

Estimating HIV incidence from case-report data: method and an application in Colombia

Juan Fernando Vesga^a, Anne Cori^a, Ard van Sighem^b and Timothy B. Hallett^a

Objective: Quantifying HIV incidence is essential for tracking epidemics but doing this in concentrated epidemic can be a particular challenge because of limited consistent high-quality data about the size, behaviour and prevalence of HIV among key populations. Here, we examine a method for estimating HIV incidence from routinely collected case-reporting data.

Methods: A flexible model of HIV infection, diagnosis and survival is constructed and fit to time-series data on the number of reported cases in a Bayesian framework. The time trend in the hazard of infection is specified by a penalized B-spline. We examine the performance of the model by applying it to synthetic data and determining whether the method is capable of recovering the input incidence trend. We then apply the method to real data from Colombia and compare our estimates of incidence with those that have been derived using alternative methods.

Results: The method can feasibly be applied and it successfully recovered a range of incidence trajectories in synthetic data experiments. However, estimates for incidence in the recent past are highly uncertain. When applied to data from Colombia, a credible trajectory of incidence is generated which indicates a much lower historic level of HIV incidence than has previously been estimated using other methods.

Conclusion: It is feasible, though not satisfactory, to estimate incidence using case-report data in settings with good data availability. Future work should examine the impact on missing or biased data, the utility of alternative formulations of flexible functions specifying incidence trends, and the benefit of also including data on deaths and programme indicators such as the numbers receiving antiretroviral therapy.

© 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins

AIDS 2014, **28** (Suppl 4):S489–S496

Keywords: case-reporting data, HIV, HIV incidence rate, mathematical modelling, spline function

Introduction

In any setting, one of the most important pieces of information in responding to an HIV epidemic, evaluating past efforts, and planning for the future is the time-course of the HIV incidence rate.

Direct observation of new infections, whether through cohort studies or via the use of incidence assays and algorithms [1], is infeasible in most national settings.

Mathematical modelling provides an alternative mean to infer incidence using other more readily available data. Currently, UNAIDS estimates of HIV incidence trends in generalized epidemics are derived through fitting a simple model to HIV prevalence measurement among pregnant women and prevalence measurements in national household surveys [2,3]. This approach has been less satisfying in settings with concentrated epidemics because of a combination of factors: inconsistencies in the measurements of prevalence over

^aDepartment of Infectious Disease Epidemiology, Imperial College London, London, UK, and ^bStichting HIV Monitoring, Amsterdam, the Netherlands.

Correspondence to Dr Juan Fernando Vesga, Department of Infectious Diseases Epidemiology, St. Mary's Campus, Norfolk Place, Imperial College London, London W2 1PG, UK.

E-mail: j.vesga10@imperial.ac.uk

Received: 21 August 2014; revised: 21 August 2014; accepted: 21 August 2014.

DOI:10.1097/QAD.0000000000000466

time, scarce prevalence measurements among hard-to-reach groups and lack of robust estimates on the size of these groups. For this reason, it is important to explore other means of estimating HIV incidence in these settings, which would not exclusively rely on estimates of HIV prevalence and size estimates of key populations.

Other sources of available data, which could potentially be combined with models to draw inference on incidence in concentrated epidemics, include time-series data on AIDS deaths and reported cases. The use of estimated death time series is investigated in another article in this collection (Stover *et al.* in this supplement). Here, we focus on the use of reported cases.

In many settings, especially in Latin America and Europe, systems are in place to centrally record the number of persons newly diagnosed with HIV. In many settings, it is believed that the coverage of this system (the proportion of newly diagnosed cases that are counted) could be as high as 80% [4,5]. However, the interval between an infection and a diagnosis is not known and may change over time, and this would confound all estimates of incidence unless further information could be included on the interval between testing and diagnosis. Possible candidates to inform the interval between infection and diagnosis are the clinical stage and CD4⁺ cell count [6]. Further, in the future, measurement of a well characterized biomarker of recent infection may also be available [7].

In this study, we build on prior work [8,9] to develop a modelling framework that can be used to estimate HIV incidence drawing primarily on HIV case-report data. We then use synthetic data to test the performance in recovering a wide range of possible incidence trajectories. Finally, we apply the model to Colombia to derive HIV incidence estimates, which are then compared with other estimates available for Colombia.

Methods

Mathematical model

The proposed model consists of a deterministic mathematical model to simulate the process of HIV infection and disease progression, death and diagnosis (Fig. 1).

This structure is expressed through a set of six ordinary differential equations as follows:

$$\frac{dU(t)}{dt} = \alpha(t) - U(t)(s(t) + \delta); \text{ Eq. 1 : Susceptible}$$

$$\frac{dI_1(t)}{dt} = S(t)U(t) - I_1(\delta + \gamma_1 + \rho_1(t)); \text{ Eq. 2}$$

: Acute infection

$$\frac{dI_2(t)}{dt} = I_1\gamma_1\epsilon_2 - I_2(\delta + \gamma_2 + \rho_2(t)); \text{ Eq. 3}$$

: CD4+ count > 500 cell/ml

$$\frac{dI_3(t)}{dt} = I_1\gamma_1\epsilon_3 + I_2\gamma_2 - I_3(\delta + \gamma_3 + \rho_3(t)); \text{ Eq. 4}$$

: CD4+ count 350 – 500 cell/ml

$$\frac{dI_4(t)}{dt} = I_1\gamma_1\epsilon_4 + I_3\gamma_3 - I_4(\delta + \gamma_4 + \rho_4(t)); \text{ Eq. 5}$$

: CD4 + count 200 – 350 cell/ml

$$\frac{dI_5(t)}{dt} = I_1\gamma_1\epsilon_5 + I_4\gamma_4 - I_5(\delta + \gamma_5 + \rho_5(t)); \text{ Eq. 6}$$

: CD4+ count < 200 cell/ml

Where $U(t)$ is the number of persons HIV susceptible at time t ; $I_1(t)$ is the number of infected individuals in acute infection at time t ; $I_j(t)$ ($j = 2, \dots, 5$) is the number of infected individuals in CD4⁺ cell count category j at time

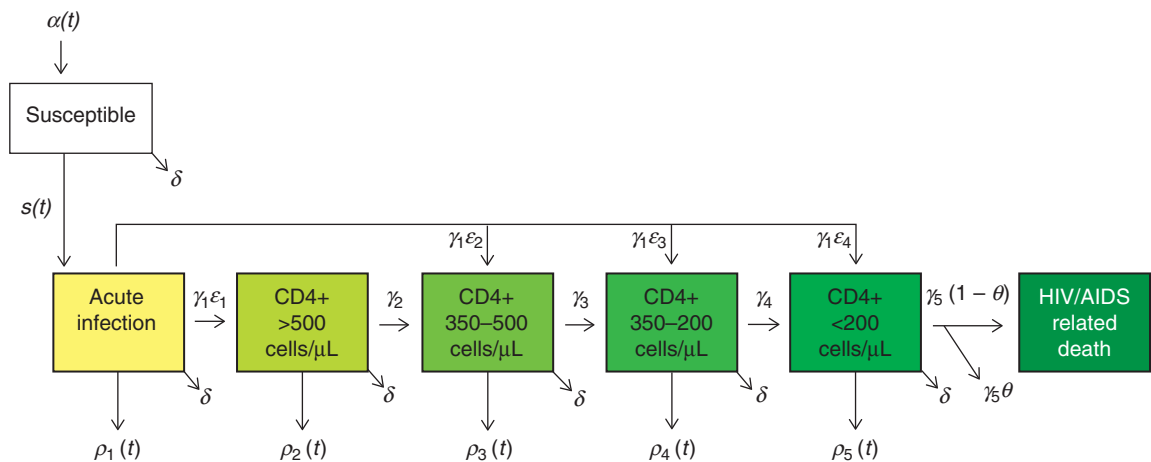


Fig. 1. Model of HIV infection and progression. Parameter values and prior distributions for calibration can be found in Table 1.

t and $s(t)$ is the time-varying rate at which the population of infected individuals transmits HIV to susceptible persons. The rate at which new infections occur in this population is given by $s(t)U(t)$. All symbols and parameters are described in Table 1.

At infection, individuals pass through a stage of acute infection following which individuals enter one of four CD4⁺ cell count categories (<200, 200–350, 350–500, 500+ cell/ μ l). Thereafter, individuals progress to lower CD4⁺ cell counts at rates which have previously been estimated from European data [10]. Death due to AIDS in the model is assumed to occur at a fixed rate only among individuals with CD4⁺ cell count below 200 cell/ μ l. The overall median survival time is 10.4 years, consistent with estimates from low-income and middle-income countries [11]. In the model framework, these parameters describing the progression of HIV infections are considered to be known perfectly.

At all times after infection, individuals can be diagnosed with HIV. At that point they transition into a ‘diagnosed’ category. The number of HIV-infected persons that transition to the diagnosed category in a given year in the model is compared with the data on the number of reported cases.

Model parameterization

HIV infection rate

Following Hogan *et al.* [2], the instantaneous hazard of infection for susceptible individuals $s(t)$ is specified as a B-spline function of time. The B-spline is parameterized by a vector of coefficients β_1, \dots, n , where $n-3$ is the number of knots of the spline which were evenly spaced on the time interval 1975–2015. This flexible functional form allows many credible incidence trajectories while constraining the trajectory to evolve smoothly with respect to time. This does not attempt to provide a mechanistic description of HIV transmission. Note that in other formulations, the spline describes a force of infection (which gives incidence when multiplied with the product of $U(t)$ and all infected individuals) rather than a hazard of infection, but this is not possible here as, for parsimony, we do not model the entire infected population.

We set $\beta_1 = \beta_2 = \beta_3 = 0$ to anchor the incidence trajectory at zero before 1975, whereas β_4 to β_6 were independently specified with uniform priors: $\beta_4 \sim U(0, 5)$, $\beta_5 \sim U(-5, 5)$ and $\beta_6 \sim U(-5, 5)$. We impose penalties [12] on the spline, in order to limit over-fitting to potentially noisy data and to represent our prior assumption that rapid large oscillations in the

Table 1. Model parameters.

Symbol	Description	Value/prior distribution	Reference
$\alpha(t)$	Birth rate	$N(t)E + (\delta)(\sum_i I_j(t) + U(t)) + I_5(t)\gamma_5$	
$N(t)$	Total population at any time t	$U(t) + \sum_j I_j$, for $j = 1, \dots, 4$	–
E	Average population growth per year	2.7%	[14]
δ	Background mortality rate	1/72 years	[14]
θ	Fraction of AIDS-related deaths detected <i>post mortem</i>	$\sim U(0, 1)$	Assumption
γ_j	Progression rates (1/years) from CD4 ⁺ cell count stage k to the subsequent inferior CD4 ⁺ cell count category	(1/yrs) $\gamma_1 = \frac{1}{0.24}$ $\gamma_2 = \frac{1}{2}$ $\gamma_3 = \frac{1}{3}$ $\gamma_4 = \frac{1}{3}$ $\gamma_5 = \frac{1}{24}$	[10,15]
ε_j	Proportion moving to each CD4 ⁺ stage after acute infection	$\varepsilon_2 = 0.58$ $\varepsilon_3 = 0.23$ $\varepsilon_4 = 0.16$ $\varepsilon_5 = 0.03$	[10,16]
ρ_5	Diagnostic rate at CD4 ⁺ <200cell/ μ L	$\rho_5 \sim U(\frac{1}{10}, \frac{1}{0.5})$	Assumption
m	Slope scaling parameter. Is the slope for the line crossing (0, ρ_5) and (0, ρ_1)	$m = \frac{0 - \rho_5(0)}{8.17 - 0}$	Assumption
ω	Relative slope variation factor.	$\omega \sim U(-1, 1)$	Assumption
$\sigma(t)$	Function for the variation of HIV detection over time	$\sigma(t) = \frac{1}{1 + e^{-r(t-t_{50})}}$, for $r = 0.003$	Assumption
t_{50}	Shape parameter in function $\sigma(t)$	$t_{50} \sim U(0.1, 1)$	Assumption
β_i	Basis coefficients for spline function $s(t)$	$\beta_i = 0$, for $i = 1, \dots, 3$ $\beta_i \sim U(0, 5)$, $i = 4$ $\beta_i \sim U(-5, 5)$, $i = 5, 6$ $\beta_i = 2\beta_{i-1} - \beta_{i-2} + \mu_i$, For $i = 7, \dots, n$	
μ_i	Deviance of the i th β from the former two coefficients β_{i-1} and β_{i-2} , for $i = 7, \dots, 10$	$\mu_i \sim N(0, \tau^2)$ $\tau^2 \sim U(0, 2)$	Assumption

HIV incidence rate are unlikely. We use second-degree difference penalties [12] expressed as:

$$\beta_i = 2\beta_{i-1} - \beta_{i-2} + \mu_i, \dots, \text{ for } i > 6$$

where the error μ_i is the amount of deviance from the trajectory defined by the two previous coefficients.

The prior for μ_i is normally distributed with parameters $N(0, \tau^2)$, where τ^2 is a hyper-parameter controlling the overall smoothness of the function; small values of τ^2 result in little deviance from the previous trajectory and large values in contrast produce more flexible curves.

HIV detection rates

The rate at which HIV-infected individuals are diagnosed is $\rho_j(t)$, and is allowed to vary by CD4⁺ cell count category j and calendar time t :

$$\rho_j(t) = ((m\omega)T_j + \rho_5(0))\sigma(t)$$

where T_j is the average time (years) to reach CD4⁺ cell count category below 200 cell/ μL ($j=5$) for those in CD4⁺ cell count category $j=1, \dots, 4$ ($T_1 = 8.17$; $T_2 = 7.93$; $T_3 = 6.74$; $T_4 = 3.74$; $T_5 = 0$).

This parameterization sets the rate of detection to vary linearly as the time to reach CD4⁺ <200 cell/ μL decreases. The slope of this line is specified by parameter ω , which allows the linear relationship to be either positive or negative. With $\omega > 0$, individuals with high CD4⁺ cell counts are more likely to be detected, which is plausible scenario if exposure to HIV is suspected or symptoms of acute infection drive testing; with $\omega < 0$, individuals with low CD4⁺ cell counts are more likely to be detected, which may be true if HIV testing is prompted by increasing experience of illness. The prior on ω reflects that both of these possibilities are equally likely: $\omega \sim U(-1, 1)$. Parameter m is a constant scaling parameter which ensures that the diagnosis rate is greater than zero for all CD4⁺ categories, irrespective of the value of ω .

Rate of testing for those in CD4⁺ cell count category below 200 cell/ μL (ρ_5) has a prior distribution such that the average interval from infection to diagnosis is uniformly distributed between 0.5 and 10 years: $\sim U(\frac{1}{10}, \frac{1}{0.5})$.

$\sigma(t)$ describes the change in the rates of diagnosis over time. Here we assume that this change can be represented by the logistic function: $\sigma(t) = \frac{1}{1+e^{-r(t-t_{50})}}$, where r is the rate of increase, and t_{50} is the time to reach half the upper limit rate. Parameter r is set at 0.003 and a prior was placed on t_{50} that allows a wide range in shapes in $\sigma(t)$ from a

stable trend over time ($t_{50} = 1$) to a sharp increase in the rate of diagnosis from close to zero at the start of the epidemic ($t_{50} = 0.1$): $t_{50} \sim U(0.1, 1)$.

Model validation and calibration procedures

The inference on the HIV incidence rate is done in a Bayesian framework. The likelihood is given by the probability of observing the number of HIV cases without AIDS reported in a year, the number of HIV cases with AIDS reported in a year (i.e. symptomatic illness or CD4⁺ cell count <200 cell/ μL) and the number of persons diagnosed with HIV after death in a given year, each modelled as Poisson process. We used the Metropolis Hastings algorithm with component-wise updating to sample from the posterior distribution of the model. We used an adaptive proposal density variance in order to achieve an acceptance rate of approximately 30%. Three chains were run in parallel for 500 000 iterations after which chains were visually inspected for convergence and 50% of the initial runs were discarded.

Models are estimated that use B-spline functions with number of knots $n=7, 8, 10$, and with hyper-prior for τ^2 variously given as $\sim U(0.001, 0.1)$, $\sim U(0.1, 0.5)$, $\sim U(0.5, 1.0)$, $\sim U(1.0, 1.5)$ or $\sim U(1.5, 2.5)$, giving a total of 20 models runs for each estimation problem. Among these, a model is chosen that has the greatest agreement to data, given its number of effective parameters, as measured by the model with smallest deviance information criterion (DIC) [13]. For that chosen model, a sample from the joint posterior distribution is formed by randomly sampling 1000 iterations across all its chains. Medians and 95% credible intervals are reported for the posterior samples.

Assessment of model performance using synthetic data

The performance of model in recovering the ‘correct’ incidence trajectory from sets of synthetic data was assessed. One major concern was that the specification of this estimation method would render it unable to correctly discern all possible trajectories with different amounts of variation over time or timing of inflections. For this reason, we created three ‘challenge datasets’ using the same assumptions as the estimation model about natural history but varying the number in the spline function that generates that incidence rate ($n=7, 8, 10$). Synthetic data were generated under the assumption of constant rates of diagnosis over time and CD4⁺ cell count.

Application of the model to Colombia

Between 1983 and 2010, Colombia reported 78 999 HIV/AIDS patients, the vast majority coming from urban areas and major cities [17]. Overall prevalence was estimated at 0.22% in 2009 in the adult population [18], but HIV prevalence appears to be substantially concentrated among MSM, among whom prevalence has been

estimated to be between 5.6 and 24.1% in different cross-sectional studies in 2010 [19]. In our method, we assumed equivalence between an AIDS diagnosis in the data and a diagnosis for a patient with a $CD4^+$ cell count below 200 in the model.

There have been official records of the number of deaths that were classified as AIDS deaths among persons who had not previously been recorded as HIV-infected (i.e. post-mortem HIV diagnosis) since 1985 [20]. No reporting bias or misclassification in this is assumed.

We used data on the total number of HIV tests reported to have been performed in Colombia to centre the prior on the trend in the rate of diagnosis over time [21]. This trend can be broadly described as a sigmoid increase with a turning point in 1996 and a plateau in the more recent years.

These data were used in the method outlined above to estimate incidence over time. The results were compared with reported estimates of incidence from UNAIDS [22], which are based on fitting a model to prevalence data and population size estimates for key populations [3].

Results

Assessment of model performance using synthetic data

We created three challenge datasets with different epidemic trajectories and the proposed method was able to successfully recover each of them (Fig. 2). A part of the method is to estimate multiple models and then select that model with the lowest DIC; as expected, the model that was favoured in each of these tests was the model which has the same number of knots as that used to generate the synthetic data. Also as expected, uncertainty becomes

greater in more recent years as the data on case reports are less informative on current incidence rates. Uncertainty for estimates in the recent years is especially great when the data indicate that trend in incidence has changed.

Estimating HIV incidence in Colombia

The model was applied to Colombian data using the proposed method. Model estimates and data on HIV cases, AIDS cases and post-mortem detection are presented in Fig. 3a–c and resultant estimates of HIV incidence are presented in Fig. 3d. For Colombia, the model formulation with the lowest DIC was one with eight knots and hyper-prior $\tau^2 \sim U(1, 1.5)$.

The model has a good fit with the observed data (Fig. 3a–c) in the early years, whereas uncertainty is widely propagated in the recent years as noted in the experiments with synthetic data. The data for post-mortem-detected cases and AIDS cases exhibit complex patterns which the model is forced to reconcile with the clear monotonic increase in reported HIV cases and with our a priori belief that the incidence trajectory should vary smoothly over time.

The resulting estimate of HIV incidence (Fig. 3d), describes an early peak in new infections in 1990 and resurgence in the epidemic since 2000. The period with the highest incidence rate is estimated to be 2008–2009.

These incidence estimates can be compared with those previously presented by UNAIDS and which are based on an entirely different methodology whereby models are fitted to observed prevalence data and estimated sizes of key population (Fig. 4). In recent years, the estimates derived through the proposed method and the UNAIDS methods are in very close agreement. However, the UNAIDS-estimated historical trajectory of incidence suggests very high-peak incidence during 1995, of 0.3 per

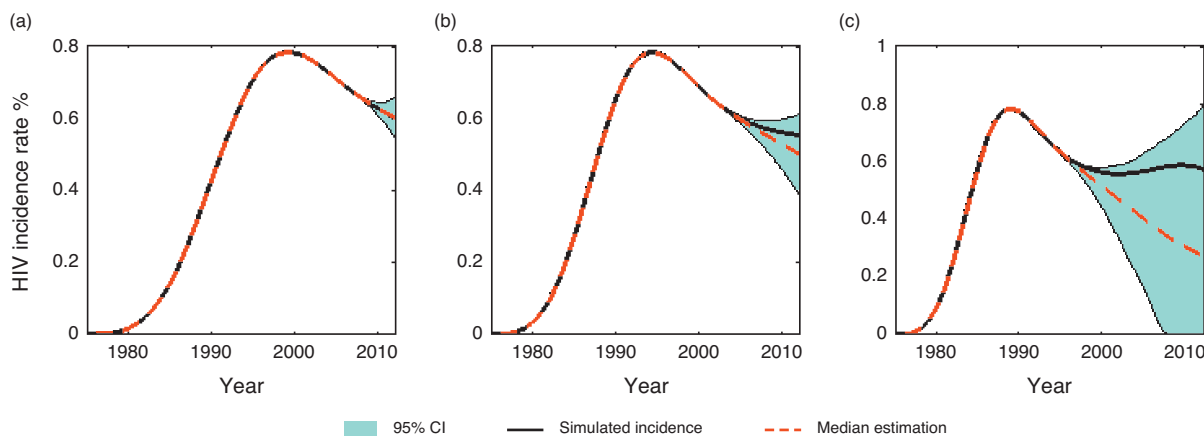


Fig. 2. Incidence estimation with three configurations of the spline function. In each panel, the black solid line is the true simulated incidence trend; the dashed red is the incidence estimation using the proposed method (median of the posterior distribution); the shaded blue region gives a 95% credible interval for incidence using the proposed method. The synthetic data were created using a model with (a) 7 knots, (b) 8 knots or (c) 10 knots.

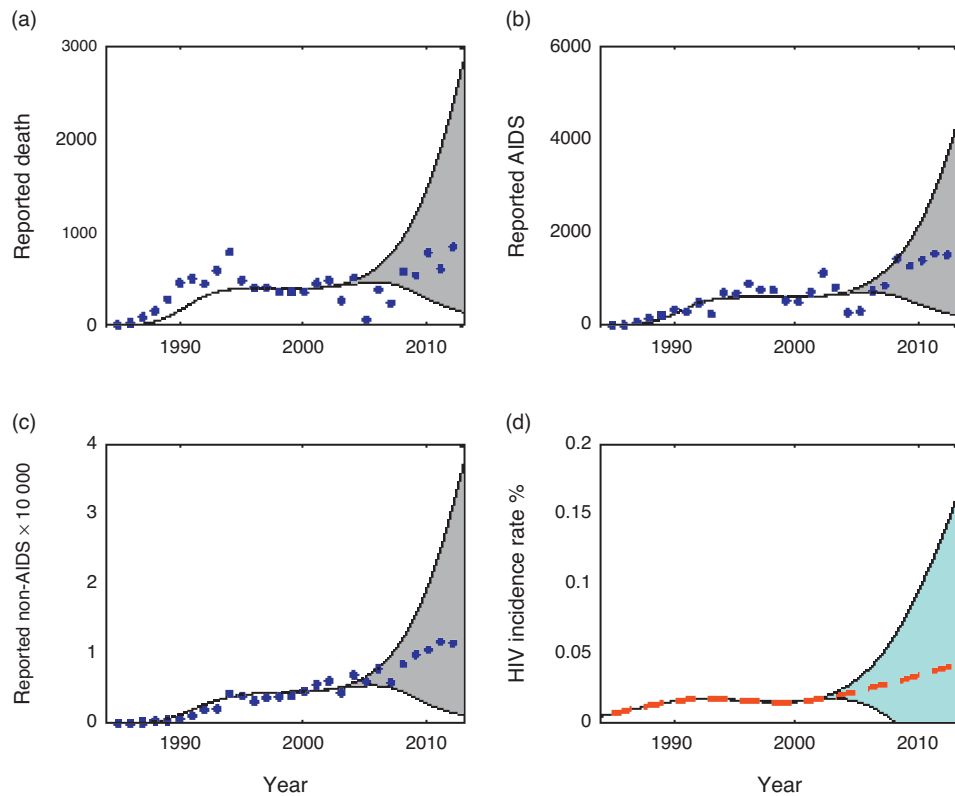


Fig. 3. Model fit to data in Colombia. Panels with data points (blue dots) as reported on a yearly basis by Colombian surveillance system, in three categories: (a) cases previously undiagnosed and only detected after death, (b) cases detected during AIDS stage and (c) cases detected and classified as non-AIDS at the time of diagnosis. In a–c, the grey shaded area are the posterior estimation from the model chosen for Colombia. (d) The resulting HIV incidence trajectory in Colombia from the proposed method (red dashed line) with 95% credible interval (shaded area).

100 person-years at risk (pyar), whereas the proposed method suggests an incidence rate much lower, at 0.03 per 100 pyar.

Discussion

We examined a method for estimating HIV incidence from case-report data. We have shown that it is capable of correctly recovering a wide range of possible incidence trajectories and have applied to data from Colombia to give complementary estimates of incidence to that derived from other methods. The present formulation gives highly uncertain estimated for most recent years, limiting its usefulness.

Although it is promising that our method can broadly identify a wide range of incidence trajectories, some small inconsistencies remained. We believe these will be because of the setting of the fixed placement of the knots and the rigid ordering of unconstrained and constrained knots. It will be possible to determine whether the evaluation of a wider range of models that incorporate these differences and model selection based

on DIC is sufficient to overcome these challenges. This would inevitably increase the computational demands of the estimation procedure and so alternative methods for estimating the model will also be explored.

It will also be useful to examine how this method works under different conditions of data availability. In the challenges with synthetic data, we assumed that all required data were present and unbiased. Although this may be a reasonable assumption for some settings, it will not be universally the case and small errors in early estimates of number of cases may propagate to large errors in the inferred incidence trajectory. We also used the same assumption for the natural history of HIV (survival rates and $CD4^+$ progression) in the method that generated the synthetic data as in the estimation model, which will inevitably flatter the performance of the model. It will be crucial to test the performance of the model when alternative assumptions are made and the model cannot be considered to have been fully tested until this step is done. A further step would be to examine the impact of the survival rates used to generate the synthetic data departing from those assumed in the estimation model, as well as propagating uncertainty in the assumptions about natural history in the estimation procedure. Finally, we did not

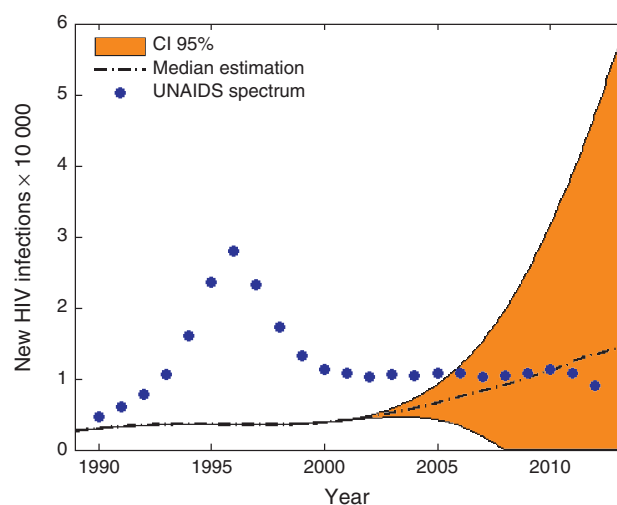


Fig. 4. Estimated number of new HIV infections in Colombia from the proposed method (red dashed lines) and current UNAIDS methods (blue dots). The orange shaded area shows the 95% credible interval of the estimates of the proposed method. Note that UNAIDS does compute uncertainty intervals for most of its statistics but intervals were not retrievable from online data sources for this statistic [22].

assess the impact of the use of inappropriate assumptions on the trend in diagnosis rates over time.

Here, we allowed HIV detection rates to vary by different $CD4^+$ cell count stage, according to a prior which reflects a wide spectrum of detection scenarios. It is possible that unlikely detection configurations are allowed within this approach, but the absence of evidence to support a more restrictive parameterization led us to opt for the more flexible assumption. Ideally, in a model-aggregation framework specific detection rates by group of transmission and stage of disease would be modelled separately, but this will only be feasible for those contexts with highly detailed case-report data and evidence of differential HIV detection and linkage to care.

This method gives uncertain results, especially in recent years, because there remains confounding by potential changes in the interval between infection and death and because numbers of cases and deaths are informative of incidence rates sometime in the past rather than current incidence. Therefore, in the current form, this method cannot be expected to reliably detect recent changes in incidence, which would be very important in monitoring a national epidemic. We expect, however, that it would be highly advantageous for this method to draw on further data. It will be possible to extend this method to incorporate data and/or estimates on deaths and numbers on antiretroviral therapy (ART), which should contribute to discriminating the earlier part of incidence trajectory in particular. Although it is not commonly used at present, the addition of measurement of biomarkers that relate to recency of infection at diagnosis

could also help discriminate current incidence levels, though this would depend on the characteristics of the biomarker and how well they are known (Bao *et al.* in this collection).

Our estimates of HIV incidence for Colombia must be interpreted with a high degree of caution and in light of our method's assumptions and present limitations. Of particular note, the estimates from the proposed model suggest that incidence was historically much lower than the levels the previous UNAIDS estimate indicate. This could stem from our assumption of complete case reporting and no misclassification of AIDS deaths: if, in fact, some diagnoses go unreported and some true AIDS deaths are misclassified, then our estimates will be too low. Nevertheless, it seems unlikely that these biases would fully explain the discrepancy. Added to this, we note that the level of incidence estimated here is in closer agreement with that recently reported by Murray *et al.* [23].

In conclusion, this method is promising technique for estimating HIV incidence trends that does not rely on using prevalence data and size estimates of key populations, and which leverages high quality routinely collected data. Future work should focus on updating the model structure to allow other forms of surveillance data, such as AIDS deaths and people on ART, and further scrutiny of model performance under circumstances of missing or biased availability of data.

Acknowledgements

T.B.H. thanks UNAIDS, The World Bank, The Rush Foundation and the Bill & Melinda Gates Foundation for funding support. A.C. thanks the NIH for funding through the NIAID cooperative agreement UM1 AI068619.

Conflicts of interest

There are no conflicts of interest.

References

1. Incidence Assay Critical Path Working Group. **More and better information to tackle HIV epidemics: towards improved HIV incidence assays.** *PLoS Med* 2011; **8**:e1001045.
2. Hogan DR, Zaslavsky AM, Hammit JK, Salomon JA. **Flexible epidemiological model for estimates and short-term projections in generalised HIV/AIDS epidemics.** *Sex Transm Infect* 2010; **86** (Suppl 2):ii84–ii92.
3. Brown T, Bao L, Raftery AE, Salomon JA, Baggaley RF, Stover J, *et al.* **Modelling HIV epidemics in the antiretroviral era: the UNAIDS Estimation and Projection package 2009.** *Sex Transm Infect* 2010; **86** (Suppl 2):ii3–ii10.
4. Centers for Disease Control and Prevention. *Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 U.S. dependent areas—2011.* Centers for Disease Control and Prevention; 2013.
5. Health Protection Agency. *HIV in the United Kingdom: 2011 Report.* London: Health Protection Agency; 2011.

6. Birrell PJ, Gill ON, Delpech VC, Brown AE, Desai S, Chadborn TR, *et al.* **HIV incidence in men who have sex with men in England and Wales 2001-10: a nationwide population study.** *Lancet Infect Dis* 2013; **13**:313–318.
7. Hall HI, Song R, Rhodes P, Prejean J, An Q, Lee LM, *et al.* **Estimation of HIV incidence in the United States.** *JAMA* 2008; **300**:520–529.
8. Sweeting MJ, De Angelis D, Aalen OO. **Bayesian back-calculation using a multistate model with application to HIV.** *Stat Med* 2005; **24**:3991–4007.
9. van Sighem A, Nakagawa F, Bezemer D, De Angelis D, Op de Coul E, Egger M, *et al.* **Estimating the Size of the Undiagnosed HIV Population in the Netherlands by Disease Stage [CROI abstract 975].** *Top Antivir Med* 2014; **22** (e-1):512–513.
10. Lodi S, Phillips A, Touloumi G, Geskus R, Meyer L, Thiébaud R, *et al.* **Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 Cells/mm³: assessment of need following changes in treatment guidelines.** *Clin Infect Dis* 2011; **53**:817–825.
11. Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, Mwita W, *et al.* **Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy.** *AIDS* 2007; **21** (Suppl 6):S55–S63.
12. Eilers PHC, Marx B. **Flexible smoothing with B-splines and penalties.** *Stat Sci* 1996; **11**:89–121.
13. Spiegelhalter DJ, Best N, Carlin B, van der Linde A. **Bayesian measures of model complexity and t.** *J R Stat Soc* 2002; **64**:583–640.
14. DANE. Estadísticas Vitales. http://www.dane.gov.co/index.php?option=com_content&view=article&id=73&Itemid=119. [Accessed 7 August 2014].
15. Hollingsworth TD, Anderson RM, Fraser C. **HIV-1 transmission, by stage of infection.** *J Infect Dis* 2008; **198**:687–693.
16. Cori A, Ayles H, Beyers N, Schaap A, Floyd S, Sabapathy K, *et al.* **HPTN 071 (PopART): a cluster-randomized trial of the population impact of an HIV combination prevention intervention including universal testing and treatment: mathematical model.** *PLoS One* 2014; **9**:e84511.
17. UNFPA. Panorama del VIH/sida en Colombia 1983-2010. Bogotá; 2012. <http://www.unfpa.org.co/uploadUNFPA/file/PANORAMA%20VIH%20SIDA%20COLOMBIA%201983-2010.pdf>. [Accessed 7 August 2014].
18. Castro J. VII Estudio nacional centinela de VIH en gestantes. Bogotá: Instituto Nacional de Salud, December 2009;1–79.
19. MPS/UNFPA. *Comportamiento sexual y prevalencia de VIH en hombres que tienen relaciones sexuales con hombres en siete ciudades de Colombia.* Bogotá: MPS/UNFPA; 2012.
20. García-Bernal R, Klaskala W, Castro J, Zhang G, Baum M. **HIV/AIDS surveillance in Colombia: regional differences in epidemic trends.** *AIDS* 1997; **11**:1297–1298.
21. Observatorio-Nacional-de-VIH-sida. *Informe de indicadores de atención integral por departamento y periodo.* Bogotá: Ministerio de la Protección Social; 2012.
22. UNAIDS. AIDSinfo Online Database. <http://aidsinfoonline.org/devinfo/libraries/asp/Home.aspx>. [Accessed 1 July 2014].
23. Evaluation IffHMa. Data Visualization Tool 2014. <http://vizhub.healthdata.org/mdg/>. [Accessed 7 August 2014].