# Peer Preprints

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

Christian L. Althaus<sup>1,†</sup>, Marc Choisy<sup>2,3</sup>, Samuel Alizon<sup>2,†</sup> and the CSF group<sup>\*</sup>

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

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

<sup>†</sup> Both authors contributed equally to this work.

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

<sup>\*</sup>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 Toulemon, Nicolas Razafindrastima, Josiane Warszawski, and also Lucette Aussel and Caroline Laborde.

# Peer Preprints

### Abstract

## NOT PEER-REVIEWED

Mathematical models are frequently used to assess the impact of control in-2 terventions for *Chlamydia trachomatis* and other sexually transmitted infections 3 (STIs). Modeling approaches that stratify the population by the number of sex Δ partners often assume the transmission risk per partner to be constant. How-5 ever, sexual behavior data suggests that people with many partners share less 6 sex acts per partner than people with fewer partners. This should lower the risk 7 of transmission per partner for highly sexually active individuals and could have 8 important epidemiological consequences for STI transmission. We devise a new 9 epidemiological model that we fit to chlamydia prevalence data from Natsal-2 10 and CSF, two population-based probability sample surveys of sexual behavior in 11 Britain and France. Compared to a standard model where the transmission risk 12 per partner is constant, a model with realistic numbers of sex acts per partner 13 provides a better fit to the data. Furthermore, the improved model provides ev-14 idence for strong assortative mixing among individuals with different numbers of 15 sex partners. Our results suggest that all chlamydia infected individuals with one 16 or more new heterosexual partners per year contribute significantly to ongoing 17 transmission, underlining that control interventions should be aimed towards all 18 sexually active young adults. 19

# Dea Introduction

## NOT PEER-REVIEWED

The spread of sexually transmitted infections (STIs) crucially depends on the number of 21 sex partners and the number of sex acts within a sexual partnership. Empirical studies 22 have shown that the number of sex partners within a given time period is a strong 23 determinant for the spread of curable STIs such as *Chlamydia trachomatis* (Fenton et al., 24 2001; Sonnenberg et al., 2013; Althaus et al., 2012c) and Neisseria gonorrhoeae (Ison 25 et al., 2013). However, the number of sex acts as an additional factor for transmission has 26 not received as much attention. In particular, mathematical models of STI transmission 27 that are frequently used in guiding public health policy decision making often assume 28 the number of sex acts per partner to be constant and independent of the number of 29 partners an individual has (Hethcote and Yorke, 1984; Anderson and May, 1991; Garnett 30 and Anderson, 1993b; Garnett et al., 1999; Keeling and Rohani, 2008). Neglecting 31 the number of sex acts per partner could lead to inaccurate estimates of the efficiency 32 of public health interventions, for instance when predicting the impact of screening 33 interventions on chlamydia spread. 34

Theoretical studies have shown that taking into account the number of sex acts in 35 models of STI transmission can dramatically alter the results (Kaplan, 1990; Røttingen 36 and Garnett, 2002). An important aspect was illustrated in the study by Nordvik 37 and Liljeros (2006), where they showed that it is not only the number of partners, 38 but also the number of sex acts per partner, that is important for the transmission of 39 an infection. Garnett and Anderson (1996) developed a generic framework to include 40 the heterogeneity in the number of sex acts between individuals and showed that it 41 can enhance the likelihood of persistence of STIs. Still, we lack a solid understanding 42 of the relationship between the number of sex acts and the number of sex partners. 43 Furthermore, it remains to be determined how this relationship can be appropriately 44 integrated into mathematical models of STI transmission that then can be validated 45 with data. 46

The results of probability sample surveys of sexual behavior allow us to formulate 47 some hypotheses on the relationship between the number of sex acts and the number 48 of sex partners. Blower and Boe (1993) analyzed data from 1770 unmarried individuals 49 aged 20–44 years collected in San Francisco (USA) from 1988–1989. They concluded that 50 it may be most appropriate to see the number of sex acts as a 'budget' split between 51 sex partners. Nordvik and Liljeros (2006) looked at the total number of sex acts in 52 relation to the total number of sex partners based on data from 1150 individuals aged 53 16–31 years collected in Sweden in 1988. The authors found that the total number of 54 sex acts remains constant in individuals with higher numbers of sex partners (it even 55 decreases in women). In principle, one can directly infer the relationship between sex 56 acts and sex partners from sexual behavior surveys, but there remain potential biases 57 due to the self-reported data. For example, it proves difficult to discriminate between 58 protected and unprotected sex acts in such surveys. A better understanding of the 59 quantitative relationship between the number of sex partners and the number of sex 60 acts could provide useful insights into the epidemiology of STIs and their control. 61

Another important aspect of taking the number of sex acts between sex partners into 62 account is its effect on the topology of the sexual contact network. If the number of sex 63 acts per sex partner is constant, each contact (edge) between susceptible and infected 64 individuals (nodes) is weighted equally, i.e., has the same transmission probability. This 65 is not the case if sex acts are explicitly accounted for (Moslonka-Lefebvre et al., 2012). 66 However, network-based models usually require computationally-intensive simulations 67

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 69 activity (Hethcote and Yorke, 1984; Garnett and Anderson, 1993b; Garnett et al., 1999). 70 These models are often referred to as 'risk class models', although the stratification is 71 based on the partner change rate that does not necessarily correlate with risk. In models, 72 the level of sexual mixing between individuals can vary from fully proportionate (to the 73 number of sex partners individuals have) to completely assortative (individuals within 74 a sexual activity or risk group mix only with themselves). While some studies based on 75 sexual behavior data indicate a certain level of assortative mixing between individuals 76 (Renton et al., 1995; Garnett and Anderson, 1993a; Garnett et al., 1996; Aral et al., 77 1999), the output of STI transmission models is often only consistent with close to 78 proportionate (or random) mixing (Garnett et al., 1999; Althaus et al., 2012a). This 79 inconsistency illustrates the challenge in estimating the degree of sexual mixing, and 80 we show that considering the number of sex acts per sex partner in models of STI 81 transmission sheds new light on this question. 82

Our goal is to gain a more detailed picture of the relationship between the number of 83 heterosexual sex acts and the number of heterosexual sex partners and study its effect on 84 the transmission of chlamydia, the most prevalent bacterial STI. To this end, we make 85 use of two datasets of sexual behavior: the British National Survey of Sexual Attitudes 86 and Lifestyles (Natsal-2) (Johnson et al., 2001), and the 'Contexte de la Sexualité en 87 France' (CSF 2006) (Bajos and Bozon, 2006, 2008). First, we directly infer the relation-88 ship between the number of heterosexual sex acts and the number of heterosexual sex 89 partners from Natsal-2. We then fit a novel epidemiological model of STI transmission 90 to chlamydia prevalence data of both datasets and indirectly infer epidemiological pa-91 rameters as well as the relationship between the number of heterosexual sex acts and 92 heterosexual sex partners. Both methods support the notion that the number of hetero-93 sexual sex acts per partner decreases with higher number of heterosexual sex partners. 94 This is also consistent with a high level of assortative mixing between individuals with 95 similar number of sex partners. Lastly, we show that taking realistic numbers of sex acts 96 into account when modeling the heterosexual transmission dynamics of chlamydia has 97 a crucial effect on the projected impact of control interventions. 98

## 99 Methods

### 100 Data

### 101 Natsal-2

Natsal-2 is a population-based probability sample survey of sexual attitudes and lifestyles 102 conducted in Britain in 1999-2001 (Johnson et al., 2001). The sample consists of 11,161 103 women and men aged 16–44 years. Urine samples for ligase chain reaction (LCR) testing 104 for chlamydia infection are available for a subset of 3569 sexually active respondents 105 aged 18–44 years (Fenton et al., 2001). We use the following variables: number of new 106 heterosexual sex partners in the last year, number of occasions of heterosexual sex in 107 last 4 weeks, chlamydia test result from urine sample. Individual weights are used for 108 all variables to adjust for unequal selection probabilities and to correct for the age and 109 gender profile in the survey population. The full dataset is available from the UK Data 110 Archive at the University of Essex (http://data-archive.ac.uk). 111

# **C**<sub>112</sub>**CSF**Preprints

'Contexte de la Sexualité en France' (CSF 2006) is a survey conducted in France among 113 12,364 randomly chosen individuals aged 18–69 years (Bajos and Bozon, 2006, 2008). 114 Urine samples testing for chlamydia infection are available for a subset of 7407 sexually 115 active respondents. The two variables we used are the number of new heterosexual sex 116 partners in the last year and the chlamydia test result from urine sample (see Supple-117 mental Files). 118

#### Sex acts and sex partners 119

We stratify the population by the reported number of new heterosexual partners in the 120 last year which results in n different sexual activity classes with  $c_i$  sex partners. The 121 average heterosexual activity of each individual during one year is assumed to be reflected 122 in the reported number of heterosexual sex acts during the last 4 weeks. For simplicity, 123 and in order to increase sample size, we pool data of women and men together. 124

The reported numbers of sex acts during the last 4 weeks are highly dispersed 125 (Fig. 1A). Therefore, we assume that the observed number of sex acts during the last 126 4 weeks (Z) for a given individual with  $c_i$  new heterosexual partners during the last 127 year follows a negative binomial distribution. This distribution is frequently used to 128 describe overdispersed data in biology and epidemiology (Lloyd-Smith, 2007; Hamilton 129 et al., 2008). In the context of data on the number of sex acts, it can be interpreted 130 as a process of sequential Bernoulli trials to reach 'sexual satisfaction' with the param-131 eters (k,p): for each of Z sex acts there is a probability p that it will not be sexually 132 satisfying and individuals engage in new sex acts until satisfaction has been reached in 133 k of them. The negative binomial distribution can also be interpreted as a mixture of 134 Poisson distributions where the rates at which different individuals have sex are drawn 135 from a gamma distribution with shape parameter k and scale parameter p/(1-p). More 136 generally, a single Poisson distribution is obtained for  $k \to \infty$  and the negative binomial 137 distribution is reduced to a geometric distribution for k = 1. 138

We assume that the mean of the negative binomial distribution is given by the 139 following functional relationship: 140

$$m_i = d_1 + d_2 \ c_i^{d_3},\tag{1}$$

where  $m_i$  and  $c_i$  are the average number of sex acts and the number of sex partners for 141 individuals of sexual activity class i, respectively.  $m_i$  can then be scaled to obtain the 142 expected number of heterosexual sex acts per year  $(M_i)$ . 143

We construct a series of nested models by fixing some of the parameters to specific 144 values. For example, if  $d_2 = 0$ , the total number of sex acts does not increase with 145 higher numbers of sex partners. For  $d_3 = 1$ , the number of sex acts changes linearly 146 with higher numbers of partners. The nested models then allow to perform hypothesis 147 testing about different functional relationships between the number of sex acts and sex 148 partners. 149

#### Epidemiological model 150

We assume the population to be stratified into n different sexual activity classes  $x_i$ 151 with  $c_i$  new heterosexual partners per year (Hethcote and Yorke, 1984; Garnett and 152 Anderson, 1993b).  $x_i$  denotes the proportion of individuals in sexual activity class i. On 153 average, individuals change their sexual activity after one year, and are proportionally 154

155 redistributed 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 156

and men. If  $y_i$  is the proportion of infected individuals in sexual activity class *i*, an SIS 157 (susceptible-infected-susceptible) model (i.e., with no immunity) can then be written as 158 follows: 159

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = \mu \sum_{j=1}^n x_j y_j + (1 - y_i) c_i \sum_{j=1}^n b_{ij} \rho_{ij} y_j - (\gamma + \mu) y_i, \tag{2}$$

where  $1/\gamma$  and  $1/\mu$  denote the average duration of infection and the average time spent 160 in one sexual activity class, respectively.  $\rho_{ij}$  represents the elements of the mixing matrix 161 that, following Garnett et al. (1999), can be defined as 162

$$\rho_{ij} = \epsilon \delta_{ij} + (1 - \epsilon) \frac{c_j x_j}{\sum_{l=1}^n c_l x_l},\tag{3}$$

where  $\delta_{ij}$  denotes the Kronecker delta (it is equal to 1 if i = j and to 0 otherwise). 163 Mixing can be varied between proportionate ( $\epsilon = 0$ ) and fully assortative ( $\epsilon = 1$ ). 164

The per partnership transmission probability  $b_{ij}$  can be considered as a Bernoulli 165 process of repeated transmission probabilities per sex act: 166

$$b_{ij} = 1 - (1 - \beta)^{a_{ij}} \tag{4}$$

where  $\beta$  denotes the transmission probability per sex act and  $a_{ij}$  the number of sex acts 167 in a partnership between an individual of sexual activity class i and j. To construct the 168 matrix  $a_{ij}$ , we use the following algorithm to distribute the number of sex acts between 169 partners: 170

- 1. Define  $s_i = \frac{M_i}{c_i}$  as the average number of sex acts per partner for an individual of sexual activity class *i*. 171 172
- 2. Set i = n, i.e., start distributing the number of sex acts from individuals of the 173 highest sexual activity class. 174
- 3. The number of sex acts from an individual of sexual activity class i will be shared 175 with individuals of the same and lower sexual activity classes j = [1, i]. Hence, we 176 distribute the number of sex acts  $s_i$  proportional to the probability of a contact 177 between an individual of class i and class j  $(\rho_{ij})$  and to what the partners are 178 willing to share  $(s_i)$ . 179
- 4. The number of sex acts in individuals of sexual activity class j = [1, i] that remain 180 to be shared  $(s_i)$  need to be updated to only those sex acts that have not been 181 shared with individuals of sexual activity class i or higher. 182

5. If i > 1, set i = i - 1 and go to step 3. Otherwise, stop. 183

We calculate the basic reproduction number,  $R_0$ , for a given sexual activity class as 184 follows:  $\frac{c_i \sum_{j=1}^n (b_{ij} \rho_{ij})}{\gamma}$ .  $R_0$  for the entire population is calculated using the next generation 185 method (Diekmann et al., 1990; Diekmann and Heesterbeek, 2000). An overview of all 186 parameters and variables to describe the epidemiological model as well as the sexual 187 behavior is given in Table 1. 188

# **Pee Table 1:** Parameters and variables used to describe the sexual behavior and the epidemiolog-

Notation	Description
$\overline{n}$	Number of host classes with different sexual activity
$c_i$	Number of new heterosexual sex partners in the last year for individuals of host class $i$
$m_i$	Average number of heterosexual sex in the last 4 weeks for individuals of host class $i$
$M_i$	Average number of heterosexual sex in the last year for individuals of host class $i$
$d_1, d_2, d_3$	Parameters describing the functional relationship between $c_i$ and $m_i$
$x_i$	Proportion of the host population in sexual activity class $i$
$X_i$	Number of individuals in sexual activity class $i$
$y_i$	Proportion of infected individuals in sexual activity class $i$
$\mu$	Rate at which individuals change their sexual activity
$\gamma$	Recovery rate of the infection
$ ho_{ij}$	Mixing matrix between individuals of sexual activity classes $i$ and $j$
$\epsilon$	Sexual mixing coefficient (assortative index)
$\delta_{ij}$	Kronecker delta (a function equal to 1 if $i = j$ , and to 0 otherwise)
$s_i$	Average number of sex acts per sex partner for an individual of host class $i\left(\frac{M_i}{c_i}\right)$
$a_{ij}$	Number of sex acts in a partnership between individuals of sexual activity classes $i$ and $j$
$\beta$	Transmission probability per sex act
$b_{ij}$	Transmission probability per partnership between individuals of sexual activity classes $i$ and $j$

### 189 Maximum likelihood estimation

### <sup>190</sup> 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 mle2 function from the package bbmle (Bolker, 2008) for the R software environment for statistical computing (R Development Core Team, 2009).

### <sup>196</sup> Indirect method

<sup>197</sup> We fit the epidemiological model to chlamydia prevalence data to obtain maximum <sup>198</sup> likelihood estimates of behavioral and/or infection parameters (Bolker, 2008). Given <sup>199</sup> a model-predicted prevalence  $y_i$  for sexual activity class *i*, the log-likelihood to find  $k_i$ <sup>200</sup> positive tests in a sample of  $X_i$  individuals over all *n* sexual activity classes is

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

We obtain the model predicted prevalences for each sexual activity class,  $y_i$ , by 201 numerically simulating equation 2 to its steady-state using the function ode from the R 202 package deSolve (Soetaert et al., 2010). The parameter inference was performed using 203 the function mle2 from the package bbmle (Bolker, 2008). We first used the minimization 204 algorithm by Nelder and Mead (1965) that is very robust in finding local optima. For 205 those fits where different starting values of the parameters resulted in multiple optima, 206 we decided to use the method SANN. SANN is a variant of simulated annealing (Bélisle, 207 1992) and usually performs better on rough likelihood surfaces at the cost of being 208 computationally expensive. 209

# 210 Results

#### Sex acts and sex partners (direct method) 211

The reported numbers of heterosexual sex acts during the last 4 weeks in Natsal-2 show 212 an increasing trend for higher number of new heterosexual partners during the last year 213 (Fig. 1A). Fitting a series of nested models to the data suggests that a *linear* relationship 214 between the number of sex acts,  $m_i$ , and the number of sex partners,  $c_i$ , describes the 215 data best (Table 2 and Fig. 1B, solid line). The model where the number of sex acts 216 for individuals with different numbers of sex partners is *constant* also fits the data 217 well (Fig. 1B, dashed line). In contrast, the commonly used assumption where the total 218 number of sex acts is strictly *proportional* to the number of sex partners is not consistent 219 with the data (Fig. 1B, dotted line). We also fit an exponential relationship between the 220 number of sex acts and sex partners (as suggested by Garnett and Anderson (1996)) to 221 the data, but this resulted in a poorer fit (AIC = 17380.1) compared to the models in 222 Table 2. 223



**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.

#### Parameter inference (indirect method) 224

We perform two series of model fitting to indirectly infer parameters from chlamydia 225 prevalence data. First, we only consider epidemiological parameters: the per sex act 226 transmission probability  $\beta$ , the infectious duration  $1/\gamma$  and the sexual mixing coefficient 227  $\epsilon$ . Second, we focus on inferring the shape of the functional relationship between the 228 number of sex acts and the number of sex partners (equation 1) and estimate  $d_1, d_2, d_3$ 229 and  $\epsilon$ . For the second approach, we assume that  $\beta$  and  $\gamma$  are fixed and informed by the 230 literature. The rationale for choosing two separate approaches is that estimating all six 231 parameters simultaneously would result in overfitting of the model. 232

**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).

Model	Parameters				No. of free	-2LogLik	$\Delta AIC$
	$d_1$	$d_2$	$d_3$	k	parameters		
Linear	6.64(6.20, 7.09)	$0.17 \ (0.01, \ 0.33)$	[1.00]	$0.65\ (0.61,\ 0.69)$	3	16821.9	0.0
Full	6.77 (6.29, 7.25)	0.07 (-0.14, 0.28)	1.34(0.31, 2.36)	$0.65\ (0.61,\ 0.69)$	4	16821.4	1.6
Exponent	5.77(5.40, 6.15)	[1.00]	0.39(0.13, 0.65)	0.65(0.61, 0.69)	3	16824.3	2.4
No intercept	[0.00]	$6.81 \ (6.42, \ 7.20)$	0.06(-0.01, 0.13)	0.65(0.61, 0.69)	3	16825.6	3.7
Constant	7.01(6.67, 7.34)	[0.00]	_	0.65(0.61, 0.69)	2	16828.4	4.6
Proportional	[0.00]	5.08(4.81, 5.35)	[1.00]	0.54(0.51, 0.57)	2	17296.0	472.2

### 233 Epidemiological parameters

We fit two models to the Natsal-2 and CSF chlamydia prevalence data to estimate the 234 epidemiological parameters. In one model, the average number of sex acts for individuals 235 with a specific number of sex partners is given by the best fit model from the direct 236 method (*Linear*). In the other model, we use the classical assumption where the number 237 of sex acts increases strictly *proportional* with the number of sex partners. Both models 238 fit the data well (Fig. 2) but vary in their estimates of the sexual mixing coefficient (Table 239 3). While the proportional model suggests that sexual mixing is proportionate ( $\epsilon = 0$ ), 240 the *linear* model estimates the mixing coefficient to be more assortative and significantly 241 different from zero ( $\epsilon = 0.83$  (95% CI: 0.46–0.96) and  $\epsilon = 0.53$  (95% CI: 0.38–0.68) for 242 Natsal-2 and CSF, respectively). The estimated recovery rates are consistent with an 243 average infectious duration of about 1 year (Althaus et al., 2010; Price et al., 2013; 244 Davies et al., 2014). The per sex act transmission probabilities for the *linear* model 245 are somewhat lower than previous estimates (9.5%, interquartile range: 6.0%-16.7%)246 (Althaus et al., 2012b). This could be explained by the fact that the transmission 247 probabilities in our model include protected and unprotected sexual intercourse. In 248 contrast, the estimates of the per sex act transmission probability for the *proportional* 249 model are very low, indicating that the number of sex acts are not realistically taken 250 into account. 251

**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.

Model	Data	Parameters			No. of free	-2LogLik	AIC
		$\beta$	$1/\gamma$	$\epsilon$	parameters		
Linear	Natsal-2	3.0% (1.9%, 4.1%)	1.12 y (0.83, 1.42)	$0.83 \ (0.46, \ 0.96)$	3	39.3	45.3
Linear	CSF	3.4% (NA, NA)	0.96 y (NA, NA)	$0.53 \ (0.38, \ 0.68)$	3	85.2	91.2
Proportional	Natsal-2	$0.7\% \ (0.3\%, \ 1.1\%)$	1.16 y (0.44, 1.88)	$0.00 \ (0.00, \ 0.00)$	3	41.0	47.0
Proportional	CSF	0.4%~(0.2%,~0.5%)	1.04 y (0.45, 1.63)	0.00(0.00, 0.00)	3	105.7	111.7



**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.

### <sup>252</sup> Functional relationship

We infer the functional relationship of the number of sex acts and sex partners by fitting 253 the model to the Natsal-2 and CSF chlamydia prevalence data simultaneously (Table 4). 254 To this end, we set the per sex act transmission probability and the infectious duration 255 to values similar as estimated in Table 3 ( $\beta = 3\%$  and  $1/\gamma = 1$  year). We find that 256 the model where the number of sex acts does not increase with the number of partners 257 (Constant) describes the data best. As in section, we again find evidence for assortative 258 mixing ( $\epsilon = 0.60, 95\%$  CI: 0.28–0.85). The estimates of  $d_2$  or  $d_3$  are close to zero for the 259 model with no intercept, the linear model and the full model. This indicates that the 260 number of sex acts does not substantially increase with higher numbers of partners. The 261 proportional model clearly provides the worst fit to the data, and is the only model that 262 suggests random sexual mixing. Interestingly, the number of estimated sex acts in the 263 last four weeks is slightly higher than the estimate from the direct method (10.6 vs. 7.0 264 for the constant model, respectively). This could either reflect an under-estimation of 265 the per sex transmission probability (set to  $\beta = 3\%$ ) or an under-reporting of sex acts 266 in Natsal-2. 267

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 IEWED 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  $\epsilon = 0.5$ .

Model	Parameters				No. of free	-2LogLik	AIC
	$d_1$	$d_2$	$d_3$	$\epsilon$	parameters		
Constant	10.6 (8.54, 12.9)	[0.0]	[1.0]	$0.60 \ (0.28, \ 0.85)$	2	125.4	129.4
No Intercept	[0.0]	$10.4 \ (6.44, \ 15.2)$	5.1e-4 (0, 1.46)	0.63 (0.16, 0.94)	3	125.4	131.5
Linear	$10.0\ (7.92,\ 12.4)$	4.6e-3 (0, 0.15)	[1.0]	0.69(0.34, 0.89)	3	125.8	131.8
Full	6.72(2.54, 12.8)	3.97(0.59, 10.2)	1.1e-4 (0, 0.17)	0.59(0.31, 0.82)	4	125.4	133.4
Proportional	[0.0]	1.28(1.23, 1.34)	[1.0]	7.6e-3 (9.8e-5, 0.38)	2	153.2	157.2

#### Chlamydia transmission and control 274

We now investigate the effects of different assumptions about the number of sex acts 275 between partners on chlamydia transmission and control. We focus on the general pop-276 ulation in Britain (Natsal-2) and use the *linear* relationship for the number of sex acts 277 as a function of sex partners from Table 2. This 'realistic' model is contrasted to the 278 classical assumption where the number of sex acts increases strictly proportionally with 279 the number of sex partners. For both models, we use the best-fit estimates of the per 280 sex act transmission probability  $\beta$ , the recovery rate  $\gamma$  and the sexual mixing coefficient 281  $\epsilon$  from Table 3. 282

For the *linear* model, distributing the number of sex acts to sex partners of different 283 sexual activity classes generates the sex acts matrix  $a_{ij}$  (Fig. 3A). From  $a_{ij}$ , we can then 284 calculate the per partnership transmission probabilities  $b_{ij}$  (equation 4), which range 285 from 19.8% to 93.9% for partnerships between individuals with 10 and one partner(s), 286 respectively (Fig. 3B). In contrast, the proportional model with a constant number of 287 sex acts per partner results in a single per partnership transmission probability of 37.0%, 288 which is close to what has been estimated for mid-risk populations Althaus et al. (2012b). 289

Differences in the per partnership transmission probability affect the transmission 290 potential of individuals from different sexual activity classes, as reflected in the basic 291 reproduction number  $(R_0)$  of each sexual activity class. In the classical scenario, where 292 the number of sex acts per partner is constant and the sexual mixing between partners is 293 fully proportional,  $R_0$  is a linearly increasing function of the number of new heterosexual 294 partners per year (Fig. 4A, gray diamonds). The proportional model further suggests 295 that only individuals with three or more new heterosexual partners per year, or 21.2%296 of the individuals who had a new partner in the previous year, will infect more than 297 one person  $(R_0 > 1)$ . In the more 'realistic' *linear* model, the value of  $R_0$  exceeds the 298 threshold of one for all individuals with one or more new heterosexual partners in the 299 last year, and saturates around three for higher number of partners (Fig. 4A, black 300 squares). The differences between the models is also reflected in the value of  $R_0$  for the 301 entire population. In the classical model, the  $R_0$  for chlamydia is 1.26 while it is 1.12 302 for the more 'realistic' model. 303

Screening for and treating of asymptomatic chlamydia infection is the primary strat-304 egy to prevent disease transmission (Althaus et al., 2014). In England, screening for 305 women and men is recommended annually and after a change of sexual partner (De-306 partment of Health, 2004). Ideally, screening will decrease  $R_0$  below the threshold of 307

## NOT PEER-REVIEWED



Peer Preprints

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.

one. The classical model predicts that low screening rates are sufficient to prevent trans-308 mission in individuals with low numbers of partners but that more than three tests per 309 year on average would be necessary to prevent transmission in individuals with many 310 partners (right panel of Fig. 4, gray diamonds). Conversely, the 'more realistic' model 311 illustrates that substantial screening (up to once every year) is already necessary to pre-312 vent transmission in individuals with low numbers of partners per year (right panel of 313

#### Discussion 315

This study provides insights into the relationship between the number of sex acts and 316 the number of heterosexual partners, and how it affects the transmission dynamics of 317 chlamydia and the projected impact of control interventions. We use direct and indi-318 rect methods to infer this relationship from two population-based probability sample 319 surveys. Using a direct method based on self-reported values, we find evidence that the 320 number of sex acts increases only marginally with the number of sex partners. This 321 finding is corroborated using an indirect method based on chlamydia prevalence data, 322 which allows us to show that a standard model where the number of sex acts increases 323 strictly proportionally with the number of sex partners is less consistent than most other 324 models. Noticeably, the most 'realistic' model, where the number of sex acts are taken 325 into account, provides strong evidence for assortative mixing between individuals with 326 different numbers of heterosexual partners. The improved model also suggests that all 327 individuals with one or more new heterosexual partners in the last year contribute to 328 chlamydia transmission  $(R_0 > 1)$ . 329

A major strength of our approach is that our model formulation allows us to exploit 330 two large datasets of sexual behavior. This gives us more power to perform model 331 comparison and reject those models that are not consistent with the data. It is also 332 encouraging that the two methods for inferring the relationship between the number of 333 sex acts and the number of sex partners (one based on chlamydia prevalence data and the 334 other on self-reported number of sex acts) are in general agreement with one another. 335 Note that an advantage of the indirect method is that it only estimates potentially 336 infectious, or unprotected, sex acts. 337

There are several limitations to our study. First, we pool heterosexual women and 338 men together, and do not take age-specific sexual behavior and different ethnic groups 339 into account. While the average number of heterosexual partners in women and men 340 should be the same, men typically show a higher variability in the number of sex partners 341 compared to women (Gouveia-Oliveira and Pedersen, 2009). Our simplifying assumption 342 is driven by the aim to keep the number of parameters small and the number of data 343 points large enough to be able to statistically compare different models. This is also the 344 reason why our analysis is focused on heterosexuals only. Data from larger probability 345 sample surveys could potentially improve these limiting aspects. However, stratifying 346 the population into women and men in the analysis using the direct method does not 347 reveal significant differences between the two sexes (results not shown). Some other 348 known important factors that we do not include in the model are concurrency of sexual 349 partnerships, condom use and potential sex-specific differences in infection parameters. 350

We extend a commonly used STI transmission model, that stratifies the population 351 into different sexual activity classes, with a sex acts matrix that describes the average 352 numbers of sex acts between individuals of different sexual activity classes. The entries 353 of our sex act matrix are either directly or indirectly inferred from data. Others have 354 used similar models but could not derive the number of sex acts per partner in such de-355 tail Regan et al. (2008); Garnett et al. (2008). Somewhat different modeling frameworks 356 that investigate the importance of sex acts by means of weighted contact networks have 357 been developed by Britton et al. (2007) and Moslonka-Lefebvre et al. (2012). These mod-358 els are arguably more realistic than the more commonly used transmission models that 359 stratify the population into different sexual activity classes. However, network models 360

<sup>361</sup> have two major disadvantages. First, the heterosexual contact networks for large popu-<sup>362</sup> lations 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 368 mixing between heterosexual individuals of different sexual activity classes in the gen-369 eral population. Using the *linear* relationship between the number of sex acts and sex 370 partners from the direct method, we estimate a sexual mixing coefficient of  $\epsilon = 0.83$  for 371 Natsal-2 (Table 3). An early study based on partner notification data indicated that a 372 proportional mixing model is not an adequate description of the observed sexual mixing 373 pattern Granath et al. (1991). Later, Renton et al. (1995) indicated that individuals 374 with high rates of sexual partners preferentially select other individuals with high rates 375 as partners. Garnett et al. (1996) also found that sexual mixing was weakly assortative 376 in patients attending sexually transmitted diseases (STDs) clinics. In contrast, Stoner 377 et al. (2000) did not find evidence for assortative mixing in members of gonorrhea and 378 chlamydia networks. Finally, it is worth noting that we do not include age as an addi-379 tional factor for assortative mixing. Young adults typically have more new sex partners 380 than older individuals, and people prefer sex partners not too different in age. This au-381 tomatically introduces a level of assortative mixing in sexual behavior. This illustrates 382 the difficulties in quantifying the degree of sexual mixing, and that the estimates are 383 likely to depend on the population studied. 384

Our estimates of the basic reproduction number  $(R_0)$  for chlamydia are consistent 385 with values that were investigated in a recent theoretical study that introduced the 386 concepts of case and partnership reproduction numbers Heijne et al. (2013). Another 387 study based on contact tracing data showed that individual reproduction numbers for 388 chlamydia rarely exceed 3.0 Potterat et al. (1999). Interestingly, this is in line with what 389 we find for the host-specific basic reproduction numbers from the *linear* model (Fig. 4A, 390 black squares), where  $R_0$  saturates around 3.0 for those individuals with a high number 391 of sex partners. 392

Finally, we show that taking into account the number of sex acts between part-393 ners with different levels of sexual activity can be important in guiding public health 394 interventions, such as chlamydia screening programs. Indeed, a high level of assorta-395 tive mixing between individuals of different heterosexual activity classes means that 396 chlamydia can easily persist in those sub-populations that are difficult to reach through 397 screening. Furthermore, our results also underline that chlamydia control interventions 398 should be aimed towards all sexually active young adults as they are likely to contribute 399 to ongoing transmission. 400

## **401** Competing interests

<sup>402</sup> We declare no competing interests.

# **Peas Authors' contributions**

404 CLA and SA designed the study, carried out the analysis, interpreted the results and 405 drafted the manuscript. MC contributed to the analysis. All authors contributed to 406 revising the manuscript and gave final approval for publication.

## 407 Acknowledgments

408 We thank Nathalie Bajos for helping with the CSF data.

# 409 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.

## 413 **References**

Althaus, C., Turner, K., Mercer, C., Auguste, P., Roberts, T., Bell, G., Herzog, S.,
Cassell, J., Edmunds, J., White, P., Ward, H., and Low, N. (2014). Effectiveness and
cost-effectiveness of traditional and new partner notification technologies for curable
sexually transmitted infections: observational study, systematic reviews and mathematical modelling. *Health Technol Assess*, 18(2).

Althaus, C. L., Heijne, J. C., Herzog, S. A., Roellin, A., and Low, N. (2012a). Individual
and population level effects of partner notification for Chlamydia trachomatis. *PLoS*One, 7(12):e51438.

Althaus, C. L., Heijne, J. C., Roellin, A., and Low, N. (2010). Transmission dynamics
of Chlamydia trachomatis affect the impact of screening programmes. *Epidemics*,
2(3):123 - 131.

Althaus, C. L., Heijne, J. C. M., and Low, N. (2012b). Towards More Robust Estimates
of the Transmissibility of Chlamydia trachomatis. Sex Transm Dis, 39(5):402–4.

Althaus, C. L., Turner, K. M. E., Schmid, B. V., Heijne, J. C. M., Kretzschmar, M., and
Low, N. (2012c). Transmission of Chlamydia trachomatis through sexual partnerships:
a comparison between three individual-based models and empirical data. J R Soc
Interface, 9(66):136–146.

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

Aral, S. O., Hughes, J. P., Stoner, B., Whittington, W., Handsfield, H. H., Anderson,
R. M., and Holmes, K. K. (1999). Sexual mixing patterns in the spread of gonococcal
and chlamydial infections. Am J Public Health, 89(6):825–33.

Bajos, N. and Bozon, M. (2006). Enquête Contexte de la Sexualité en France, 2006.
Technical report, INSERM-INED.

#### Bajos, N. and Bozon, M., editors (2008). Enquête sur la sexualité en France: pratiques, 438 genre et santé. La Découverte, Paris. 439

Bélisle, C. J. P. (1992). Convergence Theorems for a Class of Simulated Annealing 440 Algorithms on Rd. Journal of Applied Probability, 29(4):pp. 885–895. 441

Blower, S. M. and Boe, C. (1993). Sex acts, sex partners, and sex budgets: implications 442 for risk factor analysis and estimation of HIV transmission probabilities. J Acquir 443 Immune Defic Syndr, 6(12):1347–52. 444

Bolker, B. M. (2008). *Ecological models and data in R.* Princeton University Press, 445 Princeton, NJ. 446

Britton, T., Nordvik, M. K., and Liljeros, F. (2007). Modelling sexually transmitted 447 infections: the effect of partnership activity and number of partners on R0. Theor 448 Popul Biol, 72(3):389–99. 449

Davies, B., Anderson, S.-J., Turner, K. M. E., and Ward, H. (2014). How robust are 450 the natural history parameters used in chlamydia transmission dynamic models? A 451 systematic review. Theor Biol Med Model, 11:8. 452

Department of Health (2004). National Chlamydia Screening Programme (NCSP) in 453 England: Programme overview; core requirements; data collection. London, 2nd edi-454 tion. 455

Diekmann, O., Heesterbeek, J. A., and Metz, J. A. (1990). On the definition and the 456 computation of the basic reproduction ratio R0 in models for infectious diseases in 457 heterogeneous populations. J Math Biol, 28(4):365–82. 458

Diekmann, O. and Heesterbeek, J. A. P. (2000). Mathematical Epidemiology of Infectious 459 Diseases: Model Building, Analysis, and Interpretation. John Wiley, Chichester. 460

Fenton, K. A., Korovessis, C., Johnson, A. M., McCadden, A., McManus, S., Wellings, 461 K., Mercer, C. H., Carder, C., Copas, A. J., Nanchahal, K., Macdowall, W., Ridgway, 462 G., Field, J., and Erens, B. (2001). Sexual behaviour in Britain: reported sexually 463 transmitted infections and prevalent genital Chlamydia trachomatis infection. Lancet, 464 358(9296):1851-1854.465

Garnett, G. P. and Anderson, R. M. (1993a). Contact tracing and the estimation of 466 sexual mixing patterns: the epidemiology of gonococcal infections. Sex Transm Dis, 467 20(4):181-91.468

Garnett, G. P. and Anderson, R. M. (1993b). Factors controlling the spread of HIV in 469

heterosexual communities in developing countries: patterns of mixing between differ-470 ent age and sexual activity classes. Philos Trans R Soc Lond B Biol Sci, 342(1300):137-471 59. 472

- Garnett, G. P. and Anderson, R. M. (1996). Sexually transmitted diseases and sexual 473 behavior: insights from mathematical models. J Infect Dis, 174 Suppl 2:S150-61. 474
- Garnett, G. P., Hughes, J. P., Anderson, R. M., Stoner, B. P., Aral, S. O., Whittington, 475 W. L., Handsfield, H. H., and Holmes, K. K. (1996). Sexual mixing patterns of patients 476 attending sexually transmitted diseases clinics. Sex Transm Dis, 23(3):248–57. 477

Garnett, G. P., Mertz, K. J., Finelli, L., Levine, W. C., and St Louis, M. E. (1999). The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from Newark, New Jersey. *Philos Trans R Soc Lond B Biol Sci*,

 $_{481}$  354(1384):787–97.

482 Garnett, G. P., White, P. J., and Ward, H. (2008). Fewer partners or more condoms?
483 Modelling the effectiveness of STI prevention interventions. Sex Transm Infect, 84
484 Suppl 2:ii4–11.

Gouveia-Oliveira, R. and Pedersen, A. G. (2009). Higher variability in the number of
sexual partners in males can contribute to a higher prevalence of sexually transmitted
diseases in females. J Theor Biol, 261(1):100–106.

Granath, F., Giesecke, J., Scalia-Tomba, G., Ramstedt, K., and Forssman, L. (1991).
Estimation of a preference matrix for women's choice of male sexual partner according
to rate of partner change, using partner notification data. *Math Biosci*, 107(2):341–8.

Hamilton, D. T., Handcock, M. S., and Morris, M. (2008). Degree distributions in sexual
networks: a framework for evaluating evidence. Sex Transm Dis, 35(1):30–40.

Heijne, J. C. M., Herzog, S. A., Althaus, C. L., Low, N., and Kretzschmar, M. (2013).
Case and partnership reproduction numbers for a curable sexually transmitted infec-*L* Theor. Biol. 221:28, 47

495 tion. J Theor Biol, 331:38-47.

Hethcote, H. W. and Yorke, J. A. (1984). Gonorrhea transmission dynamics and control.
Lecture Notes in Biomathematics. Springer-Verlag, Berlin.

Ison, C. A., Town, K., Obi, C., Chisholm, S., Hughes, G., Livermore, D. M., Lowndes, C. M., and GRASP collaborative group (2013). Decreased susceptibility to
cephalosporins among gonococci: data from the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP) in England and Wales, 2007-2011. Lancet
Infect Dis, 13(9):762-8.

Johnson, A. M., Mercer, C. H., Erens, B., Copas, A. J., McManus, S., Wellings, K., Fenton, K. A., Korovessis, C., Macdowall, W., Nanchahal, K., Purdon, S., and Field, J. (2001). Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours.

- 506 Lancet, 358(9296):1835–1842.
- Kamp, C., Moslonka-Lefebvre, M., and Alizon, S. (2013). Epidemic spread on weighted
   networks. *PLoS Comput Biol*, 9(12):e1003352.

Kaplan, E. H. (1990). Modeling HIV infectivity: must sex acts be counted? J Acquir
Immune Defic Syndr, 3(1):55-61.

Keeling, M. J. and Rohani, P. (2008). Modeling Infectious Diseases in Humans and
 Animals. Princeton University Press, Princeton.

Lloyd-Smith, J. O. (2007). Maximum likelihood estimation of the negative binomial
dispersion parameter for highly overdispersed data, with applications to infectious
diseases. *PLoS One*, 2(2):e180.

Moslonka-Lefebvre, M., Bonhoeffer, S., and Alizon, S. (2012). Weighting for sex acts to
understand the spread of STI on networks. J Theor Biol, 311:46–53.

### <sup>518</sup> Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The EWED* <sup>519</sup> *Computer Journal*, 7(4):308–313.

Nordvik, M. K. and Liljeros, F. (2006). Number of sexual encounters involving intercourse and the transmission of sexually transmitted infections. Sex Transm Dis,
33(6):342–349.

Potterat, J. J., Zimmerman-Rogers, H., Muth, S. Q., Rothenberg, R. B., Green, D. L.,
Taylor, J. E., Bonney, M. S., and White, H. A. (1999). Chlamydia transmission:
concurrency, reproduction number, and the epidemic trajectory. Am J Epidemiol,
150(12):1331–9.

- Price, M. J., Ades, A. E., Angelis, D. D., Welton, N. J., Macleod, J., Soldan, K., Turner,
  K., Simms, I., and Horner, P. J. (2013). Mixture-of-exponentials models to explain
  heterogeneity in studies of the duration of Chlamydia trachomatis infection. *Stat Med*,
  32(9):1547–60.
- R Development Core Team (2009). R: A Language and Environment for Statistical
   Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Regan, D. G., Wilson, D. P., and Hocking, J. S. (2008). Coverage is the key for effective
  screening of Chlamydia trachomatis in Australia. J Infect Dis, 198(3):349–358.
- Renton, A., Whitaker, L., Ison, C., Wadsworth, J., and Harris, J. R. (1995). Estimating
  the sexual mixing patterns in the general population from those in people acquiring gonorrhoea infection: theoretical foundation and empirical findings. *J Epidemiol Community Health*, 49(2):205–13.
- Røttingen, J.-A. and Garnett, G. P. (2002). The epidemiological and control implications
  of HIV transmission probabilities within partnerships. Sex Transm Dis, 29(12):818–27.
- Soetaert, K., Petzoldt, T., and Setzer, R. W. (2010). Solving differential equations in
  R: package deSolve. *Journal of Statistical Software*, 33(9):1–25.
- Sonnenberg, P., Clifton, S., Beddows, S., Field, N., Soldan, K., Tanton, C., Mercer,
  C. H., da Silva, F. C., Alexander, S., Copas, A. J., Phelps, A., Erens, B., Prah, P.,
  Macdowall, W., Wellings, K., Ison, C. A., and Johnson, A. M. (2013). Prevalence,
  risk factors, and uptake of interventions for sexually transmitted infections in Britain:
  findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). Lancet,
  382(9907):1795–806.
- 549 Stoner, B. P., Whittington, W. L., Hughes, J. P., Aral, S. O., and Holmes, K. K. (2000).
- 550 Comparative epidemiology of heterosexual gonococcal and chlamydial networks: im-
- plications for transmission patterns. Sex Transm Dis, 27(4):215–23.