



Published in final edited form as:

Sex Transm Infect. 2010 June ; 86(3): 169–174. doi:10.1136/sti.2009.037341.

A Bayesian approach to uncertainty analysis of sexually transmitted infection models

Leigh F. Johnson¹, Leontine Alkema², and Rob E. Dorrington¹

¹Centre for Actuarial Research, University of Cape Town, South Africa

²Department of Statistics and Applied Probability, National University of Singapore

Abstract

Objectives—To propose a Bayesian approach to uncertainty analysis of sexually transmitted infection (STI) models, that can be used to quantify uncertainty in model assessments of policy options, estimate regional STI prevalence from sentinel surveillance data and make inferences about STI transmission and natural history parameters.

Methods—Prior distributions are specified to represent uncertainty regarding STI parameters. A likelihood function is defined using a hierarchical approach that takes account of variation between study populations, variation in diagnostic accuracy as well as random binomial variation. The method is illustrated using a model of syphilis, gonorrhoea, chlamydial infection and trichomoniasis in South Africa.

Results—Model estimates of STI prevalence are in good agreement with observations. Out-of-sample projections and cross-validations also show that the model is reasonably well calibrated. Model predictions of the impact of interventions are subject to significant uncertainty: the predicted reductions in the prevalence of syphilis by 2020, as a result of doubling the rate of health seeking, increasing the proportion of private practitioners using syndromic management protocols, and screening all pregnant women for syphilis, are 43% (95% CI: 3–77%), 9% (95% CI: 1–19%) and 6% (95% CI: 4–7%) respectively.

Conclusions—This study extends uncertainty analysis techniques for fitted HIV/AIDS models to models that are fitted to other STI prevalence data. There is significant uncertainty regarding the relative effectiveness of different STI control strategies. The proposed technique is reasonable for estimating uncertainty in past STI prevalence levels and for projections of future STI prevalence.

Keywords

sexually transmitted infection; mathematical model; uncertainty analysis

Correspondence to: Leigh Johnson, Centre for Actuarial Research, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa, Tel: +27 21 650 5761, Fax: +27 21 650 5937, Leigh.Johnson@uct.ac.za.

CONTRIBUTORS

LJ and LA developed the statistical method. LJ created the mathematical model and collated the STI prevalence data. The manuscript was written by LJ, with contributions from LA and RD.

COMPETING INTERESTS: none

COPYRIGHT

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in STI and any other BMJPG products to exploit all subsidiary rights, as set out in their licence.

INTRODUCTION

Mathematical models of sexually transmitted infections (STIs) serve a number of purposes.¹ However, the results of mathematical models can only be accepted as an adequate approximation to reality if their key assumptions are supported by empirical data. This requirement is difficult to meet when modelling STIs, as there is much uncertainty regarding the natural history and efficiency of transmission for most STIs, and parameters such as the average duration of untreated infection and the probability of transmission per contact are very difficult to estimate reliably.^{2, 3}

Several modellers have attempted to demonstrate the veracity of their models by showing that the model outputs are consistent with observed levels of STI prevalence in the population being modelled.⁴⁻⁷ However, this only partially resolves the problem of parameter uncertainty, as there will usually be many different combinations of parameters that give a similar degree of correspondence to observed prevalence, and these different parameter combinations – although they might be considered equally plausible *a priori* – do not necessarily produce the same conclusions for users of the model outputs. In addition, it is possible that the true prevalence of the STI in the population may differ from that observed due to bias in the sample or imperfect sensitivity and specificity of the test used to detect STIs, or random sampling error. Demonstration of similarity between model estimates and observations therefore does not guarantee that all model outputs are realistic, nor does dissimilarity between model estimates and observations necessarily imply that model outputs are unrealistic.

A more formal approach to dealing with parameter uncertainty is to use uncertainty analysis techniques such as Latin hypercube sampling or Monte Carlo simulation.⁸⁻¹⁰ This involves specifying probability densities to represent uncertainty regarding key model parameters, randomly sampling from these densities and running the model for each sampled parameter combination in order to estimate the range of uncertainty around the model outputs. When applying such techniques to models of STIs, one might wish to consider uncertainty subject to certain output constraints being met – typically, the constraint is that model estimates of STI prevalence must be roughly consistent with observed STI prevalence. These constraints are usually specified in terms of likelihood functions or ‘sum of squares’ criteria, and several approaches have been proposed in the case of HIV modelling.¹¹⁻¹⁵ However, relatively little work has been done in assessing uncertainty in models of STIs other than HIV. Although it would be possible to apply the techniques developed for HIV to other STIs, STI prevalence data are typically more limited than HIV prevalence data, particularly in developing countries, and this makes it necessary to combine STI data from multiple sources.^{16, 17} In addition, the sensitivity and specificity of the tests used for other STIs are typically much lower than those used for HIV,^{18, 19} and variability in diagnostic accuracy is therefore an important factor to consider when defining the likelihood function.

This paper proposes a Bayesian approach to uncertainty analysis of models of STIs other than HIV. The proposed technique is applied to a model of syphilis, gonorrhoea, chlamydial infection and trichomoniasis in South Africa. It is shown that the method can be used to assess uncertainty with respect to STI model predictions, to estimate STI prevalence at a national level from limited sentinel surveillance data, and to make inferences about STI natural history and transmission parameters.

METHOD

The Bayesian approach described below follows three steps: (1) specification of prior distributions to represent uncertainty regarding input parameters, (2) specification of a likelihood function to represent ‘goodness of fit’ to observed levels of STI prevalence, for a

given parameter combination, and (3) calculation of the posterior distribution. The technique is applied to a deterministic model of STIs in South Africa, which produces estimates of STI prevalence in the South African national population, as well as in various subpopulations (pregnant women, women attending family planning clinics, commercial sex workers, and men and women in households). The model assumptions about sexual behaviour and HIV have been fixed at the mean values estimated in a previous analysis,^{20, 21} and the STI prevalence data used to define the likelihood function are presented in a review of South African STI prevalence data.²² A summary description of the model and the data is provided in the online appendix.

Prior distributions on input parameters

Prior distributions are specified to represent uncertainty around the key model input parameters. These key parameters, represented by parameter vector ϕ , include the probability of STI transmission in a single act of unprotected sex with an infected partner, the proportion of cases that become symptomatic, the average length of time spent in different infection states (if the STI is not treated), the average duration of immunity following recovery from infection, and the proportion of STI cases that were correctly treated prior to the adoption of syndromic management protocols. Beta distributions are specified for probabilities and proportions, while gamma distributions are specified for average durations, since beta and gamma distributions relate to variables defined on the ranges $[0, 1]$ and $[0, \infty)$ respectively. The means and 95% confidence intervals of these prior distributions are shown in Table 1 (for gonorrhoea, chlamydial infection and trichomoniasis) and Table 2 (for syphilis). The sources on which these prior distributions are based are referenced in the online appendix.

Likelihood for observed STI prevalence

To define a likelihood function, it is necessary to allow for differences between observations in terms of sample type (antenatal clinic, family planning clinic, sex worker or household survey), diagnostic type and time. It is also necessary to allow for variation in prevalence between communities, since all observations are specific to particular communities and none of the data are nationally representative. Suppose that Y_i is the number of individuals testing positive in the i^{th} study. It is assumed that Y_i is binomially distributed with parameters n_i and θ_i , where n_i is the number of individuals tested in the i^{th} study, and θ_i is the prevalence that one would expect to observe in the i^{th} study. To allow for additional variation between observations due to variation in prevalence between communities and variation in diagnostic accuracy, it is assumed that $\theta_i \sim \text{beta}(\alpha_i, \beta_i)$. The number of individuals testing positive in the i^{th} study therefore follows a beta-binomial distribution, i.e.

$$Y_i \mid \alpha_i, \beta_i \sim \text{Beta-bin}(n_i, \alpha_i, \beta_i) \quad (1)$$

Parameters α_i and β_i are estimated using the method of moments, by first estimating the mean and variance of θ_i . Suppose that $M(s_i, t_i, \phi)$ represents the model estimate of the prevalence of the STI in sample type s_i (pregnant women, women attending family planning clinics, commercial sex workers, men in households or women in households), in year t_i , given a set of input parameters which is denoted by the parameter vector ϕ . If ρ_i is the true prevalence of the STI in the i^{th} study population, then a plausible model for the distribution of ρ_i might be

$$\ln\left(\frac{\rho_i}{1 - \rho_i}\right) = \ln\left(\frac{M(s_i, t_i, \phi)}{1 - M(s_i, t_i, \phi)}\right) + b_i, \quad (2)$$

where it is assumed that $b_i \sim N(0, \sigma_b^2)$. The b_i term represents the difference (on the logit scale) between the prevalence in the i^{th} study population and the prevalence that would be expected in a nationally representative sample. The logit transformation is introduced to normalize the variance of the b_i terms. Since the standard deviation of the study effects, σ_b , is unknown, a prior distribution is placed on this parameter. A gamma prior is used for σ_b , with its mean and standard deviation based on the standard deviations of the logit-transformed STI prevalence levels prior to fitting the model to the data. The same mean (0.30) and standard deviation (0.15) are used for all STIs (see Tables 1 and 2).

Equation (2) can be re-expressed as follows:

$$\rho_i = \left(1 + \left(\frac{1 - M(s_i, t_i, \boldsymbol{\varphi})}{M(s_i, t_i, \boldsymbol{\varphi})} \right) \exp(-b_i) \right)^{-1} \quad (3)$$

θ_i is then calculated, taking into account the sensitivity and specificity of the test used in the i^{th} study, Se_i and Sp_i respectively:²³

$$\theta_i = (Se_i + Sp_i - 1)\rho_i + 1 - Sp_i \quad (4)$$

The mean and variance of θ_i are then obtained by noting that

$$\begin{aligned} E(\theta_i) &= E(E(\theta_i | b_i)) \\ &= E(\rho_i) (E(Se_i) + E(Sp_i) - 1) + 1 - E(Sp_i) \end{aligned} \quad (5)$$

and

$$\begin{aligned} \text{Var}(\theta_i) &= E(\text{Var}(\theta_i | b_i)) + \text{Var}(E(\theta_i | b_i)) \\ &= \text{Var}(Se_i) \left(\text{Var}(\rho_i) + E(\rho_i)^2 \right) + \text{Var}(Sp_i) \left(\text{Var}(\rho_i) + (1 - E(\rho_i))^2 \right) + (E(Se_i) + E(Sp_i) - 1)^2 \text{Var}(\rho_i) \end{aligned} \quad (6)$$

The means and variances of the Se_i and Sp_i terms are estimated based on reviews of the sensitivity and specificity of various diagnostic tests,^{18, 24–26} and are shown in the online appendix. For a given set of input parameters $\boldsymbol{\varphi}$, a third-order Taylor approximation to ρ_i (about the point $b_i = 0$) is used to approximate the mean and variance of ρ_i . Having obtained the mean and variance of θ_i using equations (5) and (6), the α_i and β_i parameters are calculated using the method of moments, and the beta-binomial likelihood for the i^{th} observation is then computed.

Posterior calculation

By Bayes' theorem, the posterior density is proportional to the product of the joint prior density and the likelihood function. Since the distribution of this product cannot be evaluated analytically, the posterior distribution is evaluated numerically using Sampling Importance Resampling (SIR).²⁷ This was implemented by randomly sampling 20,000 parameter combinations from the joint prior, running the model for each parameter combination and calculating the associated likelihood, and then drawing a resample of 500 parameter combinations from the initial set of 20,000, using the likelihood values as sample weights. This sample of 500 parameter combinations is a sample from the posterior distribution and is used to generate the results presented.

Validation

To assess the validity of the model predictions and the model estimates, the model is (a) refitted after having omitted the most recent observations, and (b) refitted after having randomly omitted half of the observations. The resulting estimates of STI prevalence in 2005 are compared with those obtained in the initial analysis. In addition, the omitted observations are compared with the distributions of estimated prevalence levels, expressing the former as percentiles of the latter and plotting a histogram of the percentiles, similar to the verification rank histogram.²⁸ If the model has good validity, the histogram of percentiles should be approximately uniform. The model fitting procedure is also validated by comparing the model estimates of syphilis prevalence with nationally representative survey results,²⁹ which have been excluded from the likelihood calculation so that they can be used for independent validation.

RESULTS

The model estimates of trends in STI prevalence in the 15–49 age group are shown in Figure 1, together with observed levels of STI prevalence in ‘low risk’ groups (antenatal attenders, family planning clinic attenders and men and women in households). Model estimates are generally consistent with observations. However, the 95% confidence intervals around the model estimates are narrower than the ranges of variation in observed prevalence levels, since the former reflect only the uncertainty around the population mean, while the latter reflect the variability of the study effects, the variation in diagnostic accuracy, differences in sample type, as well as random binomial variation. Model estimates of STI prevalence in women tend to be similar when comparing pregnant women, family planning clinic attenders and women in households, but are substantially higher in commercial sex workers (see Table 12 in the online appendix).

The observations used in defining the likelihood (Figure 1) do not suggest any clear trend in STI prevalence over time, but the model fits a downward trend as a result of the model assumption that condom usage and use of syndromic management protocols have both increased since the mid-1990s. In the case of syphilis, it is possible to validate this modelled decline in prevalence by comparison with levels of syphilis prevalence measured in the nationally representative antenatal surveys.²⁹ Figure 1(a) shows that the modelled decline in syphilis prevalence in women is consistent with that observed in the antenatal clinic surveys (open diamonds), although the nationally representative antenatal data have not been used in calculating the likelihood function.

Tables 1 and 2 show the posterior means and 95% confidence intervals for the parameters for which prior distributions were specified. In most cases, prior and posterior means are similar, but the posterior distributions tend to have a lower variance, especially in the case of the standard deviation of study effects and the average duration of immunity parameters. For syphilis and gonorrhoea, the posterior mean female-to-male transmission probability is substantially greater than the prior mean, and for trichomoniasis and gonorrhoea, the posterior mean male-to-female transmission probability is greater than the prior mean. The posterior mean of the standard deviation of study effects is substantially greater than the prior mean in the case of trichomoniasis, indicating that there is relatively high inter-study variation in trichomoniasis prevalence.

Figure 2 compares the estimates of STI prevalence in 2005 for three analyses: (1) using all the available data, i.e. the same results as presented in Figure 1; (2) omitting the most recent data; and (3) randomly omitting half of the data. The model estimates in the second and third analyses are not significantly different from those in the first analysis. The distribution of the omitted

observations, expressed as percentiles of the model estimates, is also not significantly different from uniform in either analysis (results not shown).

To illustrate the importance of uncertainty analysis to users of STI model outputs, we consider three possible strategies for reducing the prevalence of syphilis: doubling the rate at which individuals with syphilis symptoms seek treatment (from initial rates of 0.23 per week in women and 0.57 per week in men); increasing the proportion of private practitioners using syndromic management protocols (from 45% to 80%); and increasing the proportion of pregnant women who are screened and treated for syphilis (from 60% to 100%). If the syphilis parameters are set at the prior means in Table 2 (which produce estimates of syphilis prevalence consistent with the observations shown in Figure 1(a)), the projected reduction in syphilis prevalence by 2020, for the three strategies, is 12%, 4% and 6% respectively. However, using the 500 parameter combinations in the posterior sample, the estimated average reductions are 43% (95% CI: 3–77%), 9% (95% CI: 1–19%) and 6% (95% CI: 4–7%) respectively. Although improved antenatal screening is the least effective strategy on average, it is the most effective strategy for 7% of the posterior parameter combinations and is more effective than expanding syndromic management in the private health sector for 23% of the posterior parameter combinations. The absolute and relative effectiveness of the different strategies are therefore both subject to significant uncertainty.

DISCUSSION

This paper extends uncertainty analysis techniques for fitted HIV/AIDS models to models that are fitted to other STI prevalence data. The proposed approach differs from approaches that have been described for HIV models^{12–14} in two important respects. Firstly, allowance has been made for uncertainty regarding diagnostic accuracy in the definition of the likelihood function; this is necessary in view of the variable sensitivity and specificity of the tests commonly used to detect STIs. Secondly, data from a number of sources have been used in defining the likelihood function: antenatal clinics, family planning clinics, community household surveys and samples of commercial sex workers. Since STI prevalence data may be limited in many developing countries, it is necessary to make use of these multiple data sources when fitting the model. Ideally the STI model should be capable of estimating different prevalence levels for each of these sample types, in order to control for the biases associated with different sample types, and this is a valuable feature of the model that has been used in this analysis. Nationally representative data could also be incorporated into the analysis, by setting the b_i term in respect of the national data to zero, by setting the $\text{Var}(\rho_i)$ term in equation (6) to zero, and by making appropriate adjustment to the n_i term in equation (1) to take account of the survey design.

The Bayesian approach that we have proposed has three key uses. Firstly, it can be used to assess uncertainty regarding STI model predictions. For example, it has been shown that the relative effectiveness of different syphilis treatment strategies is subject to significant uncertainty, and considering only a single combination of syphilis parameters that appears to give outputs consistent with survey data may lead to a false sense of confidence in the relative effectiveness of different strategies. Although cost-effectiveness analyses of STI interventions have typically considered uncertainty regarding cost and efficacy parameters,^{30–33} this analysis shows that uncertainty regarding STI transmission and natural history parameters may also contribute significantly to the overall uncertainty regarding the cost-effectiveness of a particular intervention.

The second use of the proposed Bayesian approach is the estimation of STI prevalence at a national level from limited sentinel surveillance data. The informal validation of the model in the case of syphilis (Figure 1(a)) confirms that model estimates of trends in syphilis prevalence

are reasonably consistent with nationally representative data. The more formal validation results suggest that the method is reasonable for estimating past STI prevalence levels and for short-term projections of future STI prevalence (<5 years beyond the most recent observation), but for longer-term projections there are insufficient data to validate the method. Although other studies have estimated STI prevalence at national or regional levels from sentinel surveillance data,^{16, 17} these analyses have not been based on dynamic models of STI transmission and have therefore been forced to assume that STI prevalence levels are stable over time. In addition, these studies have not allowed for variability in diagnostic accuracy. It would be useful to assess whether the method we have proposed produces plausible results when applied in other countries, such as China, where sentinel surveillance data have suggested increases in syphilis prevalence over the last decade.³⁴

The third use of the proposed Bayesian approach is the estimation of STI transmission and natural history parameters, which are notoriously difficult to estimate empirically. Although the posterior distributions are generally not substantially different from the prior distributions, and generally have high variance, there are a few cases in which the STI prevalence data are informative and induce a posterior distribution that differs substantially from the prior. For example, the unusually high observed prevalence of trichomoniasis in South African women (Figure 1(g)) leads to a substantially higher posterior estimate on the male-to-female transmission probability than was assumed *a priori*, on the basis of a model fitted in other settings.⁵

A potential criticism of the model that has been used is that it assumes individuals are temporarily immune to reinfection following spontaneous resolution of gonorrhoea and trichomoniasis, contrary to expert opinion that such immunity does not exist.^{35–37} Attempts were made to fit the model to the gonorrhoea, trichomoniasis and chlamydial infection prevalence data on the assumption of no immunity following recovery, but this produced significantly poorer fits (results not shown). This points to a source of uncertainty not explicitly addressed in this analysis: uncertainty regarding the choice of model structure. Further work is required to evaluate more formally the appropriateness of different model structures, using techniques such as Bayesian Model Averaging.³⁸

A further limitation of this analysis is that it does not allow for correlation between measurements in the same geographical location. Although it is unrealistic to assume that prevalence measurements in the same location would be independent of one another, regressions of observed STI prevalence levels on various covariates did not reveal residual patterns that differed significantly between locations. This suggests that modifying the method to allow for correlation between measurements in the same location would probably not change results substantially. A related concern is that sentinel surveillance systems are often not geographically representative. In South Africa, for example, there is a strong surveillance bias towards urban areas and towards the KwaZulu-Natal and Gauteng provinces,²² which could potentially bias model estimates of STI prevalence. There is thus a need to improve the collection of STI prevalence data, with a particular focus on surveillance in regions that have previously been under-sampled.

In conclusion, the Bayesian approach to uncertainty analysis of STI models provides a useful means of integrating prior knowledge about STI epidemiology and STI prevalence data, allowing for more reliable estimation of STI prevalence levels and better evaluation of the impact of STI control strategies.

Key messages

- Model estimates of the impact of STI control programmes are subject to significant uncertainty, which needs to be assessed through formal uncertainty analysis.
- The uncertainty analysis approach proposed here allows for estimation of STI prevalence at a national level from sentinel surveillance data obtained using different laboratory methods.
- The approach can be used to integrate data from different sample types (antenatal clinics, family planning clinics, households and surveys of sex workers).
- The Bayesian approach to uncertainty analysis allows for integration of prior knowledge about STI epidemiology into model parameter estimation.

Acknowledgments

We thank Debbie Bradshaw for helpful comments on earlier versions of this paper.

FUNDING

This work was partially funded by a seed grant from the Center for Statistics and the Social Sciences at the University of Washington, and by the National Institute of Child Health and Development through grant no. R01 HD054511. The funders did not participate in the study design, analysis and interpretation of data, writing of the paper, or decision to submit the paper for publication.

REFERENCES

1. Anderson, RM.; May, RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press; 1992.
2. Garnett GP, Bowden FJ. Epidemiology and control of curable sexually transmitted diseases: opportunities and problems. *Sex Transm Dis* 2000;27(10):588–599. [PubMed: 11099074]
3. Korenromp EL, Sudaryo MK, de Vlas SJ, et al. What proportion of episodes of gonorrhoea and chlamydia becomes symptomatic? *Int J STD AIDS* 2002;13:91–101. [PubMed: 11839163]
4. Turner KM, Garnett GP, Ghani AC, Sterne JA, Low N. Investigating ethnic inequalities in the incidence of sexually transmitted infections: mathematical modelling study. *Sex Transm Infect* 2004;80(5):379–385. [PubMed: 15459406]
5. White RG, Orroth KK, Korenromp EL, et al. Can population differences explain the contrasting results of the Mwanza, Rakai, and Masaka HIV/sexually transmitted disease intervention trials?: A modeling study. *J Acquir Immun Defic Syndr* 2004;37(4):1500–1513.
6. Boily M, Lowndes C, Alary M. The impact of HIV epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sex Transm Infect* 2002;78 Suppl i:i78–i90. [PubMed: 12083451]
7. Korenromp EL, Bakker R, de Vlas SJ, et al. HIV dynamics and behaviour change as determinants of the impact of sexually transmitted disease treatment on HIV transmission in the context of the Rakai trial. *AIDS* 2002;16:2209–2218. [PubMed: 12409743]
8. Blower SM, Dowlatabadi H. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int Stat Rev* 1994;62(2):229–243.
9. Iman RL, Conover WJ. A distribution-free approach to inducing rank correlation among input variables. *Commun Statist Simula Computa* 1982;11(3):311–334.
10. Cacuci DG, Ionescu-Bujor M. A comparative review of sensitivity and uncertainty analysis of large-scale systems - II: Statistical methods. *Nucl Sci Eng* 2004;147(3):204–217.
11. Alkema L, Raftery AE, Clark SJ. Probabilistic projections of HIV prevalence using Bayesian melding. *Ann Appl Stat* 2007;1(1):229–248.
12. Morgan M, Walker N, Gouws E, Stanecki KA, Stover J. Improved plausibility bounds about the 2005 HIV and AIDS estimates. *Sex Transm Infect* 2006;82 suppl 3:iii71–iii77. [PubMed: 16735297]
13. Grassly NC, Morgan M, Walker N, et al. Uncertainty in estimates of HIV/AIDS: the estimation and application of plausibility bounds. *Sex Transm Infect* 2004;80 Suppl 1:i31–i38. [PubMed: 15249697]

14. Salomon JA, Murray CJ. Modelling HIV/AIDS epidemics in sub-Saharan Africa using seroprevalence data from antenatal clinics. *Bull WHO* 2001;79(7):596–607. [PubMed: 11477962]
15. Johnson LF, Dorrington RE, Matthews AP. An investigation into the extent of uncertainty surrounding estimates of the impact of HIV/AIDS in South Africa. *S Afr J Sci* 2007;103:135–140.
16. Gerbase AC, Rowley JT, Heymann DHL, Berkley SFB, Piot P. Global prevalence and incidence estimates of selected curable STDs. *Sex Transm Infect* 1998;74 Suppl 1:S12–S16. [PubMed: 10023347]
17. Looker KJ, Garnett GP, Schmid GP. An estimate of the global prevalence and incidence of herpes simplex virus type 2 infection. *Bull WHO* 2008;86:805–812. [PubMed: 18949218]
18. Orroth KK, Korenromp EL, White RG, et al. Comparison of STD prevalences in the Mwanza, Rakai, and Masaka trial populations: the role of selection bias and diagnostic errors. *Sex Transm Infect* 2003;79:98–105. [PubMed: 12690128]
19. Paz-Bailey G, Rahman M, Chen C, et al. Changes in the etiology of sexually transmitted diseases in Botswana between 1993 and 2002: implications for the clinical management of genital ulcer disease. *Clin Infect Dis* 2005;41(9):1304–1312. [PubMed: 16206106]
20. Johnson LF, Dorrington RE, Bradshaw D, Pillay-Van Wyk V, Rehle TM. Sexual behaviour patterns in South Africa and their association with the spread of HIV: insights from a mathematical model. *Demographic Res* 2009;21:289–340.
21. Johnson, LF. Doctoral thesis. Cape Town: University of Cape Town; 2008. The interaction between HIV and other sexually transmitted infections in South Africa: a model-based evaluation. Available: http://www.commerce.uct.ac.za/Research_Units/CARE/2009/ThesisLJfinal.pdf
22. Johnson LF, Coetzee DJ, Dorrington RE. Sentinel surveillance of sexually transmitted infections in South Africa: a review. *Sex Transm Infect* 2005;81:287–293. [PubMed: 16061532]
23. Kelsey, JL.; Thompson, WD.; Evans, AS. *Methods in Observational Epidemiology*. New York: Oxford University Press; 1986.
24. Cook RL, Hutchison SL, Østergaard L, Braithwaite RS, Ness RB. Systematic review: noninvasive testing for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Ann Intern Med* 2005;142(11):914–925. [PubMed: 15941699]
25. Wiese W, Patel SR, Patel SC, Ohl CA, Estrada CA. A meta-analysis of the Papanicolaou smear and wet mount for the diagnosis of vaginal trichomoniasis. *Am J Med* 2000;108:301–308. [PubMed: 11014723]
26. Patel SR, Wiese W, Patel SC, Ohl C, Byrd JC, Estrada CA. Systematic review of diagnostic tests for vaginal trichomoniasis. *Infect Dis Obstet Gynecol* 2000;8(5–6):248–257. [PubMed: 11220487]
27. Smith AFM, Gelfand AE. Bayesian statistics without tears - a sampling resampling perspective. *Am Stat* 1992;46(2):84–88.
28. Anderson JL. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J Climate* 1996;9:1518–1530.
29. Department of Health. National HIV and Syphilis Prevalence Survey: South Africa 2006. Pretoria: 2007. Available: <http://www.doh.gov.za/docs/reports-f.html>.
30. Gilson L, Mkanje R, Grosskurth H, et al. Cost-effectiveness of improved treatment services for sexually transmitted diseases in preventing HIV-1 infection in Mwanza Region, Tanzania. *Lancet* 1997;350(9094):1805–1809. [PubMed: 9428251]
31. Terris-Prestholt F, Watson-Jones D, Mugeye K, et al. Is antenatal syphilis screening still cost effective in sub-Saharan Africa. *Sex Transm Infect* 2003;79(5):375–381. [PubMed: 14573832]
32. White RG, Orroth KK, Glynn JR, et al. Treating curable sexually transmitted infections to prevent HIV in Africa: still an effective control strategy? *J Acquir Immun Defic Syndr* 2008;47(3):346–353.
33. Marseille E, Kahn JG, Billingshurst K, Saba J. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Soc Sci Med* 2001;52(1):135–148. [PubMed: 11144911]
34. Lin CC, Gao X, Chen XS, Chen Q, Cohen MS. China's syphilis epidemic: a systematic review of seroprevalence studies. *Sex Transm Dis* 2006;33(12):726–736. [PubMed: 16755273]
35. Barnes RC, Holmes KK. Epidemiology of gonorrhoea: current perspectives. *Epidemiol Rev* 1984;6:1–30. [PubMed: 6436045]

36. Hedges SR, Mayo MS, Mestecky J, Hook EW, Russell MW. Limited local and systemic antibody responses to *Neisseria gonorrhoeae* during uncomplicated genital infections. *Infect Immun* 1999;67(8):3937–3946. [PubMed: 10417159]
37. Petrin D, Delgaty K, Bhatt R, Garber G. Clinical and microbiological aspects of *Trichomonas vaginalis*. *Clin Microbiol Rev* 1998;11(2):300–317. [PubMed: 9564565]
38. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian Model Averaging: A tutorial. *Stat Sci* 1999;14(4):382–417.

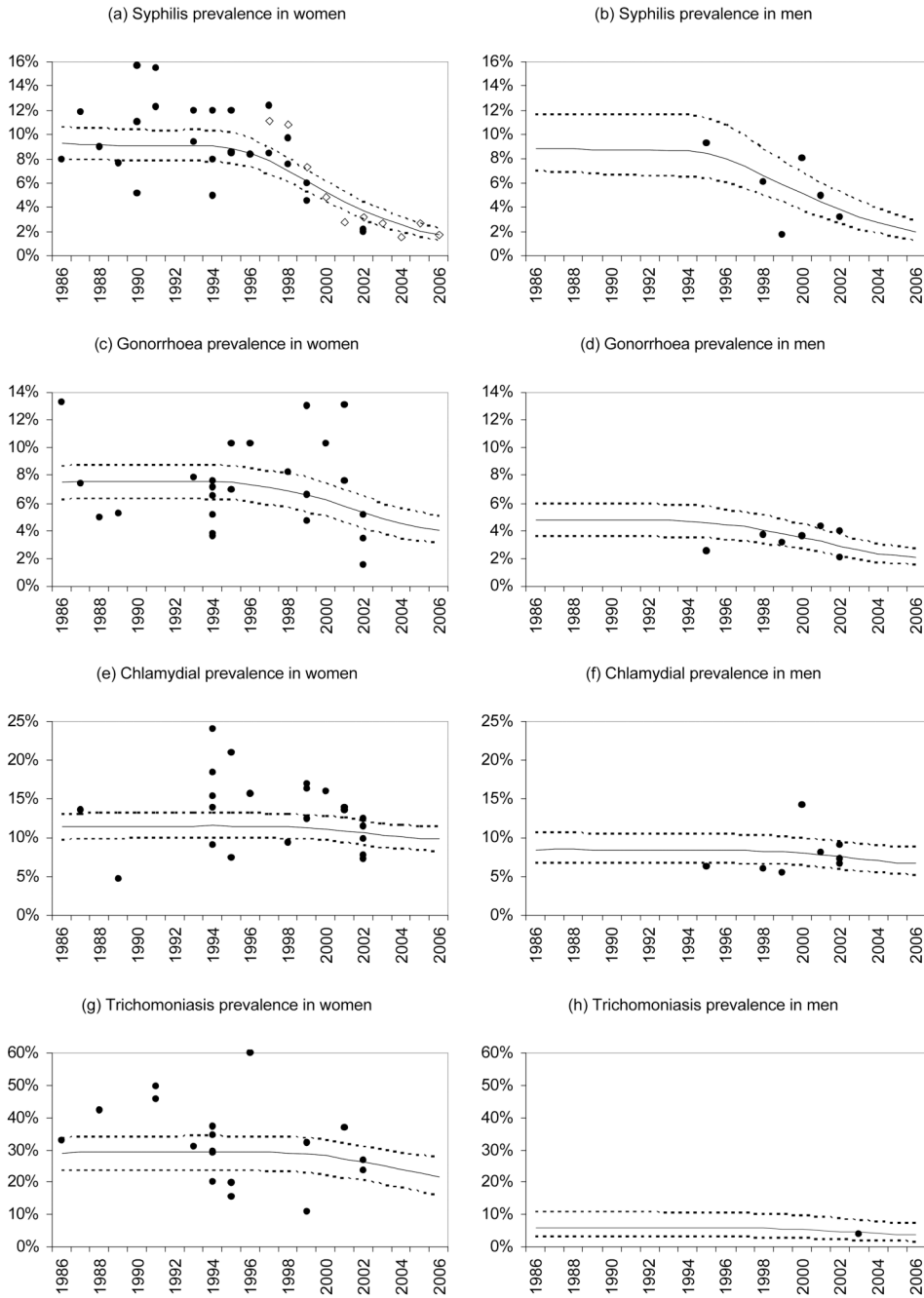


Figure 1. Comparison of observed and estimated prevalence levels

Dots represent observed STI prevalence in household surveys and surveys of antenatal clinic and family planning clinic attenders (after adjustment for expected sensitivity and specificity). Open diamonds represent syphilis prevalence levels observed in nationally representative antenatal clinic surveys (not included in likelihood calculation). Solid black lines represent posterior mean estimates of prevalence in population aged 15–49, and dashed lines represent corresponding 95% confidence intervals (2.5 and 97.5 percentiles of the posterior sample of estimates).

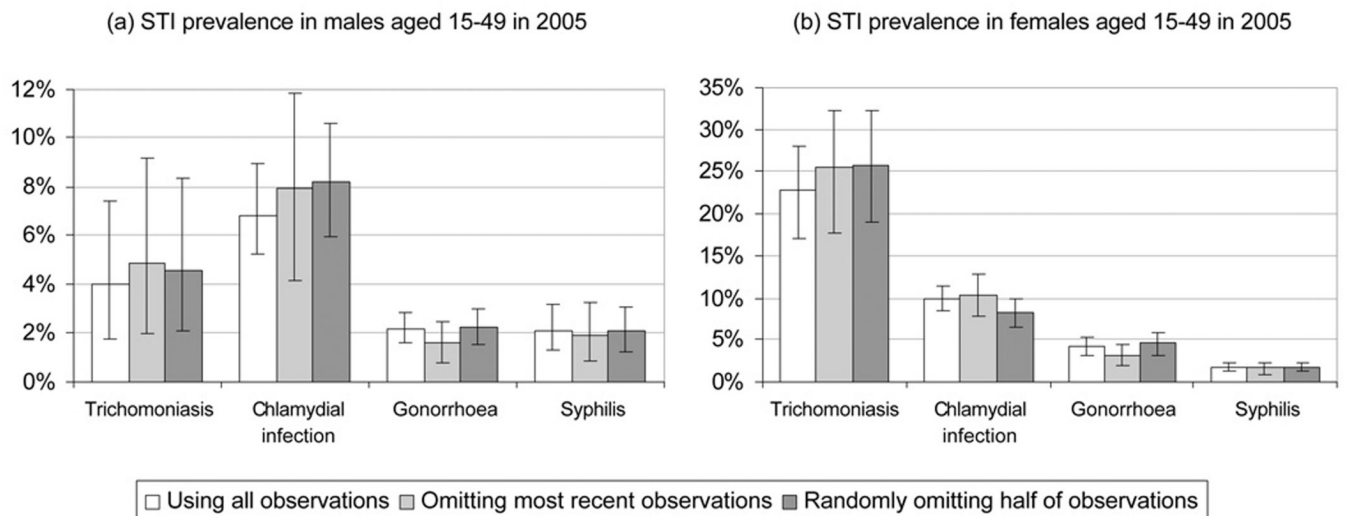


Figure 2. Comparison of model prevalence estimates when using different sets of observations to fit the model

Bars represent average levels of STI prevalence obtained from the posterior sample, and error bars represent 95% confidence intervals (2.5 and 97.5 percentiles of the distribution of STI prevalence levels in the posterior sample).

Table 1

Prior and posterior distributions for gonorrhoea, chlamydia infection and trichomoniasis

Parameter	Prior distribution	Gonorrhoea		Chlamydia infection		Trichomoniasis	
		Prior (mean, 95% CI)	Posterior (mean, 95% CI)	Prior (mean, 95% CI)	Posterior (mean, 95% CI)	Prior (mean, 95% CI)	Posterior (mean, 95% CI)
Male-to-female transmission probability	Beta	0.40 (0.21–0.60)	0.46 (0.30–0.62)	0.16 (0.02–0.40)	0.18 (0.08–0.31)	0.15 (0.03–0.34)	0.22 (0.10–0.40)
Female-to-male transmission probability	Beta	0.20 (0.11–0.31)	0.23 (0.16–0.32)	0.12 (0.03–0.26)	0.11 (0.03–0.31)	0.04 (0.01–0.09)	0.05 (0.02–0.10)
Proportion of cases becoming symptomatic:							
Males	Beta	0.90 (0.78–0.97)	0.88 (0.78–0.96)	0.30 (0.06–0.63)	0.33 (0.07–0.60)	0.40 (0.21–0.60)	0.38 (0.20–0.58)
Females	Beta	0.40 (0.13–0.71)	0.35 (0.12–0.64)	0.15 (0.03–0.34)	0.17 (0.04–0.31)	0.30 (0.13–0.51)	0.29 (0.12–0.48)
Average duration of untreated infection (weeks)							
Symptomatic males	Gamma	15.0 (6.9–26.3) ^a	17.8 (10.0–27.3) ^a	16.0 (4.4–35.1) ^b	17.0 (4.5–35.1) ^b	2.0 (0.9–3.6)	2.0 (0.8–3.7)
Symptomatic females	Gamma	15.0 (6.9–26.3) ^a	14.8 (10.1–21.2) ^a	-	-	15.0 (6.9–26.3)	14.9 (6.7–26.7)
Asymptomatic males	Gamma	-	-	90 (63–122) ^b	89 (66–117) ^b	20.0 (8.7–35.9)	21.2 (11.2–34.4)
Asymptomatic females	Gamma	-	-	-	-	150 (69–263)	157 (87–247)
Average duration of immunity (weeks)	Gamma	52 (14–114)	68 (40–113)	520 (205–979)	490 (277–773)	52 (14–114)	63 (22–121)
Proportion of cases immune after treatment	Uniform	0	-	0.50 (0.03–0.98)	0.72 (0.21–0.99)	0	-
Proportion of cases correctly treated prior to introduction of syndromic management	Beta	0.70 (0.49–0.87)	0.71 (0.53–0.86)	0.70 (0.49–0.87)	0.71 (0.51–0.88)	0.40 (0.13–0.71)	0.41 (0.14–0.68)
Standard deviation of study effects	Gamma	0.30 (0.08–0.66)	0.30 (0.17–0.44)	0.30 (0.08–0.66)	0.33 (0.21–0.47)	0.30 (0.08–0.66)	0.53 (0.33–0.75)

^aDistribution applies to both symptomatic and asymptomatic cases.^bDistribution applies to both males and females.

Table 2

Prior and posterior distributions for syphilis

Parameter	Prior distribution	Prior (mean, 95% CI)	Posterior (mean, 95% CI)
Male-to-female transmission probability	Beta	0.18 (0.09–0.29)	0.17 (0.12–0.23)
Female-to-male transmission probability	Beta	0.15 (0.07–0.26)	0.19 (0.13–0.27)
Average time (in weeks) from			
Primary to secondary	Gamma	6.6 (3.3–11.1)	6.1 (3.0–10.7)
Secondary to latent	Gamma	15.6 (8.8–24.4)	14.9 (9.1–22.1)
Latent to spontaneous resolution	Gamma	520 (269–853)	572 (355–883)
Recovery in early disease to seronegativity	Gamma	26.0 (12.8–43.9)	23.0 (12.8–42.6)
Recovery in latent infection to seronegativity	Gamma	52.0 (25.5–87.7)	51.5 (25.5–84.4)
Proportion of primary cases seronegative immediately after successful treatment	Beta	0.40 (0.21–0.60)	0.42 (0.24–0.60)
Proportion of cases correctly treated prior to syndromic management	Beta	0.70 (0.49–0.87)	0.76 (0.62–0.90)
Standard deviation of study effects	Gamma	0.30 (0.08–0.66)	0.38 (0.28–0.50)