## Supplementary material

**Supplementary material 1 - HIV/STI model equations:**

The model is represented by a set of deterministic ordinary differential equations, which are solved numerically using the C programming language. The equations are summarised as follows:

Equation 1

where:

* – low-risk population of gender *k*, HIV status *h* and HSV-2 status *s* (where *h*=0 and *s*=0 represent uninfected individuals).
* – high-risk population of gender *k* and behavioural stratum *i*, HIV status *h* and HSV-2 status *s*. Thus *k*=1 represent clients with behavioural strata *i*=1-8; *k=*2 represent FSWs with behavioural strata *i*=1-36.
* – low-risk population of gender *k*, with syphilis stage *v*.
* – high-risk population of gender *k* and behavioural stratum *i*, with syphilis stage *v* (where *v*=0 represents uninfected individuals).
* *Bk* is the entry rate into the sexually active population. Newly sexually active individuals are assumed to be uninfected.
* *μ­k*is the gender-specific rate of leaving the sexually active population for non-HIV-related reasons.
* are per-capita rates of progression to AIDS for the syphilis system of equations to ensure that the numbers of individuals in each behavioural compartment is the same for the syphilis and HIV/HSV-2 systems.
* *εik* is the rate of becoming involved with commercial sex (becoming a FSW/client).
* ( for syphilis) is the rate of ceasing involvement with commercial sex.
* and are the force of infection terms due to HIV for the low-risk and high-risk populations respectively. These terms are zero unless *h*=0 or 1. They are modified by coinfection with HSV-2 by the presence of cofactors for increasing susceptibility:
* and are the force of infection terms due to HSV-2 for the low-risk and high-risk populations respectively. These terms are zero unless *s*=0 or 1.
* and are the force of infection terms due to syphilis for the low-risk and high-risk populations respectively. These terms are zero unless *v*=0 or 1.
* and are terms describing the progression of HIV of all individuals from one stage to the next for each behavioural compartment:

Equation 2

where is the average rate of progression of HIV from stage *h* to stage *h*+1.

* and describe the progression of HSV-2 of all individuals of a given type from one stage to the next:

Equation 3

where is a factor to account for the fact that HSV-2 recurrences are more frequent in HIV positive individuals, and represents the average rate of progression from stage *s* (*s*=3 represents recurrence of symptoms).

* and describe the progression of syphilis from one stage to the next:

Equation 4

and represent the rate of recovery from primary (*v*=1) and secondary (*v*=2) syphilis in low- and high-risk individuals of gender *k*. represents the rate of progression from stage *v* to *v*+1 (*v*=1-3) or rate of loss of immunity (*v*=4). represents the rate of recurrence of symptoms from latent (*v*=3) syphilis.

* is the rate at which FSWs (*k*=2) become more consistent condom users. There is no corresponding term for clients (*k*=1), in other words =0. The rate is calculated to ensure condom use always matches the modelled level.
* represents the rate at which high-risk individuals change duration class. The transition from duration class *n*-1 to class *n* (corresponding to moving from behavioural class *i’* to *i* is given by the formula = (1/*dn-1*)× - (1/*dn*)× where *dn* is the length of the duration class *n.*

The force of infection can be generically written as:

Equation 5

where *i* is the type of individual being considered (e.g. home-based FSWs of duration 0-1 years, who consistently uses condoms with occasional clients), *k* is their gender, *x* is their coinfection status for HIV/HSV-2 if relevant, *x’* is the stage of infection of the partner (including coinfection status for HIV/HSV-2 so that for HIV or HSV-2 *x’* represents the pair of indices (*h,s*), while for syphilis it represents *v*), *cik* is the number of partners per unit time of individuals of type (*i,k*),and *ρikxi’k’x’* is a random mixing factor of the form:

Equation 6

The per-partnership transmission probabilities can be written as:

Equation 7

where is the probability of transmission per act depending on the gender of the person *k* and the stage of infection of the partner (which is included in the index *x’*); *χxx’* is a cofactor (for HIV or HSV-2) for infection depending on the coinfection status (included in indices *x* and *x’*) of both partners for other STIs; *N* is the number of acts in the partnership while the partner is infected in stage *x’*, *κ*is the effectiveness of condoms per act at preventing infection against the STI in question; *π* is the proportion of acts for which a condom is used, depending on the type of partnership and the level of condom use of the sex worker if the partnership is commercial.

### Supplementary material 2 – Detailed model structure

The following behavioural stratifications were used in the model: FSWs were stratified by typology (whether they solicit from home, brothels or public places), by duration of time since beginning to sell sex (0-1, 2-4, 5-9, 10+ years), and by their reported consistency of condom use with occasional clients (‘consistently’, ‘sometimes/often’, ‘never’); clients were stratified by duration of time since beginning to buy sex (0-1, 2-4, 5-9, 10+ years) and by level of sexual activity, defined by whether their number of visits to FSWs per month was higher (more active) or lower (less active) than the median. Upon ceasing to sell/buy sex FSWs and clients return to the general population at rates dependent on their typology (for FSWs) or activity level (for clients) and being replaced by uninfected new FSWs/clients from the general population to maintain the proportion of the general population who are high-risk.

This complex structure was found to be necessary as simpler models tested during the development stage were found to possess insufficient heterogeneity to reproduce the rapid initial rise in HIV prevalence among FSWs.

The HIV model has high viraemia phases during initial infection and shortly before developing AIDS, and infectees progress from primary to asymptomatic to pre-AIDS stages at rates from literature1 2. HSV-2 coinfection is modelled dynamically with a short initial infection, a long asymptomatic phase (incorporating low-level infectious shedding) and infectious symptomatic recurrences3. There are cofactors facilitating HIV and HSV-2 acquisition and transmission, reflecting the synergy between them4-6. The natural history of syphilis is modelled with primary and secondary infection stages followed by a latent phase with infrequent recurrences of secondary syphilis7 8. It is assumed that all individuals receive treatment eventually during this secondary stage, and so do not develop tertiary syphilis. Upon treatment they gain temporary immunity. Individuals may also be treated during the primary and secondary stages, although in this case they will not gain temporary immunity, but will instead become susceptible once more. Syphilis is also modelled dynamically, but separately from HIV and HSV-2, using the same behavioural structure so as to give an average cofactor for increasing HIV susceptibility for each behavioural compartment. Within the model, transmission probabilities were modified using risk ratios to account for the more infectious disease stages, and differences between males and females were incorporated similarly.

The effectiveness of condoms in preventing infection is introduced at the per-act level when calculating the transmission hazard, and is different for each STI. The proportion of acts between FSWs and clients which are protected depends on the condom use stratum to which the FSW belongs (using condoms “consistently”, “sometimes”, or “never” with occasional clients). The percentage of acts for which a condom is used in each category was determined from IBBA survey data, and is given in Table 3. From data from the GPS survey, a small proportion of all acts between long-term partners are protected by condoms.

It is assumed that at the beginning of the HIV epidemic the percentage of FSWs consistently using condoms with clients was between 0-10%. Each parameter set thus contained an initial fraction of consistent condom users, andthe fractionat the IBBA round 2, plus at two time-points in between which were chosen so as to best model the data from the given hypothesis. The proportion of consistent condom users amongst FSWs is then modelled to vary linearly between these times. After the second IBBA, condom use was taken to be constant due to the absence of any further data, as well as potential saturation of coverage. The fraction of FSWs “sometimes” using condoms with occasional clients was found to be low in the IBBA; it was taken to increase linearly to the IBBA round 2 value. Constraints were introduced when sampling to ensure that proportions summed to one.

Relationships between FSWs and clients are either occasional (assumed to last 1-3 sex acts) or long-term. There is little data on the behaviour of long-term partners of FSWs, but as their risk of infection is high they are included with clients. Proportionate mixing for each type of partnerships is used, with the number of FSWs visited per month by clients dependent on both the duration and activity level of the client; for FSWs the number of clients entertained per month is governed by the typology of the FSW and the duration of time for which they have been engaged in sex work. In both cases these choices of dependence were made to maximise behavioural heterogeneity between groups9. FSWs determine the number of partnerships available.

New individuals enter the susceptible population at a recruitment rate equivalent to the sum of those leaving the model through ceasing sex, mortality and migration plus the growth rate of the population. Migration is modelled assuming that migrant FSWs keep the same behaviour (rate of partner change, etc.) but that their clients have up to a 2-fold higher prevalence in the sites they migrate to.

### Supplementary material 3 - Detailed fitting procedure and model run characteristics

Once sampled, each parameter set was first pre-screened to reduce the computational time spent on parameter sets which would produce highly unrealistic prevalences. This was achieved by running the model with a constant condom use set at IBBA round 2 levels for HSV-2 and syphilis, but without HIV, until these prevalences reached equilibrium, to simulate the level of these STIs in the absence of HIV. Runs which had STI prevalences so far from the IBBA R2 levels that they would still not fit the prevalence data if even in the presence of HIV were thus screened out (condition for screening out: <1% HSV-2 or >33% syphilis in FSWs). If the parameter set passed this initial pre-screening, the system was then re-set with condom use as determined by the scenario under investigation with 10 HIV positive FSWs in total seeded proportionately across all FSW behavioural classes as initial conditions.

The start of the epidemic is itself a sampled parameter, and the model was run from that time until 1987 when systematic HIV measurement began in India10. At this point a further screening was carried out to determine if the parameter set was likely to produce a realistic epidemic or not. Relaxed criteria for exclusion were chosen – that HIV in FSWs had to be less than 43% for Mysore and less than 75% for Belgaum in 1987 respectively – and these were determined by examining fits from a small sub-sample of 20,000 runs. Parameter sets which passed these screening steps were then run until 2010.

Analysis of the results showed that the client IBBA was an important constraining factor in fitting: a comparable number of fits was obtained for all hypotheses in each districts when fitted to just the round 1 FSW data (results not shown). The addition of the client data helped distinguished between hypotheses because 1) the client IBBA provided information on prevalences in clients, which was an important constraint in itself; and 2) the client IBBA was carried out later than the FSW round 1 IBBA, thus providing second time point and giving indirect information on time trends in prevalence.

Investigation of individual unfitted runs under H1 (condom use reconstruction) for Belgaum and Mysore also shows that the accepted model fits exhibit a wide range of behaviours. The majority of unfitted runs (80% For Belgaum under H0/H1, 69% for Mysore under H0/H1; 67.2% for Mysore under H2) increase rapidly, and have plateaued by 1990. Those runs which rise more quickly plateau at higher prevalences. Those which take off slowly may continue increasing until after 2000. A small fraction of runs have prevalence <1% in FSWs (3% for Belgaum under H0/H1, 9% for Mysore under H0/H1 and H2), while the majority of unfitted runs (84% for Belgaum under H1; 74% for Mysore under H0/H1 and H2) have prevalences > 50% in FSWs. By 2004, all epidemics under hypothesis H1 are decreasing. 18% of all (unfitted) runs in Belgaum predict HIV prevalence >80% in FSWs at IBBA round 2 (2008), while 7% of all runs predict <10%. In Mysore, 8% of all runs predict prevalence to be >80% in 2008, while 13% predict prevalence < 10%.

Looking at stratified prevalence in FSWs, it is possible for FSW HIV prevalence to reach 100% in certain groups after 2000, especially in FSWs who have been in sex work for longer and in groups who have large numbers of clients such as brothel-based FSWs. In Mysore prevalence saturates in the longest duration brothel-based FSWs who do not use condoms. In Belgaum prevalence saturates in brothel and public place-based FSWs who have been in sex work for ten or more years and who do not always use condoms with occasional clients. In reality, as well as in the model, these groups are extremely small, and the empirical IBBA data available shows that new FSWs have lower HIV prevalences than old FSWs, while brothel-based FSWs typically have high prevalence. Thus this model saturation is relatively consistent with data, and because of the small size of the group in the model it will have only a small effect on the overall FSW prevalence.

**Supplementary material 4 – Figure 1:** Comparison of HSV-2 and syphilis prevalence model estimates with data. Graph shows FSW prevalence data at the time of rounds 1 and 2 of the FSW IBBA, and prevalence in clients at the time of the client IBBA. Round 2 HSV-2/syphilis data was not used in fitting or model validation.For the model results, median (middle line inside boxes), 25/75 percentiles (limits of boxes), and 2.5/97.5 (error bars) are shown. For IBBA data, 95% confidence intervals are shown.

a. Mysore



b. Belgaum *(no data on FSW round 2 HSV-2 available)*



**Supplementary material 5 – Figure 2:** Validation of model using FSW HIV prevalence by typology from the two most likely hypotheses for Mysore (figure a) and Belgaum (figure b). Also shown is IBBA prevalence data. For the model results, median (middle line inside boxes), 25/75 percentiles (limits of boxes), and 2.5/97.5 (error bars) are shown. For IBBA data, 95% confidence intervals are shown. In Mysore only 2 brothel-based FSWs were sampled in the IBBA R1, so no prevalence data is shown in this case.

a. Mysore



b. Belgaum



**Supplementary material 6 – Comparison of model projections of HIV prevalence over time with data for FSWs and clients for hypothesis H1 in Mysore**

a. and b. Mysore FSWs and clients under hypothesis H1

****

1. Wawer MJ, Gray RH, Sewankambo NK*, et al.* Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 2005;**191**(9):1403-9.

2. Boily MC, Baggaley RF, Wang L*, et al.* Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *Lancet Infect Dis* 2009;**9**(2):118-29.

3. Foss AM, Vickerman PT, Chalabi Z*, et al.* Dynamic modeling of herpes simplex virus type-2 (HSV-2) transmission: issues in structural uncertainty. *Bull Math Biol* 2009;**71**(3):720-49.

4. Foss AM, Vickerman P, Mayaud P*, et al.* Exploring the interactions between HSV-2 and HIV: model predictions for a southern Indian setting. *XVII International AIDS Conference*; 3-8 August 2008; Mexico City. <http://www.aids2008.org/>

5. Schacker T, Zeh J, Hu HL*, et al.* Frequency of symptomatic and asymptomatic herpes simplex virus type 2 reactivations among human immunodeficiency virus-infected men. *J Infect Dis* 1998;**178**(6):1616-22.

6. Conant MA, Schacker TW, Murphy RL*, et al.* Valaciclovir versus aciclovir for herpes simplex virus infection in HIV-infected individuals: two randomized trials. *Int J STD AIDS* 2002;**13**(1):12-21.

7. Garnett GP, Aral SO, Hoyle DV*, et al.* The natural history of syphilis. Implications for the transmission dynamics and control of infection. *Sex Transm Dis* 1997;**24**(4):185-200.

8. Vickerman P, Foss AM, Pickles M*, et al.* Is the Indian HIV epidemic driven by commercial sex? A modelling analysis from South India. *Sex Transm Infect* 2009(*This issue*).

9. Deering KN, Blanchard J, Moses S*, et al.* Characterising the factors associated with the numbers of client partners of female sex workers in south India: a geographic comparison. *18th ISSTDR*; 28 June - 1 July 2009; London. <http://www.isstdrlondon2009.com>.

10. Bollinger RC, Tripathy SP, Quinn TC. The human immunodeficiency virus epidemic in India. Current magnitude and future projections. *Medicine (Baltimore)* 1995;**74**(2):97-106.