preventive measures are in place could be used in guiding which measures are most effective.

To summarize, more work is thus needed for realistic modelling of the spread of STI's in a community, and more data sets should be analysed to see if the observed features in the present analysis hold also there. Still, we believe that the main finding of the present work, namely that neglecting the empirical observation that individuals with many partners tend to have fewer sex acts per partner will over-estimate R_0, will still hold true under more realistic models. Having said this, this effect might be dominated by other effects when considering more complex and realistic models. Making more realistic models is hence important for determining if our conclusions apply in real world situations.

Acknowledgments

Tom Britton is supported by the Swedish Research Council and Fredrik Liljeros and Monica K. Nordvik are supported by the Swedish National Institute of Public Health and the European research NEST project DYSONET 012911. All three authors acknowledge Stockholm Group for Epidemic Modelling (www.s-gem.se).

References

Andersson, H., 1999. Epidemic models and social networks. Math. Scientist 24, 128–147.

Andersson, H., Britton, T., 2000. Stochastic epidemic models and their statistical analysis. Springer Lecture Notes in Statistics, vol. 151. Springer, New York.

Blower, S.M., Boe, C., 1993. Sex acts, sex partners, and sex budgets: implications for risk factor analysis and estimation of HIV transmission probabilities. J. AIDS 6, 1347–1352.

Bollobás, B., Janson, S., Riordan, O., 2007. The phase transition in inhomogeneous random graphs. Rand. struct. alg., to appear.

Britton, T., Deijfen, M., Martin-Löf, A., 2006. Generating simple random graphs with prescribed degree distribution. J. Stat. Phys. 124, 1377–1397.

Colgate, S.A., Stantley, E.A., Hyman, S.P., Layne, S.P., Qualls, L., 1989. Risk behavior-based model of the cubic growth of acquired

immunodeficiency syndrome in the United States. PNAS 86 (12), 4793–4797.

Diekmann, O., De Jong, M.C.M., Metz, J.A.J., 1998. A deterministic epidemic model taking into account of repeated contacts between the same individuals. J. Appl. Probab. 35, 448–462.

Diekmann, O., Dietz, K., Heesterbeek, J.A.P., 1991. The basic reproduction ratio for sexually transmitted diseases: I. Theoretical considerations. Math. Biosci. 107, 325–339.

Eames, K.T.D., Keeling, M.J., 2004. Monogamous networks and the spread of sexually transmitted diseases. Math. Biosci. 189, 115–130.

Giesecke, J., Scalia-Tomba, G., Göthberg, M., Tüll, P., 1992. Sexual behavior related to the spread of sexually transmitted disease—a population-based survey. Int. J. STD AIDS 3, 255–260.

Hethcote, H.W., Yorke, J.A., 1984. Gonorrhea transmission dynamics and control. Springer Lecture Notes in Biomathematics, vol. 56. Springer, New York.

Jagers, P., 1975. Branching Processes with Biological Applications. Wiley, London.

Katz, B.P., Caine, V.A., Jones, R.B., 1990. In: International Symposium on Human Chlamydial Infections Chlamydial Infections Proceeds of the Seventh International Symposium on Human Chlamydial Infections. Cambridge University Press, New York.

Kretzschmar, M., Jager, J.C., Reinking, D.P., van Zessen, G., Brouwers, H., 1994. The basic reproduction ratio R_0 for a sexually transmitted disease in a pair formation model with two types of pairs. Math. Biosci. 124, 181–205.

Leynaert, B., Downs, A.M., de Vinzenzi, I., 1998. Heterosexual transmission of human immunodeficiency virus. Am. J. Epidemiol. 148, 88–96.

Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Aberg, Y., 2001. The web of human sexual contacts. Nature 411, 907–908.

Lycke, E., Lowhagen, G.-B., Hallgagen, G., et al., 1980. The risk of transmission of genital Chlamydia trachomatis infection is less than that of genital Neisseria gonorrhoeae infection. Sex. Transm. Dis. 7, 6–10.

May, R.M., Anderson, R.M., 1987. Transmission parameters of HIV infection. Nature 326, 137–142.

Newman, M.E.J., 2002. Spread of epidemic disease on networks. Phys. Rev. E 66, 016128.

Nordvik, M.K., Liljeros, F., 2006. Transmission of sexually transmitted infections. Sex. Transm. Dis. 33, 342–349.

Pastor-Satorras, R., Vespignani, A., 2002. Epidemic dynamics and endemic states in complex networks. Phys. Rev. E 63, 066117.

Rottingen, J.A., Garnett, G.P., 2002. The epidemiological and control implications of HIV transmission probabilities within partnerships. Sex. Trans. Dis. 29, 818–827.

Yorke, J.A., Hethcote, H.W., Nold, A., 1978. Dynamics and control of the transmission of gonorrhea. Sex. Trans. Dis. 5, 51–56.