Published in final edited form as:
J R Stat Soc Ser A Stat Soc. 2011 October ; 174(4): 975–989. doi:10.1111/j.1467-985X.2011.00691.x.

A Stochastic Model for Assessing *Chlamydia trachomatis* Transmission Risk Using Longitudinal Observational Data

Wanzhu Tu, **Pulak Ghosh**, and **Barry P. Katz**

Summary

Bacterium *Chlamydia trachomatis* causes genital chlamydia infection. Yet little is known about the transmission efficiency of this organism. Ethical constraint against exposing healthy subjects to infected partners precludes the possibility of quantifying transmission risk through controlled experiments. This research proposes an alternative strategy that relies on observational data. Specifically, we present a stochastic model that treats longitudinally observed infection states in a group of young women as a Markov process. The proposed model explicitly accommodates the parameters of *C. trachomatis* transmission, including per-encounter sexually transmitted infection (STI) acquisition risks, with and without condom protection, and the probability of antibiotic treatment failure. The male-to-female transmission probability of *C. trachomatis* is then estimated by combining the per-encounter disease acquisition risk and the organism's prevalence in the male partner population. The proposed model is fitted in a Bayesian computational framework.

Keywords

Binary outcome; Bacterial infection; Transmission probability; Longitudinal study; Markov model; MCMC; Observational data

1 Introduction

Chlamydia is a common sexually transmitted infection (STI) caused by the bacterium *Chlamydia trachomatis*. It is among the most prevalent STIs in the United States (CDC 2007a). It is estimated that approximately 2.3 million non-institutionalized US civilians ages 14–39 are infected with *C. trachomatis* each year (CDC 2007b; Datta *et al*. 2007), but prevalence rates are known to be substantially higher among the young (ages 14–25 years) and inner city residents (Miller *et al*. 2004; Weinstock *et al*. 2004). Repeated and prolonged infections induce substantial morbidity and detrimental sequelae including pelvic inflammatory disease, ectopic pregnancy, and tubal infertility in women and epididymis and reactive arthritis (Reiter's Syndrome) in men, although complications among men are less common (Cates and Wasserheit 1991; Hillis *et al*. 1997; Westrom 1999; Fung *et al*. 2007). Recent research also suggests that chlamydial infections increase susceptibility to human immunodeficiency virus (HIV) thus heightening the risk of HIV transmission (Fleming and Wasserheit 1999).

In the US, genital chlamydial infection is primarily transmitted through sexual intercourse among heterosexual partners, although intercourse with an infected partner does not automatically result in infection. Despite the routine use of antibiotics to treat identified cases, persistently high prevalence rates suggest substantial risk of disease transmission (Xu

Correspondence: Wanzhu Tu, Division of Biostatistics, Indiana University School of Medicine, 410 West 10th Street, Suite 3000, Indianapolis, IN 46202, wtu1@iupui.edu.

et al. 2000; Lamontagne *et al*. 2007). Given the health consequences and the role that transmission plays in the spread of these diseases, it is surprising how little is known about *C. trachomatis* transmissibility. Herein, we define transmissibility as the probability of acquiring chlamydia through one exposure with an *infected* partner.

The primary difficulty in assessing the organism-specific transmission probability is that the quantity cannot be directly obtained from *controlled experiments*, due to ethical constraints against knowingly exposing human subjects to infectious pathogens. In the 1970's, investigators examined the risk of men acquiring gonorrhea from infected female sex workers, or FSW (Holmes *et al*. 1970). In the study, uninfected crew members on board a US navy ship were allowed to visit FSW during a foreign port visit. Using the estimated prevalence of *Neisseria gonorrhoeae* among the FSW and the reported number of sexual contacts between the crew and the FSW, the investigators were able to estimate the female-to-male *N. gonorrhoeae* transmission probability. Despite concerns about the accuracy of the *N. gonorrhoeae* prevalence estimate for the FSW due to the limited sampling frame and potential bias associated with unregistered FSW (Felton 1973; Hooper *et al*. 1978; Baumgarten 1980), the estimate from this experiment remains the only accepted measure of the gonorrhea transmission probability. Studies such as this, however, are no longer recognized as ethical. Although human challenge studies injecting infectious pathogens have been used to study host immune responses to organisms such as *N. gonorrhoeae* and *Haemophilus ducreyi* (Cohen and Cannon 1999; Spinola *et al*. 2002), the methods of infection are artificial, thus not truly relevant to *sexual* transmission.

A logical alternative to experiment-based studies is to estimate transmission risk using observational data. Along this line, Katz (1992) has proposed a method that uses cross-sectional data from contact tracing programs to estimate the chlamydia transmission probability among couples. However, since contact tracing data rarely provide information on coital frequency for each individual pair of partners, estimates obtained essentially represent transmission risks associated with some "average" number of encounters among all couples. The per exposure transmission risk for *C. trachomatis* remains unknown.

In this paper, we propose an alternative strategy: Using observational data collected from a longitudinal cohort of young women, we estimate the per-encounter *C. trachomatis* acquisition risk (p), i.e., the probability of a female subject being infected with *C. trachomatis* through one unprotected sexual encounter with a male partner. It should be noted that the per-encounter *C. trachomatis* acquisition risk p reflects not only the transmissibility of the *C. trachomatis* organism, but also the organism's prevalence in the partner population. Clearly, the more transmissible the organism, the higher the risk of disease acquisition. Similarly, the more prevalent the organism in the partner population, the higher the per-encounter infection risk. Therefore, with the estimated *C. trachomatis* prevalence in the male partner population, we could accurately quantify the male-to-female *C. trachomatis* transmission probability from the following relationship: Per-encounter Acquisition Risk=Transmission Probability × Partner Prevalence.

To estimate the per-encounter *C. trachomatis* acquisition risk (p), we propose a stochastic model that describes the change of infection states within a subject over time. The proposed model allows us to examine the effect of condom use in the modification of *C. trachomatis* acquisition risk. It also allows us to assess the probability of antibiotic treatment failure. The transmission probability is achieved with the estimated per-encounter acquisition risk p, and estimated *C. trachomatis* prevalence rates in various male partner populations. Methodologically, the proposed approach is a novel application of Markov stochastic methods in longitudinally observed infection data. The model explicitly accommodates the most important infection parameters. For convenience, we carry out the model computation

in a Bayesian Markov Chain Monte Carlo framework. To the best of our knowledge, this is the first attempt to use observational data to quantify per-encounter chlamydia transmission risk among heterosexual partners.

It should be noted that there is a body of literature about HIV/AIDS transmission risk (Padian *et al.* 1990, 1991, and 2007; European Study Group on Heterosexual Transmission of HIV, 1992). However, fundamental differences between viral and bacterial STI demand a new statistical approach for the latter. For example, unlike HIV/AIDS, infections with bacteria such as *C. trachomatis* are curable. Identified infections are routinely treated. Antimicrobial treatment often effectively aborts an infection episode, causing a shift in infection state and posing greater methodological challenges for the quantification of transmission risks. In this exercise, frequently measured infection data are essential for the capture of the changing infection states. To our knowledge, not many studies are designed to generate such data. In the next section, we will describe an epidemiological study that provides ideal data for our research.

2 Description of Data Source

The Young Women's Project (YWP) is an epidemiological study of recurrent STI in adolescent women. The original purpose of the investigation was to examine the risk factors associated with STIs in young women. The YWP study design and data collection, however, provide a platform for the estimation of the *C. trachomatis* transmission probability. The basic YWP data collection scheme is depicted in Figure 1.

Briefly, young woman between the ages 14 and 17 years, able to understand English, without serious psychiatric disturbances or mental handicaps, and attending one of three participating primary-care clinics were eligible for enrollment into the YWP. At these clinics, young women who met the enrollment criteria, which were independent of prior sexual experience, were approached by research staff. Those who agreed to participate were approached and recruited at their scheduled clinical visits. Informed consent and parental permission were obtained at enrollment.

Upon enrollment, all study participants received enrollment interviews and underwent a pelvic exam, where a cervical swab for STI testing was collected and analyzed with nucleic acid amplification tests for *C. Trachomatis*. Infected participants were treated while at the clinics or shortly after test results became available. Enrolled YWP participants have quarterly clinical visits for up to 54 months. During the study period, STI testing and treatment were repeated at each quarterly visit. See Figure 1 for a schematic depiction of the study plan. In Figure 1, STI tests are represented as blue circles; treatments are represented as green circles. All test results and treatments were noted in the participant's medical record, including antimicrobial agents used, prescription dose and regimen, and route of delivery. Also at each quarterly visit, participants participated in a face-to-face interview, in which they were queried about numbers of coital events, with and without condom protection, since the previous quarterly visit.

In this research, we focus on the analysis of the first 200 participants who have completed at least two follow-up visits after the enrollment interview. We used this subset of data because this is the group of young women whose test results are available at the time of analysis.

3 Method

3.1 Notation

We let Y_{ij} indicate the infection state of the ith participant at the jth visit, $i = 1, 2,\ldots, I$, and $j = 0, 1, 2,\ldots, J_i$, where $j = 0$ represents the enrollment visit: $Y_{ij} = 1$ if the participant is infected, $Y_{ij} = 0$ otherwise. Similarly, we let Z_{ij} be the treatment indicator: $Z_{ij} = 1$ if the participant is treated, $Z_{ij} = 0$ otherwise. We denote the treatment failure probability as q. For self-reported sexual behaviors, we let n_{ij} and m_{ij} be the total number of coital events and the number of coital events protected by condom, in the three month period preceding the jth visit, respectively.

3.2 A Markov Model for *C. trachomatis* Infections

The approach that we use for the modeling of chlamydia infection is conceptualized in Figure 2.

Herein, we assume that the sequence of infection states from the ith subject form a stochastic process that satisfies the Markov property, i.e., the infection state at time j depends only on the infection state at time $j-1$, the treatment that the participant had received at $j - 1$, and her sexual behaviors between the two time points, $j - 1$ and j. In other words, we assume that the *C. trachomatis* infection state at any given visit Y_{ij} is a stochastic function of the previous infection status Y_{ij-1}, the presence or absence of antibiotic treatment Z_{ij-1}, and her possible exposure to the organism (characterized by n_{ij} and m_{ij}) between the visits. Our objective is to estimate the per-encounter risk of *C. Trachomatis* from the observed data (**Y**,**Z**, **n**,**m**).

For simplicity, we let p_i be the *C. trachomatis* acquisition risk in an unprotected sexual encounter for the ith participant. Similarly, we let p_i^c be the risk associated with a condom protected encounter, where $c > 0$ represents the effect of a condom in the modification of the per-encounter risk p_i. Specifically, if condoms are indeed effective in protecting against *C. trachomatis* acquisition, we will have $p_i > p_i^c$ or $c > 1$. Since the numbers of coital events with and without condom protection, m_{ij} and $n_{ij} - m_{ij}$, are known for the three month interval between $j - 1$ and j, the probability for the ith participant *not* to acquire infection through behavior in the interval can be calculated as $\Delta_{ij}=(1 - p_i)^{n_{ij}-m_{ij}}(1 - p_i^c)^{m_{ij}}$.

It should be noted that the current formulation is extendable. For example, one could let the per-encounter disease acquisition risk to be event-specific rather than subject-specific. Such an extension may be useful for the assessment of event-specific modifiers of transmission risk, if event-level (or partner-specific) covariates are available. We will briefly discuss that extension in Section 4.1. Herein, we assume the disease acquisition probability to be subject-specific because the current scientific understanding is that once exposed to *C. trachomatis*, whether a woman develops an infection or not is primarily influenced by her own immune response to the bacterium rather than who is the source of the pathogen, or the fashion in which the exposure occurs. Therefore, once we control for the disease prevalence in the partner population, the disease acquisition probability could be considered a subject-specific risk.

With this, all consecutive infection states in the stochastic process can be characterized in four scenarios: (a) no infection; (b) successfully treated infection; (c) newly acquired infection; and (d) recurrent infection. These scenarios are graphically depicted in Figure 3.

Analytically, the scenarios can be described as follows:

1. No infection detected between two consecutive visits. Since the participant tested negative at Time $j-1$, there would not be any treatment; since the participant tested negative at Time j, there must not be a disease acquisition via behavior. Therefore, we have $P(Y_{ij}=0|Y_{ij-1}=0)=\Delta_{ij}$, as illustrated by Figure 3(a).
2. Successfully treated infection. If the participant tested positive at Time $j-1$ but tested negative three months later at Time j, the participant must have received treatment, the treatment must have been successful, and she must not have acquired a new infection through sexual behavior. Therefore, we have conditional probability $P(Y_{ij}=0|Y_{ij-1}=1)=Z_{ij-1}(1-q)\Delta_{ij}$.
3. Newly acquired infection. If the participant tested negative for *C. trachomatis* at Time $j-1$ but tested positive at Time j, the participant must have acquired new infection. In this case the conditional probability is simply: $P(Y_{ij}=1|Y_{ij-1}=0)=1-\Delta_{ij}$.
4. Repeated infections. Finally, we consider the case of repeated infections. If the participant tested positive at two consecutive time points, there could be a number of possibilities: the first is the infection has not been treated; the second is the treatment has not been successful; the third is a new infection has been acquired. Therefore the conditional probability can be written as $P(Y_{ij}=1|Y_{ij-1}=1)=(1-Z_{ij-1})+Z_{ij-1}q+Z_{ij-1}(1-q)(1-\Delta_{ij})$.

Using the conditional probability formula, it is straightforward to derive the following unconditional probability

$$P(Y_{ij}=y_{ij})=P(Y_{ij-1}=0)\left[1-\Delta_{ij}\right]^{y_{ij}}\Delta_{ij}^{1-y_{ij}}+P(Y_{ij-1}=1)\left[(1-Z_{ij-1})+Z_{ij-1}\{q+(1-q)[1-\Delta_{ij}]\}\right]^{y_{ij}}\left[Z_{ij-1}(1-q)\Delta_{ij}\right]^{1-y_{ij}}, \quad (3.1)$$

where $\Delta_{ij}=(1-p_i)^{n_{ij}-m_{ij}}(1-p_i^c)^{m_{ij}}$ for $j=1,2,\ldots,J_i$.

Separately, Equation (3.1) implies

$$\begin{aligned} P(Y_{ij}=0) &= P(Y_{ij-1}=0)\Delta_{ij}+P(Y_{ij-1}=1)Z_{ij-1}(1-q)\Delta_{ij}, \\ P(Y_{ij}=1) &= 1-P(Y_{ij-1}=0)\Delta_{ij}+P(Y_{ij-1}=1)Z_{ij-1}(1-q)\Delta_{ij}. \end{aligned}$$

Through an iterative formula, Equation (3.1) provides a probability description of the infection process. Since Δ_{ij} represents the probability that the ith participant does not acquire infection via sexual intercourse between the $j-1$th and jth visits, the above probability statements are intuitively interpretable.

It should be noted that the above model could be further extended to accommodate situations of *C. trachomatis* spontaneous clearance. Cases of *C. trachomatis* self clearance have been documented (McCormack *et al.* 1979, Golden *et al.* 2000, Morre *et al.* 2002, Molano *et al.* 2005). Specifically, we could extend scenario 2 to include a possibility of self clearance ξ. Under this extension, Scenario 2 could be described by $P(Y_{ij}=0|Y_{ij-1}=1)=Z_{ij-1}(1-q)\Delta_{ij}+(1-\pi)\xi$, where ξ is the probability of spontaneous clearance. We do not consider this extension in the current research because our study protocol requires all identified infections be treated at the clinical visit (Tu *et al.* 2009). Additionally, symptomatic infections are also treated between visits (Batteiger *et al.* 2010). Such frequent testing and treatment have practically eliminated the possibility of observing spontaneous clearance from quarterly infection data.

4 Bayesian Estimation: Likelihood, Priors and Posterior

4.1 The Likelihood function

We write the density function of Y_{ij} as $f(y_{ij}|p_i, c, q) = P(Y_{ij} = 1)^{Y_{ij}} [1 - P(Y_{ij} = 1)]^{1-Y_{ij}}$, where $P(Y_{ij} = 1)$ is given by (3.1). Under this formulation, the parameters of interest p_i and c are linked to $P(Y_{ij} = 1)$ through $\Delta_{ij}=(1 - p_i)^{n_{ij}-m_{ij}}(1 - p_i^c)^{m_{ij}}$.

In this research, we assume that the infection acquisition probability associated with one sexual encounter is a subject-specific quantity. For the *i*th participant, this probability is p_i for coitus without condom protection. For convenience, we express p_i as a logistic function

$$p_i=\exp(\beta+U_i)/(1+\exp(\beta+U_i)) \tag{4.1}$$

where $U_i \sim G(\theta)$, and $G(\theta)$ denotes some parametric distribution with parameter θ.

As mentioned in Section 3.2, the model could be expanded to include event-level covariates $\mathbf{x}_{it}$ via $p_{it} = \exp(\beta\mathbf{x}_{it} + U_i)/(1 + \exp(\beta\mathbf{x}_{it} + U_i))$. The expansion makes the model particularly useful for the assessment of event-level factors that have the ability to modify the transmission risk. For example, in future topical microbicide studies, this model could be useful in assessing the timing, application method, and various characteristics of the microbicides. Along that vein, one could incorporate a random partner effect to accommodate the potential correlation structure of events within the same partner if partner changes are common. In our data set, a relatively small fraction of subjects (12%) reported multiple partners during the quarterly observation period. And, as previously discussed, *C. trachomatis* transmission risk was primarily influenced by the subject's immune response to the invading bacteria, which justifies the subject-specific risk assumption. But for other organisms, the modeling structure is certainly expandable to accommodate additional considerations.

Under the current formulation, while it is a common practice to assume normal distributions for random subject effects, Wang and Louis (2003) noted that under the usual normal assumption for U_i, the marginal regression model of binary response no longer had a logistic form when the random effects were integrated out. To retain the logistic structure and odds ratio interpretation, Lin *et al* (2009) recommend the use of a bridge distribution with parameter ϕ ($0 < \phi < 1$):

$$f_U(u_i/\phi)=\frac{1}{2\pi}\frac{\sin(\phi\pi)}{\cosh(\phi u_i)+\cos(\phi\pi)};\ -\infty<U_i<\infty, \tag{4.2}$$

where ϕ represents the attenuation factor of the marginal odds ratio.

The bridge density in (4.2) is symmetric around zero, and has $\sigma_U^2=\pi^2(\phi^{-2} - 1)/3$. It has a slightly heavier tail and is more peaked than the normal density.

Under this formulation, the distribution can then be written as

$$f(y_{ij}|\phi, c, \beta, q) \propto f(y_{ij}|\beta, c, q, u_i)f(u_i|\phi).$$

The likelihood then follows

$$\ell(\phi, \beta, c, q) = \prod_{i=1}^{I} \prod_{j=0}^{J_i} f(y_{ij} | \phi, c, \beta, q). \tag{4.3}$$

4.2 Prior Distributions

With the assumed bridge distribution for U_i, the marginal probability of STI acquisition (after integrating out the unobservable random effects U_i) can be written as $E_U[p_i(U_i)] = e^{\tau}/[1 + e^{\tau}]$, where $\tau = \phi\beta$ and E_U is the expected value with respect to the density of U_i. To estimate τ, we use the Markov Chain Monte Carlo (MCMC) method to approximate the values of ϕ and β iteratively. Within each iteration, the value of τ is generated from $\tau = \phi\beta$. And τ becomes the parameter of interest.

Following Lin *et al* (2009), we assume a Beta(1, 1) (uniform) prior for ϕ. For β, we specify a moderately diffused normal prior centered around zero, $\beta \sim$Normal(0, 4). We use a zero mean prior because we are uncertain of the expected value of β and $\beta = 0$ gives equal chances to infection and non-infection. A similar approach has been used in mixed effects model settings (Natarajan and McCulloch, 1998). The prior variance of 4 for the logistic regression parameter gives a relatively wide range for β, which translates into the full range of probability measure, (0, 1). The relationship between the prior variance and resulting odds ratio in logistic regression has been discussed by Spiegelhalter et al (2004). The p^c is the per encounter STI risk when a condom is used. For practical purposes, we restrict the support of c to positive values, i.e., $c > 0$. Here we assume that c follows an exponential distribution $\pi(c) \sim \exp(\lambda)$ where, $\lambda \sim$ Gamma(1, 5). Thus, the mean value of λ is 5 and variance is 25.

For treatment failure probability q, clinical trials have shown that azithromycin and doxycycline, the antibiotics that we used for *C. trachomatis* treatment in the study, are very effective (Lau and Qureshi 2002). But the effectiveness of these treatments in a real world clinical setting, often referred to as the use-effectiveness, is less well defined. Use-effectiveness of antibiotics could be affected by a number of factors, including the clinical venue (i.e., STD clinics versus primary care), patient population, route of administration, and most importantly, patient adherence. For the study population, our best guess is that the treatment failure probability q is likely to be less than 20%. Thus, we assume $\pi(q) \sim$ Beta(0.1, 0.9). This prior expects a 10% chance of a failure probability and it has great uncertainty as the effective sample size corresponding to this prior is (0.1+0.9) = 1. Among all beta priors with the same mean, those with larger effective sample sizes are considered more informative. See Thall and Wathen (2007) for a more detailed discussion on effective sample size.

4.3 Posterior Distribution and Inference

The joint posterior distribution of the parameters of the models conditional on the data is obtained by combining the likelihood in (4.3) and the prior densities using Bayes' theorem:

$$f(\beta, q, c, \phi) \propto \prod_{i=1}^{I} \prod_{j=0}^{J_i} f(y_{ij} | \theta, c, \beta, q) \times \pi(\beta) \times \pi(q) \times \pi(c) \times \pi(\phi). \tag{4.4}$$

The computations for the posterior distributions of the parameters were performed via Markov Chain Monte Carlo (MCMC) methods. The conditional posterior distributions are log-concave, and thus the MCMC can be implemented using standard algorithms. The Gibbs sampling and the associated Metropolis-Hastings algorithm were implemented using readily

available free software, WinBUGS (Spiegelhalter *et al*. 2005). We omit the explicit calculation of the conditional distribution of each parameter given the rest of the parameters because WinBUGS does not require the specification of the full conditional distribution. We use 25,000 iterations with an initial burn-in of 5,000. Convergence of the generated samples was assessed using standard tools within WinBUGS software, viz., trace plots, ACF plots as well as Gelman-Rubin convergence diagnostics. The initial values for the parameters were selected by starting with prior means and covering ±3 standard deviations. We used three different MCMC chains with three different initial values to get reasonable convergence.

5 Data Analysis and Results

We analyze 1173 quarterly test results from 200 young women who have completed at least two follow-up visits. The study subjects were recruited from an inner city population that was at increased STI risk. Each subject contributed an average 5.86 visits, ranging from 2 to 18. The mean enrollment age is 15 (standard deviation 1.1). The study sample is dominantly African American, 91%. At the time of enrollment, the subjects reported an average of 2.3 sexual partners during their lifetime (median 2). The quarterly average number of condom protected and unprotected sexual encounters were 5.8 and 14.3, respectively. *C. trachomatis* was detected from 208 of the quarterly swab samples (17.7%). Of all quarterly intervals, approximately 66% were of scenario 1 (negative-negative), 16% were of scenario 2 (positive-negative), 15% were of scenario 3 (negative-positive), and 3% were of scenario 4 (positive-positive). All infected subjects received antibiotic prescription upon diagnosis.

From the estimated value of τ, we calculate the per-encounter probability of acquiring *C. trachomatis* to be $p=\frac{\exp(-4.699)}{\exp(-4.699)+1}=0.009$, which implies that the probability of a young woman in our study population acquiring *C. trachomatis* infection by having one unprotected intercourse with a male partner is less than 1%. It should be noted that the per-encounter probability of acquiring *C. trachomatis*, p, is a population-specific estimate because it reflects not only the organism's transmissibility, but also the infection risk that the corresponding male partner population presents toward the young women. Therefore, the low infection acquisition risk estimate ($p = 0.009$) should not be surprising: If only a small number of male partners have chlamydia, the probability that a female subject getting infected through one sex encounter with a male partner will also be small!

In fact, using the estimated *C. trachomatis* prevalence in the corresponding male partner population and the relationship Per-encounter Acquisition Risk=Transmission Probability × Partner Prevalence, we will be able to calculate the *transmission probability* associated with one unprotected encounter. By transmission probability, we are referring to the probability of a young woman becoming infected by having one unprotected sexual encounter with an infected male partner. For example, if the *C. trachomatis* prevalence rate is 5% in the male partner population, then the transmission probability will be 18.0%. Figure 3 illustrates the relationship between transmission probability and *C. trachomatis* prevalence in the partner population for the point estimate of the per encounter probability and the 95% credible limits.

In our study, the exact prevalence of *C. trachomatis* infection in the male partner population is unavailable. A recent study of chlamydia prevalence based on the test results from over 23,000 men in Baltimore, Denver, San Francisco, and Seattle, found that the overall prevalence of *C. trachomatis* infection was 7%, but the rate tended to vary among cities (Schillinger *et al*. 2005). Using this overall prevalence estimate, we calculate the transmission probability as 0.129 with a 95% credible interval (0.096, 0.206).

It should also be noted that the condom effect parameter estimate is 1.6 with the estimated 95% credible interval (1.12, 2.42) not containing 1. This suggests a significant condom modification of the transmission probability. In other words, the per-encounter risk of disease acquisition with condom protection is $p^c = 0.009^{1.6} = 0.00053$, drastically lower than the per-encounter risk without condom use. Again assuming 7% prevalence in the male partner population, we estimate the transmission probability associated with one condom protected encounter to be 0.0076, which is drastically lower then the transmission risk without condom protection. While this finding is consistent with the observations about condom efficacy in protection against STI, our estimate directly quantifies the condom's role in modifying the *C. trachomatis* transmission probability.

The estimated treatment failure probability $q = 9.6\%$ is slightly higher than the known treatment failure rates from clinical trials (Lau and Qureshi, 2002). This perhaps is not surprising because this estimate is not obtained from controlled experiments but represents the treatment failure rate in the the real world. Many practical factors, such as suboptimal medication adherence, could contribute to the increased failure of antibiotic treatment in aborting an infection episode.

6 Discussion

In this research, we presented a method for the assessment of the per-encounter male-to-female *C. trachomatis* transmission probability. The primary contribution of the method is that it provides a modeling tool for epidemiologists and clinical investigators to assess the transmissibility of a disease pathogen using observational data without resorting to controlled experiments. By relying on observational data, the proposed estimation scheme has helped to overcome the ethical constraints against exposing human subjects to infectious pathogens, significantly enhancing our ability to assess the transmission efficiency of bacterial organisms.

Applying the proposed method to observational data collected from a cohort of urban adolescent women, and using the published prevalence of *C. trachomatis* infection among men from urban areas, we have obtained an estimate of the *C. trachomatis* transmission probability. To the best of our knowledge, this is the first time that the male-to-female per-exposure *C. trachomatis* transmission probability has been successfully quantified. Viewed within the context of adolescent sexual behavior, the estimated 0.129 per-exposure transmission probability represents a substantial infection risk. At the same time, the research also demonstrates a drastic reduction of infection risk, from 0.129 to 0.0076, when a condom is used. These findings point to two potentially important areas of intervention: prophylactic measures aimed at lowering the *C. trachomatis* transmission probability, and more aggressive partner notification and treatment programs aimed at reducing *C. trachomatis* prevalence among the partners.

The method that we presented in the current paper is likely to have a significant impact on future studies of STI transmission. It could also influence the development of future intervention strategies. Although the current data example examines the per-exposure risk of male-to-female transmission of *C. trachomatis*, the same method could be used to study female-to-male transmissions if similar observational data exist for male subjects. The directionality of STI transmission between heterosexual partners and the magnitude of transmission risk associated with each direction are of particular interest because of their implications on the design of intervention strategies. Also with this method, studies comparing the transmission efficiencies of different STI pathogens become feasible. These future comparative studies will help to explain, at least to some extent, the observed

differences in the scope of different STI epidemics. Finally, the method also provides an important starting point for simulation studies of STI transmission dynamics.

Methodologically, this research represents a new development in the modeling of infectious disease outcomes. The stochastic based modeling approach and Bayesian computation techniques have made the proposed method easily implementable in a variety of computing platforms. As shown in section 4.1, the method can be extended to incorporate important behavioral and clinical factors as covariates that potentially influence the transmission process. One possible way to accommodate the covariates **X** is to introduce them through the logistic structure $p=\frac{\exp(\mathbf{x}\beta)}{\exp(\mathbf{x}\beta)+1}$. Although the focus of the current research is to quantify the per-encounter transmission probability, extensions to incorporate covariates could further enrich the analytical tool set available to infectious disease investigators and make it possible to assess factors directly related to the risk of transmission, instead of the static outcome of infection. Such an extension may be more useful for the assessment of exposure factors that have the potential to modify the transmission efficiency of the organism. For example, one could use the method to evaluate the effect of topical microbicides, which are usually applied around the time of exposure, on transmission risk. Additionally, as briefly stated in Section 3, event-specific information about the sexual partner could also be examined if partner information and characteristics are available.

It should be noted, however, the proposed method has a few implicit assumptions that could potentially limit its use in practice. First, we assume knowledge of *C. trachomatis* prevalence in the male partner population. This assumption is essential for the conversion of infection acquisition risk into the transmission probability. Second, we assume that the probability of having an infected partner remains fixed for all sexual encounters. As a result, the method provides no explicit accommodation of an individual partner's contribution to the infection risk. In other words, we will not be able to differentiate the infection risk associated with coital events with different partners within the subject, i.e., the potential clustering effect. This assumption allows us to circumvent the difficulty of not having partner data. For the current application, such an approach is perhaps not unreasonable because after an exposure to the disease pathogen, whether a woman develops a clinical infection probably has more to do with her own immune response than with the source of pathogen. This said, we are exploring an extended Bayesian modeling structure that assumes a prior distribution on partner prevalence in a new study where limited partner information is available. In the current research, we are unable to pursue this extension due to the lack of partner data. Third, the validity of our estimates depends on the accuracy of STI testing as well as the behavioral reporting. While we have put in place rigorous laboratory procedures to ensure the accuracy of the testing data, we could not rule out the possibility of misreporting on the part of the study subjects. The accuracy of the behavioral data could certainly affect our estimates. For example, systematic under-reporting of coital events by the study subjects could result in an overestimate of the STI acquisition risk. To alleviate the concerns about reporting bias, we have performed a detailed data examination by comparing self reported coital counts ascertained at quarterly visits with those obtained from subjects' daily diaries in the same period. We note that over 90% of the reported counts from the two sources are in exact agreement. Although this still does not prove the accuracy of behavioral reports, it does give added assurances about the consistency of these reports. Finally, in our research, the random subject effect is assumed to have a parametric bridge distribution. Since it is hard to validate the distributional assumption, a broader class of distributions such as Dirichlet processes, might represent a viable alternative to protect the model from misspecification.

Not withstanding these limitations, we present a flexible and expandable platform for investigating various aspects of the transmission of infectious pathogens.

References

1. Batteiger BE, Tu W, Ofner S, Van Der Pol B, Stothard DR, Orr DP, Katz BP, Fortenberry JD. Repeated Chlamydia trachomatis Genital Infection in Adolescent Women. J Infect Dis. 2010; 201(1):42–51. [PubMed: 19929379]
2. Baumgarten RK. Re: "Cohort study of venereal disease. I: The risk of gonorrhea transmission from infected women to men". Am J Epidemiol. 1980 Jan; 111(1):130–131. [PubMed: 7352454]
3. Cates W Jr, Wasserheit JN. Genital chlamydial infections: epidemiology and reproductive sequelae. Am J Obstet Gynecol. 1991 Jun; 164(6 Pt 2):1771–1781. [PubMed: 2039031]
4. Centers for Disease Control and Prevention (CDC). Atlanta, GA: US Dept of Health and Human Services, CDC; 2007. Trends in Reportable Sexually Transmitted Diseases in the United States, 2006.
5. Centers for Disease Control and Prevention (CDC). CDC Fact Sheet: Chlamydia. Atlanta, GA: US Dept of Health and Human Services, CDC; 2007 December.
6. Cohen MS, Cannon JG. Human experimentation with Neisseria gonorrhoeae: progress and goals. J Infect Dis. 1999 Mar.179 Suppl 2:S375–S379. [PubMed: 10081510]
7. Datta SD, Sternberg M, Johnson RE, et al. Gonorrhea and chlamydia in the United States among persons 14 to 39 years of age, 1999 to 2002. Ann Intern Med. 2007 Jul 17; 147(2):89–96. [PubMed: 17638719]
8. Division of STD prevention, Centers for Disease Control and Prevention. [Accessed on September 10, 2007] Trends of Reportable Sexually Transmitted Diseases in the United States 2005: National Surveilance Data for Chlamydia, Gonorrhea, and Syphilis. http://www.cdc.gov/std/stats/05pdf/trends-2005.pdf
9. European Study Group on Heterosexual Transmission of HIV. Comparison of female to male and male to female transmission of HIV in 563 stable couples; British Medical Journal. 1992 Mar 28. p. 809-813.
10. Felton WF. Contrasting views on the infectivity of gonorrhoea. Br J Vener Dis. 1973 Apr; 49(2): 151–154. [PubMed: 4704867]
11. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect. 1999 Feb; 75(1):3–17. [PubMed: 10448335]
12. Fung M, Scott KC, Kent CK, Klausner JD. Chlamydial and gonococcal reinfection among men: a systematic review of data to evaluate the need for retesting. Sex Transm Infect. 2007 Jul; 83(4): 304–309. [PubMed: 17166889]
13. Golden MR, Schillinger JA, Markowitz L, St Louis ME. Duration of untreated genital infections with chlamydia trachomatis: a review of the literature. Sex Transm Dis. 2000 Jul; 27(6):329–337. [PubMed: 10907908]
14. Hillis SD, Owens LM, Marchbanks PA, Amsterdam LF, Mac Kenzie WR. Recurrent chlamydial infections increase the risks of hospitalization for ectopic pregnancy and pelvic inflammatory disease. Am J Obstet Gynecol. 1997 Jan; 176(1 Pt 1):103–107. [PubMed: 9024098]
15. Holmes KK, Johnson DW, Trostle HJ. An estimate of the risk of men acquiring gonorrhea by sexual contact with infected females. Am J Epidemiol. 1970 Feb; 91(2):170–174. [PubMed: 5416250]
16. Hooper RR, Reynolds GH, Jones OG, et al. Cohort study of venereal disease. I: the risk of gonorrhea transmission from infected women to men. Am J Epidemiol. 1978 Aug; 108(2):136–144. [PubMed: 707474]
17. Katz BP. Estimating transmission probabilities for chlamydial infection. Stat Med. 1992 Mar; 11(5):565–577. [PubMed: 1594800]
18. Lamontagne D, Baster K, Emmett L, et al. Incidence and reinfection rates of genital chlamydial infection among women aged 16–24 years attending general practice, family planning and

genitourinary medicine clinics in England: a prospective cohort study by the Chlamydia Recall Study Advisory Group. Sex Transm Infect. 2007; 83:292–303. [PubMed: 17050567]

19. Lau C-Y, Qureshi AK. Azithromycin versus doxycycline for genital chlamydial infections. Sex Transm Dis. 2002; 29:497–502. [PubMed: 12218839]
20. Lin L, Bandyopadhyay D, Lipsitz SE, Sinha D. Association Models for Clustered Data with Binary and Continuous Responses. Biometrics. 2009 (online).
21. McCormack WM, Alpert S, McComb DE, Nichols RL, Semine DZ, Zinner SH. Fifteen-month follow-up study of women infected with Chlamydia trachomatis. N Engl J Med. 1979 Jan 18; 300(3):123–125. [PubMed: 758599]
22. Miller WC, Ford CA, Morris M, et al. Prevalence of chlamydial and gonococcal infections among young adults in the United States. JAMA. 2004 May 12; 291(18):2229–2236. [PubMed: 15138245]
23. Molano M, Meijer CJ, Weiderpass E, et al. The natural course of Chlamydia trachomatis infection in asymptomatic Colombian women: a 5-year follow-up study. J Infect Dis. 2005 Mar 15; 191(6): 907–916. [PubMed: 15717266]
24. Morre SA, van den Brule AJ, Rozendaal L, et al. The natural course of asymptomatic Chlamydia trachomatis infections: 45% clearance and no development of clinical PID after one-year follow-up. Int J STD AIDS. 2002 Dec.13 Suppl 2:12–18. [PubMed: 12537719]
25. Natarajan R, McCulloch CE. Gibbs Sampling With Diffuse Proper Priors: A Valid Approach to Data-Driven Inference? Journal of Computational and Graphical Statistics. 1998; 7:267277.
26. Padian NS, Shiboski SC, Jewell NP. The effect of number of exposures on the risk of heterosexual HIV transmission. J Infect Dis. 1990 May; 161(5):883–887. [PubMed: 2324540]
27. Padian NS, Shiboski SC, Jewell NP. Female-to-male transmission of human immunodeficiency virus. JAMA. 1991 Sep 25; 266(12):1664–1667. [PubMed: 1886189]
28. Padian NS, van der Straten A, Ramjee G, et al. Diaphragm and lubricant gel for prevention of HIV acquisition in southern African women: a randomised controlled trial. Lancet. 2007 Jul 21; 370(9583):251–261. [PubMed: 17631387]
29. Schillinger JA, Dunne EF, Chapin JB, et al. Prevalence of Chlamydia trachomatis infection among men screened in 4 U.S. cities. Sex Transm Dis. 2005 Feb; 32(2):74–77. [PubMed: 15668611]
30. Spiegelhalter, D.; Abrams, K.; Myles, JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation (Statistics in Practice). Wiley; 2004.
31. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual, Version 1.4.2, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology & Public Health, Imperial College School of Medicine. 2005 available at http://www.mrc-bsu.cam.ac.uk/bugs.
32. Spinola SM, Bauer ME, Munson RS. Immunopathogenesis of Haemophilus ducreyi Infection (Chancroid). Infection and Immunity. 2002; 70(4):1667–1676. [PubMed: 11895928]
33. Thall PF, Wathen JK. Practical Bayesian adaptive randomisation in clinical trials. Eur J Cancer. 2007; 43:859–866. [PubMed: 17306975]
34. Tu W, Batteiger BE, Wiehe S, et al. Time from first intercourse to first sexually transmitted infection diagnosis among adolescent women. Arch Pediatr Adolesc Med. 2009; 163(12):1106–1111. [PubMed: 19996047]
35. Wang Z, Louis TA. Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. Biometrika. 2003; 90:765–775.
36. Weinstock H, Berman S, Cates W Jr. Sexually transmitted diseases among American youth: incidence and prevalence estimates, 2000. Perspect Sex Reprod Health. 2004 Jan-Feb; 36(1):6–10. [PubMed: 14982671]
37. Westrom, LE. D. Pelvic inflammatory disease. In: Holmes, KMPA., editor. Sexually Transmitted Diseases. 3rd ed.. New York, NY: McGraw-Hill; 1999. p. 783-809.
38. Xu F, Schillinger JA, Markowitz LE, Sternberg MR, Aubin MR, St Louis ME. Repeat Chlamydia trachomatis infection in women: analysis through a surveillance case registry in Washington State, 1993–1998. Am J Epidemiol. 2000 Dec 15; 152(12):1164–1170. [PubMed: 11130622]

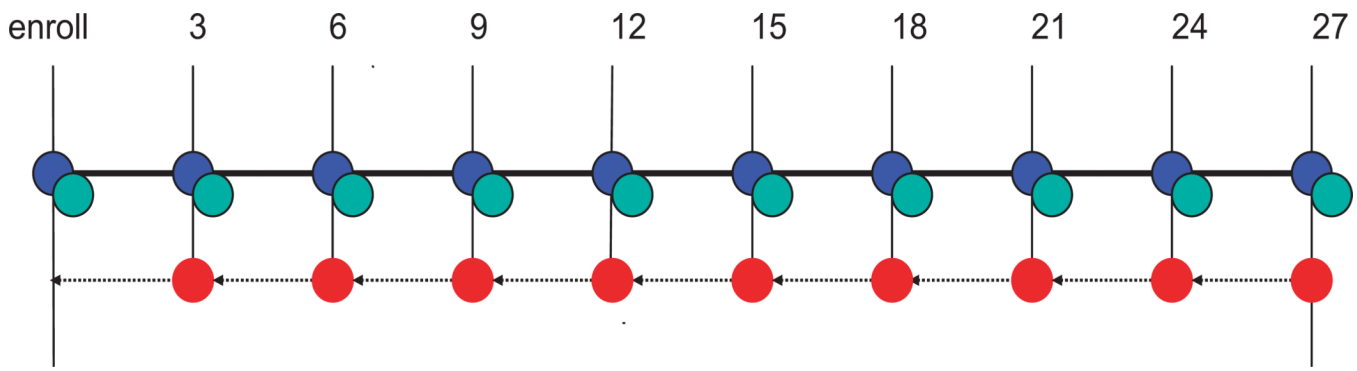

Figure 1.
Data collection scheme of the Young Women's Project: Sexually transmitted infection (STI) states were determined at enrollment and all subsequent quarter visits (blue circles represents determination of infection status). Infected subjects were treated at these visits (green circles represents antibiotic treatment). Also at each quarterly visit, participants participated in face-to-face interviews, providing numbers of coital events, with and without condom protection, in the previous quarter (represented by red circles).

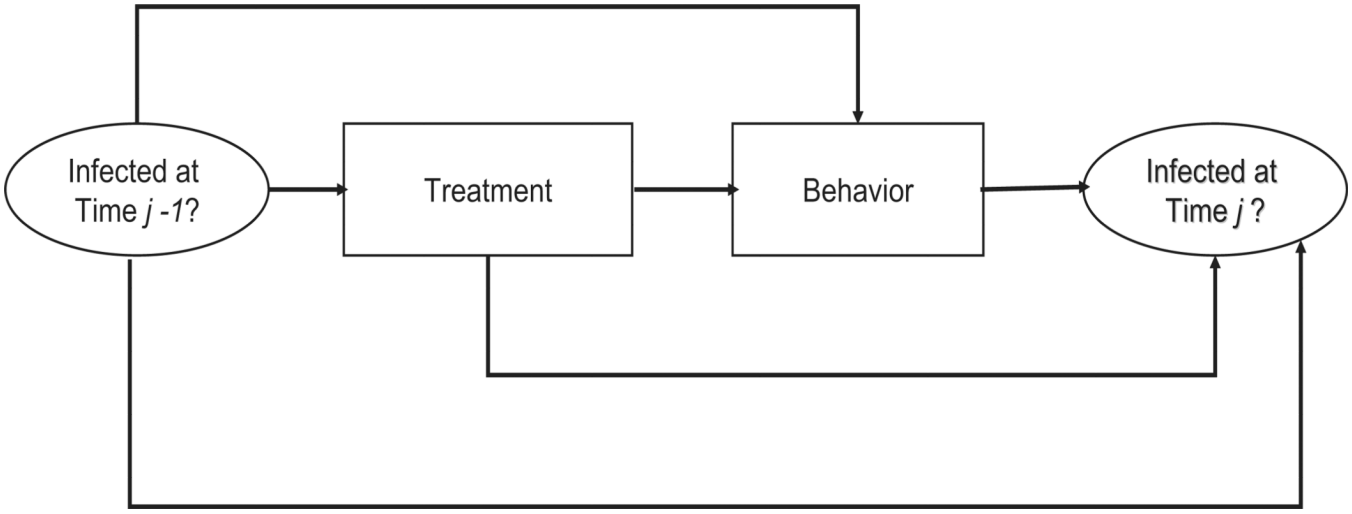

Figure 2.

A conceptual model of recurrent sexually transmitted infections (STI):We consider the infection history of a participant as a stochastic process where the current infection state depends on the infection state at the previous visit, the treatment received at the previous visit, and STI related sexual behaviors between the two visits.

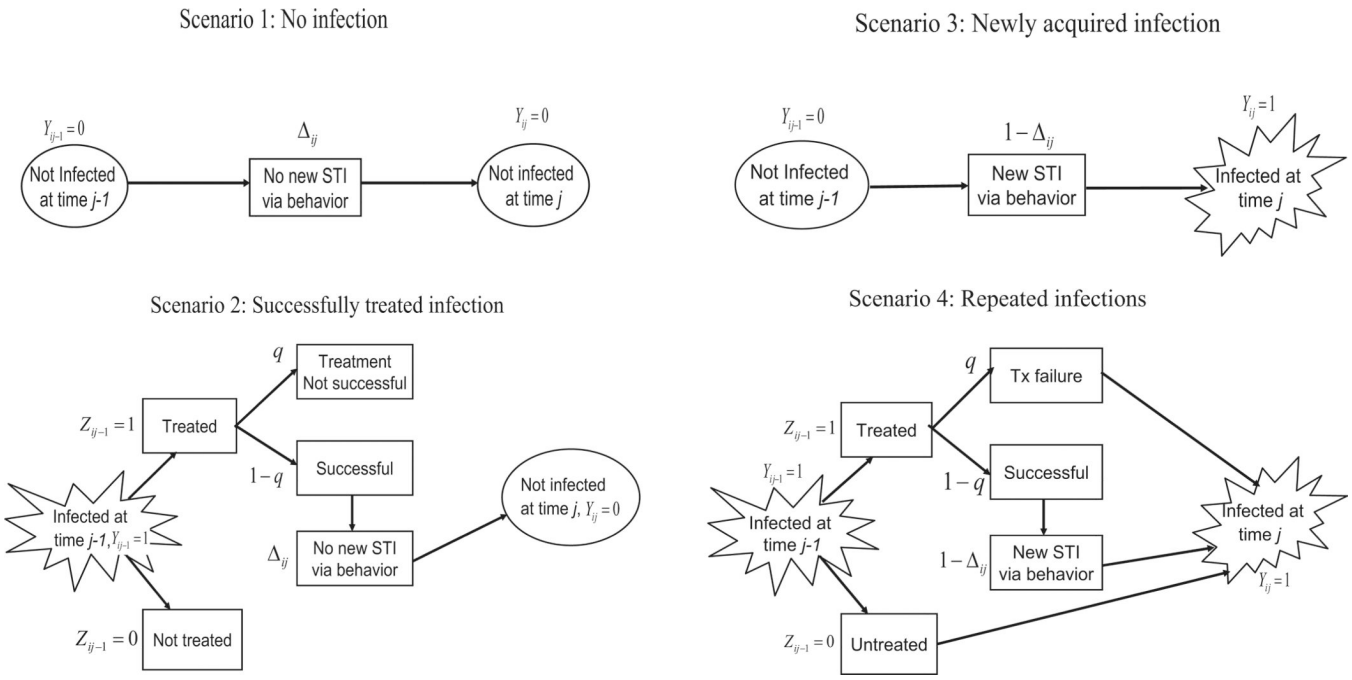

Figure 3.

Four scenarios of infection history between two consecutive visits: (1) no infection detected; (2) successfully treated infection; (3) newly acquired infection; and (4) recurrent infections.

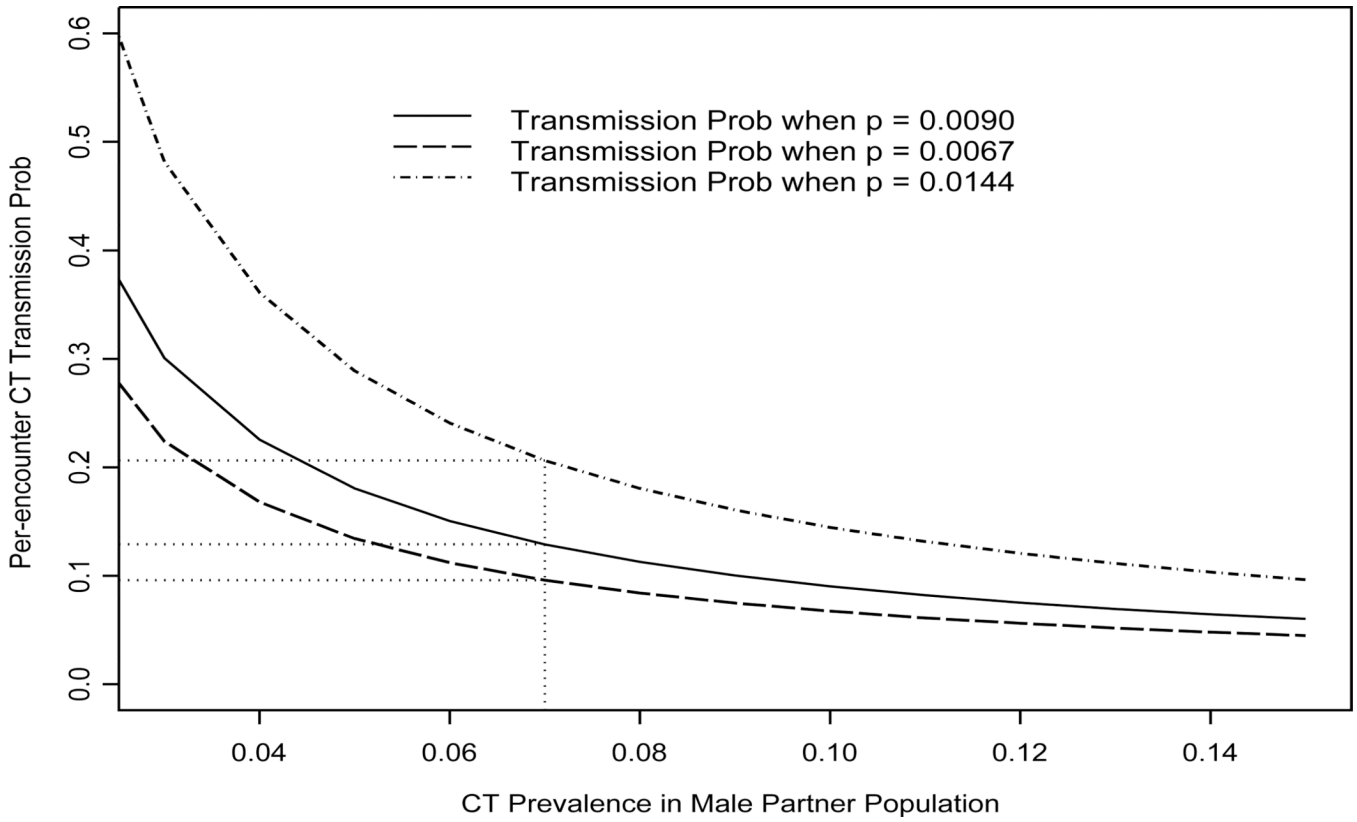

Figure 4.

Estimated *C. trachomatis* transmission probability at different levels of partner prevalence, based on the point estimate ($\hat{p}$ =0.90%), and 95% lower and upper credible limits of the per-encounter STI acquisition risk (0.67% and 1.44%). Assuming 7% *C. trachomatis* prevalence among male partners, we calculate the male-to-female transmission probability to be 12.86% with a 95% credible interval of (9.57%, 20.57%)

Table 1

Parameter Estimates

(a) Markov Model Parameters

Parameter	Posterior Mean	Posterior Median	95% C.I.
τ	−4.699	−4.733	(−4.997, −4.223)
q	0.0961	0.0981	(0.081, 0.3221)
c	1.6	1.65	(1.12, 2.423)
p	0.0091	0.0088	(0.0068, 0.0147)

(b) Transmission Probability: Without Condom

Prevalence rates among partners	Estimated value of transmission prob	95% C.I.
0.03	0.300	(0.223, 0.048)
0.04	0.225	(0.168, 0.036)
0.05	0.180	(0.134, 0.288)
0.06	0.150	(0.112, 0.240)
0.07	0.129	(0.096, 0.206)
0.08	0.113	(0.084, 0.180)
0.09	0.100	(0.074, 0.160)

(c) Transmission Probability: With Condom

Prevalence rates among partners	Estimated value of transmission prob	95% C.I.
0.03	0.018	(0.0111, 0.0377)
0.04	0.013	(0.0083, 0.0283)
0.05	0.010	(0.0066, 0.0226)
0.06	0.009	(0.0055, 0.0188)
0.07	0.008	(0.0047, 0.0161)
0.08	0.007	(0.0042, 0.0141)
0.09	0.006	(0.0037, 0.0126)