

# Evaluating large-scale HIV prevention interventions: study design for an integrated mathematical modelling approach

M-C Boily, C M Lowndes, P Vickerman, L Kumaranayake, J Blanchard, S Moses, B M Ramesh, M Pickles, C Watts, R Washington, S Reza-Paul, A C Labbe, R M Anderson, K N Deering, M Alary, on behalf of the CHARME-India team

*Sex Transm Infect* 2007;**83**:582–589. doi: 10.1136/sti.2007.027516



Supplementary tables and figures can be found on the website <http://sti.bmj.com/>

See end of article for authors' affiliations

Correspondence to:  
Dr Marie-Claude Boily,  
Department of Infectious  
Disease Epidemiology,  
Faculty of Medicine, Imperial  
College, Norfolk Place, W2  
1PG, London, UK; mc.  
boily@ic.ac.uk

Accepted 12 October 2007

**Background:** There is an urgent need to evaluate HIV prevention interventions, thereby improving our understanding of what works, under what circumstances and what is cost effective.

**Objectives:** To describe an integrated mathematical evaluation framework designed to assess the population-level impact of large-scale HIV interventions and applied in the context of Avahan, the Indian AIDS Initiative, in southern India. The Avahan Initiative is a large-scale HIV prevention intervention, funded by the Bill & Melinda Gates Foundation, which targets high-risk groups in selected districts of the six states most affected by the HIV/AIDS epidemic (Maharashtra, Karnataka, Tamil Nadu, Andhra Pradesh, Nagaland and Manipur) and along the national highways.

**Methods:** One important component of the monitoring and evaluation of Avahan relies on an integrated mathematical framework that combines empirical biological and behavioural data from different subpopulations in the intervention areas, with the use of tailor-made transmission dynamics models embedded within a Bayesian framework.

**Results:** An overview of the Avahan Initiative and the objectives of the monitoring and evaluation of the intervention is given. The rationale for choosing this evaluation design compared with other possible designs is presented, and the different components of the evaluation framework are described and its advantages and challenges are discussed, with illustrated examples.

**Conclusions:** This is the first time such an approach has been applied on such a large scale. Lessons learnt from the CHARME project could help in the design of future evaluations of large-scale interventions in other settings, whereas the results of the evaluation will be of programmatic and public health relevance.

International agencies have committed significant resources to implement large-scale HIV/AIDS interventions. Only a small fraction of the resources is, however, allocated to the evaluation of these interventions.<sup>1,2</sup> Understandably, one dilemma is whether resources are best invested to evaluate or implement interventions that “should” work. Nevertheless, in order not to jeopardise resources on a large scale, there is an urgent need to evaluate HIV prevention interventions objectively to obtain a better understanding of what works, when and how.

This paper describes an integrated mathematical evaluation framework, designed to assess the population-level impact of large-scale HIV interventions, applied in the context of Avahan, the Indian AIDS Initiative, in southern India. We first give an overview of Avahan, the objectives of the evaluation and the rationale for choosing this design. We describe the different components of the evaluation framework and discuss its advantages and challenges.

## THE INDIAN AIDS INITIATIVE, AVAHAN

The HIV epidemic in India is highly heterogeneous. HIV prevalence is highest in two northern states (Nagaland and Manipur), where transmission is mainly via injection drug use, and in four southern states (Maharashtra, Karnataka, Tamil Nadu and Andhra Pradesh), where transmission is mainly sexual and concentrated among high-risk groups.<sup>3–5</sup>

The Avahan Initiative is based on the theory of core group transmission and on the assumption that India's HIV epidemic depends on transmission from and within high-risk groups.<sup>6–9</sup> It aims to reduce HIV incidence and prevalence in high-risk

groups in order to limit HIV transmission to and within the general population. Avahan is a large-scale, US\$258 million, HIV prevention intervention funded by the Bill & Melinda Gates Foundation. The intervention targets high-risk groups in 83 selected districts of the six states most affected by the HIV/AIDS epidemic (above), and along the national highways.<sup>7</sup> In approximately 53 of these districts, Avahan and its partners are the sole or major implementers of prevention services for high-risk populations. In the southern states, the intervention focuses on male and female commercial sex workers (MFSW), men who have sex with men (MSM), and clients of MFSW. In the northeast states, the focus is mainly on injection drug users. The different components of the Avahan intervention in the southern states focus on improving the availability and quality of services for sexually transmitted infections (STI), unlimited free distribution and promotion of condoms to MFSW and MSM, expansion of retail outlets for condom social marketing, encouraging behaviour change and reducing vulnerability.<sup>7</sup> The Avahan programme, which started in 2004, attempts to take core group interventions to scale by providing services to more than 80% of known high-risk groups in the selected districts, within a relatively short time-frame.<sup>7</sup>

**Abbreviations:** C-RCT, community randomised controlled trial; CrI, credibility interval; FSW, female sex worker; GPS, general population survey; IBBA, integrated behavioural and biological assessment; M&E, monitoring and evaluation; MFSW, male and female commercial sex worker; MIS, management information system; MSM, men who have sex with men; SBS, special behavioural survey; STI, sexually transmitted infection

## Objectives of the evaluation

The overarching evaluation objective is to assess the population-level impact of a large-scale targeted intervention in a concentrated epidemic. The specific objectives are: (1) to obtain a better understanding of the local HIV transmission dynamics and the factors that determine the intervention's impact; (2) to assess the effectiveness of the intervention in high-risk groups and the general population; and (3) to estimate the cost and cost effectiveness of the intervention and its different components (namely STI and condom components) in different districts with heterogeneous sociodemographic and epidemiological contexts.

One important component of the overall monitoring and evaluation (M&E) of Avahan relies on an integrated mathematical framework, which combines serial cross-sectional biological and behavioural data from different subpopulations in the intervention areas, with the use of tailor-made transmission dynamics models embedded within a Bayesian framework. This component of the evaluation, described here, covers only the four southern states where Avahan is implemented, and represents a small fraction of the resources allocated to the intervention.<sup>7</sup>

## RATIONALE

The rationale for incorporating modelling as part of the M&E is based on the following considerations.

### Factors considered

Despite strong theoretical evidence<sup>6, 8, 9</sup> and the existence of many intervention projects targeted at high-risk groups in different countries, there is limited empirical evidence of the effectiveness and cost effectiveness of core group interventions.<sup>10-12</sup> There is thus a broad interest in understanding the extent to which the Avahan core group intervention will impact on HIV transmission in high-risk groups and the general population. The evaluation needs to be large scale and to assess the intervention impact at the population level (ie both direct and indirect effects). Impact depends on intervention efficacy, as well as on the interaction between coverage and intensity. The magnitude of impact and how fast a given change in the primary outcomes of interest (eg prevalence, incidence) can be achieved depend on the natural history of HIV/AIDS and on the epidemiological context. As a result of the long incubation period of HIV/AIDS, the full potential of an intervention on prevalence may take decades before it is achieved, especially in established/mature epidemics (fig 1).

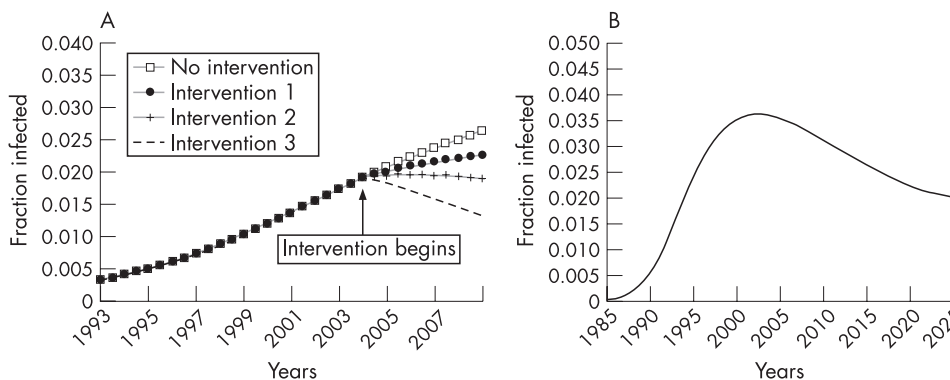
Given the goals of the evaluation, the strength of evidence needed for decision making (adequacy, plausibility or probability as defined in table 1)<sup>13</sup> and the relevant indicators (provision, utilisation, coverage, or health impact) must be specified. Given the scale and importance of Avahan, its evaluation should aim to demonstrate strong plausibility.

The external validity of the evaluation results for a wide range of contexts is particularly important in a setting such as India, with heterogeneous sociodemographic and HIV epidemiological contexts. An adequate design needs to assess the impact of the intervention on HIV trends objectively and to minimise subjective interpretation and evaluator biases. Impact assessment should not delay the scaling up of the intervention and should be logistically, programmatically and ethically feasible and affordable.<sup>1, 2</sup>

## Possible evaluation designs

Theoretically, community randomised controlled trials (C-RCT) are the gold standard for evaluating the population-level effectiveness of an intervention and making probability statements about impact (table 1).<sup>13-16</sup> In practice, they are very expensive and logistically difficult to conduct. This is especially true in the Avahan context, because of the heterogeneity of the Indian epidemic and the broad geographical scope of the Avahan programme. It would also be difficult to establish control districts in this context, as pre-existing programming and migration of high-risk groups would probably be significant sources of contamination between intervention and non-intervention communities. C-RCT, even stepped-wedge design,<sup>15, 16</sup> may be deemed unethical (table 1). The external validity of C-RCT is often limited to the location and population where it is conducted. To date, C-RCT of HIV interventions have been of short duration, included few communities (approximately 10) and produced ambiguous results.<sup>17-20</sup> Mathematical modelling has been used to help interpret inconclusive or contradictory trial results.<sup>19, 20</sup>

The evaluation of large-scale interventions is often based on data from surveillance<sup>21</sup> of relevant process and/or health indicators in target populations over time, to assess the performance and coverage of the programme, as well as levels and spread of the health indicator.<sup>1, 2, 13</sup> Surveillance data alone, however, do not permit researchers to assess the extent of an intervention's effectiveness objectively. Changes in the desired direction (eg decline in prevalence) may also be caused by the natural transmission dynamics of HIV/STI epidemics, or other external influences (eg other programmes; fig 1).<sup>6, 19</sup>



**Figure 1** (A) Theoretical setting A: Simulated concentrated HIV epidemic with and without intervention based on different frequency of condom use (1: 25%, 2: 50%, 3: 100%). The figure clearly illustrates the slow changes in prevalence over time after the intervention and that HIV prevalence can continue to increase over time, albeit more slowly, in the presence of an intervention that is effective (ie lower prevalence) when compared with the non-intervention control group. This clearly highlights how models can be used to interpret trends, by taking the natural dynamics of the epidemic into account. Using trends alone would be misleading because the prevalence carries on increasing for a few years. (B) Theoretical setting B: HIV epidemic in the general population in the absence of intervention. HIV prevalence declined after 2000 as a result of saturation and the depletion of high-risk groups caused by AIDS-related mortality.

**Table 1** Summary of important factors to consider when choosing an evaluation design (details in supplementary table S1, available on the *STI* website: <http://sti.bmj.com/>)

Strength of evidence <sup>*13</sup>	Design category	Description	Comments
<b>Was the intervention effective? → Strong causality statement</b>			
Probability: <sup>13-15</sup> Demonstrate, with a high degree of certainty, if the intervention was a causal determinant in the improvement of the primary indicators	Experimental design: Community-based randomised controlled trials <sup>14</sup>	Parallel design: Communities are randomised and allocated at the start of the trial between intervention and control arms Stepped-wedge design: <sup>15 16</sup> Each community receives the control and the intervention sequentially, at randomly allocated time points during the trial	Empirical estimates of incidence needed  High rates of loss to follow-up among high-risk cohorts, especially with long follow-up Large cohorts needed to measure differences in incidence in the general population Intervention less likely to be “real world” May be unethical as it delays the roll-out of the intervention to the control group May still be unethical if it (stepped-wedge design) increases the trial duration and slows the scale-up of the intervention
<b>Did the programme seem to have an impact? → Medium to weak causality statement</b>			
Plausibility: <sup>13 21 22</sup> Demonstrate, with a certain level of uncertainty, whether the programme may have had an effect above and beyond other external influences	Quasi-experimental design: Non-randomised valid control group to assess what might have happened in absence of the intervention	Internal control group: Population at baseline (pretest–posttest type design) <sup>21 22</sup> External control group: From areas where the programme has not been implemented Multiple baseline interrupted time series: Pretest–posttest with more than two communities repeatedly assessed over time (ideally >50 time points), before and after the (non-randomised) intervention Internal control group: Sub-groups of the population receiving the intervention who have remained completely or partly unexposed Simulated control group: Use transmission dynamics model to simulate control group under same conditions as in target population, but in absence of the intervention, using data collected at the start of the intervention	No randomisation Intervention more likely to be “real world”  More validity threats (eg selection biases, different sample characteristics, etc) than with experimental design  Do not take into account the transmission dynamics of infection  Stronger causality statement if results of intervention impact can be compared across many communities Logistically difficult if multiple time points or communities are used Additional considerations: Assess individual-level impact only Additional considerations: Estimates of the overall population-level impact of behavioural modifications on HIV rates after the intervention Estimates of the impact of the intervention, and of other contributing factors (see supplementary fig S2) Impact assessment takes into account the transmission dynamics of the epidemic Stronger causality statement
<b>Did the expected change occur? → Weak—no causality statement</b>			
Adequacy: <sup>13 21 22</sup> Assess if changes in the expected direction in primary indicators have occurred	Observational: No control group per se	Surveillance of health indicators over time among the appropriate target populations	Data necessary although mainly descriptive Can only demonstrate that the trend is going in the desired direction Intervention more likely to be “real world”

\*According to Habicht *et al.*<sup>13</sup>.

Supplementary fig S2 is available on the *STI* website (<http://sti.bmj.com/>).

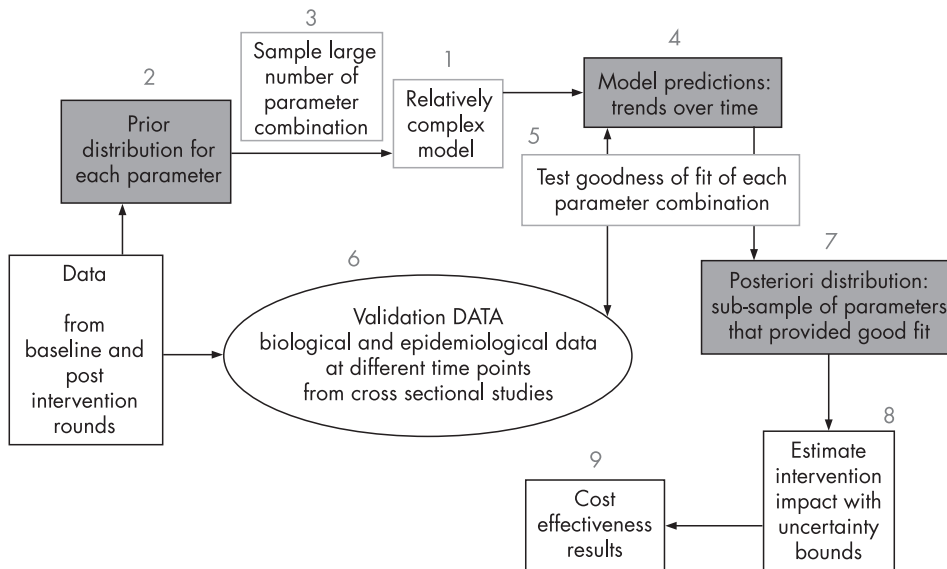
To demonstrate plausibility, a valid control group must be defined to assess what might have happened in the absence of the intervention (table 1).<sup>13 22 23</sup> Comparisons of HIV/STI trends using internal control groups from population subgroups, completely or partly unexposed, can only assess intervention effects at the individual level. External control groups from areas where the programme was not implemented, or baseline pre-intervention data from internal control groups, can yield population-level estimates. These comparisons are, however, prone to biases and cannot account for changes caused by natural infection dynamics. Simulated control groups based on transmission dynamics models may also suffer from the same internal validity threats as quasi-experimental designs, but have the added advantage of providing a framework that takes into account changes caused by natural infection dynamics and that can be used to estimate potential sources of biases as a result of changes unrelated to the intervention (details are

given in supplementary table S1, available on the *STI* website: <http://sti.bmj.com/>).<sup>17 19</sup>

## AN INTEGRATED MATHEMATICAL EVALUATION FRAMEWORK

Mathematical models are often used *a posteriori* to help interpret epidemiological trends and assess the likely impact of interventions in an explanatory fashion, or for cost-effectiveness analyses.<sup>6 9 19 20 24 25</sup> In contrast, our integrated framework has been rigorously designed *a priori* as an integral part of the Avahan M&E, and is based on a set of well-defined procedures to maximise objectivity and to take into account uncertainties in estimates. The framework was designed within the context of the intervention planned by Avahan.<sup>7</sup>

The framework is based on empirical biological and behavioural data collection and HIV/STI transmission dynamics modelling embedded within a Bayesian framework (fig 2).<sup>26–28</sup>



**Figure 2** Flowchart of the different stages of the mathematical framework described in the main text. Baseline and postintervention rounds of integrated behavioural and biological assessments, special behavioural surveys and general population surveys (see supplementary fig S1) as well as complementary data from literature reviews or other sources (table 1) will help to define biological, demographic and intervention before parameter distribution (stage 2) and to validate model results (stage 6). Ideally, baseline data are collected just before the intervention starts. In districts where these data are not available, the first round of data, even if collected after the start of the intervention, will be used as baseline for the main impact assessment, even if this leads to more conservative estimates. The magnitude of underestimation will, however, be assessed after defining a proxy baseline dataset based on literature reviews of published reports, scientific papers of representative samples of high-risk and general populations, surveys of non-governmental organisations, which gathered information on the level of condom use and sexual behaviour at the beginning of the programme, and early independent surveys in different districts. These will be complemented by information on behaviour among unexposed individuals in the first round survey (see supplementary fig S2). Supplementary figures are available on the STI website: <http://sti.bmj.com/>.

The evaluation is taking place at the district level over seven years, which is longer than most HIV trials, to maximise the chance of observing changes in HIV prevalence.

### Sources of data

The primary data for defining the model structure, informing model parameters, validating the model and defining attribution are collected within the context of the overall Avahan M&E programme by numerous agencies responsible for implementing interventions and collecting evaluation data.<sup>7</sup> Serial cross-sectional biobehavioural surveys (integrated behavioural and biological assessments; IBBA) are carried out in target high-risk groups in selected districts of the four southern states covered by Avahan. The surveys are repeated at two or three different time points (table 2). In addition, special behavioural surveys (SBS) among high-risk groups, as well as biobehavioural general population surveys (GPS), are being carried out in selected IBBA districts to validate and complement IBBA data and collect general population data. The questionnaires have been specifically designed to obtain detailed information required for epidemiological and modelling analyses. Methods to minimise social desirability biases are also used to elicit more accurate reporting of high-risk behaviour (table 1).<sup>7 29 30</sup>

Within Avahan, a management information system (MIS) consisting of a series of routine monthly indicators (eg number of condoms distributed, estimated size of core groups, number of syndromes treated) obtained from implementing partners is used to estimate intervention coverage and intensity. Costing studies of Avahan districts over time, involving a mixture of detailed bottom-up costing methods and the use of routine financial and project data,<sup>31 32</sup> are also taking place (table 2).

### Transmission dynamics model (1)

At the core of the framework is a tailor-made deterministic transmission dynamics model of HIV/AIDS and STI. Ideally, the model being developed will be as parsimonious as possible,

while including the complexity that matters.<sup>33</sup> The model structure and level of complexity will depend on the characteristics, STI levels, and risk factors observed in the population of interest and on the nature of the intervention in each district. Analyses of the baseline data will identify important sources of heterogeneity that should be included in the model. We focus on four southern states of India where HIV transmission is mostly sexual. The model will take into account age, and district-specific male and female commercial sex work, mixing patterns, STI prevalence, as well as other relevant characteristics such as migration and patterns of intravenous drug use. A series of preliminary modelling studies, such as studies on HIV and herpes simplex virus type 2 interaction and the impact of seasonal migration,<sup>34 35</sup> will help to determine which aspects of the natural history of STI/HIV and population characteristics will least influence our impact projections, and can therefore be ignored. Modelling studies assessing the simplest, yet adequate, way to model the general population are ongoing.

For each of the four main parameter categories (demographic, natural history of HIV and STI, behavioural and intervention exposure), prior parameter distributions (defined in the next section) will be specified for each round of data collection for the different risk groups. The distribution for the natural history parameters will be based on published literature. The IBBA and SBS will be used to inform behavioural and intervention exposure parameters. Demographic parameters will be based on official sources (eg census) and our different surveys. MIS, GPS and IBBA and other complementary data sources will provide estimates of the size of high-risk groups and coverage of the intervention (FMSW, MSM, clients; table 2).

### Bayesian framework (2–7)

The transmission dynamics model will be used within a Bayesian framework (fig 2).<sup>26–29</sup> Initially, available data will be used to specify what is known about each parameter by

**Table 2** Sources of data available from Karnataka, Maharashtra, Tamil Nadu and Andhra Pradesh districts where the Avahan intervention is taking place<sup>7</sup> (details in supplementary table S2, available on the STI website: <http://sti.bmj.com/>)

Surveys (responsibility)	Description	Indicators
IBBA, FTFI (KHPT, FHI + ICMR)	FSW in 27 districts; MSM in 17 districts; clients in 17 districts Repeated two or three times (the first round = baseline) Sample size ~400	HIV, CT, NG, HSV-2 and syphilis prevalence Sociodemographic and mobility Sexual behaviour and STI history
SBS: FTFI, ICVI* (CHARME)	MSM and FSW in 7 IBBA districts Repeated at least twice Sample size ~200-400 More detailed sexual behaviour data than in the IBBA (No biological data)	Exposure to the intervention (coverage and intensity) Estimates of size of high-risk groups <sup>3, 5, 7</sup> SBS and GPS: Detailed questions on different partner types, sexual mixing patterns, concurrency, mobility
GPS: FTFI, PBS* (CHARME)	General population household-based surveys in 4 districts Repeated once or twice Sample size ~4000-5000	
Costing studies <sup>31, 32</sup> (CHARME)	Primary cost data combined with available programme monitoring information. Costing by 75 districts through time, with a mixture of detailed bottom-up costing methods, and the use of routine financial and project data of Avahan Detailed costing activity: 16 intervention districts in Karnataka, 2 IBBA districts in among the three other states	At least quarterly information on spending across major budget categories such as staff, consumables, facilities/maintenance
MIS (implementing NGO)	Data on process indicators of coverage and intensity, enumeration of high-risk groups by state/district  For each partner by NGO or by district and by risk population Most indicators are reported on a monthly basis; some are reported quarterly, annually, or updated as needed <sup>7</sup>	Standardised process indicators: number of individuals receiving STI consultations, number of FSW sites covered, number of (free) condoms distributed, and many others Categories of information: infrastructure, capacity building, STI services, HIV/STI prevention communication, community mobilisation/enabling environment
Condom outlets availability (CHARME)	Surveys on early condom distribution among NGO in different districts at the beginning of the programmes Delphi surveys among NGO when necessary	Condom use and sexual activity before the intervention
Complementary sources of data		
Census		Demographic data
Sentinel HIV surveillance data from antenatal clinics		HIV prevalence over time
Literature review of published studies/reports		Pre-intervention data on HIV/STI prevalence, condom use, STI treatment, rate of partner changes and model parameters

CT, *Chlamydia trachomatis*; FHI, Family Health International; FSW, female sex worker; FTFI, face-to-face interview, used for the main questionnaire, are one-to-one meetings between the respondent and the interviewer; GPS, general population survey; HSV-2, herpes simplex virus type 2; IBBA, integrated behavioural and biological assessment; ICMR, Indian Council for Medical Research; ICVI, informal confidential voting interviews<sup>7, 30</sup> are interviewer-administered questionnaires that incorporate confidential self-completion methods by study participants who are screened from the interviewer; KHPT, Karnataka Health Protection Trust; MIS, management information system; MSM, men who have sex with men; NG, *Neisseria gonorrhoea*; NGO, non-governmental organisation; PBS, polling booth surveys<sup>7, 29</sup> are anonymous group interviews conducted with eight to 10 individuals separated from one another by a private booth. Questions are read by the interviewer. Participants reply by placing a non-identified card (with question number but no ID number) in a box in their private booth. Data only analysed at the aggregate level; SBS, special behavioural survey; STI, sexually transmitted infection.

\*Methods to minimise social desirability biases for most sensitive subset of questions.

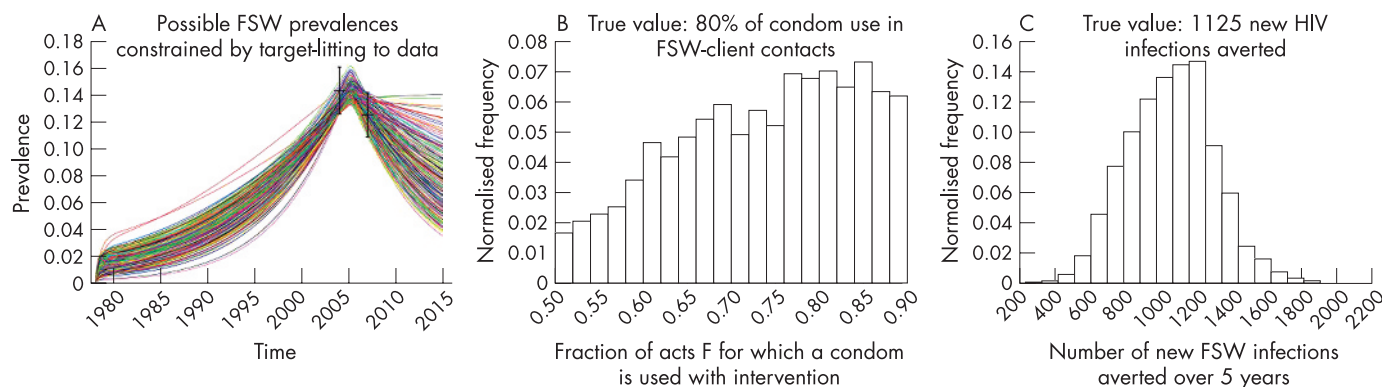
defining a plausible range of values (the prior distribution) for each one. Then, the prior distribution will be sampled many times in order to test a large number of parameter combinations (>>>10 000) against the empirical validation data used to fit the model (eg HIV and STI prevalence). The goodness of fit of each parameter set will be assessed by comparing prevalence estimates predicted by the model at specific time points with corresponding empirical estimates. The model will be fitted to two or three rounds of age-specific HIV and STI prevalence data by risk group in each district. The model will be simultaneously fitted to data from different districts and states to constrain biological parameters (fig 3A). Only the subset of parameter sets that fits the empirical data well (posterior distribution) will be kept for further analyses (fig 3B). The advantage of this approach is to produce point estimates and credibility intervals (CrI) that will reflect the uncertainty in parameter assumptions on model predictions (fig 3C).<sup>26-29</sup> This is necessary because our model will be relatively complex with many uncertain parameters, which means that more than one set of parameters could produce an equally good fit to epidemiological trends. Using only one parameter set could lead to biased estimates of intervention impact. Together, the point and CrI estimates will provide information to judge if the intervention is sufficiently effective and/or cost effective to be

declared of public health use. Ideally, the public health criteria should be determined by public health authorities and stakeholders, before evaluation.

As a result of the large quantity of data (>60 datasets per study round) and the extensive fitting procedure, the process needs to be automated and rigorous. As existing fitting methods have not been tested for very complex HIV dynamic models, our choice will be based on the results of ongoing independent validation studies (conducted before impact assessment) comparing the precision, validity of impact estimates, and computing time needed by the different "Bayesian style" methods. The procedures will be evaluated by using them to fit a range of models to "fitting" data generated by a model with the same or a more complex structure.<sup>28</sup> The latter analyses will give pointers as to whether the complexity of the model can be reduced while conserving its ability to produce adequate impact estimates. The procedures explored will include different search algorithms (Markov chain, Monte Carlo)<sup>27, 28</sup> or Latin hypercube sampling combined with likelihood methods or a more heuristic target fitting method.<sup>9, 26</sup>

### Model predictions and impact estimation (8)

The main model outcomes will include age-specific HIV/STI incidence and prevalence, and numbers of new STI/HIV infections averted over a fixed time period and by district. To



**Figure 3** An illustrative theoretical example that shows how the mathematical framework will be used to fit the model, estimate an intervention parameter and estimate the impact of the intervention with uncertainty (fig 2). A transmission dynamics model was used to simulate HIV prevalence data (“the truth”) among female sex workers (FSW) and clients in 2004 and 2007. Condom use is assumed to increase from 22% to 80% after the introduction of an intervention in 2005. (A) Results of the model fit for FSW only (clients not shown) using a target method: the prior distributions of each of the 16 model parameters were defined using a reasonable range. A total of 500 000 parameter sets were sampled and tested. Model prediction for each set was compared with the “true” prevalence; 2061 parameter sets fitted the “true” HIV prevalence well, ie fell within the confidence interval (assuming a sample of 400) of the true data points. (B) Resulting posterior distribution of plausible condom use parameter values (representing fraction of sex acts protected) suggested by the fitting procedure, condom use estimates vary between 50% and 90%, with values between 75% and 90% being more likely. The “true” value is 80%. (C) Resulting posterior distribution of new HIV infections averted over five years after the intervention using the 2061 parameter sets that fitted the data well. Most likely model estimates are between 800 and 1400 new infections averted. The true value is 1125. If the validation data (ie HIV prevalence) were more precise (narrower confidence intervals) the fitting procedure would produce more precise parameter and impact estimates. By using priors reflecting different beliefs on parameter or model assumptions, it is also possible to test different hypotheses of prior beliefs in a scientific and objective fashion. The posterior distributions resulting (through the fitting procedure) from the different priors will produce new impact estimates and permit assessment of the sensitivity of our results to different assumptions.

produce estimates and CrI of intervention impact, parameter sets from the posterior parameter distribution will first be used to simulate the different health outcomes of interest in the intervention group over time. Then, the same health outcomes will be simulated in a matched control group using the same parameter sets but with the intervention parameters (eg coverage, condom use, STI treatment) reset to pre-intervention levels, thus providing population-level impact estimates that take into account the transmission dynamics of infection. For each district, estimates of the main model outcomes in different subpopulations will be obtained by comparing predicted STI/HIV infections in the presence and absence of the intervention. Then, the district-specific model predictions will help to understand the influence of the different epidemiological contexts and local transmission dynamics of HIV/STI infections on the population-level effectiveness of the intervention, and to improve prevention strategies. The primary estimation of intervention impact with the “full” model will occur at the end of the seven years.

#### *Attribution of intervention impact to Avahan*

To estimate the fraction of new infections prevented by the Avahan intervention and its different components, extra steps are needed to reflect the uncertainty surrounding the simulated control groups (see supplementary fig S2, available on the STI website: <http://sti.bmj.com/>). First, simulated control groups using baseline data will allow the estimation of the overall impact of any changes in high-risk or treatment-seeking behaviours after the intervention (as a result of the intervention or any other causes) on HIV, independently of the transmission dynamics. Second, IBBA, SBS, GPS and MIS data will be used to estimate plausible ranges for the fraction of individuals exposed to any intervention, or to the Avahan intervention specifically, and the incremental level of behavioural modification observed among those exposed to any intervention, or Avahan intervention only (when they are the only provider), compared with those not exposed (simulated control group). MIS process indicators of programme adequacy will be used for validation of coverage and intensity, and for estimating the

improvement in STI services. Together, this will permit an estimation of the fraction of new HIV infections prevented by any intervention to which Avahan contributed (contribution) or by Avahan only (attribution), while taking into account uncertainties about risk behaviours of the control group.

#### **Cost-effectiveness analyses (9)**

By merging modelled effectiveness estimates with empirical costing studies the cost effectiveness of the programme and its different components at the district level will be derived.<sup>9 31 32</sup> Modelled projections of HIV/STI cases averted will be combined to obtain estimates of the overall cost per disability-adjusted life-year saved, and then compared with the empirical cost data to obtain district, state and programme-wide cost-effectiveness estimates (measured as cost per HIV and STI cases averted and cost per disability-adjusted life-year saved). The models will be used to explore how the projected cost effectiveness changes if the intervention (and the resulting patterns of behaviour change) is sustained for different lengths of time, using district-based cost data. To estimate the uncertainty in the cost-effectiveness ratio, the uncertainty of costs, based on data from 75 districts, will be combined with the uncertainty in the impact estimates at the sampling stage.

#### **DISCUSSION**

The integrated mathematical framework presented has been designed for the rigorous evaluation of a large-scale HIV intervention, and to minimise limitations and validity threats associated with uncertainty in parameter assumptions, model specification, and the non-experimental nature of the design (table 1). Serial cross-sectional studies have been designed to collect detailed modelling data, using techniques to minimise reporting biases. The uncertainty in model structure will be studied before impact assessment to validate models of reduced complexity. The extensive Bayesian fitting procedure will take into account the uncertainty in model parameters on impact estimates, and permit the testing of different hypotheses on previous beliefs in a scientific and objective manner. Despite residual uncertainty about the simulated control group,

## Key messages

- We present a novel mathematical framework combining empirical biological and behavioural data collection with HIV/STI transmission dynamics modelling, embedded within a Bayesian framework, to provide quantitative intervention impact estimates that take into account changes caused by the transmission dynamics of HIV infection.
- This is the first time such an approach has been applied on such a large scale.
- The integrated mathematical framework, combined with the high-quality second-generation surveillance data being collected through the overall Avahan M&E, will help achieve a higher level of certainty in conclusions about intervention effectiveness than analysis of epidemiological trends alone.
- Lessons learnt from the CHARME project could help the design of future evaluations of large-scale interventions in other settings, whereas the results of the current evaluation of the Avahan initiative will be of programmatic and public health relevance.

mathematical modelling remains the only way to provide quantitative impact estimates that take into account changes caused by the transmission dynamics of the infection; this cannot be done on the basis of observed epidemiological trends alone. The combination of impact estimates with costing data will provide estimation of the cost effectiveness of the intervention, an issue of ever-increasing importance in a context of limited funding and competing public health priorities. The added value of defining the plan of analysis before the evaluation is to minimise observer biases. This is important given that the evaluation occurs outside the context of a blinded experimental design. As the intervention will be evaluated in a larger number of sites than would be possible with C-RCT, it will be possible to make an overall assessment of the intervention impact in different epidemiological contexts, hopefully improving external validity.

This is the first time such an approach has been applied on such a large scale. The framework needs to deal with an unprecedented quantity of data for an HIV/AIDS modelling study, and a substantial amount of programming needs to be carried out before impact assessment results can be produced. The integrated mathematical framework, combined with the high-quality second-generation surveillance data being collected through the overall Avahan M&E, will help achieve a higher level of certainty in conclusions than analysis of epidemiological trends alone. Lessons learnt from the CHARME project could help the design of future evaluations of large-scale interventions in other settings, whereas the results of the evaluation will be of programmatic and public health relevance.

## ACKNOWLEDGMENTS

The authors would like to thank Gina Dallabetta and Padma Chandrasekaran for very useful discussion and suggestions.

## Authors' affiliations

**M-C Boily, M Pickles, R M Anderson**, Department of Infectious Diseases Epidemiology, Imperial College, London, UK

**M Alary**, Centre Hospitalier affilié Universitaire de Québec (CHA), Québec, Canada

**J Blanchard, S Moses, R Washington, S Reza-Paul, K N Deering**,

University of Manitoba, Manitoba, Canada

**C M Lowndes**, Health Protection Agency, London, UK

**P Vickerman, L Kumaranayake, C Watts**, London School of Hygiene & Tropical Medicine (LSHTM), London, UK

**B M Ramesh**, Karnataka Health Promotion Trust (KHPT), Bangalore, India

**R Washington**, St. John's Medical College, Bangalore, India

**A C Labbe**, University of Montreal, Montreal, Canada

Funding: Support for this research was provided by the Bill & Melinda Gates Foundation through Avahan, its India AIDS Initiative.

Competing interests: None.

Contributions: MCB, CML, PV, LK, JB, SM, BMR, CW, RW, ACL, RMA, SRP and MA contributed to the different aspects of the study design. KD and MP designed the mathematical models used to produce simulation results and figures. MP helped with the design of the Bayesian framework. MCB wrote the first draft of the manuscript with the help of CML. KD, MP, MA, PV, LK, JB, SM, and MA contributed to the different drafts of the manuscript.

The views expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Bill & Melinda Gates Foundation and Avahan.

CHARME-India is the CHA—HIV/AIDS Research, Monitoring and Evaluation Project, India. The CHARME-India team includes the investigators RMA, MA, JB, MCB, LK, ACL, CML, SM, BMR, SRP, PV, CW, RW and co-workers S Chandrashekar, KN Deering, A Foss, K Gurav, AA Jayachandran, S Joseph, B Mahapatra, A Phillips and M Pickles. The main collaborating institutions include CHA, University of Manitoba, KHPT, St John's Medical College, Imperial College, LSHTM; collaborating institutions for field work include the Tata Institute of Social Sciences, Mumbai, and the Centre for Media Studies, Hyderabad.

## REFERENCES

- 1 **Bennett S**, Boerma JT, Brugha R. Scaling up HIV/AIDS Evaluation. *Lancet* 2006;**367**:79–82.
- 2 **Bertozi S**, Padian NS, Wegbreit J, et al. HIV/AIDS prevention and treatment. In: *Disease control priorities in developing countries*, 2nd edn. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, et al, editors. New York: Oxford University Press, 2006.
- 3 **Chandrasekaran P**, Dallabetta G, Loo V, et al. India: the unfinished agenda. *Lancet Infect Dis* 2006;**8**:508–21.
- 4 **Solomon S**, Chakraborty A, D'Souza Yephthomi R. A review of the HIV epidemic in India. *AIDS Educ Prevent*, 2004;**16**(Suppl A), 155–69.
- 5 **Moses S**, Blanchard JF, Kang H, et al. *AIDS in South Asia: understanding and responding to a heterogeneous epidemic*, World Bank report. Washington: World Bank, 2006;**113**:15–77.
- 6 **Boily MC**, 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 1), i78–90.
- 7 **Priya N**, Loo V. Evaluation of the Avahan program: framework and general design issues. New Delhi, India: Avahan program, 2007.
- 8 **Yorke JA**, Hethcote HW, Nold A. Dynamics and control of the transmission of gonorrhoea. *Sex Transm Dis* 1978;**5**:51–6.
- 9 **Vickerman P**, Terris-Prestholt F, Delany S, et al. Are targeted HIV prevention activities still cost-effective in high prevalence settings? Results from an STI treatment intervention for sex workers in Hillbrow, South Africa. *Sex Transm Dis*, 2006;**33**(10 Suppl), S122–32.
- 10 **Sangani P**, Rutherford G, Wilkinson G. Population-based interventions for reducing sexually transmitted infections, including HIV infection. *Cochrane Database Syst Rev* 2004;(2):CD001220.
- 11 **Creese A**, Floyd K, Alban A, et al. Cost-effectiveness of HIV/AIDS interventions in Africa: a systematic review of the evidence. *Lancet* 2002;**359**:1635–43.
- 12 **Ngugi EN**, Chakkalackal M, Sharma A, the Kibera HIV Study Group, et al. Sustained changes in sexual behavior by female sex workers after completion of a randomized HIV prevention trial. *J Acquir Immune Defic Syndr* 2007;**45**:588–94.
- 13 **Habicht JP**, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 1999;**28**:10–18.
- 14 **Donner A**, Klar N. Design and analysis of cluster randomised trials in health research. London: Arnold; New York: Oxford University Press, 2000;**178**:11–13.
- 15 **Brown CA**, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006;**6**:54.
- 16 **Hussey MA**, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007;**28**:182–91.
- 17 **Alary M**, Lowndes CM, Boily MC. Community randomized trials for HIV prevention: the past, a lesson for the future? *AIDS* 2003;**17**:2661–3.
- 18 **Wawer MJ**, Sewankambo NK, Serwadda D, the Rakai Project Study Group and Gray RH, et al. Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomised community trial. *Lancet* 1999;**353**:525–35.

- 19 **Boily MC**, Lowndes CM, Alary M. Complementary hypothesis concerning the community sexually transmitted disease mass treatment puzzle in Rakai, Uganda. *AIDS* 2000;**14**:2583–92.
- 20 **Korenromp EL**, Van Vliet C, Grosskurth H, *et al*. Model-based evaluation of single-round mass treatment of sexually transmitted diseases for HIV control in a rural African population. *AIDS* 2000;**14**:573–93.
- 21 **Biglan A**, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci* 2000;**1**:31–49.
- 22 **Shadish WR**, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin, 2002:1–313.
- 23 UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. Guidelines for second generation HIV surveillance for HIV: the next decade. pp 8–37. [http://www.who.int/hiv/pub/surveillance/en/cds\\_edc\\_2000\\_5.pdf](http://www.who.int/hiv/pub/surveillance/en/cds_edc_2000_5.pdf) (accessed 29 Oct 2007).
- 24 **Foss AM**, Watts CH, Vickerman P, *et al*. Could the CARE-SHAKTI intervention for injecting drug users be maintaining the low HIV prevalence in Dhaka, Bangladesh? *Addiction* 2007;**102**:114–25.
- 25 **Korenromp EL**, White RG, Orroth KK, *et al*. Determinants of the impact of sexually transmitted infection treatment on prevention of HIV infection: a synthesis of evidence from the Mwanza, Rakai, and Masaka intervention trials. *J Infect Dis*, 2005;**191**(Suppl 1), S168–78.
- 26 **Van de Velde N**, Brisson M, Boily MC. Modeling human papillomavirus vaccine effectiveness: quantifying the impact of parameter uncertainty. *Am J Epidemiol* 2007;**165**:762–75.
- 27 **Cancre N**, Tall A, Rogier C, *et al*. Bayesian analysis of an epidemiologic model of plasmodium falciparum malaria infection in Ndiop, Senegal. *Am J Epidemiol* 2000;**152**:760–70.
- 28 **Pickles M**, Vickerman P, Alary M, *et al*. What is the most important data to collect for dynamical models? 17th Meeting of the International Society for Sexually Transmitted Diseases Research (ISSTD). Seattle, WA, USA, 28 July–1 August 2007.
- 29 **Kang H**. Study brief for polling booth surveys as part of Bagalkot Demonstration Project. Bangalore, India: India–Canada Collaborative HIV/AIDS Project, 2005.
- 30 **Gregson S**, Zhuwau T, Ndlovu J, *et al*. Methods to reduce social desirability bias in sex surveys in low-development settings: experience in Zimbabwe. *Sex Transm Dis* 2002;**29**:568–75.
- 31 **Kumaranayake L**, Chandrashekar S, Alary M. Economic Analysis of Avahan Interventions in Southern India: A Methodology. XVIIIth Annual Conference of Karnataka Association of Community Health. JN Medical College, Belgaum, Karnataka, India, 10–11 June 2006.
- 32 **Kumaranayake L**, Pepperall J, Goodman H, *et al*. Costing guidelines for HIV/AIDS prevention strategies. UNAIDS best practice collection – key materials. Geneva: UNAIDS, 2000.
- 33 **Koopman JS**. Modeling infection transmission – the pursuit of complexities that matter. *Epidemiology* 2002;**13**:622–4.
- 34 **Foss A**, Vickerman P, Chalabi Z, *et al*. Dynamic modelling of HSV-2 transmission: issues in structural uncertainty. Royal Statistical Society Conference. University of York, York, UK, 16–20 July 2007.
- 35 **Deering K**, Boily MC, Ramesh BM, *et al*. Exploring the effects of seasonal migration of high-risk groups in rural southwest India: a mathematical modelling simulation. 17th Meeting of the International Society for Sexually Transmitted Diseases Research (ISSTD). Seattle, WA, USA, 28 July–1 August 2007.

### Access a vast information database with *Toll-Free linking*

“Toll-free” linking gives you immediate access to the full text of many of the cited articles in a paper’s reference list—FOR FREE. With the support of HighWire’s vast journal catalogue, a huge reference library is now open to you. If HighWire hosts the journal, you can view the full text of the referenced article, completely free of charge by following the Free Full Text links in the references.