# A Bayesian approach to estimate changes in condom use from limited human immunodeficiency virus prevalence data

J. Dureau and K. Kalogeropoulos,

*London School of Economics and Political Science, UK*

P. Vickerman

*London School of Hygiene and Tropical Medicine, UK*

and M. Pickles and M.-C. Boily

*Imperial College London, UK*

**Summary.** Evaluation of large-scale intervention programmes against human immunodeficiency virus (HIV) is becoming increasingly important, but impact estimates frequently hinge on knowledge of changes in behaviour such as the frequency of condom use over time, or other self-reported behaviour changes, for which we generally have limited or potentially biased data. We employ a Bayesian inference methodology that incorporates an HIV transmission dynamics model to estimate condom use time trends from HIV prevalence data. Estimation is implemented via particle Markov chain Monte Carlo methods, applied for the first time in this context. The preliminary choice of the formulation for the time varying parameter reflecting the proportion of condom use is critical in the context studied, because of the very limited amount of condom use and HIV data available. We consider various novel formulations to explore the trajectory of condom use over time, based on diffusion-driven trajectories and smooth sigmoid curves. Numerical simulations indicate that informative results can be obtained regarding the amplitude of the increase in condom use during an intervention, with good levels of sensitivity and specificity performance in effectively detecting changes. The application of this method to a real life problem demonstrates how it can help in evaluating HIV interventions based on a small number of prevalence estimates, and it opens the way to similar applications in different contexts.

*Keywords*: Bayesian inference; Condom use; Epidemic modelling; Human immunodeficiency virus infections; Particle Markov chain Monte Carlo methods; Time varying parameter

## 1. Introduction

Significant resources are being committed to implement large-scale interventions against infectious diseases such as the human immunodeficiency virus (HIV), that killed an estimated 2 million individuals in 2008 (UNAIDS and World Health Organization, 2009). Increasing attention is given to the evaluation of these large-scale intervention programmes to understand what still needs to be done to control the epidemic and eventually to achieve elimination.

Even if antiretroviral therapy has become an important component of large-scale prevention interventions, the use of condoms and circumcision remain important strategies for reducing

HIV transmission. Although there are difficulties in estimating condom use trends accurately, due to biases that are inherent in self-reported behaviour (Turner and Miller, 1997; Zenilman *et al.*, 1995; Hanck *et al.*, 2008), its average level closely determines the spread of HIV (Boily *et al.*, 2007). Thus, it is important to assess whether trends in epidemiological data such as HIV prevalence can be used to infer the effect of interventions on risk behaviours that are susceptible to self-reported bias. This is furthermore motivated by the fact that directly observed quantities such as HIV prevalence do not provide straightforward indications of the effect of an intervention. Indeed, even if an efficient intervention is introduced early, an epidemic has an intrinsic dynamic that can still lead to a growth in prevalence. Alternatively, in a mature epidemic the prevalence can decrease even though on-going interventions are inefficient (Boily *et al.*, 2002). Hence, it is the trajectory of the underlying condom use over time, and especially since the beginning of a prevention programme, that can shed light on the effect of the intervention and on the future trajectory of the epidemic. In this perspective, we develop a Bayesian approach to evaluate the Avahan programme, which is a large-scale HIV and acquired immune deficiency syndrome intervention targeted to high-risk groups, through the estimation of the evolution of condom use from the available HIV prevalence data.

The Avahan intervention was motivated by high levels of HIV prevalence among high-risk groups observed in southern India (typically greater than 20%) (Ramesh *et al.*, 2008), which lead to concerns about infections bridging to their long-term partners and the general population. The programme was launched by the Bill and Melinda Gates Foundation in 2003 (Bill and Melinda Gates Foundation, 2008) and has targeted high-risk groups for HIV infection, in particular female sex workers (FSWs), by promoting and distributing free condoms. Several studies have been conducted to examine the effect of the Avahan programme (Boily *et al.*, 2007, 2008; 2013; Deering *et al.*, 2008; Lowndes *et al.*, 2010; Pickles *et al.*, 2010, 2013), and to learn from it to inform future large-scale interventions. A key part of such evaluations is examining how risk behaviours, chiefly the use of condoms, defined as the proportion of sex acts protected by condoms at a given time, have changed over the course of the intervention. However, this can be difficult to measure in practice. Baseline condom use may be difficult to record when an intervention needs to be implemented rapidly, as happened with the Avahan programme, or may be recorded only on few occasions. Although those targeted by the intervention may be asked about their condom use history (Lowndes *et al.*, 2010), their answers may be subject to social desirability and recall biases. In principle, the total number of condoms sold or distributed can be enumerated (Bradley *et al.*, 2010), but accurate records may not be available, condoms may be used for family planning by lower-risk individuals and the distribution of condoms is not a guarantee of their correct usage (Bradley *et al.*, 2010; Kumar *et al.*, 2011). Thus, in addition to direct approaches through quantitative behavioural surveys or records of condom availability, model-based methods can be used to infer unobserved quantities of interest, such as the use of condoms, and to complete the partial information that is available from observed quantities such as HIV prevalence, using knowledge of the dynamics of large-scale epidemics.

A first study in the context of the Avahan programme was presented in Pickles *et al.* (2010, 2013) and Boily *et al.* (2013). In this work, a deterministic dynamic model for HIV or sexually transmitted infection was formulated on the basis of a compartmental representation incorporating heterogeneous sexual behaviour. The main aim of this approach was to shed light on the partially observed trajectory of condom use on the basis of different sources of data, with their own potential biases. This initial inference proceeded by setting three different hypothesized scenarios for the condom use evolution over time, before and after the introduction of the Avahan programme in 2003, that were estimated by using either self-reported condom use data

from a number of serial cross-sectional surveys on FSWs (Lowndes *et al.*, 2010) or using condom distribution and condom sales data. The three hypotheses alternatively made the following assumptions.

(a) The use of condoms initially increased according to the reported FSWs survey data until the Avahan initiative started in 2003 and was assumed to remain constant afterwards, despite evidence of an increase from FSW survey data. This was to acknowledge that FSWs may overreport the use of condoms after being exposed to the intervention.

(b) The use of condoms increased as in scenario (a) up to the Avahan initiative, but use was assumed to increase from 2003 onwards in accordance with self-reported FSWs survey data.

(c) Scenario (c) is the same as scenario (b) but the increase after 2003 was in accordance with the condom distribution and sales data.

The HIV–sexually transmitted infection model was then used with each of these scenarios, which were therefore assessed on the basis of the fit to the available prevalence observations.

The work that we present in this paper operates in a similar context to that in Pickles *et al.* (2010) but, rather than making explicit assumptions about the evolution of condom use based on additional sources, we aim to develop a Bayesian inference framework exploring the entire space of the condom use trajectories without any reference to the 2003 Avahan intervention, FSWs surveys or condom distribution data. By assessing the effect of the Avahan intervention on condom use under less stringent assumptions, this approach provides additional and stronger evidence of the effect of the Avahan programme on condom use, beyond the three scenarios above. From a methodology perspective, the model formulation can be put in a state space setting where an underlying latent process (condom use trajectory) is observed through the prevalence data, and the link between these quantities is given by the deterministic model for the HIV infections. Inference in this context is a challenging task given the limited amount of HIV prevalence data aside from initial conditions (four observations) that are concentrated over a period of 6 years and are utilized to estimate a 25-years-long trajectory. Various models for the condom use trajectories were considered, including smooth and non-differentiable (yet continuous) choices. In the remainder of this paper, the term 'trajectory prior' is used to refer to these models to avoid confusion with the deterministic HIV model. We also present a general and efficient computational scheme using Markov chain Monte Carlo (MCMC) techniques based on the particle MCMC (PMCMC) algorithm (Andrieu *et al.* (2010); see Dureau *et al.* (2013) for an application in a different context). The focus is on estimating the amplitude of the change in the use of condoms since 2003 (the start of the Avahan programme) to assess the effect of the Avahan intervention on condom use. In addition to the scientific interest of this application, there are various methodological challenges given the limited amount of data. For this reason, the properties of the estimators arising from the MCMC method are studied via simulations, and the performance is assessed from a decision-making perspective through their sensitivity and specificity in detecting changes in the use of condoms.

The next section presents the models that are introduced in this paper, the data that are typically available for such studies and the way in which prior information is incorporated. The computational techniques, mainly the PMCMC algorithm, are also presented. The methodology that is employed to compare the performance of the proposed trajectory priors is presented in Section 3, and the results from this study are introduced in Section 4 along with an application to real data from the Indian acquired immune deficiency syndrome initiative Avahan. Finally, Section 5 concludes with some relevant discussion.

## 2. Models and methods

Our model specification consists of two parts. The first part is based on the formulation of Pickles *et al.* (2010) for the HIV transmission population dynamics. Our proposed methodology targets mainly the second part, where the HIV transmission model adopted is extended by considering a parametric time varying version of the condom use parameter. In what follows we first introduce and describe the HIV transmission model (Section 2.1), before getting into our proposed extension of it (Section 2.2), its prior parameterization (Section 2.3) and its MCMC implementation (Section 2.4).

### 2.1. Human immunodeficiency virus transmission model for female sex workers

In this section, we present the part of our modelling framework that aims to capture the dynamics of HIV transmission in an environment such as the Mysore district, in southern India. The model is based on that of Pickles *et al.* (2010), focusing only on HIV transmission between FSWs and clients. Importantly, the model retained the essential heterogeneity in sexual risk behaviour that is observed in the data by stratifying FSWs in high and low volume (numbers of clients seen in a year) sex workers. A stable but open population is considered, consisting of FSWs and their clients. In line with Vickerman *et al.* (2010), we assume that the remaining members of the population (e.g. companions of clients or sex workers), who are not involved directly in sex work, have little influence on the dynamics of the epidemic and can therefore be ignored. The model structure groups FSW into high-risk, $H$, and low-risk, $L$, sex workers, who may have different numbers of clients, $C$. Each individual in these three groups can either be susceptible, $S$, to HIV infection, infectious, $I$, or retired either because of death or ceasing commercial sexual activity. Moreover, individuals who either decease or stop being involved in commercial sex are assumed to be replaced by susceptible individuals, such that the size of the risk population remains constant. A graphical representation of the model is given in Fig. 1 via a flow diagram for each of these three population groups. A key component of the epidemic model is the force of infection $\beta(t)$ which is assumed to be the product between the number of clients and the probability of HIV transmission in at least one out of $a$ sexual acts per client encounter. We denote respectively by $U(t)$ and $e$ the proportion of FSWs' sex acts that are protected by condoms and the efficacy of condoms in protecting against transmission of HIV per sex act. A sexual act can lead to HIV transmission when either no condom is being used, occurring with probability $1 - U(t)$, or when a condom is not being used effectively, occurring with probability $U(t)(1 - e)$. Hence, the combined probability of a sexual act that can lead to HIV transmission, which is also referred to as a risky sexual act, simplifies to $1 - e\,U(t)$. Consequently, the respective forces of infection of high-risk FSWs, low-risk FSWs and clients can be defined as

$$\left.\begin{aligned}
\beta_H(t) &= C_H(1 - [1 - p_S\{1 - e\,U(t)\}]^a), \\
\beta_L(t) &= C_L(1 - [1 - p_S\{1 - e\,U(t)\}]^a), \\
\beta_C(t) &= \frac{C_H + C_L}{2}\,\frac{N_L + N_H}{N_C}(1 - [1 - p_C\{1 - e\,U(t)\}]^a),
\end{aligned}\right\} \tag{1}$$

where

(a) $C_H$ and $C_L$ are the numbers of clients of FSWs per month, which differ for high-risk and low-risk FSWs and

(b) $N_C$, $N_L$ or $N_H$ are the total numbers of clients, low-risk and high-risk sex workers respectively, considered to be constant over time. By construction, $N_L$ and $N_H$ are equal.
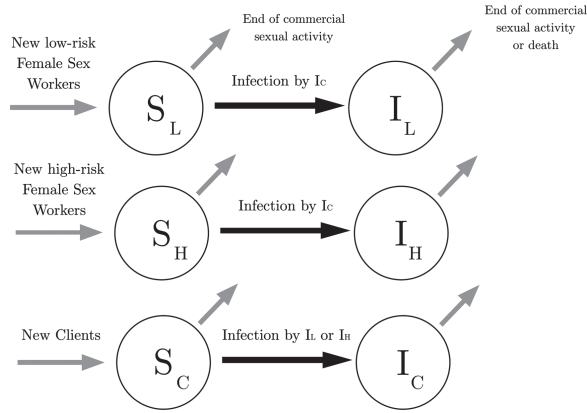
**Fig. 1.** Flow diagram of the model

(c) $p_S$ or $p_C$ are the probability of HIV transmission from client to sex worker or sex worker to client respectively during an unprotected sex act.

Additionally, the transmission dynamics of HIV will depend on the following lengths of time.

(a) $\mu_S^{-1}$ or $\mu_C^{-1}$ are the average lengths of sexual activity as a sex worker or client respectively.
(b) $\alpha^{-1}$ is the average life expectancy with HIV.

Information on these quantities is available from various serial cross-sectional biobehavioural surveys conducted in the Mysore district during the intervention (Ramesh *et al.*, 2008). However, there is still uncertainty around these biological and behavioural parameters, which is reflected in the estimates of condom use by using a Bayesian approach (De Angelis *et al.*, 2008).

The complete transmission dynamics model can be defined via the following set of differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}S_H}{\mathrm{d}t}(t) &= -\beta_H(t)\,S_H(t)\frac{I_C(t)}{N_C} + (\mu_S+\alpha)\,I_H(t), \\
\frac{\mathrm{d}I_H}{\mathrm{d}t}(t) &= \beta_H(t)\,S_H(t)\frac{I_C(t)}{N_C} - (\mu_S+\alpha)\,I_H(t), \\
\frac{\mathrm{d}S_L}{\mathrm{d}t}(t) &= -\beta_L(t)\,S_L(t)\frac{I_C(t)}{N_C} + (\mu_S+\alpha)\,I_L(t), \\
\frac{\mathrm{d}I_L}{\mathrm{d}t}(t) &= \beta_L(t)\,S_L(t)\frac{I_C(t)}{N_C} - (\mu_S+\alpha)\,I_L(t), \\
\frac{\mathrm{d}S_C}{\mathrm{d}t}(t) &= -\beta_C(t)\,S_C(t)\left\{\frac{C_H N_H}{C_H N_H + C_L N_L}\frac{I_H(t)}{N_H} + \frac{C_L N_L}{C_H N_H + C_L N_L}\frac{I_L(t)}{N_C}\right\} + (\mu_C+\alpha)\,I_C(t), \\
\frac{\mathrm{d}I_C}{\mathrm{d}t}(t) &= \beta_C(t)\,S_C(t)\left\{\frac{C_H N_H}{C_H N_H + C_L N_L}\frac{I_H(t)}{N_H}\frac{C_L N_L}{C_H N_H + C_L N_L}\frac{I_L(t)}{N_L}\right\} - (\mu_C+\alpha)\,I_C(t).
\end{aligned}
\tag{2}
$$

In these equations, the state of the epidemic over time is described by the number of susceptible individuals among clients $S_C$, among the low-risk FSWs, $S_L$, and among the high-risk FSWs,

$S_H$, as well as the corresponding numbers of infected individuals, $I_C$, $I_H$ and $I_L$. Three types of constant parameters are involved: the initial prevalence among the various groups of interest at time $t_0 = $ January 1st, 1985 ($\theta_{t_0} = \{S_H(t_0), I_H(t_0), S_L(t_0), I_L(t_0), S_C(t_0), I_C(t_0)\}$), time invariant parameters describing the biological and behavioural determinants of HIV transmission ($\theta_b = \{\mu_S, \mu_C, \alpha, C_H, C_L, p_S, p_C, a, e\}$) and the parameters that play a role in the trajectory priors for condom use between 1985 and 2010 ($\theta_c$). All three components $\theta_{t_0}$, $\theta_b$ and $\theta_c$ are integrated into a global vector of constant parameters, which is denoted by $\theta$. The trajectory $\mathbf{X}_{0:N}$ of the space vector $\mathbf{X}(t) = \{S_H(t), I_H(t), S_L(t), I_L(t), S_C(t), I_C(t)\}$ at times $\{t_0, \ldots, t_N\}$ is then defined as a deterministic function of $\theta$ and $U$ ($\mathbf{X}_{0:N} = f(\theta, U)$) through the HIV transmission model, and the resulting HIV prevalence points are compared with the available observations $y_{t_1}, y_{t_2}, \ldots, y_{t_N}$, denoted by $\mathbf{y}_{1:N}$. The function $f(\cdot)$ is not available in closed form but can be obtained given the trajectory of $U$ by solving the above ordinary differential equations. More specifically, we introduce a time discretized version of $U$, with equidistant points in time of step $\delta$, $U^d_{0:N} = \{U(t_0), U(t_0 + \delta), U(t_0 + 2\delta), \ldots, U(t_N)\}$. The partition of the condom use trajectory can be made arbitrarily fine through the user-specified parameter $\delta$ to limit the approximation error that is induced by the time discretization. For ease of notation, the superscript $d$ will be omitted in the remainder of the paper.

Assigning a model for the observation error provides the likelihood of the observation $\mathbf{y}_{1:N}$ conditional on the $U$-trajectory $p(\mathbf{y}_{1:N}|\theta, U)$. The data consist of HIV prevalence estimates obtained from random samples of FSWs and clients in the Mysore district obtained at different time points. The binomial distribution is parameterized at each $t_i$ by $\{I_H(t_i) + I_L(t_i)\}/(N_H + N_L)$ for prevalence observations among FSWs, and by $I_C(t_i)/N_C$ for prevalence observations among clients.

Overall, the model appearing in Fig. 1 and equations (1) and (2) contains fewer parameters than the model in Pickles *et al.* (2010). This was done mainly for parsimony; models of increased complexity can be used, provided that there is adequate prior information on their parameters. More details on informative priors are provided in Section 2.3.

## 2.2. Trajectory priors for condom use

From a modelling perspective, the main contribution of this paper lies in the introduction of various models for the evolution of the use of condoms, aiming to explore the complete space of trajectories rather than restricting the analysis to a limited set of scenarios. First, we assign a trajectory prior aiming to impose little constraint on the trajectory of condom use over time. We proceed by assigning Brownian motion to a transformed version of $U$ such that it takes values on the real line. An alternative similar formulation could use integrals of Brownian motion. Loosely speaking, these models can be linked with a smoothing splines approach (Wahba, 1990). The second objective aims at a slightly more informative formulation regarding the evolution of condom use and considers sigmoid-shaped growth curves. Such growth curves, which were also mentioned in Pickles *et al.* (2010), provide natural models for related quantities and have been used in various contexts such as biology (Zwietering *et al.*, 1990), marketing (Lessne and Hanumura, 1998) and epidemiology (Omran, 1971). Hence, the second trajectory prior is based on the generalized deterministic Bertalanffy–Richards (DBR) model (see for example Garcia (1983) and Yuancai *et al.* (1997)). To enrich this context and to address estimation issues that can be encountered with the DBR model (Lei and Zhang, 2004), we also consider an alternative empirical sigmoid curve model. In what follows, we denote with $x$ the latent process that drives the condom use trajectory, which in turn provides the link with the prevalence observations through the model in Section 2.1.

### 2.2.1. *Brownian motion*

The first formulation assigns Brownian motion to a transformed version of the condom use trajectory. As the latter must be constrained in the [0,1] region, we work with the logit transformation of $U(t)$, which is denoted by

$$U(t) = \frac{\exp\{x(t)\}}{1 + \exp\{x(t)\}},$$

$$\mathrm{d}x(t) = \sigma\,\mathrm{d}B_t. \tag{3}$$

The use of diffusion processes to describe time varying quantities in contexts that are associated with uncertainty has been explored in the context of epidemiology; see for example Cazelles and Chau (1997), Cori *et al.* (2009) and Dureau *et al.* (2013). It can also be seen as a prior according to which $x(t)$ is a random walk with continuous, yet non-differentiable, trajectories. It is used here in an attempt to incorporate a limited amount of prior information on the shape of the trajectory. It can also be used as an exploration tool for potential modelling–remodelling steps towards more informative formulations. In fact, inspection of the output from this model motivated the choice of sigmoid-shaped growth curves, which are described in the next section. Variations of this formulation may include smoother diffusion models, by taking integrals of the Brownian motion, or alternative transformations such as the probit link. We note at this point that very little information is available on the volatility in model (3), which is determined mostly by its prior. More details are provided in Sections 2.3 and 3.

### 2.2.2. *Deterministic Bertallanfy–Richards function*

Qualitatively, condom use trends reconstructions by alternative methods (Lowndes *et al.*, 2010; Bradley *et al.*, 2010) suggest that condom use was quite low in 1985 and has grown over recent years. The above motivated the use of a growth curve parametric model as an alternative to the Brownian motion diffusion model. We use the generalized Bertalanffy–Richards family (Richards, 1959; Garcia, 1983) that can be written as

$$U(t) = \eta\{1 - B\exp(-kt)\}^{1/(1-m)},$$

or alternatively, in a differential equation framework,

$$U(t) = \{(1-m)\,x(t) + \eta^{1-m}\}^{1/(1-m)}, \tag{4}$$

$$\mathrm{d}x(t) = -k\,x(t)\,\mathrm{d}t. \tag{5}$$

The growth curve can be parameterized by four quantities: the initial value of $U_0$, the time of inflection $\tilde{t}$ (at which $U''(\tilde{t}) = 0$), the value of condom use after an infinite time $\eta$ (also termed the asymptote) and the 'shape' or 'allometric' parameter $m$. The time of inflection can be related to the parameter $k$ by

$$k\tilde{t} = \log\left(\frac{B}{1-m}\right).$$

This family contains various growth curves, including the logistic ($m = 2$) and Gompertz ($m \rightarrow \infty$) functions. Furthermore, this definition implies that the initial value $U_0$ is lower than $m^{1/(1-m)}\eta$. To focus on sigmoid-shaped growth curves, we restrict our attention to cases where $m \geqslant 1$ (Yuancai *et al.*, 1987).

### 2.2.3. *Deterministic empirical sigmoid curve*

An empirical sigmoid model is also considered, to address the potential difficulties that can

arise with the parameterization of the DBR model. First, inference on the allometric parameter $m$ in the DBR has been shown to be problematic (Lei and Zhang, 2004). In addition, since growth models are used to study intrinsically growing objects, trajectories that are inexplicably stable for a long period of time and that eventually start to pick at a rapid pace are not typical under the DBR formulation. The latter may lead to underestimating the amplitude of a shift in condom use under the potential extrinsic influence of the Avahan intervention. For this reason, we also consider an alternative sigmoid curve, defined in the following way:

$$U(t) = a + \frac{b}{1 + c\, x(t)},$$

$$\mathrm{d}x(t) = -k\, x(t)\, \mathrm{d}t. \tag{6}$$

Here the model is parameterized by its baseline $U_0$, its asymptote $\eta$, its time of inflection $\tilde{t}$ and the increase rate $k$, from which $a$, $b$ and $c$ can be computed:

$$\left.\begin{aligned} a &= \eta - b, \\ b &= \left(1 + \frac{1}{c}\right)(\eta - U_0), \\ c &= \exp(k\tilde{t}). \end{aligned}\right\} \tag{7}$$

### 2.2.4. Stochastic growth curves

It is also possible to combine the Brownian motion and the growth curve approaches by using diffusion processes. Stochastic extensions of the DBR and deterministic sigmoid model can be considered, in which the mean behaviour remains intact although some random perturbations are introduced through a stochastic differential equation. To ensure positivity, restrict $U(t)$ below 1 and to retain the link with the DBR curve, geometric Brownian motion can be used to replace equations (5) and (6):

$$\mathrm{d}x(t) = -k\, x(t)\, \mathrm{d}t + \sigma\, x(t)\, \mathrm{d}B_t. \tag{8}$$

The stochastic growth curve that is defined by equations (4) and (8) was also mentioned in Garcia (1983). A convenient feature for both stochastic extension of the DBR and the deterministic sigmoid models is the fact that, since

$$x(t) = \frac{1}{1 - m}\{U(t)^{1-m} - \eta^{1-m}\},$$

and $x(t)$ is strictly negative, the resulting condom use trajectory is maintained strictly below $\eta$. Given the limited data at our disposal, these models can hardly be fitted in the context of this paper. Nevertheless, they may be helpful in cases where more observations are available.

### 2.3. Priors

The parameters contained in $\theta_{t_0}$ and $\theta_b$ cannot be identified from the prevalence observations only, so we assign informative priors on them. These are summarized in Table 1 and are similar to the priors that were used in Pickles *et al.* (2010, 2013). The priors relative to general quantities as transmission probability for unprotected acts or life expectancy with HIV were based on previous literature. Cross-sectional individual-based surveys conducted in Mysore were used to inform the priors regarding quantities that are more sociologically and geographically specific.

Although there is some information in the data regarding $\theta_c$, the parameters that are used to describe the condom use trajectory, their posterior will depend on the prior to a large extent. As

**Table 1.**  Table of priors for the different components of $\{\theta_{t_0}, \theta_b, \theta_c\}$

| Definition | Notation | Range of uniform priors for the district of Mysore (Pickles et al., 2010) |
|---|---|---|
| *HIV transmission model parameters* | | |
| Probability of transmission from client to sex worker per act | $p_S$ | 0.0006–0.0055 |
| Probability of transmission from sex worker to client per act | $p_C$ | 0.0001–0.007 |
| Condom efficacy per act | $e$ | 80–95% |
| Mean number of acts per client | $a$ | 1–2 |
| Mean number of clients per high-risk FSW | $C_H$ | 46.6–54.0 clients per month |
| Mean number of clients per low-risk FSW | $C_L$ | 20–23.7 clients per month |
| Total number of FSWs | $N_H + N_L$ | 1943 |
| Clients/FSW population ratio | $\dfrac{N_C}{N_H + N_L}$ | 7–19 |
| Mean length of sexual activity as FSW | $\mu_S^{-1}$ | 45–54 months |
| Mean length of sexual activity as client | $\mu_C^{-1}$ | 154–191 months |
| Mean life expectancy after infection with HIV | $\alpha^{-1}$ | 87–138.5 months |
| Initial proportion of infected FSWs in 1985 | $\dfrac{I_H + I_L}{N_H + N_L}$ | 0–5% |
| Initial proportion of infected clients in 1985 | $\dfrac{I_C}{N_C}$ | 0–5% |
| *Condom trajectory priors parameters* | | *Prior* |
| Allometric parameters (DBR) | $m$ | $\mathcal{N}(1, 10^6)\mathbb{I}_{]1,\infty[}$ |
| Growth rate (deterministic sigmoid) | $k$ | $\mathcal{N}(0, 10^6)\mathbb{I}_{]0,\infty[}$ |
| Asymptote (DBR, deterministic sigmoid) | $\eta$ | Unif$(0, 1)$ |
| Initial value (all trajectory priors) | $U_0$ | Unif$(0, 1)$ |
| Time of inflection (DBR, deterministic sigmoid) | $\tilde{t}$ | Unif$(1985, 2012)$ |
| Allometric parameters, initial conditions and asymptote (DBR) | $(U_0, \eta, m)$ | 0 if $U_0 \geqslant m^{1/(1-m)}\eta$ |
| Volatility (DBM) | $\sigma$ | Unif$(0, 2.1)$ |

mentioned earlier, there is very little information on the volatility parameter of the deterministic Brownian motion (DBM) formulation. Throughout this paper we used a uniform prior between 0 and 2.1. As explained in more detail in Section 3.1, the parameter of main interest in this study is the quantity $\Delta U = U(2009) - U(2003)$. Simulations suggest that the combination of the DBM hypothesis with a Unif$(0, 1)$ prior for $U_0$ (condom use on January 1st, 1985) results in a symmetric prior on $\Delta U$ that is centred near 0 with 2.5% and 97.5% points at $\pm 0.6$ respectively. We considered it as a reasonably vague prior for $\Delta U$ and evaluated the performance of the resulting model via the simulation experiments of Section 3. More diffuse priors can also be used by setting a larger value for the upper limit of the uniform prior for $\sigma$. The sensitivity of our analysis to these assumptions is explored in Section 4.2. Regarding the parameters of the sigmoid curves, we used vague priors that are also shown in Table 1.

## 2.4.  Computational schemes for implementation

The joint posterior distribution can be calculated up to proportionality through the HIV transmission model of Section 2.1, which links the prevalence observations to the condom use trajectories, the trajectory priors of Section 2.2 and the remaining priors of Section 2.3. For the DBR and deterministic sigmoid trajectory priors, the model can be put in a non-linear regres-

**Table 2.** Algorithm 1: particle smoother algorithm

---

With $J$ being the number of particles and $n$ the number of observations, initialize $L^0(\theta) = 1$, $W_0^j = 1/J$ and
   sample $(U_0^j)_{j=1,\dots,J}$ from $p(U_0|\theta)$:
*for* $i = 0$ to $n - 1$ *do*
   *for* $j = 1$ to $J$ *do*
      sample $(\tilde{U}(t_i, t_i + \delta, \dots, t_{i+1})^j)$ from $p\{U(t_i, t_i + \delta, \dots, t_{i+1})|\theta, U(t_i)^j\}$
      calculate the resulting prevalence $h\{\tilde{X}(t_{i+1})\}^j$ by solving the ordinary differential equation (e.g. with the
         Euler step)
      set $\alpha^j = p[y_{i+1}|h\{\tilde{X}(t_{i+1})\}^j]$
   *end for*
   set $W_{i+1}^j = \alpha^j / \Sigma_{k=1}^J \alpha^k$, and $L^{i+1}(\theta) = L^i(\theta)(1/J)\Sigma\alpha^j$
   resample $(\tilde{U}(t_1, t_1 + \delta, \dots, t_{i+1})^j, \tilde{X}(t_1, t_1 + \delta, \dots, t_{i+1})^j)_{j=1,\dots,N}$ according to $(W_{i+1}^j)$,
*end for*

---

**Table 3.** Algorithm 2: PMCMC algorithm (particle marginal Metropolis–Hastings version)

---

With $M$ being the number of iterations, set the current $\theta$-value $\tilde{\theta}$ to an initial value, use the particle smoother
   according to algorithm 1 to compute $\hat{p}(\mathbf{y}_{1:N}|\tilde{\theta}) = L(\tilde{\theta})$ and sample $\tilde{U}(t_1, t_1 + \delta, \dots, t_n)^{\theta}$ from
   $p\{U(t_1, t_1 + \delta, \dots, t_n)|\mathbf{y}_{1:N}, \tilde{\theta}\}$
*for* $i = 1$ to $M$ *do*
   sample $\tilde{\theta}^*$ from $Q(\tilde{\theta}, \cdot)$
   use the particle smoother to compute $L(\tilde{\theta}^*)$ and sample $\tilde{U}(t_1, t_1 + \delta, \dots, t_n)^{\tilde{\theta}^*}$ from $\hat{p}\{U(t_1, t_1 + \delta, \dots, t_n)$
      $|\mathbf{y}_{1:N}, \tilde{\theta}^*\}$
   set $\tilde{\theta} = \tilde{\theta}^*$ and $\tilde{U}^{\tilde{\theta}}(t_1, t_1 + \delta, \dots, t_n) = \tilde{U}^{\tilde{\theta}^*}(t_1, t_1 + \delta, \dots, t_n)$ with probability $1 \wedge L(\tilde{\theta}^*)Q(\tilde{\theta}^*, \tilde{\theta})/\{L(\tilde{\theta})Q(\tilde{\theta}, \tilde{\theta}^*)\}$
   record $\tilde{\theta}$ and $\tilde{U}^{\tilde{\theta}}(t_1, t_1 + \delta, \dots, t_n)$
*end for*

---

sion framework, with the non-linear function being the solution of the ordinary differential equation and can therefore be implemented with standard software such as WinBUGS through `WBDiff` (Lunn, 2004). However, this approach cannot be extended to the DBM case where more involved techniques are required. Since the posterior probability density function is intractable, a data augmentation scheme can be utilized. This inference problem poses some challenges due to the high dimension of the discretized representation of $U_{0:N}$ and to the strong correlation between $U_{0:N}$ and the vector of constant parameters, $\theta$. This correlation leads to poor mixing and convergence properties, when applying a Gibbs scheme to $\theta$ and $U_{0:N}$. The PMCMC algorithm (see Andrieu *et al.* (2010)) can be used to circumvent these difficulties by updating the two components jointly, thus reducing the problem to a small dimensional MCMC algorithm on $\theta$ that takes advantage of the state space structure of the model. The Metropolis–Hastings step of this algorithm is determined by the estimates of the likelihood $\hat{p}(y|\theta)$ provided by a particle smoother.

The particle smoother and the PMCMC algorithm are described respectively in algorithms 1 and 2 (Tables 2 and 3), in terms of the quantities that were introduced in the previous sections. More details about this algorithm and its practical implementation can be found in Dureau *et al.* (2013) through an application in a similar context, and in Andrieu *et al.* (2010).

Regarding the choice of $Q(\cdot)$ in algorithm 2, we use a random-walk Metropolis–Hastings algorithm in a transformed parameter space (log- or logit) to ensure boundedness constraints. Each iteration of the MCMC algorithm requires an execution of the particle smoother, which

induces substantial computational cost if the importance sampling covariance matrix $\Sigma$ is ill adapted. Adaptive approaches (Roberts and Rosenthal, 2009) can be used to tune $\Sigma$ but they require lengthy explorations of the target space. We propose to speed up this process by pre-exploring a proxy posterior density relying on a Gaussian approximation of the dynamic system and the extended Kalman filter methodology (Dureau *et al.*, 2013). A simple bootstrap version of the particle smoother is used as it is not straightforward to consider data-driven transition proposals given the complex observation regime of our model. Note that, given the short length of the observed time series, simpler alternatives to PMCMC sampling may perform reasonably well. For example, the resampling step can be omitted and the particle smoother output can be used to approximate $p(U_{0:N}|\mathbf{y}_{1:N}, \theta)$. However, in our application, it appears that a substantial number of additional particles are needed, thus not offering a great reduction in the computational cost when compared with PMCMC sampling. We therefore suggest the use of PMCMC sampling as a robust computational tool that still does not require a large amount of computation time in applications of this type. To meet the computational requirements of the ensemble simulations in the present paper, and to facilitate future applications of the methodology that we are proposing, the calculations were made by using the *SSM* inference package (`https://github.com/JDureau/ssm`). This package, targeted to time series analysis via state space models, generates executables for likelihood optimization and Bayesian inference algorithms in parallelizable C for any stochastic compartmental model expressed in a high-level modelling grammar (see Dureau *et al.* (2014) for more information).

## 3. Evaluation methodology based on ensemble simulations

Given the limited amount of information in the data that were available (five prevalence observations, including initial conditions), it is very likely that the posterior output will be influenced substantially by the choice of $U$-priors and their parameters. In this section we explore the performance of the proposed inferential mechanisms via simulation-based experiments designed to mimic the behaviour of data sets that are typically encountered in the context of the motivating application. Clearly, the approach of this paper heavily relies on the HIV infection model and the results will be dependent on its specification. We therefore set up the simulation experiments under the assumption that the model of Section 2.1, parameterized according to the priors of Section 2.3, is correct. The focus is on quantities that are related with the condom use trajectories, under the different choices of Section 2.2, that can be estimated from the samples of the posterior distribution provided by the MCMC algorithms of Section 2.4. We also provide some discussion regarding the static parameters appearing in the condom use trajectory priors.

### 3.1. Parameter of interest

By fitting each of the previously introduced models, we obtain samples from the marginal posterior density $p^M\{U(t)|y\}$ ($M \in \{\text{DBR}, \text{deterministic sigmoid}, \text{DBM}\}$). However, our interest mainly lies in the amplitude of the shift in condom use between 2003 and April 2009, which is henceforth denoted by $\Delta U$. The posterior draws of condom use trajectories can be transformed to provide samples from the posterior of this parameter of interest. The samples can then be used to form an estimator $\hat{\Delta}U^M$ of $\Delta U$ such as the posterior median of $\hat{p}^M(\Delta U|y)$. In what follows we explore the frequentist properties of this estimator, derived from each of the trajectory priors.

It may also be of interest to assess the estimating capabilities, given the limited amount of data, for the hyperparameters of the various $U$-priors ($U_0$, $\eta$, $r$, $m$, $\tilde{t}$ and $\sigma$). As it turns out, there is information for some of them ($U_0$, $\eta$ and $\tilde{t}$), whereas some others are difficult to estimate and are determined mostly by their prior ($r$, $m$ and $\sigma$). Nevertheless, from a subject matter

point of view, interest lies mainly on $\Delta U$, whereas the remaining quantities (in the condom use priors) can be regarded as nuisance parameters. Another appealing feature of $\Delta U$ is that it appears in all models and therefore provides an omnibus quantity for comparison. Hence, inference properties of these parameters ($U_0, \eta, r, m, \tilde{t}$ and $\sigma$) are only studied indirectly through inference properties of $\Delta U$.

### 3.2. Measures of performance

The performance of each estimator $\hat{\Delta} U^M$ in estimating $\Delta U$ is evaluated from the following criteria (where $L = 1000$, the number of simulations):

$$\text{Bias}^M = (1/L) \sum_i (\hat{\Delta} U_i^M - \Delta U_i),$$

$$\text{MSE}^M = (1/L) \sum_i (\hat{\Delta} U_i^M - \Delta U_i)^2,$$

$$\text{Std}^M = \sqrt{\{\text{MSE}^M - (\text{bias}^M)^2\}}.$$

In addition to these quantities, we are also interested in assessing the discriminative ability of each model in detecting increases in condom use. The focus is given to increases in condom use that are at least as high as a prespecified threshold $\Gamma$, which reflects the minimum practical increase. When analysing the data a researcher may decide that condom use did increase more than $\Gamma$ if the value of the estimator $\hat{\Delta} U^M$ is higher than a user-specified threshold $\gamma$. Each decision mechanism may lead to different types of error and is therefore associated with a particular sensitivity and specificity. More specifically we can define the rate of true and false positive results in the following way:

(a)  sensitivity (true positives rate) for $\gamma$, $\sharp(\hat{\Delta} U^M > \gamma, \Delta U > \Gamma)/\sharp(\Delta U > \Gamma)$;
(b)  specificity ($1 -$ false positives rate) for $\gamma$, $\sharp(\hat{\Delta} U^M < \gamma, \Delta U < \Gamma)/\sharp(\Delta U < \Gamma)$.

We proceed by first reporting the sensitivity and specificity corresponding to the case $\gamma = \Gamma$. This corresponds to saying that $\Delta U$ is higher than $\Gamma$ if its estimator is higher than $\Gamma$. We then use a range of $\gamma$s and obtain the sensitivity–specificity pairs that correspond to each of them. A lower detection threshold $\gamma$ will increase the sensitivity of the method, but it also increases the risk for false positive results, and vice versa. These pairs are combined to form the receiver operating characteristics curve by plotting sensitivity *versus* $1 -$ specificity. The area under the receiver operating characteristics curve, AUC, provides an overall measure of discriminatory power as it reflects the probability of correctly classifying a randomly chosen positive instance as higher than a randomly chosen negative one (Fawcett, 2006). For example, an AUC-value of 50% indicates no power (i.e. random choice). This detailed procedure is repeated to assess the ability to detect two different levels of increase in the use of condoms, with $B$ set to 20%, 30% and 40%.

### 3.3. Simulation procedure

The performance of the estimators that are derived from the various models is measured by using a set of simulated experiments where condom use trajectories are sampled from a given growth curve model (deterministic sigmoid), and parameters from $\theta$ are sampled following their prior distributions. To maximize the utility of this test procedure, and to serve future application of this method in different districts targeted by the Avahan programme, only plausible and realistic condom use trajectories are considered: cases with prevalence in 2010 between 2% and 40% and

with condom use shifts that occurred after 1995. Furthermore, the test trajectories have been sampled so that $\Delta U$ regularly spans the [0;0.9] interval. Alternatively, a second set of simulations has been generated based on stepwise condom use trajectories (constant $U$ up to $\tilde{t}$, and then constant $U$ after $\tilde{t}$) to explore the effect of model misspecification and to assess the sensitivity of results to the underlying model used for simulations (deterministic sigmoid).

For each of these experiments, an epidemic is simulated to provide observations $y_i^s$ replicating the observation framework applied in Mysore: three prevalence estimates among FSWs and one among clients, concentrated during the period of the intervention. From these observations, the MCMC algorithm is applied to each method $M$ to sample from $p\{U^M(t)|y_{1:N}^s\}$. Then, given the posterior $U$-samples, the estimators $\hat{\Delta}U^M$ can be computed and compared with their true counterparts $\Delta U$ through the measures of performance that were introduced in the previous Section 3.2.

## 4. Results

### 4.1. Comparison of the condom use trajectory models from ensemble simulations

This section contains the results of the simulation experiments that were conducted as described in Section 3.3. In all cases each model was fitted to each of the 1000 simulated data sets and the estimator is formed via the posterior median at the points of interest. These samples are used to assess the bias, standard deviation and mean-squared error MSE of these estimators. Moreover, in Table 4, we examine the ability of each model to classify the simulated instances of the $\Delta U$-parameter, and we assess the risk of overstating *versus* understating. In other words, we assess the ability of model-driven estimators to address questions such as 'was the shift in condom use during the intervention over 0.2, 0.3 or 0.4?', by looking at their sensitivity and specificity as well as the resulting AUC.

We begin with results corresponding to cases where the data were simulated from the deterministic sigmoid model. Table 4 reports bias, standard deviation and MSE for each model-driven estimator. It suggests that all models underestimate $\Delta U$. More precisely, the DBR model tends to understate strongly the shift in amplitude, by 0.33 on average. The bias of the deterministic sigmoid model is only slightly smaller ($-0.31$), and optimal results are obtained with the Brownian motion model ($-0.26$). Similarly, in terms of MSE, the performance of the DBM model is better. Fig. 2 plots the bias against the true value of $\Delta U$ for each model and reveals an increasing association. Note that the ranking of the models is consistent across the entire range of $\Delta U$-values. If, for example, the true shift is between 40% and 50%, it is on average underestimated by 0.21 with the best method (DBM) and more than 0.30 points with the DBR and deterministic sigmoid methods.

**Table 4.** Frequentist properties of the various estimators of the amplitude of the shift in condom use during the intervention, estimated from 1000 simulations from the deterministic empirical sigmoid model

| Statistic | Results for the DBR model | Results for the deterministic sigmoid model | Results for the DBM model |
|---|---|---|---|
| Bias | $-0.33$ | $-0.31$ | $-0.26$ |
| Error standard deviation | 0.21 | 0.20 | 0.19 |
| MSE | 0.15 | 0.14 | 0.10 |

**Fig. 2.**    Bias of each model as a function of the true amplitude of the shift in use of condoms, estimated from 1000 simulations: ■, DPR model; ■, deterministic sigmoid model; ■, DBM model

**Table 5.**    General distinctive power (AUC) of the median estimator of the shift, and specific sensitivity and specificity when answering the question is the shift in condom use during the intervention stronger than 0.2, than 0.3 and than 0.4?†

| Question | Parameter | Results for the DBR model | Results for the deterministic sigmoid model | Results for the DBM model |
|---|---|---|---|---|
| $\Delta U > 0.2$? | AUC | 0.93 | 0.93 | 0.92 |
|  | Sensitivity | 26% | 29% | 53% |
|  | Specificity | 100% | 100% | 100% |
| $\Delta U > 0.3$? | AUC | 0.92 | 0.92 | 0.89 |
|  | Sensitivity | 5% | 14% | 29% |
|  | Specificity | 100% | 100% | 100% |
| $\Delta U > 0.4$? | AUC | 0.90 | 0.89 | 0.87 |
|  | Sensitivity | 0% | 7% | 14% |
|  | Specificity | 100% | 100% | 100% |

†These quantities were estimated from 1000 simulations from the deterministic empirical sigmoid model.

The bias in estimating $\Delta U$ under each of the $U$-priors can be attributed to a large extent to the prior that is implied by each formulation on $\Delta U$. As mentioned in Section 2.3 the DBM approach results in a symmetric prior on $\Delta U$ that is centred near 0 with 2.5% and 97.5% points at $\pm 0.6$ respectively. The posterior median is therefore pulled towards 0, thus resulting in conservative estimates. The amount of shrinkage depends on the upper limit of the uniform prior on $\sigma$. The corresponding priors under the DBR and deterministic sigmoid formulations result in priors for $\Delta U$ that put more mass around zero, which heavily depends on the priors that are placed on $k$ and $m$. It would perhaps be interesting in the future to consider alternative priors for $k$ and $m$ to reduce the bias in $\Delta U$ although it is not clear whether this should be the main optimization criterion. Regarding the DBM model, the uniform prior assigned on $\sigma$ appears to work reasonably well.

Table 5 suggests that all estimators based on the median of the posterior density of $p(\Delta U | \mathbf{y}_{1:N})$ have good distinguishing power: the AUC is between 0.87 and 0.93 in all cases. In line with the
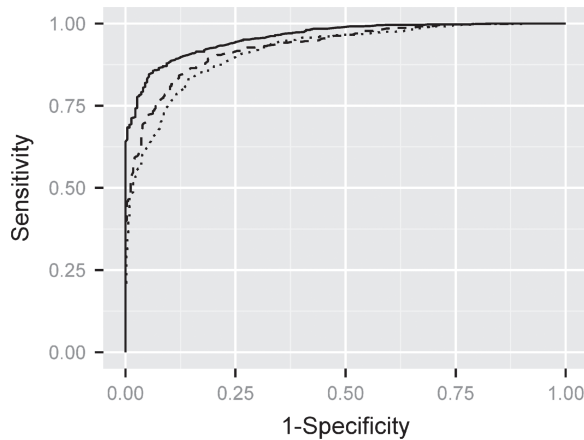
**Fig. 3.** Receiver operating characteristics curve when testing for $\Delta U > 0.2$ ( ——— ), $\Delta U > 0.3$ (– – –) and $\Delta U > 0.4$ (· · · · · ·), under a Brownian motion trajectory prior: the curves were estimated from 1000 simulations; very similar shapes are obtained for the alternative trajectory priors

**Table 6.** Frequentist properties of the various estimators of the amplitude of the shift in condom use during the intervention, estimated from 1000 simulations from stepwise condom use trajectories (constant $U$ until $\bar{t}$, and constant $U$ after $\bar{t}$)

| Statistic | Results for the DBR model | Results for the deterministic sigmoid model | Results for the DBM model |
|---|---|---|---|
| Bias | −0.35 | −0.33 | −0.26 |
| Error standard deviation | 0.21 | 0.20 | 0.18 |
| MSE | 0.16 | 0.15 | 0.10 |

results of Table 5, the estimates provided by the DBM model achieve better sensitivity (53%, 29% and 14%) than the other models (between 0% and 29%). The strong negative bias that is observed in Table 5 is also reflected by an optimal 100% specificity of the estimators. The performance, in terms of sensitivity, decreases as the level of $\Delta U$ increases. Fig. 3 also depicts the ROC curve for the DBM model, obtained as described in Section 3.3.

In an additional simulation experiment we examine the frequentist properties of the models in the case where underlying condom use trajectories were generated from a step function. As shown in Table 6, a very similar picture is obtained with the DBM model having the smallest bias and MSE, being slightly better than the DBR model.

Overall, the results that are presented in Tables 4–6 provide an informative qualitative assessment of the ability of the various models to capture $\Delta U$ from limited prevalence data on an important and diverse set of likely scenarios (1000 experiments with true $\Delta U$ spanning between 0 and 0.9). First of all, MSE- and AUC-figures suggest that, although the number of prevalence observations is low and some elements of the transmission process are uncertain, it is still possible to extract information on our time varying parameter and to provide estimates of $\Delta U$ during the intervention. Furthermore, there seems to be a possibility of controlling the risk of overstating these quantities by analysing the outputs of the three models, that offer different
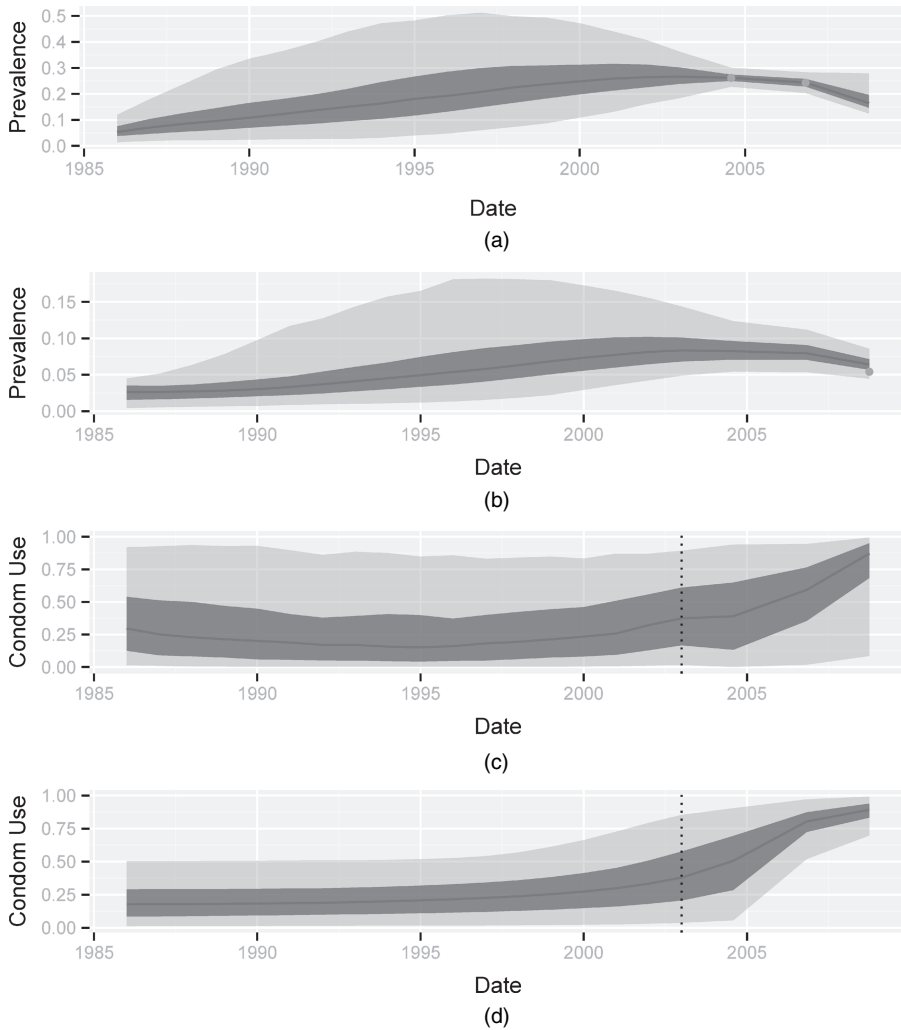
**Fig. 4.**    Estimates obtained for the Mysore district: (a) reconstructed prevalence trajectory among FSWs when condom use is modelled with Brownian motion; (b) reconstructed prevalence trajectory among clients when condom use is modelled with Brownian motion; (c) reconstructed condom use trajectory when modelled with Brownian motion; (d) reconstructed condom use trajectory when modelled with the deterministic sigmoid model

levels of compromise between sensitivity and specificity. Thus, although the procedure may fail to identify some shifts in use of condoms, there seems to be some reassurance that once a shift has been detected it will most probably be true.

Finally, note that the two models with the higher overall performance, the DBM and deterministic sigmoid model, are quite different in nature: the deterministic sigmoid trajectories are smooth, whereas under the Brownian motion prior they are non-differentiable. Hence, the choice between the two models can also be based on prior beliefs of the researcher regarding the smoothness of the condom use trajectories.

### 4.2.    Application: what can we infer on the trajectory of condom use in Mysore?
Mysore is one of the districts that were targeted by the Avahan intervention, and this programme

**Table 7.**  Estimates of the change in condom use in Mysore between 2003 and 2009

| Model | Posterior mean | Posterior median | 95% credible interval |
|---|---|---|---|
| DBR | 0.30 | 0.29 | [0.09; 0.54] |
| Deterministic sigmoid | 0.48 | 0.49 | [0.09; 0.87] |
| DBM | 0.52 | 0.54 | [0.06; 0.95] |

was the first HIV prevention intervention in this region. Four HIV prevalence estimates have been obtained between 2003 and 2009: three among FSWs, and one among clients. Results from the inference procedure using a Brownian motion model are shown in Fig. 4, suggesting a strong effect of the intervention. The purpose of this paper was to determine what level of increase of condom use between 2003 and 2009 can be inferred while controlling the risk of overstating it. As was shown in Section 4.1, deterministic sigmoid models could provide a good alternative to the DBM formulation. Hence, we also present here results obtained with this model for the Mysore data set (see Fig. 4). Note that the posterior uncertainty around the condom use trajectory is lower in the case of the deterministic sigmoid model compared with the DBM model. This is in line with our expectations given the fact that this prior imposes a particular growth structure and therefore introduces additional information. Loosely speaking this may be viewed as a bias–variance trade-off. In general, the choice of each $U$-formulation is likely to affect the inference procedure and therefore it is useful to explore its implications, e.g. via the simulation experiments of the previous section. Table 7 shows the estimates of $\Delta U$ for each of the three models presented. The results indicate a positive increase in all cases. In particular, the posterior medians for the DBM and deterministic sigmoid models are 0.54 and 0.49 whereas the 95% credible intervals are [0.06;0.95] and [0.09;0.97] respectively.

A stronger conclusion regarding a lower bound for the condom use shift between 2003 and 2009 can be made by comparing the posteriors medians with the results of Table 7. If the underlying set of simulations is to be considered realistic, an argument in favour of an increase in the use of condoms being at least 0.4 can be made. Since the posterior medians are more than 0.4 under both DBM and deterministic sigmoid models (0.54 and 0.49 respectively), Table 7 suggests that a statement for $\Delta U > 0.4$ will be correct with probability given by the specificity of each model (100% in both cases). Although more informative than the credible intervals that are obtained directly from the posterior densities (over 0.06), these numbers are heavily dependent on the assumption that the simulations of Section 3 provided an adequate approximation of the reality and should therefore be treated with caution. Fig. 4 and Table 7 show that the results that are obtained from the deterministic sigmoid and Brownian motion models strongly coincide: they suggest that condom use was stable over the period 1985–2003, remaining below 0.5, sharply increased between 2003 and 2007, and kept growing until 2009.

Finally, Table 8 summarizes the results for sensitivity analysis regarding the choice of priors for the parameters $\sigma$ (DBM model) and $\tilde{t}$ (DBR and deterministic sigmoid models). As we can see from the results of Table 8, the posterior summaries are not altered substantially for a number of prior formulations, suggesting a reasonable amount of robustness to their specification.

In summary, the results of this section support the findings of other independent modelling (Boily *et al.*, 2008; Pickles *et al.*, 2010) and empirical epidemiological studies (Lowndes *et al.*, 2010; Deering *et al.*, 2008; Bradley *et al.*, 2010) that suggested that the use of condoms has

**Table 8.**  Sensitivity analysis: effect on the prior on $\sigma$ and $\tilde{t}$ on the estimated changes in condom use

| Model | Parameter | Prior density | Posterior mean | Posterior median | 95% credible interval |
|---|---|---|---|---|---|
| DBM | $\sigma$ | Unif(0, 2.7) | 0.578 | 0.613 | [0.048;0.975] |
| | | Unif(0, 2.4) | 0.509 | 0.525 | [0.029;0.964] |
| | | Unif(0, 2.1) | 0.516 | 0.529 | [0.045;0.961] |
| | | Unif(0, 1.8) | 0.463 | 0.448 | [0.059;0.918] |
| | | Unif(0, 1.5) | 0.465 | 0.461 | [0.035;0.874] |
| DBR | $\tilde{t}$ | Unif(1985, 2012) | 0.309 | 0.299 | [0.084;0.589] |
| | | Unif(1985, 2009) | 0.307 | 0.294 | [0.081;0.593] |
| | | Unif(1988, 2012) | 0.309 | 0.299 | [0.084;0.589] |
| | | Unif(1982, 2012) | 0.309 | 0.299 | [0.085;0.589] |
| Deterministic sigmoid | $\tilde{t}$ | Unif(1985, 2012) | 0.471 | 0.487 | [0.054;0.903] |
| | | Unif(1985, 2009) | 0.462 | 0.474 | [0.057;0.903] |
| | | Unif(1988, 2012) | 0.472 | 0.487 | [0.054;0.903] |
| | | Unif(1982, 2012) | 0.472 | 0.487 | [0.054;0.903] |

increased substantially in Mysore following the start of the Avahan programme. However, all these other studies have been based on approaches that strongly rely on self-reported condom use data which can be biased. Our method here has the advantage of not imposing any strong prior on the presence and timing of consequences of the intervention. Together, this new and previous analysis support recent analysis of the effect of the Avahan programme in reducing HIV infection among FSWs and clients in Mysore (Pickles *et al.*, 2010).

## 5.  Discussion

In this paper, we presented a Bayesian approach to draw conclusions regarding the evolution of time varying behavioural parameters in the context of HIV such as the use of condoms among FSWs. Inference can be based on prevalence estimates and a substantial amount of information from additional sources can be incorporated via prior distributions. To describe the behaviour of condom use trajectories we introduced three different formulations based on Brownian motion and growth curves such as the generalized Bertalanffy–Richards and empirical sigmoid models. To our knowledge, these formulations are new in this context. The computational framework presented allows estimation of condom use trajectories as well as functionals thereof, using advanced MCMC methods and following ideas of Dureau *et al.* (2013), Cazelles and Chau (1997) and Rasmussen *et al.* (2011). Nevertheless, in comparison with these approaches, the problem of evaluating the Avahan intervention by estimating its effect on the use of condoms from prevalence estimates is of additional difficulty due to the limited amount of information; the application to Mysore district was based on three observations of prevalence among FSWs and one among clients, plus a hypothesis on the initial value of prevalence in 1985. Various simulation experiments were conducted to assess the validity of the procedure, examining the frequentist properties of the underlying estimators and the ability of the model to avoid overestimation via conducting receiver operating characteristics analysis. The evidence from the simulation experiments is encouraging, suggesting that the approach can be used in this context for making conservative estimates of changes in the use of condoms with both the Brownian motion and the deterministic sigmoid trajectory priors. However, the overall performance is bound to depend

on the deterministic HIV infection model which was parameterized on the basis of a substantial amount of prior information, as in Pickles *et al.* (2010), as well as on assumptions such as the very low HIV prevalence in 1985. Most of the prior information utilized in this study was obtained from additional data sources (individual biobehavioural surveys). In the presence of all these sources of data, it would be interesting to consider and contrast a joint inferential scheme through an evidence synthesis framework in the spirit of Goubar *et al.* (2008) and Presanis *et al.* (2011).

Although the representation of the natural history of HIV transmission in this paper is simpler in behavioural terms in comparison with the model that was presented in Pickles *et al.* (2010), the model adopted retains the heterogeneity in FSW behaviour and it is enriched by exploring the condom use trajectories space rather than working with predetermined scenarios. Nevertheless, there are reasons for a potential overestimation of the shift amplitude in this simpler model as coinfections with other sexually transmitted diseases were ignored (although higher transmission probabilities per unprotected act were allowed to compensate for them), and no acute phase was considered. However, diffusion-driven models aim at capturing and compensating for structural misspecifications while capturing the main dynamics of the system and have been shown here to provide conservative estimates. Overall it may be viewed as a different and complementary choice in the trade-off between richness and tractability of the model compared with Pickles *et al.* (2010, 2011). Lastly, this approach relies on the hypothesis that changes in transmission probabilities are solely related to changes in the use of condoms, ignoring for example potential changes in the frequency of commercial sex partnerships. This choice can be motivated by the strong focus of the Avahan intervention on prevention measures and the relative stability in the frequency of commercial sex that are exhibited by the series of cross-sectional biobehavioural surveys that were conducted during the period of the intervention.

From a wider perspective, the present paper provides a generic methodology for applications of diffusion processes coupled to mechanistic epidemiological models in an evaluation monitoring framework. Naturally, this methodology can be used to study the effect of the Avahan intervention in the other regions of India that were targeted by the programme. In a different context, the approach that was presented in this paper has been applied in the context of the 2014 Ebola virus outbreak to monitor in realtime the effect of the interventions conducted in Liberia, Sierra Leone and Guinea (Camacho *et al.*, 2014). Lastly, the applicability of the methodology presented and more generally of the use of diffusion models to capture the time varying drivers of epidemics will be facilitated by the development of a generic inference toolbox for compartmental models (Dureau *et al.*, 2014), motivated in part by the computational challenges that were encountered in the present paper. Examples of such applications are the study of the natural history of dengue or foot-and-mouth disease.

## References

Andrieu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc.* B, **72**, 269–342.

Bill and Melinda Gates Foundation (2008) Avahan—the India AIDS initiative: the business of HIV prevention at scale. Bill and Melinda Gates Foundation, Seattle. (Available from `https://docs.gates foundation.org/Documents/avahan_hivprevention.pdf`.)

Boily, M., Abu-Raddad, L., Desai, K., Masse, B., Self, S. and Anderson, R. (2008) Measuring the public-health impact of candidate HIV vaccines as part of the licensing process. *Lancet Infect. Dis.*, **8**, 200–207.

Boily, M., Lowndes, C. and Alary, M. (2002) The impact of hiv epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sexly Transmttd Infectns*, **78**, suppl. 1, i78–i90.

Boily, M.-C., Lowndes, C. M., Vickerman, P., Kumaranayake, L., Blanchard, J., Moses, S., Ramesh, B. M., Pickles, M., Watts, C., Washington, R., Reza-Paul, S., Labbe, A. C., Anderson, R. M., Deering, K. N. and Alary, M. (2007) Evaluating large-scale HIV prevention interventions: study design for an integrated mathematical modelling approach. *Sexly Transmttd Infectns*, **83**, 582–589.

Boily, M.-C., Pickles, M., Lowndes, C. M., Ramesh, B. M., Washington, R., Moses, S., Deering, K. N., Mitchell, K. M., Reza-Paul, S., Blanchard, J., Vassall, A., Alary, M. and Vickerman, P. (2013) Positive impact of a large-scale hiv prevention programme among female sex workers and clients in south India. *AIDS*, **27**, 1449–1460.

Bradley, J., Moses, S., Blanchard, J., Rajaram, S., Ramesh, B., Verma, S. and Alary, M. (2010) Assessing reported condom use among female sex workers in southern India through examination of condom availability. *Sexly Transmttd Infectns*, **86**, suppl. 1, i44–i48.

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., Pollington, T., Carney, J. R., Glover, R., Smout, R. E.,Tiffany, A., Edmunds, W. J. and Funk, S. (2014) Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLOS Curr.*, **7**.

Cazelles, B. and Chau, N. (1997) Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic. *Math. Biosci.*, **140**, 131–154.

Cori, A., Boëlle, P., Thomas, G., Leung, G. and Valleron, A. (2009) Temporal variability and social heterogeneity in disease transmission: the case of SARS in Hong Kong. *PLOS Computnl Biol.*, **5**, no. 8, article e1000471.

De Angelis, D., Gilks, W. R. and Day, N. E. (1998) Bayesian projection of the acquired immune deficiency syndrome epidemic (with discussion). *Appl. Statist.*, **47**, 449–498.

Deering, K., Vickerman, P., Moses, S., Ramesh, B., Blanchard, J. and Boily, M. (2008) The impact of out-migrants and out-migration on the HIV/AIDS epidemic: a case study from South-West India. *AIDS*, **22**, S165–S181.

Dureau, J., Ballesteros, S. and Bogich, T. (2014) SSM: inference for time series analysis with State Space Models. *Preprint arXiv:1307.5626v4.*

Dureau, J., Kalogeropoulos, K. and Baguelin, M. (2013) Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*, **14**, 541–555.

Garcia, O. (1983) A stochastic differential equation model for the height growth of forest stands. *Biometrics*, **39**, 1059–1072.

Goubar, A., Ades, A. E., De Angelis, D., McGarrigle, C. A., Mercer, C. H., Tookey, P. A., Fenton, K. and Gill, O. N. (2008) Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data (with discussion). *J. R. Statist. Soc.* A, **171**, 541–580.

Hanck, S., Blankenship, K., Irwin, K., West, B. and Kershaw, T. (2008) Assessment of self-reported sexual behavior and condom use among female sex workers in India using a polling box approach: a preliminary report. *Sexly Transmttd Dis.*, **35**, 489–494.

Kumar, R., Mehendale, S. M., Panda, S., Venkatesh, S., Lakshmi, P. V. M., Kaur, M., Prinja, S., Singh, T., Virdi, N. K., Bahuguna, P., Sharma, A. K., Singh, S., Godbole, S. V., Risbud, A., Manna, B., Thirumugal, V., Roy, T., Sogarwal, R. and Pawar, N. D. (2011) Impact of targeted interventions on heterosexual transmission of HIV in India. *BMC Publ. Hlth*, **11**, no. 1, 549.

Lei, Y. and Zhang, S. (2004) Features and partial derivatives of Bertalanffy-Richards growth model in forestry. *Nonlin. Anal. Modllng Control*, **9**, 65–73.

Lessne, G. and Hanumara, R. (1988) Growth curve analysis in marketing: method and application. *J. Marktng Res.*, **25**, 391–396.

Lowndes, C., Alary, M., Verma, S., Demers, E., Bradley, J., Jayachandran, A., Ramesh, B., Moses, S., Adhikary, R. and Mainkar, M. (2010) Assessment of intervention outcome in the absence of baseline data: reconstruction of condom use time trends using retrospective analysis of survey data. *Sexly Transmttd Infectns*, **86**, suppl. 1, i49–i55.

Lunn, D. (2004) *WinBUGS Differential Interface—Worked Examples*. London: Imperial College School of Medicine.

Omran, A. (1971) The epidemiologic transition: a theory of the epidemiology of population change. *Milbnk Mem. Fund Q.*, **49**, 509–538.

Pickles, M., Boily, M.-C., Vickerman, P., Lowndes, C. M., Moses, S., Blanchard, J. F., Deering, K. N., Bradley, J., Ramesh, B. M., Washington, R., Adhikary, R., Mainkar, M., Paranjape, S. and Alary, M. (2013) Assessment

of the population-level effectiveness of the Avahan HIV-prevention programme in South India: a preplanned, causal-pathway-based modelling analysis. *Lancet Globl Hlth*, **1**, no. 5, e289–e299.

Pickles, M., Foss, A. M., Vickerman, P., Deering, K., Verma, S., Demers, E., Washington, R., Ramesh, B. M., Moses, S., Blanchard, J., Lowndes, C. M., Alary, M., Reza-Paul, S. and Boily, M.-C. (2010) Interim modelling analysis to validate reported increases in condom use and assess HIV infections averted among female sex workers and clients in southern India following a targeted HIV prevention programme. *Br. Med. J.*, **86**, suppl. 1, i33.

Presanis, A., De Angelis, D., Goubar, A., Gill, O. and Ades, A. (2011) Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men. *Biostatistics*, **12**, 666–681.

Ramesh, B., Moses, S., Washington, R., Isac, S., Mohapatra, B., Mahagaonkar, S., Adhikary, R., Brahmam, G. N. Vf., Paranjape, R. Sg., Subramanian, T. and Blanchard, J. (2008) Determinants of HIV prevalence among female sex workers in four South Indian states: analysis of cross-sectional surveys in twenty-three districts. *AIDS*, **22**, S35.

Rasmussen, D., Ratmann, O. and Koelle, K. (2011) Inference for nonlinear epidemiological models using genealogies and time series. *PLOS Computnl Biol.*, **7**, no. 8, article e1002136.

Richards, F. (1959) A flexible growth function for empirical use. *J. Exptl Bot.*, **10**, 290–301.

Roberts, G. and Rosenthal, J. (2009) Examples of adaptive mcmc. *J. Computnl Graph. Statist.*, **18**, 349–367.

Turner, C. and Miller, H. (1997) Zenilman's anomaly reconsidered: fallible reports, ceteris paribus, and other hypotheses. *Sexly Transmttd Dis.*, **24**, no. 9, 522–527.

UNAIDS and World Health Organization (2009) *AIDS Epidemic Update*. Geneva: World Health Organization Regional Office Europe.

Vickerman, P., Foss, A.,Pickles, M., Deering, K., Verma, S., Demers, E., Lowndes, C., Moses, S., Alary, M. and Boily, M. (2010) To what extent is the HIV epidemic in southern India driven by commercial sex?: a modelling analysis. *AIDS*, **24**, no. 16, 2563–2572.

Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

Yuancai, L., Marques, C. and Macedo, F. (1997) Comparison of Schnute's and Bertalanffy-Richards' growth functions. *For. Ecol. Mangmnt*, **96**, 283–288.

Zenilman, J. M., Weisman, C. S., Rompalo, A. M., Ellish, N., Upchurch, D. M., Hook 3rd, E. W. and Celentano, D. (1995) Condom use to prevent incident stds: the validity of self-reported condom use. *Sexly Transmttd Dis.*, **22**, no. 1, 15.

Zwietering, M., Jongenburger, I., Rombouts, F. and Van't Riet, K. (1990) Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.*, **56**, 1875–1881.