

Mathematical Statistics
Stockholm University

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

Tom Britton, Monica K. Nordvik and Fredrik Liljeros

Research Report 2005:15

ISSN 1650-0377

Postal address:

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

Internet:

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

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

Tom Britton*, Monica K. Nordvik[†] and Fredrik Liljeros[‡]

December 2005

Abstract

We model a sexually transmitted infection in a population where individuals have different numbers of partners separated into long term relationships and short term relationships, where the risk of transmission is higher in long term relationships. The number of partners of the two types is specified by the degree distribution in the community. For this structured 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 relationships) 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 realistic transmission rates, the effect that so-called superspreaders have on R_0 is over-estimated when not admitting for different types of relationships.

Key words or phrases: Basic reproduction number – branching process approximation – network epidemic model – sexually transmitted infections

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

[†]Department of Sociology, Stockholm University, SE-106 91 Stockholm, Sweden.

[‡]Department of Sociology, Stockholm University, SE-106 91 Stockholm, Sweden.

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. The relevance of sexual activity and community contact patterns are clearly very important factors in determining potential spread, as are transmission parameters (e.g. Anderson and May, 1991 Ch. 11, Liljeros *et al.*, 2003).

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 new sex partners. In Diekmann *et al.* (1991) the forming and separation of partnerships is modelled dynamically in time – partners break up and new partnerships are formed with specified intensities. This leads to a Markov type model and the number of partnerships 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 related model, specifically devoted to AIDS in a heterosexual population, is presented in Mode (1991) where conclusions are made from simulations. A different approach is taken by Eames and Keeling (2004) where each individual has at most one active partnership among a set of 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 community 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 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) looks 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 the sexual relationships are modelled by a (random) network which does not change over time. However, if in reality new partnerships are formed and old break up, but in a time-stationary way, then this fits in to the present model for the partnership network by interpreting the number of partners an individual has in the model as the actual number of partners the individual has during a time period corresponding to the typical length of the infectious period of the disease in question. Empirical evidence show 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 the model allows an arbitrary but specified degree distribution. Further, in the model we incorporate two types of partnerships, long and short, having distinct transmission probabilities. This is done for two reasons. The duration of a relationship, or rather the number of sex acts, affect the probability of disease transmission. Secondly, individuals having many sex partners tend to have short relationships with fewer sex acts per partner compared to individuals with one or few partners.

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 that there are long and short relationships, simply treating all relationships equivalently, and derive R_0 under this assumption. By calibrating parameters under 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).

In Section 4 the heterosexual model is then applied to data on sexual activity collected in the island of Gotland, Sweden (Giesecke *et al.*, 1992, recently reanalysed in Nordvik & Liljeros, 2006). The data is from a representative community sample and contains the number of sex partners of the individuals and also information about the sexual activity in each relationship, thus being quite unique in its detail. For a few different scenarios of the transmission parameters we determine and compare expression for R_0 under different relationship definitions. 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 long and short relationships

Assume that there are two types of relationships, long term relationships and short term relationships, and that the probability of transmission between an infectious individual and a susceptible individual in a long relationship is p_L and p_S in a short relationship ($p_L \geq p_S$). (This model may very well be used for non-STI diseases. The important feature is that there is a social network structure having two types of relationships.)

Different individuals have different number of short and long term relationships. Let $\pi_{i,j}$ denote the proportion of individuals having i long and j short relationships during a period corresponding to the typical length of the infectious period, and categorize an individual as an (i, j) -type accordingly. Assume further that long (short) term partners are chosen randomly among the long (short) term relationships in the community. This implies that the probability that a long term partnership is with an (i, j) -individual (i.e. having i long and j short term relationships) is

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

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

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

These probabilities are said to be size-biased reflecting the fact that individuals with many long (short) relationships are more likely to be selected as a long (short) partner. An infectious individual infects his susceptible partners independently, and with probability p_L for long term partners and p_S for short term partners. An individual can only get 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 & 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.*, 2005) and extensive model analysis (cf. Bollobas *et al.*, 2005).

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 the individual was infected by, will be susceptible with large probability, and the chance that two individuals will contact the same individual is negligible which motivates why individuals infect approximately independently thus justifying the branching process approximation.

The number of individuals an infected will infect depends on the number of long and short term 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 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, L) if s/he has i long term relationships, j short term relationships and if s/he was infected by a long term relationship (of course $i \geq 1$ for this to make sense). Similarly, an individual is of type (i, j, S) if s/he has the same partner structure but was infected by a short term relationship.

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

$$\begin{aligned}\lambda_{(i,j,L)(k,l,L)} &= (i - 1)\tilde{\pi}_{k,l,LP_L} \\ \lambda_{(i,j,L)(k,l,S)} &= j\tilde{\pi}_{k,l,SP_S} \\ \lambda_{(i,j,S)(k,l,L)} &= i\tilde{\pi}_{k,l,LP_L} \\ \lambda_{(i,j,S)(k,l,S)} &= (j - 1)\tilde{\pi}_{k,l,SP_S}\end{aligned}$$

For example, in the expression for $\lambda_{(i,j,L)(k,l,L)}$ the (i, j, L) -individual has i long term partners, but s/he was infected by one of them, so there are $i - 1$ remaining susceptible long term partners, and each of them is with a (k, l) -individual with probability $\tilde{\pi}_{k,l,L}$, and any such partnership results in infection with probability p_L . 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 to 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 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 s/he is hard to say any qualitative results concerning R_0 as a function of model parameters. If the community structure, i.e. degree distributions, are kept fixed and p_L and p_S are varied, it can be shown that R_0 increases linearly if p_L and p_S increase keeping p_L/p_S fixed.

2.2 No distinction between long and short term relationships

If we had neglected that there are different types of relationships 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 relationship 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 $p_L = p_S = p$. The reason for this is that even when the transmission probabilities are identical partners are selected differently in the two models.

With this terminology it can be shown that (Andersson, 1999) the basic reproduction number equals

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

where $E(\cdot)$ and $V(\cdot)$ denote expected value and variance respectively. In 1 it is seen that for fixed average number of partners and fixed transmission probability R_0 increases with the variance of the number of partners, and if the variance is infinite – so is R_0 (Pastor-Satorras and Vespignani, 2002). 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 relationships) with the previous where we acknowledge long and short term relationships,

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 short plus the number of long partnerships. 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

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

This choice of $\{\pi_k\}$ and p make 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 long and short relationships gives a larger R_0 (half of the community have 3 long and 3 short relationships and the other half has 1 long and 1 short relationship) and simple examples going in the opposite direction (half of the community have 1 long and 3 short relationships and the other half has 3 long and 1 short relationship).

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 long and short relationships

We still assume that there are two types of contacts, long term and short term, but the probability of transmission also depends on who is infectious and who is susceptible: $p_L^{(mf)}$ is the probability that a long term relationship in which the male is infectious and the female is susceptible results in transmission, and similarly for $p_L^{(fm)}$, $p_S^{(mf)}$ and $p_S^{(fm)}$. Further, let $\pi_{ij}^{(m)}$ denote the community proportion of males having i long and j short relationships (during a period equal to a typical infectious period), and let $\pi_{i,j}^{(f)}$ denote the corresponding proportion for females. In order for the total number of long and short term relationships of men 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 long-term relationship of a female is with an (i, j) -male with probability $\tilde{\pi}_{i,j,L}^{(m)}$, the corresponding short-term relationship has

probability $\tilde{\pi}_{i,j,S}^{(m)}$, and males have corresponding relationships with specified female types with probabilities $\tilde{\pi}_{i,j,L}^{(f)}$ and $\tilde{\pi}_{i,j,S}^{(f)}$ respectively, where

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

Consider an infected (i, j) -female who was infected through a long term relationship. Let $\lambda_{(i,j,L)(k,l,L)}^{(f)}$ denote the expected number of men, having (k, l) -partners that she infects through a long-term relationship. Similarly, $\lambda_{(i,j,L)(k,l,L)}^{(m)}$ is the corresponding expected number, but for a man 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 relationship 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 relationship s/he gets infected. It follows that

$$\begin{aligned} \lambda_{(i,j,L)(k,l,L)}^{(f)} &= (i-1)\tilde{\pi}_{k,l,LP_L}^{(m)} \binom{fm}{f}, \\ \lambda_{(i,j,L)(k,l,S)}^{(f)} &= j\tilde{\pi}_{k,l,SP_L}^{(m)} \binom{fm}{f}, \\ \lambda_{(i,j,S)(k,l,L)}^{(f)} &= i\tilde{\pi}_{k,l,LP_L}^{(m)} \binom{fm}{f}, \\ \lambda_{(i,j,S)(k,l,S)}^{(f)} &= (j-1)\tilde{\pi}_{k,l,SP_L}^{(m)} \binom{fm}{f}, \\ \lambda_{(i,j,L)(k,l,L)}^{(m)} &= (i-1)\tilde{\pi}_{k,l,LP_L}^{(f)} \binom{mf}{f}, \\ \lambda_{(i,j,L)(k,l,S)}^{(m)} &= j\tilde{\pi}_{k,l,SP_L}^{(f)} \binom{mf}{f}, \\ \lambda_{(i,j,S)(k,l,L)}^{(m)} &= i\tilde{\pi}_{k,l,LP_L}^{(f)} \binom{mf}{f}, \\ \lambda_{(i,j,S)(k,l,S)}^{(m)} &= (j-1)\tilde{\pi}_{k,l,SP_L}^{(f)} \binom{mf}{f}. \end{aligned}$$

In the branching process approximation of the epidemic starting with one man say, the next generation will consist of women, the generation thereafter of men etc. This implies that the offspring distribution between generations typically differ every other generation. We therefore look at the off-spring distribution two generations later. Suppose we start with a man of type (i, j, L) , which hence had i long and j short relationships and was infected by a long term relationship. Let $\lambda_{(i,j,L)(k,l,L)}^{(2m)}$ denote the expected number of infected men of type (k, l, L) infected two generations later that are caused by our (i, j, L) -man. Then, by conditioning and summing over possible female routes, $\lambda_{(i,j,L)(k,l,L)}^{(2m)}$ satisfies

$$\lambda_{(i,j,L)(k,l,L)}^{(2m)} = \sum_{r,s} \left(\lambda_{(i,j,L)(r,s,L)}^{(m)} \lambda_{(r,s,L)(k,l,L)}^{(f)} + \lambda_{(i,j,L)(r,s,S)}^{(m)} \lambda_{(r,s,S)(k,l,L)}^{(f)} \right),$$

and similarly

$$\begin{aligned}\lambda_{(i,j,L),(k,l,S)}^{(2m)} &= \sum_{r,s} \left(\lambda_{(i,j,L)(r,s,L)}^{(m)} \lambda_{(r,s,L)(k,l,S)}^{(f)} + \lambda_{(i,j,L)(r,s,S)}^{(m)} \lambda_{(r,s,S)(k,l,S)}^{(f)} \right) \\ \lambda_{(i,j,S),(k,l,L)}^{(2m)} &= \sum_{r,s} \left(\lambda_{(i,j,S)(r,s,L)}^{(m)} \lambda_{(r,s,L)(k,l,L)}^{(f)} + \lambda_{(i,j,S)(r,s,S)}^{(m)} \lambda_{(r,s,S)(k,l,L)}^{(f)} \right) \\ \lambda_{(i,j,S),(k,l,S)}^{(2m)} &= \sum_{r,s} \left(\lambda_{(i,j,S)(r,s,L)}^{(m)} \lambda_{(r,s,L)(k,l,S)}^{(f)} + \lambda_{(i,j,S)(r,s,S)}^{(m)} \lambda_{(r,s,S)(k,l,S)}^{(f)} \right).\end{aligned}$$

If the (i, j, X) -indices are labelled in some way and we let $\Lambda^{(2m)}$ denote the matrix containing the elements above, labelled accordingly, then $\Lambda^{(2m)}$ describes the mean off-spring matrix after two generations, starting with infected men. The largest eigenvalue to this matrix specifies the growth rate after two generations, and in particular 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 1-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 a female infected, 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 in the case of a homosexual community it is hard to say any general conclusions about R_0 . 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 long and short term relationships

If we neglect that there are different types of relationships 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 relationship where the male is infectious and the female is susceptible, and $p^{(fm)}$ is the opposite transmission probability. Further, let $\pi_k^{(m)}$ denote the proportion males having k partners (during a period of a typical infectious period), and let $\pi_k^{(f)}$ be the corresponding for females. Similar to before, the probability that a female has contact with a male having k partners, equals

$$\tilde{\pi}_k^{(m)} = \frac{k\tilde{\pi}_k^{(m)}}{\sum_j j\tilde{\pi}_j^{(m)}}, \quad \text{and} \quad \tilde{\pi}_k^{(f)} = \frac{k\tilde{\pi}_k^{(f)}}{\sum_j j\tilde{\pi}_j^{(f)}}$$

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 $p_L^{(fm)} = p_S^{(fm)} = p^{(fm)}$ and $p_L^{(mf)} = p_S^{(mf)} = p^{(mf)}$ due to the difference between the two models partner selection.

As before, the early stages of the epidemic, assuming a large community, is approximated by a multitype branching process. 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, as when admitting different types of relationships, define the basic reproduction number as the square root of this 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 long and short relationships) 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

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

In order for the comparison to be fair the transmission parameters should be a weighted average of the long and short term transmission probabilities, where the weights take into account how frequent the different types of relationships are. More precisely, they should satisfy

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

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

When comparing R_0 for the two models calibrated as above it is possible to find examples going in both ways. In the next section the two models are fitted to data and computed.

4 Application to a study on sexual relationships in 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% (778 individuals). It was concluded that the material was free from systematic biases (Giesecke *et al*, 1992).

The respondents were asked not just to indicate the number of sexual partners during the last year, but also when and for how long the relationship continued, and the approximate number of sex acts they had with each partner during the relationship, thus making the study quite unique. Individuals reporting no sexual activity (21% of the women and 29% of the men) will not contribute to the spread and are left out in what follows. In the remaining data, 45.2 % were women and 54.8 % men. In Figure 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 Figure 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.

It is seen that the total number of sex acts decreases with the number of partners for women, and at least not increases for men. This implies that the average number of sex acts *per partner* decreases with the number of partners. 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.

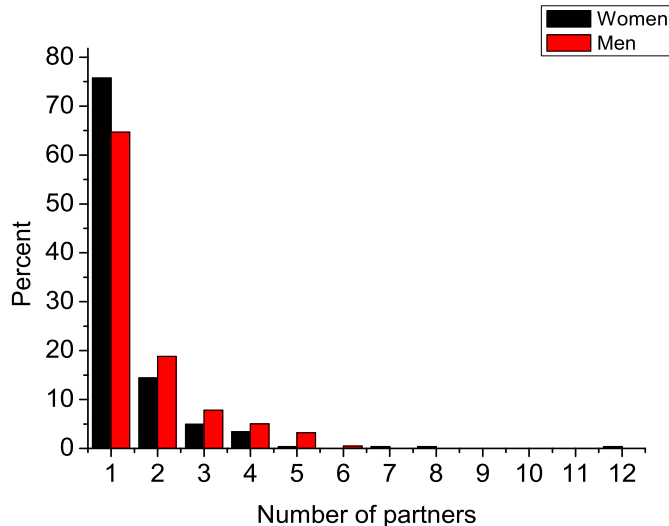


Figure 1: Distribution of the number of sex partners during one year for men and women in the Gotland data.

4.2 Estimation of R_0 from the Gotland data

The data presented in the previous section contains no information about any specific STI. Instead we use the data to fit the heterosexual network described in Section 3.1. Then we estimate R_0 for this (type of) 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, among other things, on properties of the disease, the type of sexual activity, the gender of the infector and of the susceptible, and the frequency of sexual encounters. Both practical and ethical reasons make it hard to obtain such detailed information. In the present data material we have information about the (approximate) number of sex acts in each relationship. We can thus define an infection probability which varies with the number of sex acts and also depends on the sex of the infectious individual. One natural such function is to assume that each sex act, independently and with equal probability, results in that the susceptible gets infected. The probability that a female susceptible escapes infection from an infectious male partner with 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 hence $1 - (1 - p_{mf})^k$. The

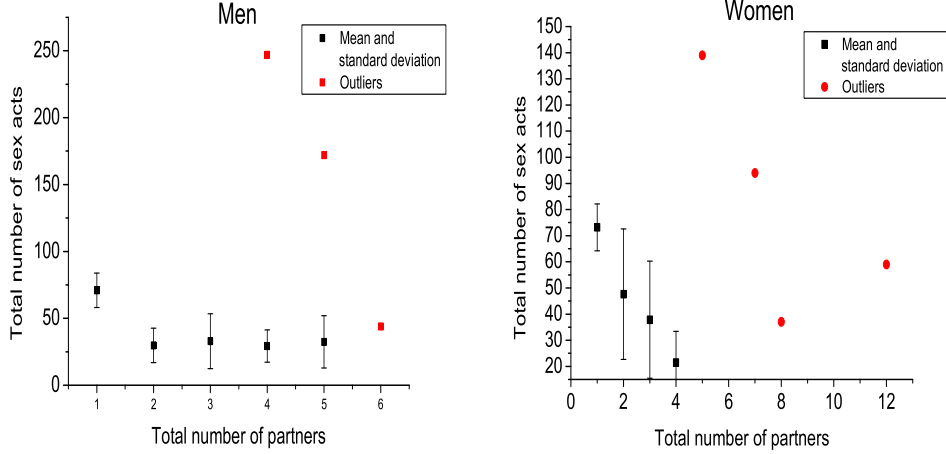


Figure 2: Average total number of sex acts per year, separated for men and women, and also depending on the number of sex partners.

corresponding probability that an infectious female infects her partner with 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 $p_{fm} = 0.5p_{mf}$, the latter being more realistic for most sexually transmitted diseases.

In the model in Section 3 we did not allow the transmission probability in a relationship to depend on the number of sex acts because such a model would make the number of different types of relationships enormous and the number of individuals in each type of relationship very small thus introducing over-parametrisation. Instead we only distinguished between long and short term relationships, where long and short refer to the number of sex acts rather than calendar time. Using our more detailed data we can separate long and short relationships in many ways. In the analysis we have performed two such separations. The first separation is where short relationship is defined by a single sex act and everything more than one sex act is a long relationship. With this definition $p_S^{(mf)} = 1 - (1 - p_{mf})^1 = p_{mf}$ (and similarly $p_S^{(fm)} = p_{fm}$) since a short relationship always has one single sex act. In order to obtain expressions for $p_L^{(mf)}$ as a function of p_{mf} we have, for each relationship in which a male has more than one sex act, computed $1 - (1 - p_{mf})^{\#\text{sex acts}}$ and we have then estimated $p_L^{(mf)}$ by the arithmetic mean of these expressions.

The quantity $p_L^{(fm)}$ is estimated similarly, only replacing p_{mf} by p_{fm} and considering averages over females in different relationships.

We have also treated a different separation in which anything less than 10 sex acts is considered a short relationship and 10 or more sex acts is defined as a long relationship. The estimation of $p_L^{(mf)}$ and $p_S^{(mf)}$ as functions of p_{mf} , as well as $p_L^{(fm)}$ and $p_S^{(fm)}$ as functions of p_{fm} is done by taking similar averages, only now over the new separations of short and long relationships.

In Figure 3 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. $p_{fm} = p_{mf}$). The curve with squares is for the separation with one sex act being a short relationship and the curve with triangles correspond to the separation with a short relationship defined as less than 10 sex acts. Also in the figure is a curve where there is no distinction between short and long relationships (corresponding to the model of Section 3.2). For this curve, $p^{(mf)}$ and $p^{(fm)}$ are estimated as above, only now the averages are taken over *all* relationships of the considered gender.

In the Figure 3 it is seen that when the per-sex-act transmission probability is smaller than ≈ 0.4 , as is usually the case, the effect of *not* distinguishing between short and long relationships (Section 3.2) is that R_0 is overestimated. An intuitive explanation to this, also taking the contact behaviour in the community into account, is the following. When all relationships are assumed to have equal transmission probability, this transmission probability is overestimated for relationships having few sex acts. In the Gotland data it was seen that relationships with few sex acts tend to be between individuals having many partners. And, when all relationships 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 overestimation 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 overestimated when all relationships are treated equally. We suspect that the difference in R_0 between treating all relationships as identical compared to when admitting different types of relationships would be even bigger if the model would have distinguished between more types of relationships.

A second question is then why the overestimation only holds when the per-sex-act transmission is low or moderate (in Figure 3 it no longer holds when p_{mf} exceeds about 0.5). This has to do with the fact, mentioned in Section 3.2, that the model with two types of relationships for the special case where both relationships have equal transmission probability does *not*

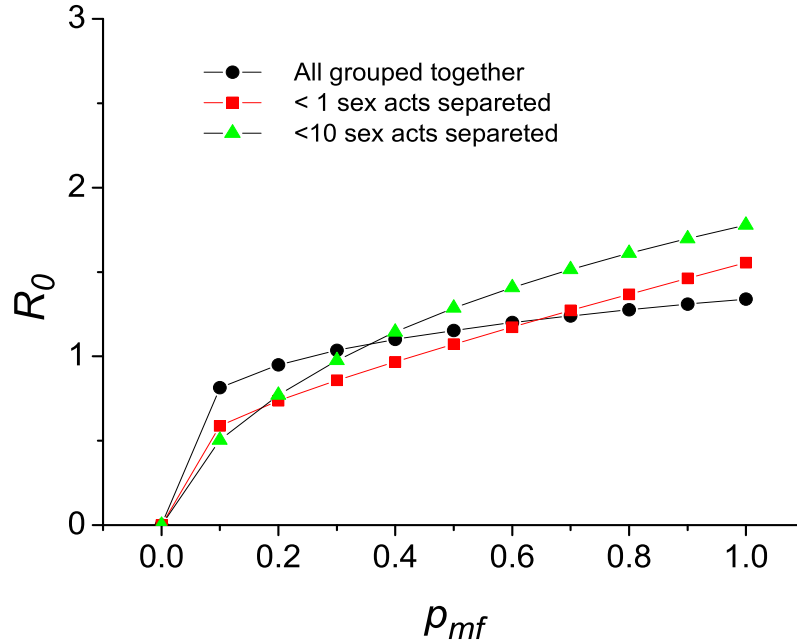


Figure 3: Plot of R_0 for the Gotland community as a function of the male-to-female per contact probability p_{mf} and under the assumption that $p_{fm} = p_{mf}$. The three curves are for the case when a short relationship is defined as one sex act, as 10 or less sex acts, and where long and short relationships are not separated.

coincide with the model having one type of relationship. A simple example illustrates this. Suppose that half of the community have 1 long and 0 short relationship and the other half of the community have 0 long and 3 short relationships (males and females are assumed to behave equivalently). Then, all individuals with one relationship will have partners also having only one relationship, and individuals with three relationships will all have partners with three relationships (Note that this is not the case if only one type of relationship is considered and 50% have one relationship and 50% have three relationships). If the transmission probability is not very different between long and short relationships (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. As an effect, R_0 will be higher with two types of relationships whenever the per-sex-act transmission probability is large enough. Of course the example

is extreme but since individuals in steady relationships tend to have fewer short relationships, the explanation should be valid in real communities.

Another observation in Figure 3 is that there seems to be little difference between the two separations between long and short relationships 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 Figure 4 below 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. The same type of qualitative results holds also under this assumption. The only qualitative difference seems to be that when the per-sex-act transmission probability is large, the models distinguishing between long and short relationships give lower R_0 estimates as compared to both per-sex-act transmission probabilities being equal.

A reasonable per-sex-act transmission is 0.2. So how can there be endemic diseases since Figure 4 indicates an R_0 estimate of about 0.6? Of course, the model does not capture all features of an STI, but even if it would, this could be explained by the disease being endemic only in a more restrictively defined core group (the data was 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 separating long and short relationships. An approximation using branching processes was derived thus giving an expression R_0 determining whether a major outbreak is possible or not. When fitted to data the main conclusion was that neglecting differences between relationships has the effect that R_0 is overestimated.

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. In real life there might be some assortative mixing meaning that promiscuous people tend to have partners with promiscuous individuals. However, to include this into the model makes the analysis harder, and secondly, to find reliable empirical information about the degree distributions of partners of interviewed individuals seems very hard, thus making assortative mixing hard to estimate.

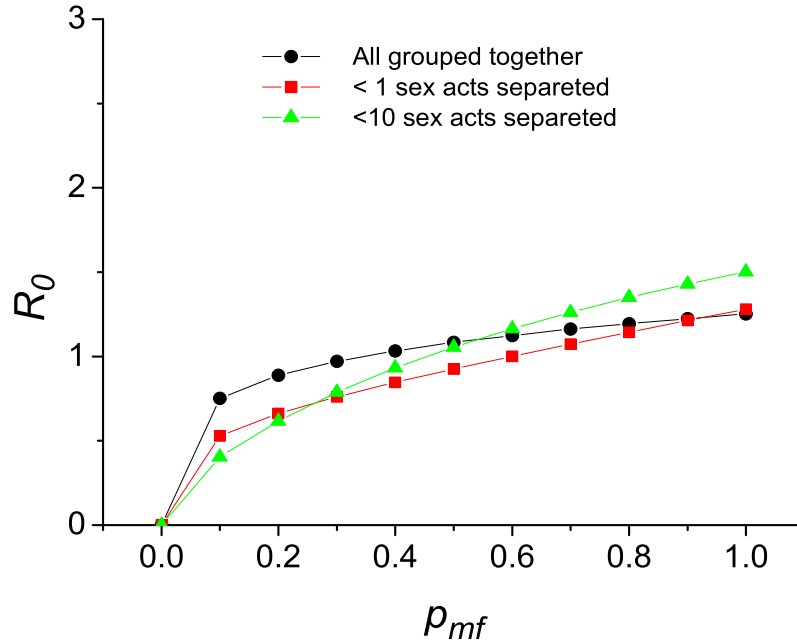


Figure 4: Plot of R_0 for the Gotland community as a function of the male-to-female per contact probability p_{mf} and under the assumption that $p_{fm} = 0.5p_{mf}$. The three curves are for the case when a short relationship is defined as one sex act, as 10 or less sex acts, and where long and short relationships are not separated.

Another assumption is that the sexual network is considered fixed. As mentioned previously, the number of partners of a randomly selected individual is said to reflect 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 of course being that the present model 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 relationships) and others have lower risk, and similarly the per-sex-act probability of infecting others may vary between individuals.

An important question not addressed in the present paper is to study the effect of different vaccination policies. To derive expressions for the repro-

duction number after vaccination as well as to compare various vaccinations strategies are important problems.

More work is thus needed for realistic modelling of the spread of STI's in a community. 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 overestimate R_0 , will still hold true under more realistic models.

Acknowledgements

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

- Anderson, R.M., and May, R.M. (1991) *Infectious diseases of humans*. Oxford UP, Oxford.
- Andersson, H. (1999) Epidemic models and social networks. *Math. Scientist*, **24**, 128-147.
- Andersson H., and Britton T. (2000). *Stochastic epidemic models and their statistical analysis*. Springer lecture notes in statistics **151**. Springer, New York.
- Bollobás, B., Janson, S., Riordan, O. (2005) The phase transition in inhomogeneous random graphs. Submitted.
- Britton, T., Deijfen, M. and Martin-Löf, A. (2005) Generating simple random graphs with prescribed degree distribution. Submitted.
- Colgate, S.A., Stantley, E.A., Hyman, S.P., Layne, S.P. and 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., Dietz, K., and 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., and 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. *International Journal of STD & AIDS*, **3**, 255-260.
- Hethcote, H.W. and Yorke, J.A. (1984) *Gonorrhea transmission dynamics and control*. Springer Lecture Notes in Biomathematics, **56**. Springer, New York.
- Jagers P. (1975) *Branching processes with biological applications*. Wiley, London.
- Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., and Aberg, Y. (2001) The web of human sexual contacts. *Nature*, **411**, 907-908.
- Liljeros, F., Edling, C.R., Amaral, L.A.N. (2003) Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and infection*, **5**, 189-196.
- May, R.M., and Anderson R.M. (1987) Transmission parameters of HIV infection. *Nature*, **326**, 137-142.
- Mode, C.J. (1991) A stochastic model for the development of an AIDS epidemic in a heterosexual population. *Math. Biosci.*, **107**, 491-520.
- Newman, M.E.J. (2002) Spread of epidemic disease on networks. *Phys. Rev. E*, **66**, 016128.
- Nordvik, M.K. and Liljeros, F. (200?) Number of sexual encounters involving intercourse and the transmission of sexually transmitted infections. To appear in *Sexually Transmitted Diseases*
- Pastor-Satorras, R. and Vespignani, A. (2002) Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, **63**, 066117.