Per-partnership transmission probabilities for Chlamydia trachomatis infection

Lewis, Joanna E A; White, Peter J. ; Price, Malcolm

DOI:
10.1093/ije/dyaa202

Document Version
Peer reviewed version

Citation for published version (Harvard):
Lewis, JEA, White, PJ & Price, M 2020, 'Per-partnership transmission probabilities for Chlamydia trachomatis infection: evidence synthesis of population-based survey data', *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyaa202

Per-partnership transmission probabilities for *Chlamydia trachomatis* infection: Evidence synthesis of population-based survey data

Joanna Lewis[1,2], Peter J. White[1,3] and Malcolm J. Price[4,5,6]

[1]MRC Centre for Global Infectious Disease Analysis and National Institute for Health Research (NIHR) Health Protection Research Unit in Modelling and Health Economics, Imperial College London School of Public Health, UK;

[2]Centre for Applied Statistics Courses, UCL Great Ormond Street Institute of Child Health, University College London, UK;

[3]Modelling and Economics Unit, National Infection Service, Public Health England, London, UK;

[4]Institute of Applied Health Research, University of Birmingham, UK

[5]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK

Corresponding Author: Joanna Lewis

Department of Infectious Disease Immunology,

Imperial College London,

St Mary's Campus,

Norfolk Place,

London W2 1PG

joanna.lewis@imperial.ac.uk

Word count (abstract plus main text): 3488

Abstract

Background: Chlamydia is the most commonly-diagnosed sexually-transmitted infection worldwide. Mathematical models used to plan and assess control measures rely on accurate estimates of chlamydia's natural history, including the probability of transmission within a partnership. Several methods for estimating transmission probability have been proposed, but all have limitations.

Methods: We have developed a new model for estimating per-partnership chlamydia transmission probabilities from infected to uninfected individuals, using data from population-based surveys. We used data on sexual behavior and prevalent chlamydia infection from the second UK National Study of Sexual Attitudes and Lifestyles (Natsal-2) and the US National Health and Nutrition Examination Surveys 2009-2014 (NHANES) for Bayesian inference of average transmission probabilities, across all new heterosexual partnerships reported. Posterior distributions were estimated by Markov chain Monte Carlo sampling using the Stan software.

Results: Posterior median male-to-female transmission probabilities per partnership were 32.1% (95% credible interval [CrI] 18.4-55.9%) (Natsal-2) and 34.9% (95%CrI 22.6-54.9%) (NHANES). Female-to-male transmission probabilities were 21.4% (95%CrI 5.1-67.0%) (Natsal-2) and 4.6% (95%CrI 1.0-13.1%) (NHANES). Posterior predictive checks indicated a well-specified model, although there was some discrepancy between reported and predicted numbers of partners, especially in women.

Conclusions: The model provides statistically rigorous estimates of per-partnership transmission probability, with associated uncertainty, which is crucial for modelling and understanding chlamydia epidemiology and control. Our estimates incorporate data from several sources including population-based surveys and use information contained in the

correlation between number of partners and the probability of chlamydia infection. The evidence synthesis approach means that it is easy to include further data as it becomes available.

Key words: chlamydia, transmission, mathematical model, Bayesian statistics, evidence synthesis, population-based survey

Key messages:

- Estimates for parameters like transmission probability are important for building models of sexually-transmitted diseases that can be used to understand their epidemiology and plan and assess control interventions.
- Average per-partnership (rather than per-sex-act) transmission probability is a particularly useful parameter because there is more and better data on numbers of partnerships than numbers of sex acts.
- We have developed a new method for estimating per-partnership chlamydia transmission probability, using data from population-level studies. We used a Bayesian approach to provide a probability distribution representing the estimate and associated uncertainty.
- We applied our method to the Second National Study of Sexual Attitudes and Lifestyles (Natsal-2) from the UK and National Health and Nutrition Examination Surveys (NHANES) from the US.

Introduction

Chlamydia is the most commonly-diagnosed sexually transmitted infection worldwide. In 2018 there were 1,382 and 3,694 chlamydia diagnoses per 100,000 15-24-year-old US men and women, respectively,(1) and 1,342 and 2,637 in England.(2) There is marked geographic variation in chlamydia burden,(3) and the effectiveness of widespread testing and/or screening in chlamydia control remains uncertain,(4,5) but the need for cost-effective control measures becomes ever-clearer as evidence for the link to pelvic inflammatory disease (PID) is strengthened(6) yet resources for sexual health services are reduced.

Mathematical models are important tools for assessing and predicting the effectiveness and cost-effectiveness of chlamydia control policies. Numerous models have been developed for these purposes(7) but a comparison of three individual-based models found they produced very different results.(8) A key parameter in any transmission-dynamic model is the transmission probability per infectious contact, where a "contact" may be defined either as a partnership or as a sex act. Transmission probability has to be estimated indirectly, as it would be unethical to conduct a study measuring it directly, and is subject to significant uncertainty. Modeling studies have used values ranging from 0.0375 to 0.154 per sex act; sometimes assuming equal male-to-female and female-to-male transmission rates, and sometimes allowing for a higher risk in the male-to-female direction.(7)

Transmission probability estimates can be based on cross-sectional concordance studies of sexual partnerships. For example, Katz used data from a US clinic to estimate the proportion of heterosexual couples forming in which the man only, the woman only, neither partner, or both are infected.(9) Using the observed proportion of couples in each state, he estimated

the male-to-female and female-to-male transmission probabilities over the time between partnership formation and observation.(9) However, concordance was observed before the partnership ended, and so the estimated transmission probabilities represented only transmission before observation – not the full per-partnership probability. Furthermore, these estimates do not allow for recovery and/or re-infection within a partnership. Althaus and colleagues proposed an alternative model based on differential equations which explicitly incorporated partnership formation and breakage, occurring with constant hazards.(10) The analysis is informative but the estimates it provides depend on values assumed for other parameters in the model, some of which are not well-defined; in particular, the duration of infection and the number of partnerships in the last six months. Finally, transmission probabilities can be estimated by calibrating a transmission model to population prevalence data.(11) With this approach, the values estimated depend on the data to which the model is calibrated, the values of other parameters, and the structural assumptions in the model.

In this paper we develop a different approach. We calculate average per-partnership chlamydia transmission probabilities from an infected man to an uninfected woman and from an infected woman to an uninfected man, using data from two population-based surveys: the 1999-2001 UK National Survey of Sexual Attitudes and Lifestyles (Natsal-2)(12) and the 2009-2014 US National Health and Nutrition Examination Surveys (NHANES)(13), synthesized with information on the clearance rate of untreated chlamydia infections. The method avoids many of the assumptions that are required for estimation within a dynamic model, and its reliance on other unknown quantities is minimal and well-described. Furthermore, because estimates are based on data from population-based surveys, the

results are directly applicable to the general population. The methods could also be applied to other sexually transmitted infections with a susceptible-infected-susceptible (SIS) model of natural history.

Methods

The aim of the study was to provide a mathematical and statistical model that can be used to infer per-partnership transmission probability from survey data. We present an overview of our methods; further details are in the Supplementary Information.

Mathematical model

We used an SIS model of infection and recovery (Figure 1). Our model considers asymptomatic infections; symptomatic infections prompt treatment seeking and are therefore short-lived and unlikely to cause onward infection or to be detected in population-based surveys.

Let each individual *j*, of sex *x*, experience a force of infection F_j. This force of infection (accounting for heterosexual transmission only) is the rate at which an individual makes contacts with infected members of the opposite sex, χ_{xj}, multiplied by the per-contact transmission probability, $\rho_{x\prime\to x}$:

$$F_j = \chi_{xj}\rho_{x\prime\to x}.$$

(*x'* denotes the opposite sex to *x*.)

Individuals' recovery rate is λ_x. The probability that individual *j* is infected at a given moment is π_j. At steady state, the number of new infections per unit time ($F_j(1-\pi_j)$) equals the number of recoveries ($\lambda_x \pi_j$,):

$$F_j(1-\pi_j) = \chi_{xj}\rho_{x\prime\to x}(1-\pi_j) = \lambda_x \pi_j$$

Hence,

$$\rho_{x\prime\to x} = \frac{\pi_j}{1-\pi_j} \times \frac{\lambda_x}{\chi_{xj}}$$

Data

We inferred parameter values in the model by synthesizing data from several sources.

Clearance of untreated chlamydia infection

Data informing the clearance rate of untreated infections came from studies in the literature synthesized in previous analyses.(14,15) Further details are provided in the original papers.(14,15)

Numbers of partners

We used data on sexual behaviour and chlamydia infection from two population-based studies: Natsal-2,(16) and the three NHANES conducted biennially between 2009 and 2014(17). We combined data from three NHANES to achieve a larger sample size than would be possible using only one.(17)*

In Natsal-2, participants reported on their number of new opposite-sex partners in the last year, and this information was used to inform a probability distribution for the number of new partners in the last year.

In NHANES, participants were asked their number of partners, and whether they had had any new partners, in the last 12 months. We used these two questions to provide a proxy for the number of new partners in the last year. Where respondents reported no new partners in the last year, we took the number of new partners to be zero; where they reported one partner and a new partner, we took the number of new partners to be one; otherwise, we assumed that all but one of their total reported partners was new. This approach is similar to the use elsewhere of "shifted negative binomial" distributions for modelling partner numbers.(18)

Infection status

The publicly-available data from both Natsal-2 and NHANES also includes chlamydia infection status, diagnosed using nucleic acid amplification tests (NAATs) on urine samples. Natsal-2 participants were eligible for a urine sample if they were aged 18-44 years and had ever had sex, and a randomly-selected half of those eligible were invited to provide samples. All NHANES participants aged 14-39 years were invited to provide a sample for testing, but the publicly-available data excludes 14-17-year-olds.

Numbers of partnerships reported by susceptible and infected men and women in Natsal-2 and NHANES are provided in Supplementary Tables S1 and S2.

Statistical model

We conducted a Bayesian evidence synthesis, using data from the sources described to construct a likelihood. Survey weights were incorporated by multiplying the relevant component of the log-likelihood by the weight. The likelihood was combined with appropriate priors to provide a joint posterior for the model parameters.

Clearance of untreated infections

The statistical model used for the clearance rates of untreated chlamydia infection is described elsewhere.(14) The model involves two courses for infection: fast- or slow-clearing. A proportion *p* of incident infections clear fast, and the remainder, 1 – *p*, clear slowly. Some of the data on chlamydia clearance came from studies using culture diagnosis methods, and the model accounts for this using a sensitivity parameter for culture diagnosis in people with a previous positive culture for that infection, ψ. In this analysis we assumed that only the slow-clearing infections last long enough to be detected in population-based studies. The clearance rate (denoted λ_x above) is therefore equal to the slow clearance rate in the clearance model, and the transmission probability we estimated is the probability that an infection is transmitted and then follows the slow-clearing course.

Partnership dynamics

We used negative binomial distributions to model the estimated numbers of new partners reported in the last year by men and women. A negative binomial distribution with size α and mean μ can arise as a mixture of Poisson distributions, where the mixing distribution for the Poisson rate is a Gamma distribution with shape α and rate $\frac{\mu}{\alpha}$.(19) In our model, the shape and rate depend on the sex of the individual, but are constrained so that the

expected number of partnerships per man must equal the expected number of partnerships per woman.

Prevalence

We used our model to calculate the probability π_j of each individual *j* being infected, given the number of partners they reported. The infection status of *j* has a Bernoulli distribution with parameter π_j:

$$P(\delta_j|\pi_j) = P_{Bernoulli}(\delta_j|\pi_j) = \begin{cases} \pi_j & \delta_j = 1 \\ 1-\pi_j & \delta_j = 0 \end{cases}$$

where

$$\delta_j = \begin{cases} 1 & \text{if } j \text{ is infected} \\ 0 & \text{if } j \text{ is uninfected} \end{cases}$$

Full likelihood

The log-likelihood of the data is given by:

$$L = L_{turnover} + L_{clearance} + L_{infection}$$

where:

- $L_{turnover}$ is the log-likelihood associated with partnership turnover (negative binomial distribution);
- $L_{clearance}$ is the log-likelihood associated with clearance, and
- $L_{infection}$ is the log-likelihood associated with the infection status of each participant at the time of testing in the survey (Bernoulli distribution).

Inference and Estimation

Priors

We used uninformative priors for all parameters except the sensitivity of chlamydia diagnosis by culture, which enters the model for chlamydia clearance. This had a $\psi \sim \text{Beta}(78,8)$ prior, based on studies comparing the performance of culture diagnosis and NAATs.(14)

Bayesian methods and sampling of posterior distribution

Estimation was carried out by sampling from the posterior using a Markov chain Monte Carlo (MCMC) algorithm implemented in the Stan software,(20) within the R environment.(21) The data, Stan model file and R scripts used for handling input and results are all available online at https://github.com/joanna-lewis/ct_transmission_probs. We ran four chains for 2000 iterations each, discarding the first 1000 "warmup" iterations of each chain. Posterior predictive checks were carried out, comparing simulated and observed partner number distributions, and prevalence in men and women reporting different numbers of partners. We also used prior distributions for the proportion of infections leading to symptoms for men and women to simulate the annual number of symptomatic infections that would have occurred under the parameter values inferred (see supplementary information).

Sensitivity Analysis

We conducted three sensitivity analyses to investigate different aspects of our model, which are described in detail in Supplementary Information. First, we relaxed the assumption of

equal average numbers of partnerships in men and women. Secondly, we constructed a model in which individuals only form partnerships with members of the opposite sex reporting the same number of partnerships. This tests two aspects of the model: (a) by imposing totally assortative mixing by number of partners, it tests the effect of assuming that partners are chosen at random from all those available; and (b) by allowing for differing force of infection in individuals reporting different numbers of partners, it tests the effect of using a single average transmission probability across all partnerships. Finally, we used data from Natsal-2 to investigate the effect of studying the number of partnerships without a condom, rather than total partnership numbers.

Results

For all parameters split $\hat{R}$ statistics for the MCMC sampling were between 0.9990 and 1.0032, indicating good convergence, and the effective sample size was greater than 0.4 per transition of the Markov chain. No transitions ended with a divergence.

In Natsal-2 the mean number of new partners per year was inferred as 0.59 (95%CrI 0.54-0.65). Overall chlamydia prevalence was 2.1% (95%CrI 1.6-2.8%) in men and 2.0% (95%CrI 1.4-2.8%) in women, compared to survey-based estimates of 2.4% (95%CI 1.5-3.6%) and 1.5% (95%CI 1.0-2.1%). In NHANES the mean number of new partners inferred was 0.92 (95%CrI 0.85-1.00). Prevalence was 1.7% (95%CrI 1.3-2.3%) in men and 3.7% (95%CrI 2.8-4.6%) in women, compared to survey-based estimates 1.9% (95%CI 1.3-2.6%) and 2.3% (95%CI 1.7-3.0%).

Figure 2 shows posterior distributions for the per-partnership transmission probabilities, derived using Natsal-2 and NHANES. Using Natsal-2, the posterior median transmission probabilities were 32.1% (95%CrI 18.4-55.9%) (male-to-female) and 21.4% (95%CrI 5.1-67.0%) (female-to-male). Using NHANES, they were 34.9% (95%CrI 22.6-54.9%) (male-to-female) and 4.6% (95%CrI 1.0-13.1%) (female-to-male). The posterior distributions for all parameters are summarized in Supplementary Table S4.

Posterior predictions for the partner number distributions generally agreed with data but there was some discrepancy, especially in women (Supplementary Figure S2). Predicted numbers of infections, by reported numbers of partners, agreed well with observations in both sexes, for both studies (Supplementary Figure S3).

For Natsal-2 we simulated median (2.5th-97.5th centile) 109,000 (25,000-327,000) symptomatic cases in men; the number of diagnoses recorded in 2000 was estimated as 30,000-41,000.(22) In women we simulated median (2.5th-97.5th centile) 46,000 (25,000-77,000) symptomatic cases; 48,000-105,000 diagnoses were recorded.(22) For NHANES, we simulated median (2.5th-97.5th centile) 397,000 (83,000-1149,000) symptomatic cases in men; the number of diagnoses recorded in 2009 was 307,000. We simulated median (2.5th-97.5th centile) 429,000 (259,000-682,000) symptomatic cases in women, and 879,000 diagnoses were recorded.

In the sensitivity analyses we found that relaxing the assumption of equal partnership numbers in men and women led to no meaningful differences in the posterior ditributions for transmission probabilities. In a model where partnerships formed only between

individuals reporting the same number of partners, we found evidence of higher transmission probabilities in couples reporting fewer partners. Our model using data on partnerships without a condom resulted in posterior distributions shifted to slightly higher transmission probabilities, but the shift was small compared with the width of the distribution.

Discussion

We have described a new statistical model for inferring the per-partnership transmission probability of a sexually transmitted infection, and have applied it to population-level data on chlamydia from the UK and the US. Our method provides its estimates with uncertainty, which is crucial for modelling and understanding chlamydia epidemiology and control. Estimates of average per-partnership (as opposed to per-sex-act) transmission probability are valuable for building predictive models of control measures, because data availability means that behavioural models can be parameterised more reliably in terms of number of partnerships than number of sex acts. Our estimates incorporate data from several sources including population based surveys and make use of information that is often disregarded, contained in the correlation between the number of partners reported and the probability of chlamydia infection.

In the UK we found a male-to-female transmission probability of 32.1% per partnership (95%CrI 18.4-55.9%), which was consistent with the corresponding US result of 34.9% (95%CrI 22.6-54.9%). The posterior for female-to-male transmission probability inferred from the UK data was much more uncertain, with posterior median 21.4% (95%CrI 5.1-

67.0%). The equivalent for the US data was lower, but with a narrower and overlapping credible interval: 4.6% (95%CrI 1.0-13.1%).

Posterior predictive checks agreed well with the original data, indicating a well-specified model. The main exception is the partnership number data in women: in both Natsal-2 and NHANES, higher partner numbers are under-reported compared to simulations. Under-reporting of partner numbers by women is a recognized phenomenon which has been widely discussed.(23) The partnership number distributions may explain the low female-to-male transmission estimated using NHANES. If NHANES respondents reported new partner(s) in the last year, and more than one partner in total, then we took the number of new partners to be one less than the total number of partners: in fact, this proxy is an upper bound, as more than one could have been an existing partner. If the number of partners and hence the contact rate is over-estimated by this proxy then there will be a corresponding reduction in the per-partnership transmission probability.

Katz estimated a male-to-female transmission probability of 39.5% (95%CI 19.3-59.7%) per partnership:(9) consistent with our estimate. Katz's estimate for female-to-male transmission probability is 32.3% (95%CI 10.0-54.6%): well within our credible interval for UK data, but barely overlapping for the US estimate. Althaus et al.'s ODE-based pair model produced a higher estimated transmission probability per partnership (55.5%, IQR 49.2-62.5%), assuming two partners every six months (four per year).(10) However, they note that their model does not account for heterogeneity in transmissibility of chlamydia, whereas ours allows for differences by sex. We also account for sex differences in chlamydia

clearance rate and heterogeneity in partnership turnover rates, which is an important feature in explaining observed partner number distributions.

Our model assumes a closed system at steady state. These assumptions are reasonable as the number of people entering and leaving the sexually-active population each year is small compared to the total population, and any changes in the model parameters are slow compared to the dynamics of the system. We have ignored the role of same-sex contacts, but their effect on our estimates is also likely to be small because only people with at least one opposite-sex partner were included in the data. We chose to include people reporting partners of both sexes in our analysis to maximise the amount of data used, and because excluding them ignores their involvement in the heterosexual network and could bias our results.

Another assumption of the analysis is that individuals choose partnerships at random from all the partnerships offered by the opposite sex. Whilst we know that sexual mixing is to some extent assortative, sensitivity analysis indicates that assortativity would not lead to greatly differing force of infection per contact in people reporting different numbers of partners (see Supplementary Information). There was some evidence from this analysis of a higher transmission probability in people reporting no new partners, particularly in the NHANES dataset. This could reflect lower condom use or longer partnerships and would be an interesting avenue for further research. However, even if there are qualitative differences between partnerships, leading to heterogeneity in transmission probabilities, this does not invalidate the concept of a single average across all partnerships, which is still a hugely useful quantity for modelling. In a further sensitivity analysis we modelled number

of partnerships without a condom, estimated using data from Natsal-2. The posterior distributions suggested that qualitative differences such as condom use may reduce population-average transmission probabilities, but to an extent that is small compared with the uncertainty in the estimates. It might be valuable for sexual behavior surveys to collect explicit information on the annual number of new partnerships without a condom for parameter inference and predictive modelling, and our sensitivity analysis suggests that our model could be used to infer transmission probabilities from such data.

The evidence synthesis approach that we used can readily incorporate further data as it becomes available, so that improved data collection would allow our analysis to be augmented to improve our estimates. For example, there is particular uncertainty in the proportion of infections that become symptomatic in each sex, and in the clearance rate of untreated infections in men; the latter limiting the precision of the female-to-male transmission probability. We have argued elsewhere that surveillance and screening programmes could be used to collect data on long-term chlamydia clearance in men to inform a more precise estimate of clearance rate(15). Additionally, it has been suggested that previous exposure to chlamydia may confer partial immunity,(24) which would reduce the transmission probability to older and/or more sexually active individuals, who would be more likely to have had a prior infection. Whilst further empirical study of chlamydia immunology is required, it is interesting that the posterior predictive checks showed that our model tends to under-predict prevalence in those reporting few partners and over-predict in those reporting several partners (Figure S4), which would be consistent with partial immunity in high-risk individuals who are more likely to have been infected before.

In conclusion, it is important to use rigorous parameter estimates in computational models, and to quantify their uncertainty and its effect on conclusions and recommendations. Our method provides such estimates for the probability of chlamydia transmission, and with appropriate data the methods described here could also be applied to other sexually transmitted infections which can be represented using the SIS model. The estimates can be used in transmission modeling to understand the effect of control policies on patterns of prevalence.

Ethics

This was a secondary analysis of publicly-available data, and no additional ethical approval was required or sought.

Funding

JL and PJW were supported by the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Modelling Methodology at Imperial College London in partnership with Public Health England (PHE) (grant number HPRU-2012-10080). PJW was also supported by the NIHR HPRU in Modelling and Health Economics, a partnership between PHE, Imperial College London and LSHTM, for funding (grant number NIHR200908). Additionally, PJW was supported by the MRC Centre for Global Infectious Disease Analysis (grant number MR/R015600/1); this award is jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO) under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union (EU). MJP was supported by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham.

This paper presents independent research, and the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed are those of the authors and not necessarily those of the Department of Health and Social Care, EU, FCDO, MRC, National Health Service, NIHR, or PHE.

Data availability

The raw data analysed in this study has been made publicly available by the researchers in question, and can be accessed as described in the References.

Conflict of interest

None declared.

References

1. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2018. Atlanta: U.S. Department of Health and Human Services. 2019.
2. . National chlamydia screening programme (NCSP): data tables. http://www.chlamydiascreening.nhs.uk/ps/data.asp Accessed 6 January 2016.
3. Lewis J, White PJ. Estimating local chlamydia incidence and prevalence using surveillance data. Epidemiology 2017; 28:492-502.
4. Gottlieb SL, Xu F, Brunham RC. Screening and treating *Chlamydia trachomatis* genital infection to prevent pelvic inflammatory disease: interpretation of findings from randomized controlled trials. Sex Transm Dis 2013; 40:97-102.
5. Lewis J, White PJ. Changes in chlamydia prevalence and duration of infection estimated from testing and diagnosis rates in England: a model-based analysis using surveillance data, 2000--15. Lancet Public Health 2018; 3:e271-78.
6. Price MJ, Ades AE, De Angelis D, et al. Risk of pelvic inflammatory disease following *Chlamydia trachomatis* infection: analysis of prospective studies with a multistate model. Am J Epidemiol 2013; 178:484-492.

7. Davies B, Anderson S, Turner KME, Ward H. How robust are the natural history parameters used in chlamydia transmission dynamic models? A systematic review. Theor Biol Med Model 2014; 11:8.

8. Althaus CL, Turner KME, Schmid BV, Heijne JCM, Kretzschmar M, Low N. Transmission of *Chlamydia trachomatis* through sexual partnerships: a comparison between three individual-based models and empirical data. J R Soc Interface 2012; 9:136-146.

9. Katz BP. Estimating transmission probabilities for chlamydial infection. Stat Med 1992; 11:565-577.

10. Althaus CL, Heijne JCM, Nicola Low M. Towards more robust estimates of the transmissibility of *Chlamydia trachomatis*. Sex Transm Dis 2012; 39:402-404.

11. Turner KM, Adams EJ, Gay N, Ghani AC, Mercer C, Edmunds WJ. Developing a realistic sexual network model of chlamydia transmission in Britain. Theor Biol Med Model 2006; 3:3.

12. Fenton KA, Korovessis C, Johnson AM, et al. Sexual behaviour in Britain: reported sexually transmitted infections and prevalent genital C*hlamydia trachomatis* infection. Lancet 2001; 358:1851-1854.

13. National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/index.htm Accessed 17 July 2020.

14. Price MJ, Ades AE, De Angelis D, et al. Mixture-of-exponentials models to explain heterogeneity in studies of the duration of *Chlamydia trachomatis* infection. Stat Med 2013; 32:1547-60.

15. Lewis J, Price MJ, Horner PJ, White PJ. Genital *C. trachomatis* infections clear more slowly in men than women, but are less likely to become established. J Infect Dis 2017; 216:237-244.

16. Johnson A, Fenton K, Copas A et al. *National Survey of Sexual Attitudes and Lifestyles, 2000-2001* [computer file]. Colchester, Essex: UK Data Archive [distributor], 2005. SN: 5223, http://doi.org/10.5255/UKDA-SN-5223-1.

17. Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data.* [computer file]. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention [distributor].

18. Handcock MS, Jones JH. Likelihood-based inference for stochastic models of sexual network formation. Theor Popul Biol 2004; 65:413-422.

19. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Chapman &Hall/CRC Press, London, third edition. 2013.

20. Stan Development Team. RStan: the R interface to Stan, Version 2.10.1. http://mc-stan.org/interfaces/rstan.html Accessed 7 June 2016.

21. R Core Team. R: A Language and Environment for Statistical Computing. 2016.

22. Chandra N, Soldan K, Dangerfield C, et al. Filling in the gaps: estimating numbers of chlamydia tests and diagnoses by age group and sex before and during the implementation of the English National Screening Programme, 2000 to 2012. Eurosurveillance 2017; 22:30453.

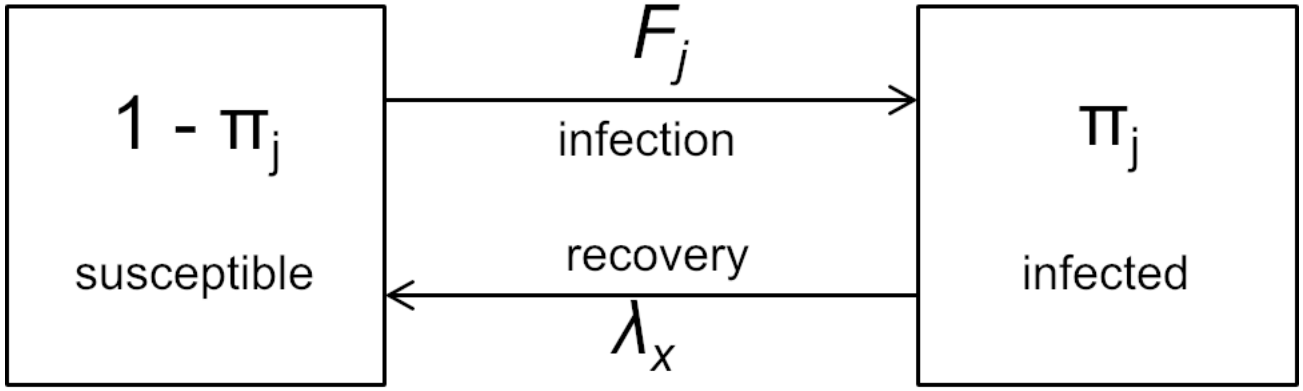
23. Morris M. Telling tails explain the discrepancy in sexual partner reports. Nature 1993; 365:437-440.

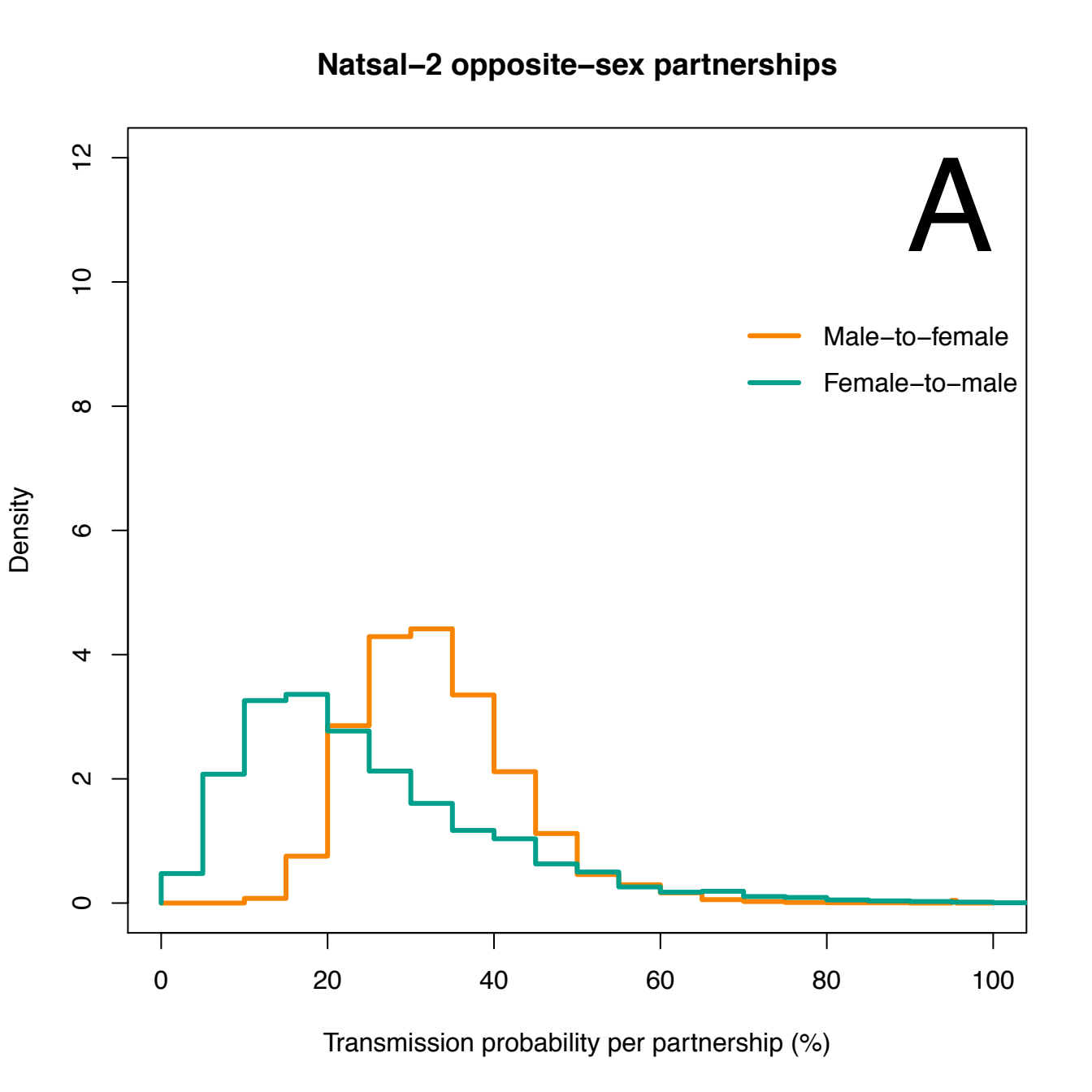
24. Batteiger BE, Xu F, Johnson RE, Rekart ML. Protective immunity to *Chlamydia trachomatis* genital infection: Evidence from human studies. Journal of Infectious Diseases 2010; 201 (Suppl 2):S178-S189.

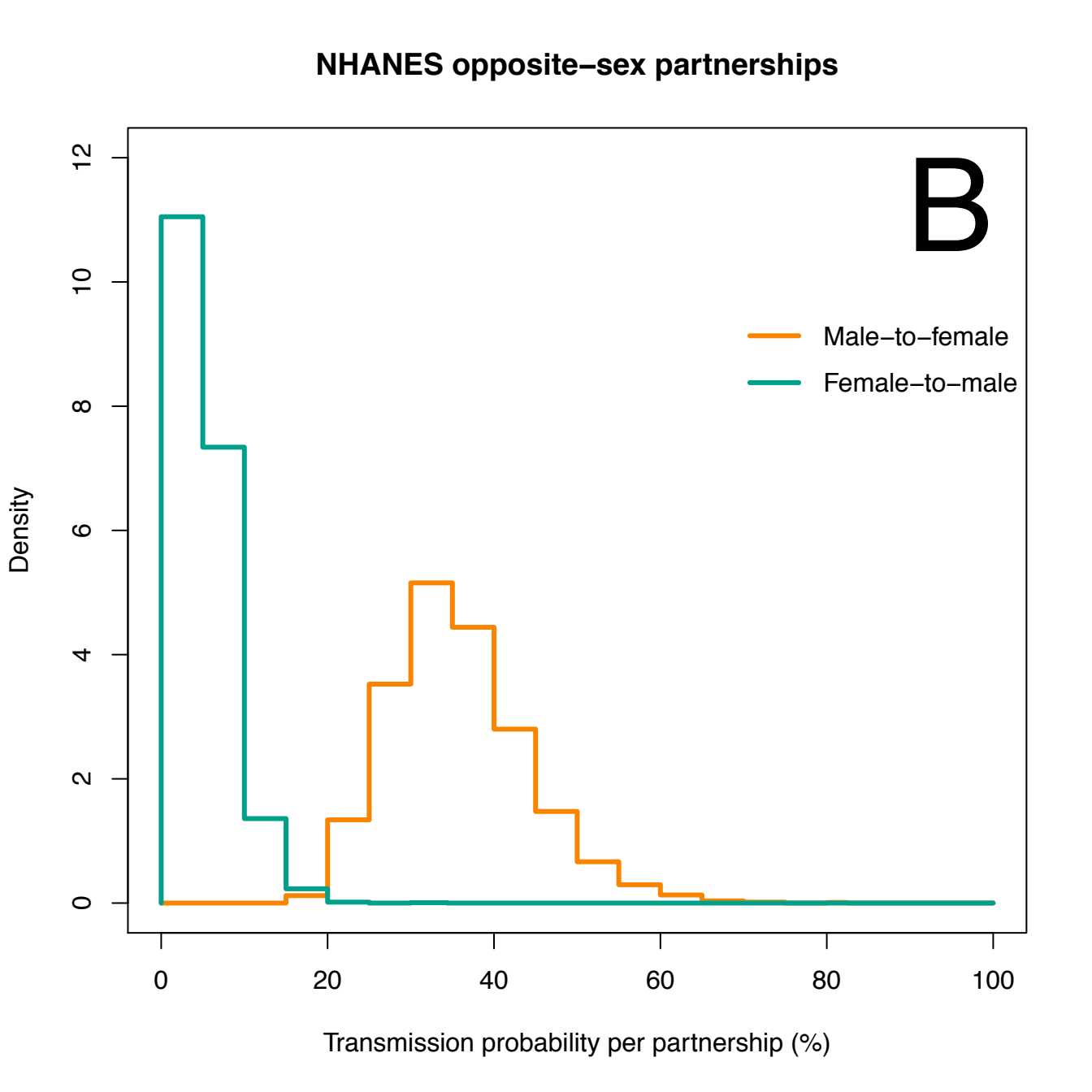
Figure Legends

Figure 1: SIS (susceptible-infected-susceptible) model of chlamydia infection and recovery for individual *j*, of sex *x*. π_j is the probability of being infected with chlamydia and $1 - \pi_j$ is the probability of being susceptible. F_j is the force of infection and λ_x is the recovery rate.

Figure 2: Posterior distributions for the per-partnership probability of chlamydia transmission, derived using number of new partners reported in (A) The second National Study of Sexual Attitudes and Lifestyles (Natsal-2), and (B) the National Health and Nutrition Examination Surveys (NHANES) 2009-2014 (all studies combined). The yellow line in each figure represents male-to-female transmission probability and the green line female-to-male.


1 - π_j
susceptible
F_j
infection
recovery
λ_x
π_j
infected


Natsal–2 opposite–sex partnerships
A
Male–to–female
Female–to–male
Density
0
2
4
6
8
10
12
0
20
40
60
80
100
Transmission probability per partnership (%)


NHANES opposite−sex partnerships
B
Male−to−female
Female−to−male
Density
0
2
4
6
8
10
12
0
20
40
60
80
100
Transmission probability per partnership (%)

Per-partnership transmission probabilities for *Chlamydia trachomatis* infection: Evidence synthesis of population-based survey data – Supplementary Information

Joanna Lewis, Peter J. White and Malcolm J. Price

Contents

1. Methods

The aim of the study is to provide a mathematical and statistical model that can be used to infer per-partnership transmission probability from survey data.

a. Mathematical model

Let each individual *j* experience a force of infection F_j, which depends on his or her rate of forming infectious contacts (partnerships). Assume that all women recover from infection at the same rate, λ_f, and all men recover at the same rate, λ_m. We use a susceptible-infected-susceptible (SIS) model of infection and recovery (Figure 1). The probability that individual *j* is infected at a given moment is π_j, and the probability that he or she is susceptible is $1 - \pi_j$.

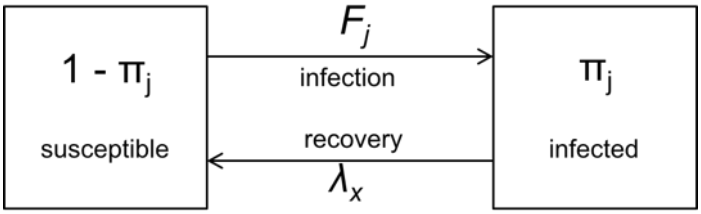


Figure S1: Susceptible-infected-susceptible (SIS) model of chlamydia infection and recovery.

Assuming only heterosexual transmission, the force of infection is the rate at which an individual makes contacts with infected members of the opposite sex, multiplied by the per-contact transmission probability. We denote the sex of individual *j* with the symbol *x*, and the opposite sex with the symbol *x'*. The rate of contacting infected members of the opposite sex is χ_{xj}, and the per-contact transmission probability from the opposite sex is $\rho_{x\prime\to x}$. Then:

$$F_j = \chi_{xj}\rho_{x\prime\to x}.$$

At steady state, the number of new infections per unit time equals the number of recoveries, so we know also that:

$$F_j(1-\pi_j) = \chi_{xj}\rho_{x\prime\to x}(1-\pi_j) = \lambda_x\pi_j$$

Hence,

$$\rho_{x\prime\to x} = \frac{\pi_j}{1-\pi_j}\frac{\lambda_x}{\chi_{xj}}$$

and

$$\frac{\pi_j}{1-\pi_j} = \frac{\chi_{xj}\rho_{x\prime\to x}}{\lambda_x}$$

The following assumptions are implicit in this argument and are discussed in the main text:

1. Closed system: the number of people entering and leaving the system is negligible.
2. Steady state: prevalence is stable, and force of infection and recovery rate do not change.
3. Identical partnerships: all partnerships have the same risk of transmission, regardless of partnership length and frequency of sex acts.

Our model considers asymptomatic infections; symptomatic infections prompt treatment-seeking and are therefore short-lived and unlikely to cause onward infection or to be detected in population-based surveys.

b. Data

We infer parameter values in the model by synthesizing data from several sources.

i. Clearance of untreated chlamydia infection

Data informing the clearance rate of untreated chlamydia infection in men and women came from studies in the literature synthesized in previous analyses.[1,2] In each study people found to be infected with chlamydia were re-tested at a later date, having remained untreated in the interim. The number who cleared their infection provides information on the clearance rate. Nine studies in women and eight in men were included, involving a total of 569 women and 165 men. Further details are provided in the original papers describing this analysis.[1,2]

ii. Partnership numbers

We used data on sexual behaviour and chlamydia infection from two population-based studies: the second National Study of Sexual Attitudes and Lifestyles (Natsal-2),[3] and the three National Health and Nutrition Examination Surveys (NHANES) conducted biennially between 2009 and 2014[4]. The ideal data to inform the sexual contact rate would be the number of new sexual partnerships formed in the last year.

In Natsal-2, participants reported their number of opposite-sex partners in the last year and were then asked:

- *Was this [woman/man] a new partner who you had sex with for the first time during the last year?* (if they had reported one partner) or
- *How many of these [women/men] were new partners who you had sex with for the first time during the last year?* (if they had reported more than one partner).

This information was used to inform the distribution of number of new partners in the last year in the Natsal-2 population.

We combined data from the three NHANES conducted between 2009 and 2014 to achieve a larger sample size than would be possible using just one study.[4] Participants were asked:

- *In the past 12 months, with how many [women/men] have you had vaginal sex?* and
- *In the past 12 months, did you have any kind of sex with a person that you never had sex with before?*

We used these two questions to provide a proxy for the number of new partners in the last year according to the following algorithm:

- If a participant stated they had had no new partners in the last year, we took the number of new partners to be zero.
- If a participant stated they had had new partner(s) in the last year, and reported one partner in total, we took the number of new partners to be one.
- If a participant stated they had had new partner(s) in the last year, and reported more than one partner in total, we took the number of new partners to be one less than the total number of partners.

This approach is similar to the use elsewhere of “shifted negative binomial” distributions for modelling partner numbers.[5]

iii. Infection status

The publicly-available data from both Natsal-2 and NHANES also includes chlamydia infection status, diagnosed using nucleic acid amplification tests (NAATs) on urine samples, which provides information on the prevalence of infection in individuals reporting different numbers of partners. Natsal-2 participants were eligible for a urine sample if they were aged 18-44 years and had ever had sex, and a randomly-

selected half of these eligible participants were invited to provide samples. All NHANES participants aged 14-39 years were invited to provide a sample for chlamydia testing, but the publicly-available data excludes 14-17-year-olds.

The raw data on numbers of partnerships reported by susceptible and infected men and women in Natsal-2 and NHANES are provided in Supplementary Tables S1 and S2.

Table S1: Raw data from the Second National Study of Sexual Attitudes and Lifestyles (Natsal-2) used to inform the model.

Number of new partners	Number of Men						Chlamydia prevalence in men (95%CI) (%)	Number of Women						Chlamydia prevalence in women (95%CI)(%)
	Unweighted			Weighted				Unweighted			Weighted			
	Uninfected	Infected	Total	Uninfected	Infected	Total		Uninfected	Infected	Total	Uninfected	Infected	Total	
0	784	9	793	1063.78	13.57	1077.35	1.3 (0.5, 2.5)	1346	17	1363	1210.19	11.38	1221.57	0.9 (0.5, 1.5)
1	243	6	249	266.15	6.57	272.72	2.4 (0.8, 5.6)	309	9	318	213.90	5.11	219.01	2.3 (1.0, 4.7)
2	98	4	102	99.39	4.91	104.30	4.7 (1.1, 12.5)	76	5	81	55.64	2.85	58.49	4.9 (1.4, 11.6)
3	51	3	54	48.32	4.62	52.94	8.7 (1.5, 25.1)	27	2	29	16.57	1.82	18.39	9.9 (1.0, 33.2)
4	18	2	20	16.79	3.05	19.84	15.4 (1.3, 50.0)	18	0	18	13.40	0	13.40	0*
5	14	1	15	13.55	0.94	14.50	6.5 (0.1, 32.8)	9	2	11	7.33	0.86	8.19	10.5 (0.9, 36.2)
6	8	1	9	8.88	1.94	10.83	18.0 (0.4, 67.7)	3	0	3	2.04	0	2.04	-
7	6	0	6	7.92	0	7.92	0*	2	0	2	1.06	0	1.06	-
8	1	0	1	1.62	0	1.62	0*	1	0	1	0.48	0	0.48	-
9	5	2	7	5.68	1.51	7.19	21.0 (2.1, 60.6)	1	0	1	1.26	0	1.26	-
10	3	1	4	4.09	0.58	4.67	12.4 (0.0, 1.0)	1	1	2	0.48	0.25	0.73	34.1‡
11	1	0	1	1.76	0	1.76	0*	0	1	1	0	0.62	0.62	100*
12	3	0	3	4.46	0	4.46	0*	0	0	0	0	0	0	-
13	1	0	1	0.99	0	0.99	0*	1	0	1	0.81	0	0.81	0*
15	1	0	1	1.16	0	1.16	0*	0	0	0	0	0	0	-
19	1	0	1	0.78	0	0.78	0*	0	0	0	0	0	0	-
20	0	0	0	0	0	0	-	2	0	2	1.91	0	1.91	0*
36	0	0	0	0	0	0	-	1	0	1	0.46	0	0.46	0*
39	0	0	0	0	0	0	-	1	0	1	0.44	0	0.44	0*
55	1	0	1	0.44	0	0.44	0*	0	0	0	0	0	0	-
Overall mean number of partners	0.85	2.41	0.89	0.71	2.07	0.74	**Overall prevalence in men:** 2.38 (1.52, 3.55)	1.03	1.51	1.04	0.95	1.31	0.96	**Overall prevalence in women:** 1.48 (1.01, 2.08)

CI: confidence interval

*Confidence intervals cannot be calculated where 0% or 100% of individuals were infected.

‡A confidence interval could not be calculated because of the small number of individuals.

Table S2: Raw data from the National Health and Nutrition Examination Surveys (NHANES; 2009-2014 combined) used to inform the model.

Number of new partners	Number of Men						Chlamydia prevalence in men (95%CI) (%)	Number of Women						Chlamydia prevalence in women (95%CI)(%)
	Unweighted			Weighted				Unweighted			Weighted			
	Uninfected	Infected	Total	Uninfected	Infected	Total		Uninfected	Infected	Total	Uninfected	Infected	Total	
0	1617	29	1646	1662.23	25.61	1687.84	1.5 (0.9, 2.3)	1901	43	1944	1898.59	34.05	1932.65	1.8 (1.2, 2.5)
1	306	12	318	330.54	10.92	341.46	3.2 (1.4, 6.3)	291	18	309	289.92	10.79	300.71	3.6 (2.0, 5.9)
2	123	4	127	109.99	2.99	112.98	2.6 (0.6, 7.1)	78	7	85	78.20	5.21	83.42	6.3 (1.8, 14.9)
3	72	2	74	61.48	0.83	62.31	1.3 (0.1, 5.5)	34	2	36	31.85	1.20	33.05	3.6 (0.1, 18.5)
4	58	2	60	55.29	0.65	55.94	1.2 (0.1, 4.9)	17	0	17	15.40	0	15.40	0*
5	19	2	21	23.46	0.72	24.18	3.0 (0.0, 92.3)	15	1	16	14.50	0.44	14.94	2.9 (0.0, 93.5)
6	23	1	24	22.14	0.55	22.69	2.4 (0.0, 19.3)	4	1	5	4.78	0.44	5.22	8.43‡
7	13	0	13	14.68	0	14.68	0*	8	0	8	8.31	0	8.31	0*
8	11	0	11	9.24	0	9.24	0*	3	1	4	5.22	1.35	6.57	20.6‡
9	19	1	20	17.35	0.62	17.97	3.5 (0.0, 98.8)	3	0	3	1.91	0	1.91	0*
10	3	0	3	2.89	0	2.89	0*	2	0	2	0.98	0	0.98	0*
11	8	0	8	6.41	0	6.41	0*	7	0	7	4.25	0	4.25	0*
12	5	0	5	4.31	0	4.31	0*	0	1	1	0	0.62	0.62	0*
13	3	0	3	3.21	0	3.21	0*	0	0	0	0	0	0	-
14	8	1	9	5.09	0.53	5.62	9.4‡	1	0	1	0.42	0	0.42	0*
17	3	1	4	2.72	0.44	3.16	13.9‡	1	0	1	0.69	0	0.69	0*
18	3	0	3	2.36	0	2.36	0*	0	0	0	0	0	0	-
19	6	0	6	3.99	0	3.99	0*	0	1	1	0	0.39	0.39	100*
21	2	0	2	2.90	0	2.90	0*	0	0	0	0	0	0	-
23	0	0	0	0	0	0	-	1	0	1	0.82	0	0.82	0*
24	3	1	4	3.36	0.52	3.89	13.5‡	2	0	2	4.13	0	4.13	0*
29	2	0	2	1.60	0	1.60	0*	1	0	1	0.60	0	0.60	0*
34	1	0	1	0.80	0	0.80	0*	1	0	1	1.24	0	1.24	0*
38	1	0	1	0.62	0	0.62	0*	0	0	0	0	0	0	-
39	1	0	1	0.36	0	0.36	0*	0	0	0	0	0	0	-
44	1	0	1	2.11	0	2.11	0*	0	0	0	0	0	0	-
49	4	0	4	5.52	0	5.52	0*	1	0	1	1.60	0	1.60	0*
78	1	0	1	0.90	0	0.90	0*	0	0	0	0	0	0	-
79	1	0	1	0.64	0	0.64	0*	0	0	0	0	0	0	-
89	0	0	0	0	0	0	-	1	0	1	2.22	0	2.22	0*
99	1	1	2	0.75	0.52	1.27	41.2‡	0	0	0	0	0	0	-
Overall mean number of partners	1.26	3.74	1.31	1.18	2.53	1.20	**Overall prevalence in men:** 1.87 (1.33, 2.56)	0.52	1.17	0.54	0.57	1.01	0.58	**Overall prevalence in women:** 2.25 (1.65, 3.00)

CI: confidence interval

*Confidence intervals cannot be calculated where 0% or 100% of individuals were infected.

‡A confidence interval could not be calculated because of the small number of individuals.

c. Statistical model

We conducted a Bayesian evidence synthesis using data from the sources described to construct a likelihood. This was combined with appropriate priors to provide a posterior distribution for the model parameters.

i. Partnership dynamics

We used negative binomial distributions to model the estimated numbers of new partners reported in the last year by men and women. A negative binomial distribution with size α and mean μ can arise as a mixture of Poisson distributions, where the mixing distribution for the Poisson rate is a Gamma distribution with shape α and rate $\frac{\mu}{\alpha}$. Formally, let the number of new partners reported by individual *j* be represented by the random variable N_j which has a Poisson distribution with rate σ_j:

$$N_j \sim Poisson(\sigma_j),$$

so that

$$P(n_j|\sigma_j) = P_{Poisson}(n_j|\sigma_j) = \frac{e^{-\sigma_j}{\sigma_j}^{n_j}}{n_j!}$$

Now, let the partner change rate be a random variable having a Gamma distribution with shape α_j and rate $\beta_j = \frac{\mu_j}{\alpha_j}$:

$$\sigma_j \sim Gamma(\alpha_j, \beta_j)$$

so that

$$P(\sigma_j|\alpha_j, \beta_j) = P_{Gamma}(\sigma_j|\alpha_j, \beta_j) = \frac{(\beta_j)^{\alpha_j}}{\Gamma(\alpha_j)}{\sigma_j}^{(\alpha_j - 1)}e^{-\sigma_j\beta_j}.$$

It can be shown[6] by integrating over the Poisson rate σ_j that N_j has a negative binomial distribution with size α_j and mean $\mu_j = {\alpha_j}/{\beta_j}$:

$$N_j \sim NB(\mu_j, \alpha_j)$$

$$P(n_j|\mu_j, \alpha_j) = P_{NB}(n_j|\mu_j, \alpha_j) = \binom{n_j + \alpha_j - 1}{n_j}\left(\frac{\mu_j}{\mu_j + \alpha_j}\right)^{n_j}\left(\frac{\alpha_j}{\mu_j + \alpha_j}\right)^{\alpha_j}$$

In our model, the shape and rate depend on the sex of the individual:

$$(\alpha_j, \beta_j) = \begin{cases} (\alpha_m, \beta_m) & \text{for men} \\ (\alpha_f, \beta_f) & \text{for women} \end{cases}$$

As we are considering heterosexual transmission, the expected number of partnerships per man must equal the expected number of partnerships per woman, so we constrain the negative binomial partnership number distributions in men and women to have the same mean:

$${\alpha_m}/{\beta_m} = {\alpha_f}/{\beta_f} = \mu.$$

The Gamma distribution is the conjugate prior for the Poisson. Given that we observe n_j new partnerships in a year in individual *j*, we can "update" our knowledge of the partner change rate in individual *j* and say that

$$\sigma_j \sim Gamma(\alpha_j + n_j, \beta_j + 1)$$

(See [6] for a full discussion of conjugate priors, including the Poisson model.)

ii. Prevalence

As described above, the probability that individual *j* is infected with chlamydia is a function of the Poisson rate of forming partnerships with infected people (χ_{xj}), the per-partnership transmission probability ($\rho_{x\prime\to x}$), and the clearance rate (λ_x):

$$\frac{\pi_j}{1-\pi_j} = \frac{\chi_{xj}\rho_{x\prime\to x}}{\lambda_x} \qquad (1)$$

The rate of individual *j* forming infectious contacts, χ_{xj}, equals the rate of forming contacts, σ_j, multiplied by the proportion of contacts offered by the opposite sex that are infectious, $\pi_p^{x\prime}$:

$$\chi_{xj} = \sigma_j \pi_p^{x\prime} \qquad (2)$$

$\pi_p^{x\prime}$ is calculated by integrating (numerically) the product of prevalence and expected number of partnerships formed, over all possible partner change rates in sex *x'*, and then dividing by the total expected number of partnerships formed, $\mu_{x\prime} = \mu$:

$$\pi_p^{x\prime} = \frac{1}{\mu}\int_{\sigma=0}^{\infty} P(\sigma|\mu,\alpha_{x\prime})\frac{\sigma\pi_p^x\rho_{x\to x\prime}}{\sigma\pi_p^x\rho_{x\to x\prime} + \lambda_x}\sigma d\sigma$$

Substituting (2) into (1), the probability that an individual *j* is infected, π_j, therefore fulfills the equality:

$$\frac{\pi_j}{1-\pi_j} = \frac{\sigma_j\pi_p^{x\prime}\rho_{x\prime\to x}}{\lambda_x}$$

$$\pi_j = \frac{1}{1+z_j}$$

where

$$z_j = \frac{\lambda_x}{\sigma_j\pi_p^{x\prime}\rho_{x\prime\to x}}$$

For individual *j*, the exact value of σ_j is not known, but the reported number of new partners, n_j, provides some information, allowing us to update our Gamma prior as described above. The expected prevalence in individuals reporting n_j partners is calculated by integrating the product of prevalence and the updated Gamma probability density for individual *j*:

$$\pi_j = \int_{\sigma=0}^{\infty} P_{Gamma}(\sigma|\alpha_j+n_j,\beta_j+1)\frac{1}{1+z(\sigma_j)}d\sigma$$

$$= \frac{(\beta_j+1)^{\alpha_j+n_j}}{\Gamma(\alpha_j+n_j)}\int_{\sigma=0}^{\infty}\frac{\sigma^{\alpha_j+n_j}e^{-(\beta_j+1)\sigma}}{\sigma+\lambda_x/\pi_p^{x\prime}\rho_{x\prime\to x}}d\sigma$$

The infection status of *j* has a Bernoulli distribution with parameter π_j:

$$P(\delta_j|\pi_j) = P_{Bernoulli}(\delta_j|\pi_j) = \begin{cases}\pi_j & \delta_j = 1\\ 1-\pi_j & \delta_j = 0\end{cases}$$

where

$$\delta_j = \begin{cases}1 & \text{if } j \text{ is infected}\\ 0 & \text{if } j \text{ is uninfected}\end{cases}$$

iii. Infection clearance rate

We modelled immunological clearance of infection using the parameter λ_x. The statistical model is described elsewhere,[1] and allows for two courses of infection: fast- or slow-clearing. A proportion *p* of

incident infections clear fast, and the remainder, 1 – *p*, clear slow. In this analysis we assume that only the slow-clearing infections last long enough to be detected in population-based studies. The clearance rate (denoted λ_x below) is therefore equal to the slow clearance rate in the clearance model, and the transmission probability we estimate is the probability that an infection is transmitted and then follows the slow-clearing course. The parameter values are inferred from published observational data in men and women[1,2].

In the absence of data on the rates of testing and treating for asymptomatic chlamydia infection at the time of Natsal-2 and NHANES, we were not able to account in our model for chlamydia clearance via treatment of asymptomatic infections. We investigated the results of this decision in our predictive checks (see below).

iv. Full likelihood

The full set of model parameters is $\{\mu, \beta, \sigma, p, \lambda, \psi, A\}$, where $A = {}^{F_j}/_{\sigma_j} = \pi_p^{x\prime}\rho_{x\prime\to x}$ for transmission from *x'* to *x* or $\pi_p^x\rho_{x\to x\prime}$ for transmission from *x* to *x'*. From these we derive the parameters $\{\alpha, \pi, \pi_p, \rho\}$. The meaning of each symbol is summarized in Table S3.

Table S3: Summary of symbols used to describe the model.

Symbol	Description
μ	Mean number of new partnerships per person.
$\beta = (\beta_m, \beta_f)$	Rate parameters for gamma distributions
$\sigma = (\sigma_1, \sigma_2, \dots)$	Poisson rates of partnership formation.
$p = (p_m, p_f)$	Proportion of infections in men and women which are fast-clearing.
$\lambda = (\lambda_m, \lambda_f)$	Clearance rate of slow-clearing infections
ψ	Sensitivity of culture diagnosis methods (for the clearance rate model).
$A = (A_{f\to m}, A_{m\to f})$	Per-partnership prevalence, multiplied by per-partnership transmission probability.
$\alpha = (\alpha_m, \alpha_f)$	Shape parameters for gamma distributions.
$\pi = (\pi_1, \pi_2, \dots)$	Expected chlamydia prevalence in each individual.
$\pi_p = (\pi_p^m, \pi_p^f)$	Proportion of all partnerships in which the man/woman is infected.
$\rho = (\rho_{m\to f}, \rho_{f\to m})$	Per-partnership transmission probability from an infected man/woman to a susceptible woman/man.

Survey weights w_i are incorporated by multiplying the relevant component of the log-likelihood by the weight. The log-likelihood of the data is given by:

$$L = L_{turnover} + L_{clearance} + L_{infection}$$

where:

- $L_{turnover}$ is the log-likelihood associated with the partnership turnover data in men and women.
$$L_{turnover} = L_{turnover}^m + L_{turnover}^f = \sum_m w_j \times P_{NB}(n_j|\alpha_m, \beta_m) + \sum_f w_j \times P_{NB}(n_j|\alpha_f, \beta_f)$$

- $L_{clearance}$ is the log-likelihood associated with the clearance data:

$$L_{clearance} = \sum_{data} P_{binomial}(r|n_{test}, \theta)$$

where n_{test} is the number of people tested for each data point, r is the number who had cleared their infection and ϑ is the proportion expected to clear the infection (full details provided elsewhere[1]).

- $L_{prevalence}$ is the log-likelihood associated with the prevalence data in men and women reporting different numbers of partners:

$$L_{prevalence} = \sum_{m} w_j P_{Bernoulli}(\delta_j|\pi_j) + \sum_{f} w_j P_{Bernoulli}(\delta_j|\pi_j)$$

d. Inference and Estimation

i. Priors

Prior distributions for the parameters were as follows:

$\mu \sim$ Exponential(0.1) (uninformative)
$\beta \sim$ Exponential(0.1) (uninformative)
$p \sim$ Beta$(1,1)$ (uninformative)
$\lambda_{slow} \sim$ Exponential(0.001) (uninformative)
$\psi \sim$ Beta$(78,8)$ (based on studies comparing test performance[7])
$A \sim$ Exponential(0.001) (uninformative)

ii. Bayesian methods and sampling of posterior distribution

Estimation was carried out by sampling from the posterior using a Markov chain Monte Carlo (MCMC) algorithm implemented in the Stan software,[8] within the R environment.[9] The data, Stan model file and R scripts used for handling input and results are all available online at https://github.com/mrc-ide/ct_transmission_prob. MCMC estimation is carried out by drawing thousands of samples from the joint posterior distribution. We ran four chains for 2000 iterations each, discarding the first 1000 "warmup" iterations of each chain. The results reported below are summary means, medians and credible intervals of the marginal distributions from this sampled joint posterior.

iii. Posterior predictive checks

We carried out graphical posterior predictive checks[6] to check the fit of the model. We simulated values for the data (number of partners and infection status for each individual), using each sample from the joint posterior distribution. The simulated data were compared to observed data to look for any systematic differences.

We expect that a proportion ϕ_x of incident chlamydia infections in sex x will cause symptoms that prompt testing and treatment, while the remaining $1 - \phi_x$ are asymptomatic. As noted above, our model considers *asymptomatic* infections, so the modelled force of infection represents the force of asymptomatic infection. The force of *symptomatic* infection is $\frac{\phi_x}{1-\phi_x}$ times the force of asymptomatic infection, and we expect to observe symptomatic diagnoses in the population at this per-person rate. We used prior distributions for ϕ ($\phi_m \sim Beta(11,5)$; $\phi_f \sim Beta(27,90)$ [7]), the posteriors for force of infection, and the population size, to simulate the annual number of symptomatic diagnoses.

2. Results

a. Posterior parameter distributions

Table S4: Summary of posterior distributions for model parameters, inferred using data from the second National Study of Sexual Attitudes and Lifestyles (Natsal-2) and National Health and Nutrition Examination Surveys (NHANES). The first six parameters were sampled directly; the last three were calculated from the first six, as described in the text.

Parameter	Natsal				NHANES			
	Men		Women		Men		Women	
	Mean	Median (95% CrI)	Mean	Median (95% CrI)	Mean	Median (95% CrI)	Mean	Median (95% CrI)
μ (Mean partnerships)	0.593	0.592 (0.545, 0.646)	Shared parameter		0.922	0.921 (0.848, 1.001)	Shared parameter	
β (Rate parameter for gamma distribution)	0.512	0.510 (0.437, 0.597)	0.366	0.364 (0.291, 0.453)	0.176	0.176 (0.156, 0.198)	0.137	0.136, (0.116, 0.160)
p (Proportion of infections fast-clearing)	0.314	0.314 (0.208, 0.423)	0.206	0.206 (0.150, 0.266)	0.316	0.316 (0.209, 0.430)	0.207	0.207 (0.152, 0.267)
λ (year^{-1}) (Slow clearance rate)	0.642	0.571 (0.144, 1.54)	0.737	0.735 (0.601, 0.884)	0.634	0.566 (0.122, 1.512)	0.735	0.733 (0.600, 0.883)
ψ (Sensitivity of culture diagnosis*)	0.911	0.912 (0.860, 0.953)	Shared parameter		0.911	0.912 (0.859, 0.954)	Shared parameter	
$A = (A_{f\to m}, A_{m\to f})$ ($\pi_p \times \rho$; see below)	0.025	0.022 (0.005, 0.062)	0.028	0.028 (0.018, 0.043)	0.014	0.012 (0.003, 0.033)	0.040	0.039 (0.028, 0.056)
α (Shape parameter for gamma distribution)	0.303	0.303 (0.262, 0.350)	0.217	0.215 (0.179, 0.261)	0.162	0.162 (0.146, 0.179)	0.126	0.126 (0.111, 0.142)
π_p (Proportion of all partnerships infected.)	0.087	0.086 (0.062, 0.115)	0.105	0.104 (0.069, 0.148)	0.114	0.113 (0.084, 0.147)	0.261	0.261 (0.207, 0.317)
$\rho = (\rho_{f\to m}, \rho_{m\to f})$ (Per-partnership transmission probability)	0.252	0.214 (0.051, 0.670)	0.334	0.321 (0.184, 0.559)	0.052	0.046 (0.010, 0.131)	0.358	0.349 (0.226, 0.549)

CrI: credible interval

*Culture sensitivity at re-testing for chlamydia clearance, in people previously diagnosed by culture.[1]

b. Posterior predictive checks

i. Partner number distributions

Figure S2 illustrates the model's agreement with partnership number data, showing the actual and simulated proportions of men and women who reported each number of partners. Transparent grey circle markers represent simulations from the posterior distributions; lines show the 50th (solid) and 2.5th/97.5th (dashed) centiles of the simulations, and red crosses show the data. For a perfect model and completely accurate reporting of the data, we would expect the dashed lines to enclose 95% of data points.

In both studies, the partnership numbers simulated in men generally agreed well with the data. The predictive properties were less good in women, with under-reporting of high partner numbers compared to simulations. If the average number of partnerships formed by men and women were allowed to differ then the agreement between simulations and data was improved and the posterior distributions for transmission probability remained similar. In our model we chose to constrain the average number in men and women to be equal because this is a necessary condition in reality.

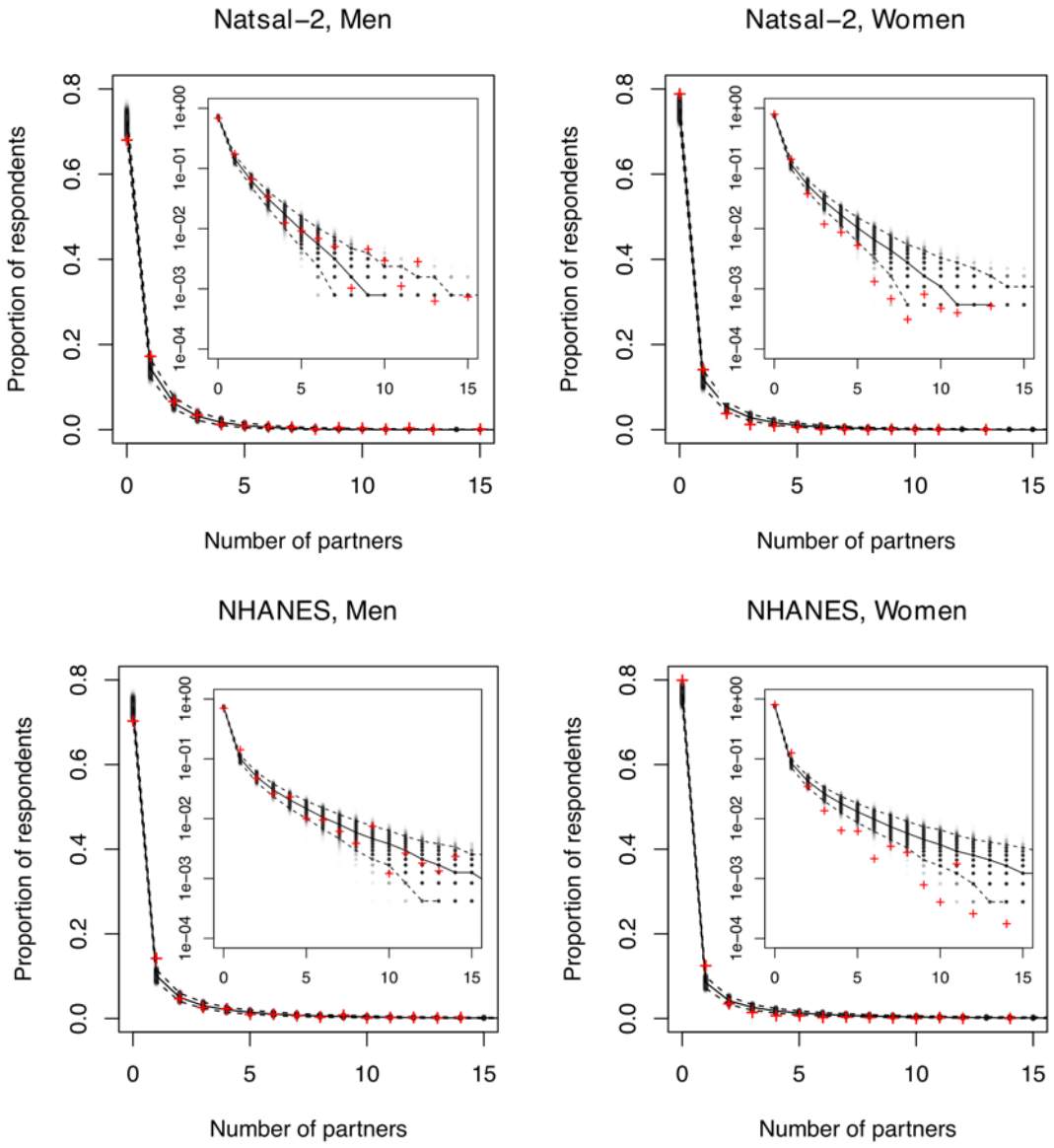


Figure S2: Simulated (grey) and observed (red) proportions of men (left) and women (right) reporting different numbers of new partners in the last year in the second National Study of Sexual Attitudes and Lifestyles (Natsal-2; top) and National Health and Nutrition Examination Studies (NHANES) 2009-2014 (bottom). The main graph in each panel uses a linear scale on the y-axis, and the inset shows the same information but on a log scale. Simulations are shown using transparent grey markers, so that several

superimposed markers appear as a darker grey. The solid and dashed lines show the 2.5th, 50th and 97.5th centiles of the simulations. The observed data shown takes into account the survey weights.

ii. Infection status

We checked the predictive properties of the infection model by using each sampled parameter set to simulate infection status in each survey participant, given their reported number of partners. In Figures S3 (Natsal-2) and S4 (NHANES), each transparent grey marker shows simulated prevalence among the participants reporting a given number of partners, which agreed well with the observed data. Only a small number of participants reported the highest numbers of partners (see bar graphs in lower panels), so only a few levels of prevalence were possible in those with several partners. For example, one man in Natsal-2 reported 19 partners, so simulated prevalence could only be 0 (one man, uninfected) or 1 (one man, infected).

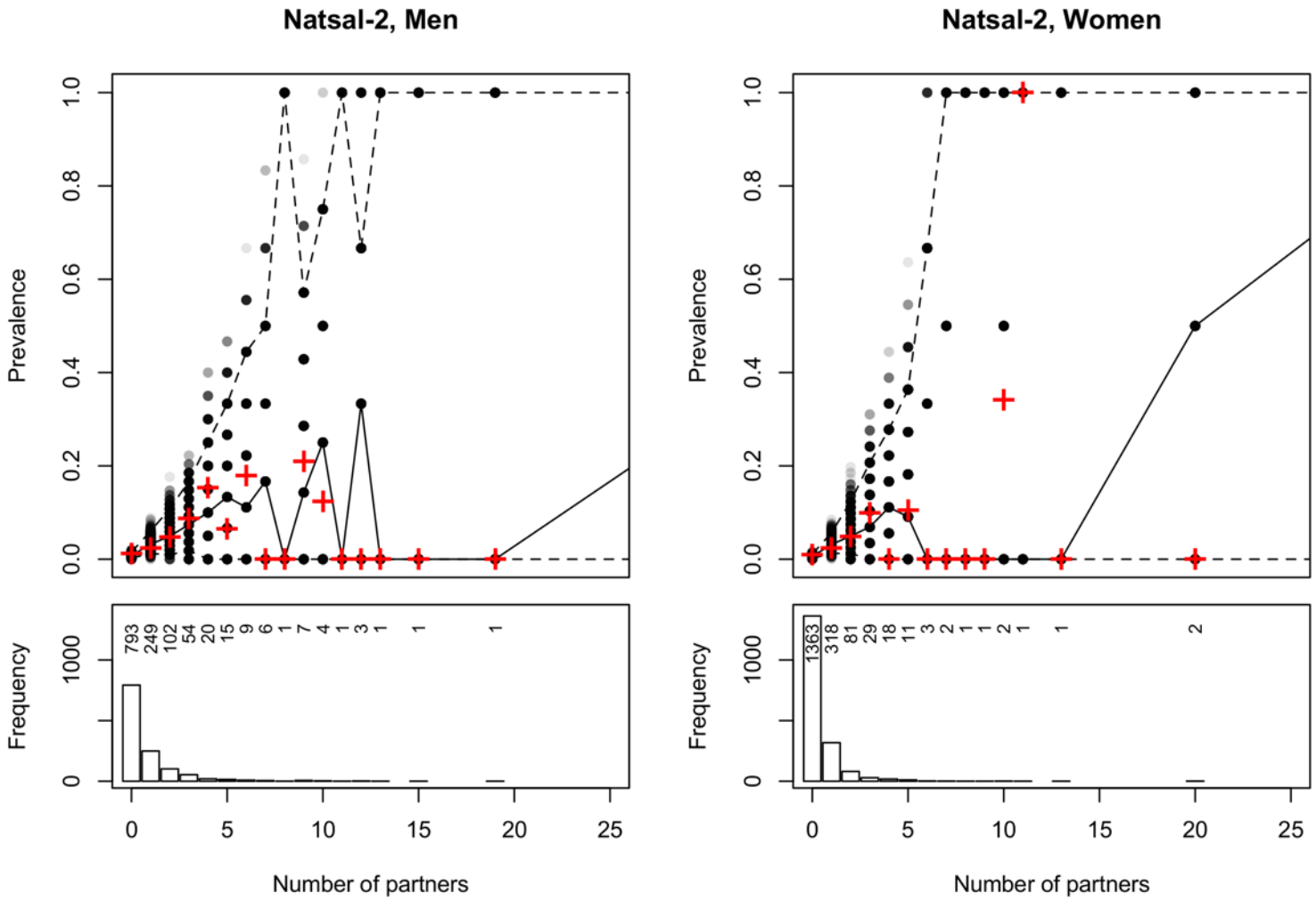


Figure S3: Simulated (grey) and observed (red) chlamydia prevalence (y-axis) in men and women reporting different numbers of new partners in the last year (x-axis) in the second National Study of Sexual Attitudes and Lifestyles (Natsal-2). Simulations are shown using transparent grey markers, so that several superimposed markers appear as a darker grey. The solid and dashed lines join the 2.5th, 50th and 97.5th centiles of the simulations. The observed data takes into account the survey weights. Bar charts below each plot show the (unweighted) number of survey participants reporting each number of partnerships.

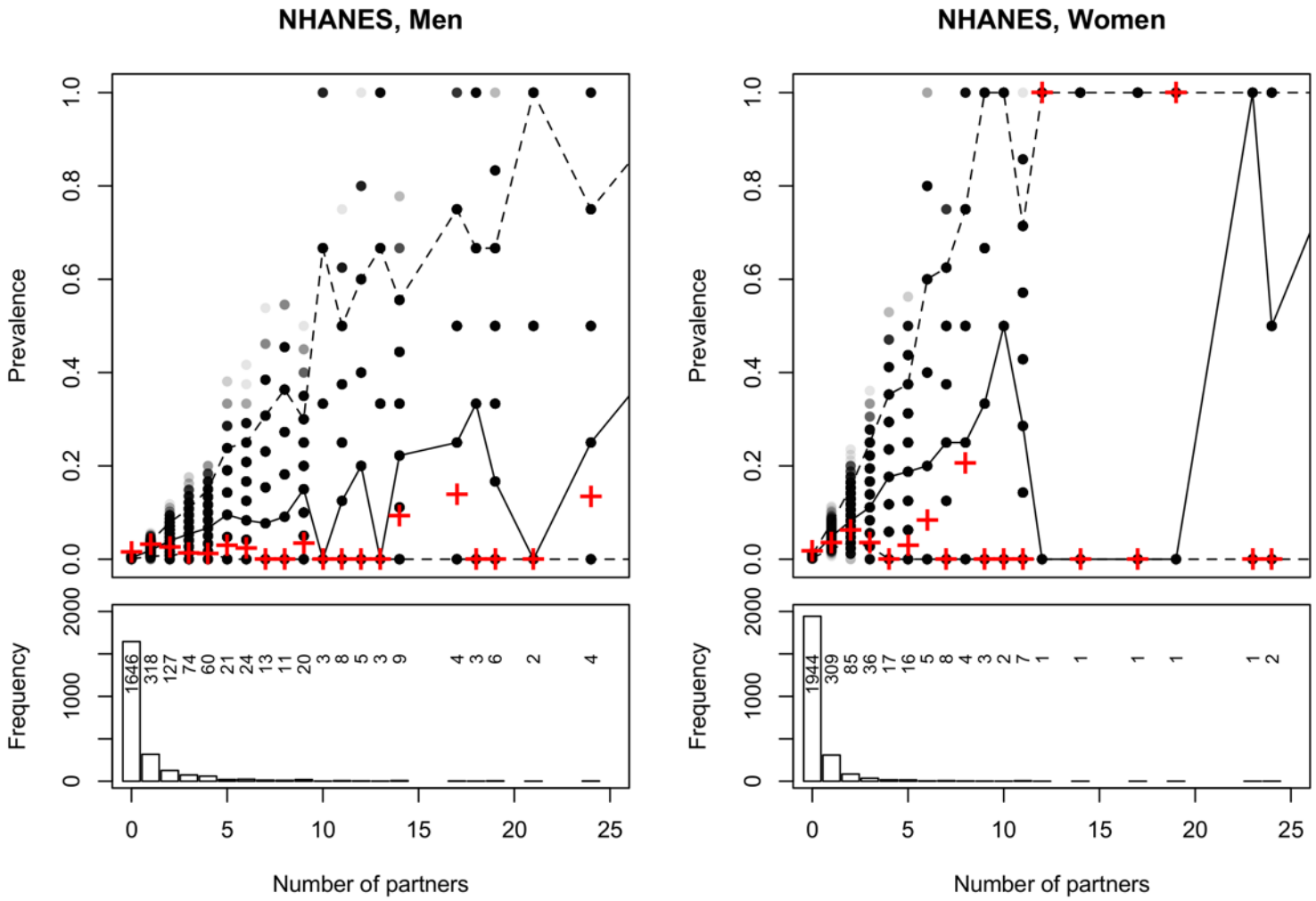


Figure S4: Simulated (grey) and observed (red) chlamydia prevalence (y-axis) in men and women reporting different numbers of new partners in the last year (x-axis) in the National Health and Nutrition Examination Studies (NHANES). Simulations are shown using transparent grey markers, so that several superimposed markers appear as a darker grey. The solid and dashed lines join the 2.5th, 50th and 97.5th centiles of the simulations. The observed data takes into account the survey weights. Bar charts below each plot show the (unweighted) number of survey participants reporting each number of partnerships.

iii. Symptomatic infections

Table S5 shows the median and central 95% range of simulated numbers of symptomatic chlamydia cases, based on our posterior distributions and the male and female populations of England aged 15-44 in 2000 (Natsal-2), or the US aged 15-39 in 2009 (NHANES). For comparison, we also report the number of diagnoses recorded in surveillance systems covering approximately the same times and locations. In men in both studies and women in Natsal-2 the range of our simulations overlapped with the range from surveillance, suggesting that most of the observed diagnoses can be accounted for by treatment-seeking in response to symptoms, and that few additional diagnoses were made as a result of asymptomatic testing. In women in NHANES, more diagnoses were observed than we expected to be sought by symptomatic cases alone, so it seems likely that there was additional testing of asymptomatic women which would merit further empirical investigation.

Table S5: Numbers of symptomatic chlamydia cases simulated using posterior parameter distributions inferred using Natsal-2 and NHANES data, and diagnoses recorded in surveillance systems covering approximately the same times and locations. For comparison to Natsal-2 we used diagnosis rate ranges in 15-44-year-olds in 2000,[10] and for NHANES we used the range of recorded diagnoses over the years 2009-2014.[11]

Survey	Group	Simulated symptomatic cases (1000s; median and 95% CrI)	Observed diagnoses (1000s)
Natsal-2	Men aged 15-44 years	109 (25-327)	30-41
	Women aged 15-44 years	46 (25-77)	48-105
NHANES	Men aged 15-39 years	397 (83-1149)	307-398
	Women aged 15-39 years	429 (259-682)	879-981

c. Sensitivity Analysis

i. Balancing partnership numbers

We tested the effect of constraining the mean numbers to be equal by repeating the analysis, relaxing the constraint of equal mean partnership number in men and women (see online code). Figure S5 illustrates this model's agreement with partnership number data. In both studies the agreement between simulations and observations is improved compared to the constrained model, especially in women, but more than 5% of observations still fell outside the 95% prediction interval. Using Natsal-2, the posterior median (95%CrI) for the mean number of new partners per year in men was 0.75 (0.67-0.83) and in women was 0.40 (0.35-0.45). Inferred transmission probabilities were 32.4% (18.4-55.5)% (male-to-female) and 26.2% (5.8-84.8)% (female-to-male). Using NHANES, the inferred mean number of partners in men was 1.10 (1.08-1.33) and in women was 0.58 (0.52-0.66). Transmission probabilities were 31.3% (20.4-48.7)% (male-to-female) and 6.3% (1.4-18.0)% (female-to-male). Therefore, constraining the mean number of partnerships to be equal did not materially change the posterior distributions for transmission probabilities.

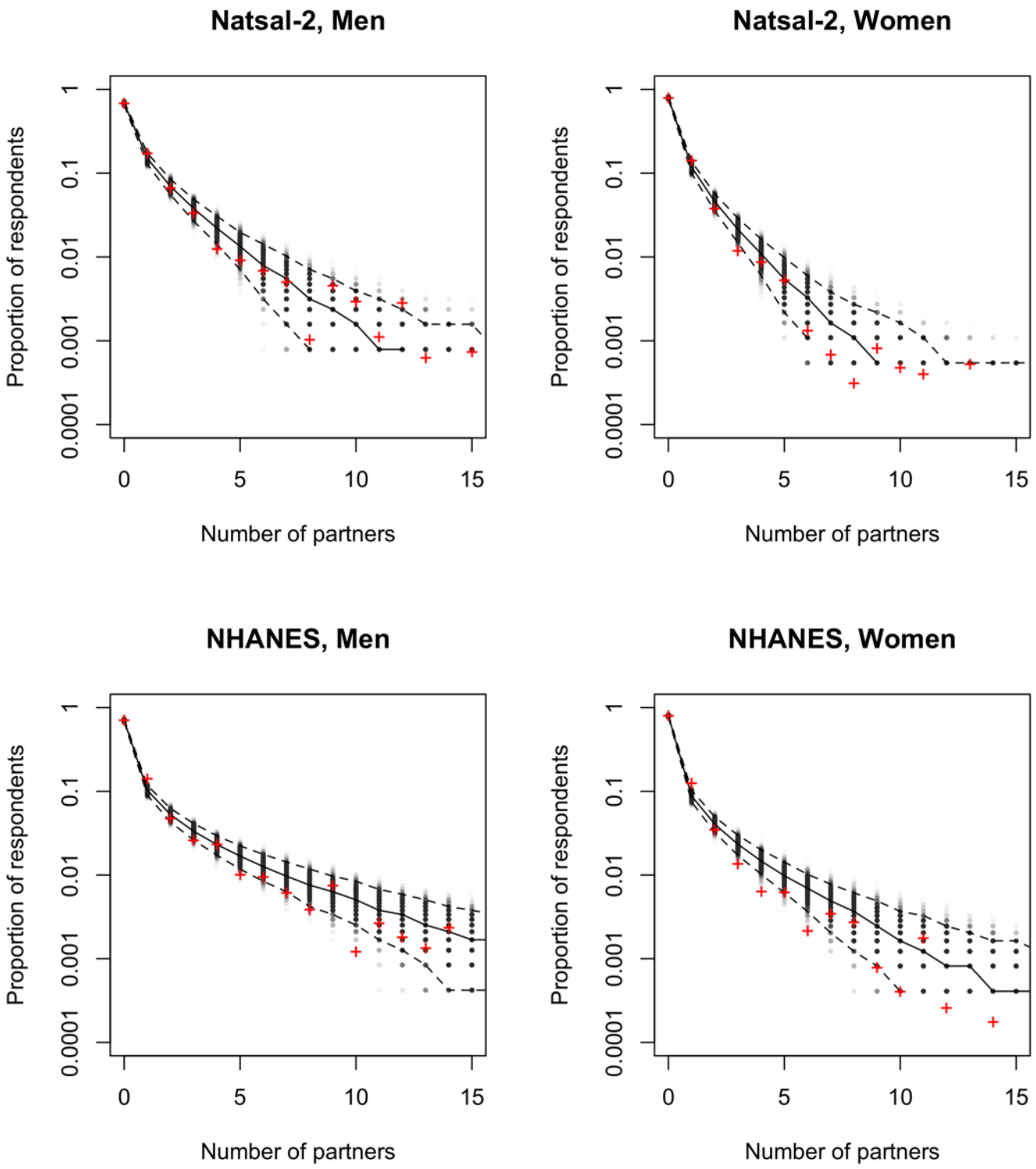


Figure S5: Simulated (grey) and observed (red) proportions of men (left) and women (right) reporting different numbers of new partners in the last year in the second National Study of Sexual Attitudes and Lifestyles (Natsal-2; top) and the National Health and Nutrition Examination Surveys (NHANES) 2009-2014 (bottom). In this model, the mean number of partnerships was not constrained to be equal between the sexes. Simulations are shown using transparent grey markers, so that several superimposed markers appear as a darker grey. The solid and dashed lines show the 2.5th, 50th and 97.5th centiles of the simulations. The observed data shown takes into account the survey weights.

ii. Condom use

In Natsal-2 participants were asked, *With how many different women/men have you had vaginal (or anal) intercourse in the past year without using a condom?* To investigate the potential effects of condom use on our estimates, we used this question to estimate the number of new partners without a condom:

- If participants reported 0 partners without a condom then we classified them as having 0 new partners without a condom.
- If participants reported the same number of partners in the last year as partners without a condom (i.e. if all partners in the last year were without a condom) then we classified the number of new partners without a condom as the same as the total number of new partners.
- If neither of these conditions applied then we classified the number of new partners without a condom as the reported number of partners without a condom.

We used the same model as in the main analysis to estimate the transmission probabilities in partnerships where condoms were not always used. Figure S6 shows the posterior distributions compared to the posteriors in the main analysis.

As expected, the posterior distributions were shifted slightly to the right, suggesting higher transmission probabilities in partnerships without a condom, but the shift was small compared to the uncertainty in the estimates. The posterior median (95% credible interval) transmission probabilities were 40.1% (21.5-72.8)% from men to women and 31.6% (7.2-96.1)% from women to men. We conclude that it might be valuable for sexual behavior surveys to collect information on the annual number of new partnerships without a condom for parameter inference and predictive modelling. In the absence of such data, however, it is more reliable to calculate an average probability across all new partnerships, and we have no reason to suppose that such an average is not valid.

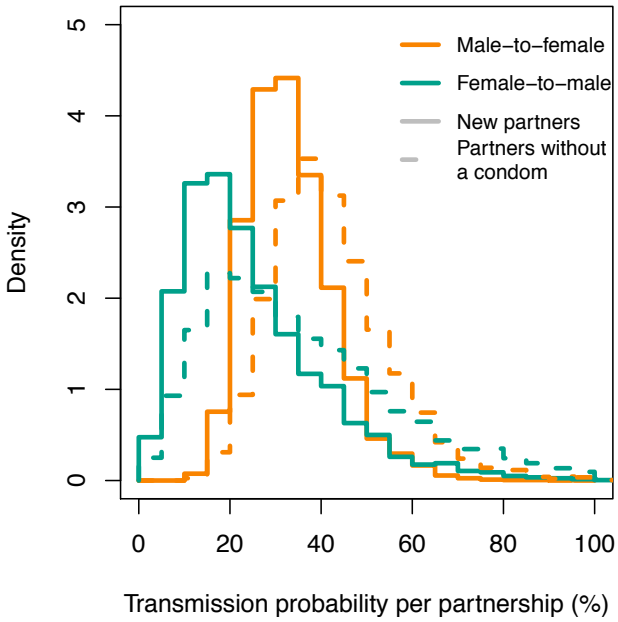


Figure S6: Posterior distributions for the per-partnership probability of chlamydia transmission, derived using data from the second National Study of Sexual Attitudes and Lifestyles (Natsal-2). The orange lines represent male-to-female transmission probability and the green lines female-to-male. The solid lines represent distributions inferred from reported numbers of new heterosexual partners, as in the main analysis. The dashed lines represent distributions inferred from the estimated number of new partners without a condom, as described in the text above.

iii. Assortative mixing

The model reported in the main text assumes random mixing between men and women – that is, that for individual *j*, the probability that a partnership they form with a member of the opposite sex is a potential source of infection does not depend on *j*'s partnership formation rate. In fact, evidence indicates that sexual mixing is assortative,[12,13] although this is difficult to quantify precisely.

To investigate the potential effects of assortative mixing in our model, we reasoned that if individuals with more partners tend to form partnerships with others who also have more partners – and therefore the partners are more likely to be infected with chlamydia – then $\pi_p^{x\prime}$ would be higher in people with more partners. If the transmission probability were the same for every partnership then we would therefore expect the product

$$A_{x\prime\to x} = \pi_p^{x\prime}\rho_{x\prime\to x}$$

to be higher in people with more partners.

We ran an adapted model which allows A to be different for men and women reporting different numbers of partners. If people with more partners are more likely to form partnerships with infected people then we would expect A to be higher in those individuals.

Figure S7 shows the posterior distributions for A that we inferred in men and women reporting different numbers of partners. For Natsal-2, although the posterior distributions for A were slightly higher in people reporting no new partners, there was considerable overlap and therefore no evidence of significantly higher prevalence in partnerships presented to individuals with high partnership formation rate than to those with low formation rate. In NHANES the posterior distributions suggested higher values for A in both men and women reporting no new partners: the opposite of what we would expect if there is assortative mixing. This pattern may arise if there is a higher transmission probability in slow-turnover partnerships, because they tend to last longer and have more sex acts during the infectious period, possibly with lower levels of condom use.

We found no evidence in either Natsal-2 or NHANES of higher A in people reporting more partners, providing confidence that the random mixing in the model has not affected our results.

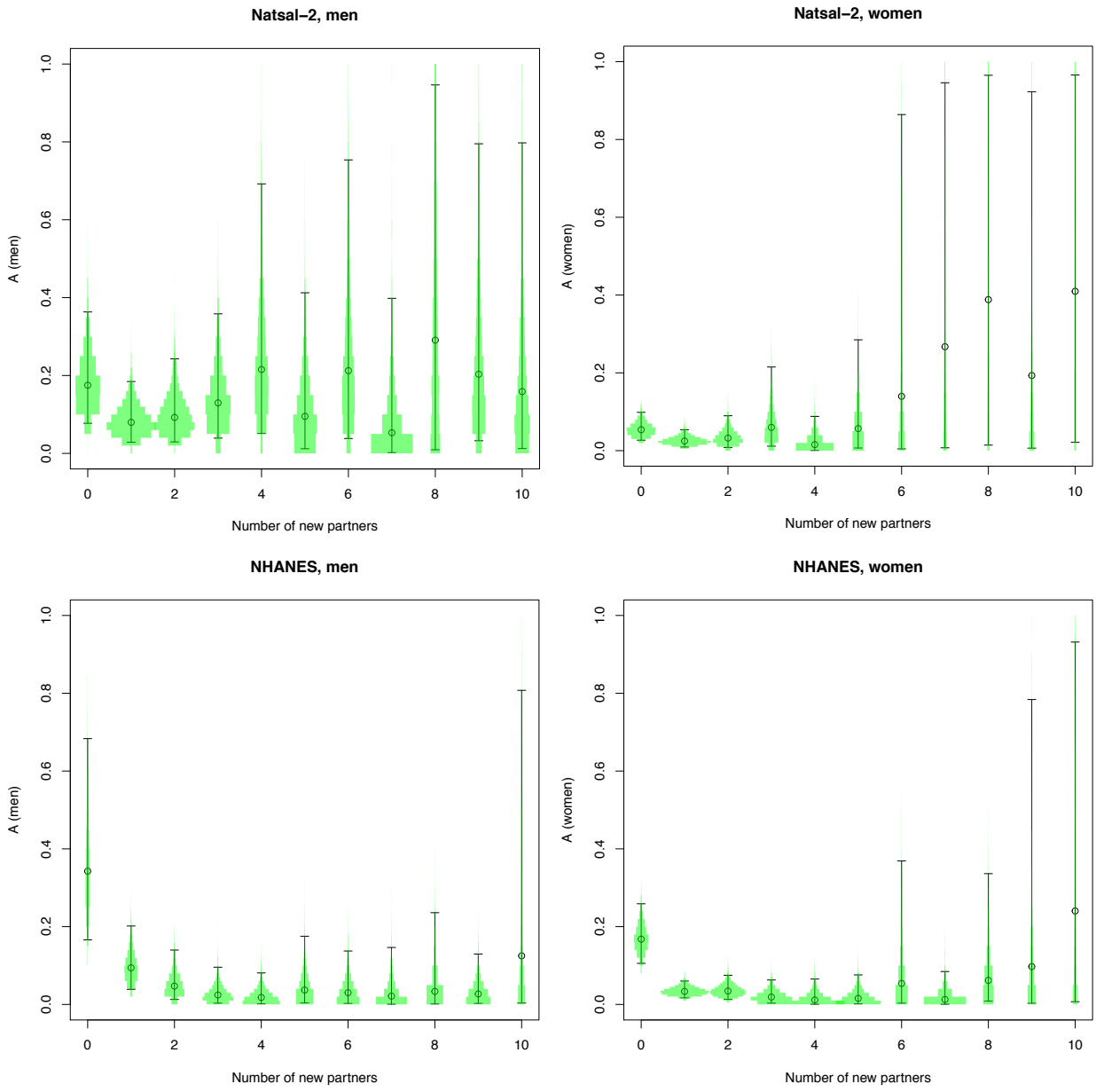


Figure S7: Posterior distributions for A inferred separately for men and women reporting different numbers of partners. Error bars show median and 95% credible interval, and green polygons are histograms of the posteriors.

3. References

1. Price MJ, Ades AE, De Angelis D, et al., Mixture-of-exponentials models to explain heterogeneity in studies of the duration of Chlamydia trachomatis infection. *Stat Med.* 2013;32:1547-60.
2. Lewis J, Price MJ, Horner PJ, et al., Genital *C. trachomatis* infections clear more slowly in men than women, but are less likely to become established. *J Infect Dis.* 2017;216:237-244.
3. Johnson A, Fenton K, Copas A, et al., *National Survey of Sexual Attitudes and Lifestyles, 2000-2001* [computer file]. Colchester, Essex: UK Data Archive [distributor], 2005. SN: 5223, http://doi.org/10.5255/UKDA-SN-5223-1.
4. Centers for Disease Control and Prevention (CDC)} and National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data.* [computer file]. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention [distributor].
5. Handcock MS and Jones JH, Likelihood-based inference for stochastic models of sexual network formation *Theor Popul Biol.* 2004;65:413-422.
6. Gelman A, Carlin JB, Stern HS, et al.. Bayesian Data Analysis. Chapman &Hall/CRC Press, London, third edition. 2013.
7. Geisler WM, Wang C, Morrison SG, et al., The natural history of untreated *Chlamydia trachomatis* infection in the interval between screening and returning for treatment. *Sex Transm Dis.* 2008;35:119-123.
8. Stan Development Team. RStan: the R interface to Stan, Version 2.10.1. http://mc-stan.org/interfaces/rstan.html Accessed 7 June 2016.
9. R Core Team. R: A Language and Environment for Statistical Computing. 2016.
10. Chandra N, Soldan K, Dangerfield C, et al., Filling in the gaps: estimating numbers of chlamydia tests and diagnoses by age group and sex before and during the implementation of the English National Screening Programme, 2000 to 2012. *Eurosurveillance.* 2017;22:30453.
11. Centers for Disease Control and Prevention. *NCHHSTP AtlasPlus*. https://www.cdc.gov/nchhstp/atlas/index.htm. Accessed 3 December 2019.
12. Aral SO, Hughes JP, Stoner B, et al., Sexual Mixing Patterns in the Spread of Gonococcal and Chlamydial Infections *Am J Public Health.* 1999;89:825-833.
13. Prah P, Copas AJ, Mercer CH, et al., Patterns of sexual mixing with respect to social, health and sexual characteristics among heterosexual couples in England: analyses of probability sample survey data *Epidemiol Infect.* 2015;143:1500-1510.