How sex acts scale with the number of sex partners: evidence from *Chlamydia trachomatis* data and implications for control

Christian L. Althaus$^{1,†}$, Marc Choisy$^{2,3}$, Samuel Alizon$^{2,†}$ and the CSF group$^*$

$^1$ Institute of Social and Preventive Medicine (ISPM), University of Bern, 3012 Bern, Switzerland
$^2$ Laboratoire MIVEGEC (UMR CNRS 5290, IRD 224, UM), 34394 Montpellier Cedex 5, France
$^3$ Oxford University Clinical Research Unit, Hanoi, Vietnam

Author for correspondence: Christian L. Althaus (christian.althaus@alumni.ethz.ch)

$^†$ Both authors contributed equally to this work.

Keywords: *Chlamydia trachomatis*, infection control, sexual mixing, mathematical model, parameter inference

---

*The CSF groups is composed of Nathalie Bajos and Michel Boson (scientific heads), Nathalie Beltez (coordinator), Armelle Andro, Michèle Ferrand, Véronique Goulet, Anne Laporte, Henri Leridon, Charlotte Le Van, Sharman Levinson, Laurent Touloumon, Nicolas Razafindrastima, Josiane Warszawski, and also Lucette Aussel and Caroline Laborde.*
Mathematical models are frequently used to assess the impact of control interventions for *Chlamydia trachomatis* and other sexually transmitted infections (STIs). Modeling approaches that stratify the population by the number of sex partners often assume the transmission risk per partner to be constant. However, sexual behavior data suggests that people with many partners share less sex acts per partner than people with fewer partners. This should lower the risk of transmission per partner for highly sexually active individuals and could have important epidemiological consequences for STI transmission. We devise a new epidemiological model that we fit to chlamydia prevalence data from Natsal-2 and CSF, two population-based probability sample surveys of sexual behavior in Britain and France. Compared to a standard model where the transmission risk per partner is constant, a model with realistic numbers of sex acts per partner provides a better fit to the data. Furthermore, the improved model provides evidence for strong assortative mixing among individuals with different numbers of sex partners. Our results suggest that all chlamydia infected individuals with one or more new heterosexual partners per year contribute significantly to ongoing transmission, underlining that control interventions should be aimed towards all sexually active young adults.
Introduction

The spread of sexually transmitted infections (STIs) crucially depends on the number of sex partners and the number of sex acts within a sexual partnership. Empirical studies have shown that the number of sex partners within a given time period is a strong determinant for the spread of curable STIs such as *Chlamydia trachomatis* (Fenton et al., 2001; Sonnenberg et al., 2013; Althaus et al., 2012c) and *Neisseria gonorrhoeae* (Ison et al., 2013). However, the number of sex acts as an additional factor for transmission has not received as much attention. In particular, mathematical models of STI transmission that are frequently used in guiding public health policy decision making often assume the number of sex acts per partner to be constant and independent of the number of partners an individual has (Hethcote and Yorke, 1984; Anderson and May, 1991; Garnett and Anderson, 1993b; Garnett et al., 1999; Keeling and Rohani, 2008). Neglecting the number of sex acts per partner could lead to inaccurate estimates of the efficiency of public health interventions, for instance when predicting the impact of screening interventions on chlamydia spread.

Theoretical studies have shown that taking into account the number of sex acts in models of STI transmission can dramatically alter the results (Kaplan, 1990; Røttingen and Garnett, 2002). An important aspect was illustrated in the study by Nordvik and Liljeros (2006), where they showed that it is not only the number of partners, but also the number of sex acts per partner, that is important for the transmission of an infection. Garnett and Anderson (1996) developed a generic framework to include the heterogeneity in the number of sex acts between individuals and showed that it can enhance the likelihood of persistence of STIs. Still, we lack a solid understanding of the relationship between the number of sex acts and the number of sex partners. Furthermore, it remains to be determined how this relationship can be appropriately integrated into mathematical models of STI transmission that then can be validated with data.

The results of probability sample surveys of sexual behavior allow us to formulate some hypotheses on the relationship between the number of sex acts and the number of sex partners. Blower and Boe (1993) analyzed data from 1770 unmarried individuals aged 20–44 years collected in San Francisco (USA) from 1988–1989. They concluded that it may be most appropriate to see the number of sex acts as a ‘budget’ split between sex partners. Nordvik and Liljeros (2006) looked at the total number of sex acts in relation to the total number of sex partners based on data from 1150 individuals aged 16–31 years collected in Sweden in 1988. The authors found that the total number of sex acts remains constant in individuals with higher numbers of sex partners (it even decreases in women). In principle, one can directly infer the relationship between sex acts and sex partners from sexual behavior surveys, but there remain potential biases due to the self-reported data. For example, it proves difficult to discriminate between protected and unprotected sex acts in such surveys. A better understanding of the quantitative relationship between the number of sex partners and the number of sex acts could provide useful insights into the epidemiology of STIs and their control.

Another important aspect of taking the number of sex acts between sex partners into account is its effect on the topology of the sexual contact network. If the number of sex acts per sex partner is constant, each contact (edge) between susceptible and infected individuals (nodes) is weighted equally, i.e., has the same transmission probability. This is not the case if sex acts are explicitly accounted for (Moslonka-Lefebvre et al., 2012). However, network-based models usually require computationally-intensive simulations.
Kamp et al. (2013), which is why mathematical models of STIs tend to adopt a simpler approach by stratifying the population into subgroups with varying levels of sexual activity (Hethcote and Yorke, 1984; Garnett and Anderson, 1993b; Garnett et al., 1999). These models are often referred to as ‘risk class models’, although the stratification is based on the partner change rate that does not necessarily correlate with risk. In models, the level of sexual mixing between individuals can vary from fully proportionate (to the number of sex partners individuals have) to completely assortative (individuals within a sexual activity or risk group mix only with themselves). While some studies based on sexual behavior data indicate a certain level of assortative mixing between individuals (Renton et al., 1995; Garnett and Anderson, 1993a; Garnett et al., 1996; Aral et al., 1999), the output of STI transmission models is often only consistent with close to proportionate (or random) mixing (Garnett et al., 1999; Althaus et al., 2012a). This inconsistency illustrates the challenge in estimating the degree of sexual mixing, and we show that considering the number of sex acts per sex partner in models of STI transmission sheds new light on this question.

Our goal is to gain a more detailed picture of the relationship between the number of heterosexual sex acts and the number of heterosexual sex partners and study its effect on the transmission of chlamydia, the most prevalent bacterial STI. To this end, we make use of two datasets of sexual behavior: the British National Survey of Sexual Attitudes and Lifestyles (Natsal-2) (Johnson et al., 2001), and the ‘Contexte de la Sexualité en France’ (CSF 2006) (Bajos and Bozon, 2006, 2008). First, we directly infer the relationship between the number of heterosexual sex acts and the number of heterosexual sex partners from Natsal-2. We then fit a novel epidemiological model of STI transmission to chlamydia prevalence data of both datasets and indirectly infer epidemiological parameters as well as the relationship between the number of heterosexual sex acts and heterosexual sex partners. Both methods support the notion that the number of heterosexual sex acts per partner decreases with higher number of heterosexual sex partners. This is also consistent with a high level of assortative mixing between individuals with similar number of sex partners. Lastly, we show that taking realistic numbers of sex acts into account when modeling the heterosexual transmission dynamics of chlamydia has a crucial effect on the projected impact of control interventions.

Methods

Data

Natsal-2

Natsal-2 is a population-based probability sample survey of sexual attitudes and lifestyles conducted in Britain in 1999-2001 (Johnson et al., 2001). The sample consists of 11,161 women and men aged 16–44 years. Urine samples for ligase chain reaction (LCR) testing for chlamydia infection are available for a subset of 3569 sexually active respondents aged 18–44 years (Fenton et al., 2001). We use the following variables: number of new heterosexual sex partners in the last year, number of occasions of heterosexual sex in last 4 weeks, chlamydia test result from urine sample. Individual weights are used for all variables to adjust for unequal selection probabilities and to correct for the age and gender profile in the survey population. The full dataset is available from the UK Data Archive at the University of Essex (http://data-archive.ac.uk).
‘Contexte de la Sexualité en France’ (CSF 2006) is a survey conducted in France among 12,364 randomly chosen individuals aged 18–69 years (Bajos and Bozon, 2006, 2008). Urine samples testing for chlamydia infection are available for a subset of 7407 sexually active respondents. The two variables we used are the number of new heterosexual sex partners in the last year and the chlamydia test result from urine sample (see Supplementary Files).

**Sex acts and sex partners**

We stratify the population by the reported number of new heterosexual partners in the last year which results in \( n \) different sexual activity classes with \( c_i \) sex partners. The average heterosexual activity of each individual during one year is assumed to be reflected in the reported number of heterosexual sex acts during the last 4 weeks. For simplicity, and in order to increase sample size, we pool data of women and men together.

The reported numbers of sex acts during the last 4 weeks are highly dispersed (Fig. 1A). Therefore, we assume that the observed number of sex acts during the last 4 weeks (\( Z \)) for a given individual with \( c_i \) new heterosexual partners during the last year follows a negative binomial distribution. This distribution is frequently used to describe overdispersed data in biology and epidemiology (Lloyd-Smith, 2007; Hamilton et al., 2008). In the context of data on the number of sex acts, it can be interpreted as a process of sequential Bernoulli trials to reach ‘sexual satisfaction’ with the parameters \((k,p)\): for each of \( Z \) sex acts there is a probability \( p \) that it will not be sexually satisfying and individuals engage in new sex acts until satisfaction has been reached in \( k \) of them. The negative binomial distribution can also be interpreted as a mixture of Poisson distributions where the rates at which different individuals have sex are drawn from a gamma distribution with shape parameter \( k \) and scale parameter \( p/(1−p) \). More generally, a single Poisson distribution is obtained for \( k \to \infty \) and the negative binomial distribution is reduced to a geometric distribution for \( k = 1 \).

We assume that the mean of the negative binomial distribution is given by the following functional relationship:

\[
m_i = d_1 + d_2 c_i^{d_3},
\]

(1)

where \( m_i \) and \( c_i \) are the average number of sex acts and the number of sex partners for individuals of sexual activity class \( i \), respectively. \( m_i \) can then be scaled to obtain the expected number of heterosexual sex acts per year \( (M_i) \).

We construct a series of nested models by fixing some of the parameters to specific values. For example, if \( d_2 = 0 \), the total number of sex acts does not increase with higher numbers of sex partners. For \( d_3 = 1 \), the number of sex acts changes linearly with higher numbers of partners. The nested models then allow to perform hypothesis testing about different functional relationships between the number of sex acts and sex partners.

**Epidemiological model**

We assume the population to be stratified into \( n \) different sexual activity classes \( x_i \) with \( c_i \) new heterosexual partners per year (Hethcote and Yorke, 1984; Garnett and Anderson, 1993b). \( x_i \) denotes the proportion of individuals in sexual activity class \( i \). On average, individuals change their sexual activity after one year, and are proportionally
distributed among all sexual activity classes. For simplicity, we assume that sexual activity and the natural history and transmission of the infection are the same in women and men. If \( y_i \) is the proportion of infected individuals in sexual activity class \( i \), an SIS (susceptible-infected-susceptible) model (i.e., with no immunity) can then be written as follows:

\[
\frac{dy_i}{dt} = \mu \sum_{j=1}^{n} x_j y_j + (1 - y_i) c_i \sum_{j=1}^{n} b_{ij} \rho_{ij} y_j - (\gamma + \mu) y_i,
\]

where \( 1/\gamma \) and \( 1/\mu \) denote the average duration of infection and the average time spent in one sexual activity class, respectively. \( \rho_{ij} \) represents the elements of the mixing matrix that, following Garnett et al. (1999), can be defined as

\[
\rho_{ij} = \epsilon \delta_{ij} + (1 - \epsilon) \frac{c_j x_j}{\sum_{l=1}^{n} c_l x_l},
\]

where \( \delta_{ij} \) denotes the Kronecker delta (it is equal to 1 if \( i = j \) and to 0 otherwise). Mixing can be varied between proportionate \( (\epsilon = 0) \) and fully assortative \( (\epsilon = 1) \).

The per partnership transmission probability \( b_{ij} \) can be considered as a Bernoulli process of repeated transmission probabilities per sex act:

\[
b_{ij} = 1 - (1 - \beta)^{a_{ij}}
\]

where \( \beta \) denotes the transmission probability per sex act and \( a_{ij} \) the number of sex acts in a partnership between an individual of sexual activity class \( i \) and \( j \). To construct the matrix \( a_{ij} \), we use the following algorithm to distribute the number of sex acts between partners:

1. Define \( s_i = \frac{M}{c_i} \) as the average number of sex acts per partner for an individual of sexual activity class \( i \).
2. Set \( i = n \), i.e., start distributing the number of sex acts from individuals of the highest sexual activity class.
3. The number of sex acts from an individual of sexual activity class \( i \) will be shared with individuals of the same and lower sexual activity classes \( j = [1, i] \). Hence, we distribute the number of sex acts \( s_i \) proportional to the probability of a contact between an individual of class \( i \) and class \( j \) \( (\rho_{ij}) \) and to what the partners are willing to share \( (s_j) \).
4. The number of sex acts in individuals of sexual activity class \( j = [1, i] \) that remain to be shared \( (s_j) \) need to be updated to only those sex acts that have not been shared with individuals of sexual activity class \( i \) or higher.
5. If \( i > 1 \), set \( i = i - 1 \) and go to step 3. Otherwise, stop.

We calculate the basic reproduction number, \( R_0 \), for a given sexual activity class as follows:

\[
R_0 = \frac{c_i \sum_{j=1}^{n} (b_{ij} \rho_{ij})}{\gamma}.
\]

\( R_0 \) for the entire population is calculated using the next generation method (Diekmann et al., 1990; Diekmann and Heesterbeek, 2000). An overview of all parameters and variables to describe the epidemiological model as well as the sexual behavior is given in Table 1.
Table 1: Parameters and variables used to describe the sexual behavior and the epidemiological model.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Number of host classes with different sexual activity</td>
</tr>
<tr>
<td>c_i</td>
<td>Number of new heterosexual sex partners in the last year for individuals of host class i</td>
</tr>
<tr>
<td>m_i</td>
<td>Average number of heterosexual sex in the last 4 weeks for individuals of host class i</td>
</tr>
<tr>
<td>M_i</td>
<td>Average number of heterosexual sex in the last year for individuals of host class i</td>
</tr>
<tr>
<td>d_1, d_2, d_3</td>
<td>Parameters describing the functional relationship between c_i and m_i</td>
</tr>
<tr>
<td>x_i</td>
<td>Proportion of the host population in sexual activity class i</td>
</tr>
<tr>
<td>X_i</td>
<td>Number of individuals in sexual activity class i</td>
</tr>
<tr>
<td>y_i</td>
<td>Proportion of infected individuals in sexual activity class i</td>
</tr>
<tr>
<td>μ</td>
<td>Rate at which individuals change their sexual activity</td>
</tr>
<tr>
<td>γ</td>
<td>Recovery rate of the infection</td>
</tr>
<tr>
<td>ρ_{ij}</td>
<td>Mixing matrix between individuals of sexual activity classes i and j</td>
</tr>
<tr>
<td>ϵ</td>
<td>Sexual mixing coefficient ( assortative index)</td>
</tr>
<tr>
<td>δ_{ij}</td>
<td>Kronecker delta (a function equal to 1 if i = j, and to 0 otherwise)</td>
</tr>
<tr>
<td>s_i</td>
<td>Average number of sex acts per sex partner for an individual of host class i (( \frac{M_i}{c_i} ))</td>
</tr>
<tr>
<td>a_{ij}</td>
<td>Number of sex acts in a partnership between individuals of sexual activity classes i and j</td>
</tr>
<tr>
<td>β</td>
<td>Transmission probability per sex act</td>
</tr>
<tr>
<td>b_{ij}</td>
<td>Transmission probability per partnership between individuals of sexual activity classes i and j</td>
</tr>
</tbody>
</table>

Maximum likelihood estimation

Direct method

We estimate the parameters \( d_1, d_2 \) and \( d_3 \) that describe the functional relationship between the number of sex acts and sex partners (equation 1) and the dispersion parameter \( k \) from the negative binomial distribution using a maximum likelihood approach and the \( \text{mle2} \) function from the package \( \text{bbmle} \) (Bolker, 2008) for the R software environment for statistical computing (R Development Core Team, 2009).

Indirect method

We fit the epidemiological model to chlamydia prevalence data to obtain maximum likelihood estimates of behavioral and/or infection parameters (Bolker, 2008). Given a model-predicted prevalence \( y_i \) for sexual activity class i, the log-likelihood to find \( k_i \) positive tests in a sample of \( X_i \) individuals over all \( n \) sexual activity classes is

\[
L(\beta, \gamma, \epsilon, d_1, d_2, d_3) = \sum_{i=1}^{n} \left( \log \left( \frac{X_i}{k_i} \right) + k_i \log y_i + (X_i - k_i) \log(1 - y_i) \right). \tag{5}
\]

We obtain the model predicted prevalences for each sexual activity class, \( y_i \), by numerically simulating equation 2 to its steady-state using the function \( \text{ode} \) from the R package \( \text{deSolve} \) (Soetaert et al., 2010). The parameter inference was performed using the function \( \text{mle2} \) from the package \( \text{bbmle} \) (Bolker, 2008). We first used the minimization algorithm by Nelder and Mead (1965) that is very robust in finding local optima. For those fits where different starting values of the parameters resulted in multiple optima, we decided to use the method SANN. SANN is a variant of simulated annealing (Bélisle, 1992) and usually performs better on rough likelihood surfaces at the cost of being computationally expensive.
Sex acts and sex partners (direct method)

The reported numbers of heterosexual sex acts during the last 4 weeks in Natsal-2 show an increasing trend for higher number of new heterosexual partners during the last year (Fig. 1A). Fitting a series of nested models to the data suggests that a linear relationship between the number of sex acts, $m_i$, and the number of sex partners, $c_i$, describes the data best (Table 2 and Fig. 1B, solid line). The model where the number of sex acts for individuals with different numbers of sex partners is constant also fits the data well (Fig. 1B, dashed line). In contrast, the commonly used assumption where the total number of sex acts is strictly proportional to the number of sex partners is not consistent with the data (Fig. 1B, dotted line). We also fit an exponential relationship between the number of sex acts and sex partners (as suggested by Garnett and Anderson (1996)) to the data, but this resulted in a poorer fit ($\text{AIC} = 17380.1$) compared to the models in Table 2.

![Figure 1: Number of heterosexual sex acts during the last 4 weeks as a function of the number of new heterosexual partners during the last year. (A) Boxplot of the reported data in Natsal-2 ($n = 2824$). The number of sex acts are positively correlated with the number of sex partners (Pearson’s $r = 0.07; p < 10^{-3}$). (B) Functional relationship between the number of sex acts and the number of sex partners together with the means and standard errors of the reported data. The best fit model (Linear, solid line) is shown together with the models assuming the total number of sex acts (Constant, dashed line) or the number of sex acts per partner (Proportional, dotted line) to be constant.](image)

Parameter inference (indirect method)

We perform two series of model fitting to indirectly infer parameters from chlamydia prevalence data. First, we only consider epidemiological parameters: the per sex act transmission probability $\beta$, the infectious duration $1/\gamma$ and the sexual mixing coefficient $\epsilon$. Second, we focus on inferring the shape of the functional relationship between the number of sex acts and the number of sex partners (equation 1) and estimate $d_1$, $d_2$, $d_3$ and $\epsilon$. For the second approach, we assume that $\beta$ and $\gamma$ are fixed and informed by the literature. The rationale for choosing two separate approaches is that estimating all six parameters simultaneously would result in overfitting of the model.
Table 2: Maximum likelihood estimates of the functional relationship between the number of sex acts and the number of sex partners (direct method). Fixed parameters are given in squared brackets and estimates are shown together with 95% confidence intervals. The six different models were sorted in increasing order of their AIC. Only models whose AIC were within 10 from the best fit (Linear) model are shown together with the classical model where the number of sex acts increases strictly proportional to the number of sex partners. Based on likelihood ratio tests, the linear model is superior to the constant and full model ($p < 0.05$).

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>No. of free parameters</th>
<th>$-2 \text{LogLik}$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>$6.64$ (6.20, 7.09)</td>
<td>$0.17$ (0.01, 0.33)</td>
<td>[1.00]</td>
<td>$0.65$ (0.61, 0.69)</td>
<td>3</td>
<td>$16821.9$</td>
<td>0.0</td>
</tr>
<tr>
<td>Full</td>
<td>$6.77$ (6.29, 7.25)</td>
<td>$0.07$ (-0.14, 0.28)</td>
<td>$1.34$ (0.31, 2.36)</td>
<td>$0.65$ (0.61, 0.69)</td>
<td>4</td>
<td>$16821.4$</td>
<td>1.6</td>
</tr>
<tr>
<td>Exponent</td>
<td>$5.77$ (5.40, 6.15)</td>
<td>[1.00]</td>
<td>$0.39$ (0.13, 0.65)</td>
<td>$0.65$ (0.61, 0.69)</td>
<td>3</td>
<td>$16824.3$</td>
<td>2.4</td>
</tr>
<tr>
<td>No intercept</td>
<td>$[0.00]$</td>
<td>$6.81$ (6.42, 7.20)</td>
<td>$0.06$ (-0.01, 0.13)</td>
<td>$0.65$ (0.61, 0.69)</td>
<td>3</td>
<td>$16825.6$</td>
<td>3.7</td>
</tr>
<tr>
<td>Constant</td>
<td>$7.01$ (6.67, 7.34)</td>
<td>[0.00]</td>
<td>–</td>
<td>$0.65$ (0.61, 0.69)</td>
<td>2</td>
<td>$16828.4$</td>
<td>4.6</td>
</tr>
<tr>
<td>Proportional</td>
<td>$[0.00]$</td>
<td>$5.08$ (4.81, 5.35)</td>
<td>[1.00]</td>
<td>$0.54$ (0.51, 0.57)</td>
<td>2</td>
<td>$17296.0$</td>
<td>47.2</td>
</tr>
</tbody>
</table>

Epidemiological parameters

We fit two models to the Natsal-2 and CSF chlamydia prevalence data to estimate the epidemiological parameters. In one model, the average number of sex acts for individuals with a specific number of sex partners is given by the best fit model from the direct method (Linear). In the other model, we use the classical assumption where the number of sex acts increases strictly proportional with the number of sex partners. Both models fit the data well (Fig. 2) but vary in their estimates of the sexual mixing coefficient (Table 3). While the proportional model suggests that sexual mixing is proportionate ($\epsilon = 0$), the linear model estimates the mixing coefficient to be more assortative and significantly different from zero ($\epsilon = 0.83$ (95% CI: 0.46–0.96) and $\epsilon = 0.53$ (95% CI: 0.38–0.68) for Natsal-2 and CSF, respectively). The estimated recovery rates are consistent with an average infectious duration of about 1 year (Althaus et al., 2010; Price et al., 2013; Davies et al., 2014). The per sex act transmission probabilities for the linear model are somewhat lower than previous estimates (9.5%, interquartile range: 6.0%–16.7%) (Althaus et al., 2012b). This could be explained by the fact that the transmission probabilities in our model include protected and unprotected sexual intercourse. In contrast, the estimates of the per sex act transmission probability for the proportional model are very low, indicating that the number of sex acts are not realistically taken into account.

Table 3: Maximum likelihood estimates of epidemiological parameters. The relationship between the number of sex acts and the number of sex partners is either based on the best fit model from the direct method (Linear) or the classical assumption where the number of sex acts increases strictly Proportional to the number of sex partners. Estimates are shown together with 95% confidence intervals.

<table>
<thead>
<tr>
<th>Model</th>
<th>Data</th>
<th>Parameters</th>
<th>$\beta$</th>
<th>$1/\gamma$</th>
<th>$\epsilon$</th>
<th>No. of free parameters</th>
<th>$-2 \text{LogLik}$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Natsal-2</td>
<td>3.0% (1.9%, 4.1%)</td>
<td>$1.12$ y (0.83, 1.42)</td>
<td>$0.83$ (0.46, 0.96)</td>
<td>3</td>
<td>$39.3$</td>
<td>45.3</td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>CSF</td>
<td>3.4% (NA, NA)</td>
<td>$0.96$ y (NA, NA)</td>
<td>$0.53$ (0.38, 0.68)</td>
<td>3</td>
<td>$85.2$</td>
<td>91.2</td>
<td></td>
</tr>
<tr>
<td>Proportional</td>
<td>Natsal-2</td>
<td>0.7% (0.3%, 1.1%)</td>
<td>$1.16$ y (0.44, 1.88)</td>
<td>$0.00$ (0.00, 0.00)</td>
<td>3</td>
<td>$41.0$</td>
<td>47.0</td>
<td></td>
</tr>
<tr>
<td>Proportional</td>
<td>CSF</td>
<td>0.4% (0.2%, 0.5%)</td>
<td>$1.04$ y (0.45, 1.63)</td>
<td>$0.00$ (0.00, 0.00)</td>
<td>3</td>
<td>$105.7$</td>
<td>111.7</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 2**: Chlamydia prevalence and model fits. For both datasets, the model where the total number of sex acts increases *linearly* with the number of sex partners (black squares) provides a better fit to the data compared to the classical model where the number of sex acts increases strictly *proportional* with the number of sex partners (gray diamonds). Data points (circles) are shown together with the 95% binomial proportion confidence intervals.

**Functional relationship**

We infer the functional relationship of the number of sex acts and sex partners by fitting the model to the Natsal-2 and CSF chlamydia prevalence data simultaneously (Table 4). To this end, we set the per sex act transmission probability and the infectious duration to values similar as estimated in Table 3 ($\beta = 3\%$ and $1/\gamma = 1$ year). We find that the model where the number of sex acts does not increase with the number of partners (*Constant*) describes the data best. As in section , we again find evidence for assortative mixing ($\epsilon = 0.60$, 95% CI: 0.28–0.85). The estimates of $d_2$ or $d_3$ are close to zero for the model with *no intercept*, the *linear* model and the *full* model. This indicates that the number of sex acts does not substantially increase with higher numbers of partners. The *proportional* model clearly provides the worst fit to the data, and is the only model that suggests random sexual mixing. Interestingly, the number of estimated sex acts in the last four weeks is slightly higher than the estimate from the direct method (10.6 vs. 7.0 for the constant model, respectively). This could either reflect an under-estimation of the per sex transmission probability (set to $\beta = 3\%$) or an under-reporting of sex acts in Natsal-2.

Overall, our indirect method to infer the functional relationship between the number of sex acts and sex partners corroborates the results from the direct method. Models where the number of sex acts is *constant* or shows a slight *linear* increase with the number of sex partners provide the best and most parsimonious fit to the data. In contrast, the classical assumption of a strict *proportionality* between the total number of sex acts and the number of sex partners is clearly at odds with the data.
Table 4: Maximum likelihood estimates of the functional relationship between the number of sex acts and sex partners (indirect method). Fixed parameters are given in squared brackets and estimates are shown together with 95% confidence intervals. The five different models are sorted in increasing order of their AIC. Likelihood ratio tests show that the constant model is superior to all models with more parameters. The linear and no intercept model are favored over the full and proportional model. The models are fitted with initial parameter values from Table 2 and $\epsilon = 0.5$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$\epsilon$</th>
<th>No. of free parameters</th>
<th>-2LogLik</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.6 (8.54, 12.9)</td>
<td>[0.0]</td>
<td>[1.0]</td>
<td>0.60 (0.28, 0.85)</td>
<td>2</td>
<td>125.4</td>
<td>129.4</td>
<td></td>
</tr>
<tr>
<td>No Intercept</td>
<td>[0.0]</td>
<td>10.4 (6.44, 15.2)</td>
<td>5.1e-4 (0, 1.46)</td>
<td>0.63 (0.16, 0.94)</td>
<td>3</td>
<td>125.4</td>
<td>131.5</td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>10.0 (7.92, 12.4)</td>
<td>[0.0]</td>
<td>[1.0]</td>
<td>0.69 (0.34, 0.89)</td>
<td>3</td>
<td>125.8</td>
<td>131.8</td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>6.72 (2.54, 12.8)</td>
<td>3.97 (0.59, 10.2)</td>
<td>1.1e-4 (0, 0.17)</td>
<td>0.59 (0.31, 0.82)</td>
<td>4</td>
<td>125.4</td>
<td>133.4</td>
<td></td>
</tr>
<tr>
<td>Proportional</td>
<td>[0.0]</td>
<td>1.28 (1.23, 1.34)</td>
<td>[1.0]</td>
<td>7.6e-3 (9.8e-5, 0.38)</td>
<td>2</td>
<td>153.2</td>
<td>157.2</td>
<td></td>
</tr>
</tbody>
</table>

Chlamydia transmission and control

We now investigate the effects of different assumptions about the number of sex acts between partners on chlamydia transmission and control. We focus on the general population in Britain (Natsal-2) and use the linear relationship for the number of sex acts as a function of sex partners from Table 2. This ‘realistic’ model is contrasted to the classical assumption where the number of sex acts increases strictly proportionally with the number of sex partners. For both models, we use the best-fit estimates of the per sex act transmission probability $\beta$, the recovery rate $\gamma$ and the sexual mixing coefficient $\epsilon$ from Table 3.

For the linear model, distributing the number of sex acts to sex partners of different sexual activity classes generates the sex acts matrix $a_{ij}$ (Fig. 3A). From $a_{ij}$, we can then calculate the per partnership transmission probabilities $b_{ij}$ (equation 4), which range from 19.8% to 93.9% for partnerships between individuals with 10 and one partner(s), respectively (Fig. 3B). In contrast, the proportional model with a constant number of sex acts per partner results in a single per partnership transmission probability of 37.0%, which is close to what has been estimated for mid-risk populations Althaus et al. (2012b).

Differences in the per partnership transmission probability affect the transmission potential of individuals from different sexual activity classes, as reflected in the basic reproduction number ($R_0$) of each sexual activity class. In the classical scenario, where the number of sex acts per partner is constant and the sexual mixing between partners is fully proportional, $R_0$ is a linearly increasing function of the number of new heterosexual partners per year (Fig. 4A, gray diamonds). The proportional model further suggests that only individuals with three or more new heterosexual partners per year, or 21.2% of the individuals who had a new partner in the previous year, will infect more than one person ($R_0 > 1$). In the more ‘realistic’ linear model, the value of $R_0$ exceeds the threshold of one for all individuals with one or more new heterosexual partners in the last year, and saturates around three for higher number of partners (Fig. 4A, black squares). The differences between the models is also reflected in the value of $R_0$ for the entire population. In the classical model, the $R_0$ for chlamydia is 1.26 while it is 1.12 for the more ‘realistic’ model.

Screening for and treating of asymptomatic chlamydia infection is the primary strategy to prevent disease transmission (Althaus et al., 2014). In England, screening for women and men is recommended annually and after a change of sexual partner (Department of Health, 2004). Ideally, screening will decrease $R_0$ below the threshold of
Figure 3: Average number of sex acts (left panel) and per partnership transmission probabilities (right panel) between individuals of sexual activity class i and j. The matrices correspond to the best-fit (Linear) model from section .

Figure 4: Host-specific basic reproduction number (left panel) and required screening rate to push $R_0$ below one (right panel). The results from the best-fit linear model (black squares) are compared to the classical model where the number of sex acts increases strictly proportional with the number of sex partners (gray diamonds). Data are from Natsal-2 only. The classical model predicts that low screening rates are sufficient to prevent transmission in individuals with low numbers of partners but that more than three tests per year on average would be necessary to prevent transmission in individuals with many partners (right panel of Fig. 4, gray diamonds). Conversely, the ‘more realistic’ model illustrates that substantial screening (up to once every year) is already necessary to prevent transmission in individuals with low numbers of partners per year (right panel of
Discussion

This study provides insights into the relationship between the number of sex acts and the number of heterosexual partners, and how it affects the transmission dynamics of chlamydia and the projected impact of control interventions. We use direct and indirect methods to infer this relationship from two population-based probability sample surveys. Using a direct method based on self-reported values, we find evidence that the number of sex acts increases only marginally with the number of sex partners. This finding is corroborated using an indirect method based on chlamydia prevalence data, which allows us to show that a standard model where the number of sex acts increases strictly proportionally with the number of sex partners is less consistent than most other models. Noticeably, the most ‘realistic’ model, where the number of sex acts are taken into account, provides strong evidence for assortative mixing between individuals with different numbers of heterosexual partners. The improved model also suggests that all individuals with one or more new heterosexual partners in the last year contribute to chlamydia transmission ($R_0 > 1$).

A major strength of our approach is that our model formulation allows us to exploit two large datasets of sexual behavior. This gives us more power to perform model comparison and reject those models that are not consistent with the data. It is also encouraging that the two methods for inferring the relationship between the number of sex acts and the number of sex partners (one based on chlamydia prevalence data and the other on self-reported number of sex acts) are in general agreement with one another. Note that an advantage of the indirect method is that it only estimates potentially infectious, or unprotected, sex acts.

There are several limitations to our study. First, we pool heterosexual women and men together, and do not take age-specific sexual behavior and different ethnic groups into account. While the average number of heterosexual partners in women and men should be the same, men typically show a higher variability in the number of sex partners compared to women (Gouveia-Oliveira and Pedersen, 2009). Our simplifying assumption is driven by the aim to keep the number of parameters small and the number of data points large enough to be able to statistically compare different models. This is also the reason why our analysis is focused on heterosexuals only. Data from larger probability sample surveys could potentially improve these limiting aspects. However, stratifying the population into women and men in the analysis using the direct method does not reveal significant differences between the two sexes (results not shown). Some other known important factors that we do not include in the model are concurrency of sexual partnerships, condom use and potential sex-specific differences in infection parameters.

We extend a commonly used STI transmission model, that stratifies the population into different sexual activity classes, with a sex acts matrix that describes the average numbers of sex acts between individuals of different sexual activity classes. The entries of our sex act matrix are either directly or indirectly inferred from data. Others have used similar models but could not derive the number of sex acts per partner in such detail Regan et al. (2008); Garnett et al. (2008). Somewhat different modeling frameworks that investigate the importance of sex acts by means of weighted contact networks have been developed by Britton et al. (2007) and Moslonka-Lefebvre et al. (2012). These models are arguably more realistic than the more commonly used transmission models that stratify the population into different sexual activity classes. However, network models...
Have two major disadvantages. First, the heterosexual contact networks for large populations are usually not known. Second, we currently lack meaningful summary statistics for weighted networks, and model results tend to rely on computationally expensive simulations. An alternative extension of our analyses would be to use a recently developed framework, that manages to derive analytical results for weighted networks by adopting a configuration network approach and using joint probability distributions of number of sex partners and number of sex acts Kamp et al. (2013).

This work also brings new elements to the ongoing debate on the level of assortative mixing between heterosexual individuals of different sexual activity classes in the general population. Using the linear relationship between the number of sex acts and sex partners from the direct method, we estimate a sexual mixing coefficient of $\epsilon = 0.83$ for Natsal-2 (Table 3). An early study based on partner notification data indicated that a proportional mixing model is not an adequate description of the observed sexual mixing pattern Granath et al. (1991). Later, Renton et al. (1995) indicated that individuals with high rates of sexual partners preferentially select other individuals with high rates as partners. Garnett et al. (1996) also found that sexual mixing was weakly assortative in patients attending sexually transmitted diseases (STDs) clinics. In contrast, Stoner et al. (2000) did not find evidence for assortative mixing in members of gonorrhea and chlamydia networks. Finally, it is worth noting that we do not include age as an additional factor for assortative mixing. Young adults typically have more new sex partners than older individuals, and people prefer sex partners not too different in age. This automatically introduces a level of assortative mixing in sexual behavior. This illustrates the difficulties in quantifying the degree of sexual mixing, and that the estimates are likely to depend on the population studied.

Our estimates of the basic reproduction number ($R_0$) for chlamydia are consistent with values that were investigated in a recent theoretical study that introduced the concepts of case and partnership reproduction numbers Heijne et al. (2013). Another study based on contact tracing data showed that individual reproduction numbers for chlamydia rarely exceed 3.0 Potterat et al. (1999). Interestingly, this is in line with what we find for the host-specific basic reproduction numbers from the linear model (Fig. 4A, black squares), where $R_0$ saturates around 3.0 for those individuals with a high number of sex partners.

Finally, we show that taking into account the number of sex acts between partners with different levels of sexual activity can be important in guiding public health interventions, such as chlamydia screening programs. Indeed, a high level of assortative mixing between individuals of different heterosexual activity classes means that chlamydia can easily persist in those sub-populations that are difficult to reach through screening. Furthermore, our results also underline that chlamydia control interventions should be aimed towards all sexually active young adults as they are likely to contribute to ongoing transmission.

Competing interests

We declare no competing interests.
Authors’ contributions

CLA and SA designed the study, carried out the analysis, interpreted the results and drafted the manuscript. MC contributed to the analysis. All authors contributed to revising the manuscript and gave final approval for publication.

Acknowledgments

We thank Nathalie Bajos for helping with the CSF data.

Funding

CLA received funding through an Ambizione grant from the Swiss National Science Foundation (project 136737). SA is funded by an ATIP-Avenir from CNRS and INSERM. SA and MC are also supported by the CNRS and the IRD.

References


