OPEN

Incorporating incidence information within the UNAIDS Estimation and Projection Package framework: a study based on simulated incidence assay data

Le Bao[a], Jingyi Ye[a] and Timothy B. Hallett[b]

Objectives: The objective of this study is to extend the UNAIDS incidence estimation model, the UNAIDS Estimation and Projection Package (EPP), so that it can incorporate data from incidence assays.

Methods: We propose combining the likelihood of the incidence assay data with the likelihood of other data, in a manner that is consistent with the biomarker-based incidence estimator using incidence assay data. Two calibrating parameters specify the performance of the incidence assays: the false recent rate and the mean duration of recent infection. We then use synthetic data, based on prevalence data obtained from antenatal clinic surveillances, and in some cases household surveys, from 24 countries, to examine the impact of including incidence assay data, under circumstances wherein the incidence assay data imply the same or a different incidence rate as that inferred from the prevalence data alone, and wherein incorrect calibrating parameters for the incidence assay data are used.

Results: Using incidence assay data, in addition to prevalence data, can improve estimate by narrowing uncertainty intervals in derived HIV incidence estimates, and by providing information on levels or trends in incidence that were not apparent in the prevalence data alone. However, the effect is relatively modest if the sample size of the incidence assay survey is small and results can be biased if the calibrating parameters for the incidence assay data are not known accurately.

Conclusion: Incorporating information from incidence assays in the manner proposed has the potential to improve estimates. Further work will examine in more detail the circumstances under which the contribution of incidence assay data would be most valuable.

AIDS 2014, **28** (Suppl 4):S515–S522

Keywords: Bayesian melding, estimation of projection package, HIV/AIDS epidemic, incidence assay

Introduction

The incidence rate of HIV, or any infectious disease, is a crucial quantity that informs how fast the epidemic is spreading, and is thus used as part of prevention programming planning, evaluation and quantifying future programme needs. The most common estimates of HIV incidence rates for countries are from UNAIDS [1]. These estimates are inferred by using observed HIV prevalence data to fit a mathematical model, called the UNAIDS Estimation and Projection Package (EPP) [2–5], which incorporates assumptions on survival times with HIV and the scale-up of antiretroviral therapy (ART). However, this is only an indirect estimate rather than a

[a]Department of Statistics, the Pennsylvania State University, University Park, Pennsylvania, USA, and [b]Department of Infectious Disease Epidemiology, Imperial College London, London, UK.

Correspondence to Le Bao, Department of Statistics, the Pennsylvania State University, 323 Thomas Building, University Park, PA 16802, USA.

Tel: +1 814 865 1348; fax: +1 814 863 7114; e-mail: lebao@psu.edu
Received: 4 August 2014; revised: 4 August 2014; accepted: 4 August 2014.

DOI:10.1097/QAD.0000000000000434

ISSN 0269-9370

direct measurement and has many limitations such as not being able to pick-up trends in HIV incidence rapidly and being vulnerable to biases wrought by inappropriate assumptions in the model.

One other method of estimating incidence is the use of 'incidence assays' [6]. These are laboratory assays that return a biomarker that is related to the time that a person has been infected. If the nature of that relationship is known, then the incidence assay can be applied to persons who are HIV-positive in a cross-sectional survey to form an estimate of HIV incidence, relating to the period of time immediately prior to the survey. The relationship between the biomarker and the time since infection can be summarized, for the purposes of estimating HIV incidence by two parameters: the false recent rate, which is the proportion of persons infected for at least 450 days that return a positive biomarker in the incidence assay; and the mean duration of recent infection, which is the mean time spent after infection for which individuals return a positive biomarker in the incidence assay [7,8]. Low false recent rates and mean duration of recent infection on the order of 1 year may be considered the most desirable properties.

Many assays have been proposed for this use [6]. One assay, the BED [9], was used in a number of surveys but was found to have a high false recent rate that could lead to erroneous estimates of incidence [10]. There are, however, newer assays, which are thought to have properties more suitable to estimating incidence in this way and the characteristics of these assays are being measured using a large range of samples (http://www.incidence-estimation.com/page/cephia) and so will be better quantified. One assay in particular – Limiting-Antigen Avidity Assay [11] – has already been used in a national survey in South Africa [12].

It would be desirable for data from incidence assays to contribute to the overall estimates of the epidemic computed by EPP. This would allow the prevalence data, assumptions about survival and data from incidence assays to join forces to produce better estimates. Here, we describe a new approach that shows how information from incidence assays can be incorporated into EPP and examine how these data may affect HIV estimates under different scenarios for the availability of data from incidence assays.

Materials and methods

The EPP fits an epidemiologic model to prevalence data obtained from sentinel surveillance systems and national population-based HIV sero-prevalence surveys. It produces the size of infected population $Y(t)$, the size of uninfected population $Z(t)$, the prevalence $\rho(t)$ and the infection rate $I(t)$. All population counts refer to adults aged 15–49 years and $N(t) = X(t) + Y(t) + Z(t)$. The rates at which the group sizes change are described by ordinary differential equations:

$$\begin{cases} \dfrac{dZ(t)}{dt} &= E(t) - \dfrac{r(t)Y(t)Z(t)}{N(t)} - \mu Z(t) - \dfrac{\alpha_{50}(t)Z(t)}{N(t)} + \dfrac{M(t)Z(t)}{N(t)}, \\ \dfrac{dY(t)}{dt} &= \dfrac{r(t)Y(t)Z(t)}{N(t)} - \text{HIV death} - \dfrac{\alpha_{50}(t)Y(t)}{N(t)} + \dfrac{M(t)Y(t)}{N(t)}, \end{cases} \quad (1)$$

where $E(t)$ is the number of new adults entering the model at time t, $a_{50}(t)$ is the number of people leaving the model because of reaching age 50, μ is the non-AIDS mortality rate and $M(t)$ is the rate of net migration into the population; all of these parameters are defined by an external life-table. The infection rate, $r(t)$, can be specified in various functional forms such as a random walk [13,14], a spline [15,16] or a systematic mean structure (r-trend model) [17]. In this article, we use the r-trend model for illustrations, but our proposed approach is generic across all formulations.

The r-trend model assumes that the epidemic starts at t_0 with an initial infection rate r_0 and that the yearly change in the infection rate $r(t)$ can be described as:

$$\log r(t+1) - \log r(t) = \theta_1 \times (\theta_0 - r(t)) + \theta_2 \rho(t) + \theta_3 \gamma(t), \quad (2)$$

where the yearly change in $r(t)$ depends on other key factors: the current infection rate $r(t)$, the prevalence $\rho(t)$ and a tendency for $r(t)$ to reach a steady state t_1 years after t_0. Informative prior distributions are placed on the parameters, which are based on posterior values obtained from previously applying the model to epidemic data in a large number of countries:

$$\begin{cases} t_0 \sim U[1970, 1990] \\ t_1 \sim N(20, 4.5) \\ \log r_0 \sim N(0.42, 0.23) \\ \theta_0 \sim N(0.46, 0.12) \\ \theta_1 \sim N(0.17, 0.07) \\ \theta_2 \sim N(-0.68, 0.24) \\ \theta_3 \sim N(-0.038, 0.009) \end{cases} \quad (3)$$

Once the above input parameters are given, the dynamic model produces the HIV prevalence and incidence at each year. The likelihood of prevalence data collected from antenatal clinics (ANCs) is then defined by a

hierarchical model that takes the clinic effects into account [18]. Moreover, one additional parameter is used to describe the bias in ANC data with respect to prevalence data from National Population Based Surveys (NPBS), for example Demographic and Health Surveys (DHS) [19]. In keeping with the current approach to including such data, the prevalence data from NPBS are viewed as representative for the general population, so that the bias parameter is only involved in the likelihood of ANC data but not NPBS data as shown below [17].

$$\Phi^{-1}(\mathrm{X}_{\mathrm{ANC,st}}) = \Phi^{-1}(\rho_{\mathrm{t}}) + \mathrm{b_s} + \varepsilon_{\mathrm{st}}, \tag{4}$$

$$\Phi^{-1}(\mathrm{X}_{\mathrm{NPBS,t}}) = \Phi^{-1}(\rho_{\mathrm{t}}) + \varepsilon_{\mathrm{NPBS,t}} \tag{5}$$

where $\mathrm{X}_{\mathrm{ANC,st}}$ is the observed prevalence from ANC clinics at time t, $\mathrm{X}_{\mathrm{NPBS,t}}$ is the observed prevalence from national population-based surveys at time t, Φ^{-1} is the inverse of standard normal cumulative distribution function, $\mathrm{b_s}$ is the ANC bias parameter and ε's are independent normal errors.

We assume that newly available incidence assay data will be collected from the same household surveys used to furnish overall prevalence estimates. The characteristics of the incidence assay are the false recent rate β_{T} and Ω_{T}, which is the mean duration of recent infection. The data provided by the use of the incidence assays at time t include:

(1) $N_S(t)$: The number of HIV-negative individuals,
(2) $N_{NR}(t)$: The number of nonrecently infected HIV-positive individuals,
(3) $N_R(t)$: The number of recently infected individuals.

Following [8], we assume that (N_S, N_{NR}, N_R) follow a trinomial distribution with the following expected values, and the time index t is ignored in the expressions for simplicity:

$$E(N_S) = N(1-\rho) \tag{6}$$

$$E(N_{NR}) = N\rho - N(1-\rho)I(\Omega_T - \beta_T T)/365 - N\beta_T\rho \tag{7}$$

$$E(N_R) = N(1-\rho)I(\Omega_T - \beta_T T)/365 + N\beta_T\rho \quad \text{(see Kassanjee } et\ al. \text{ equation 25)} \tag{8}$$

where ρ and I are prevalence and incidence produced by the EPP model, respectively.

The log likelihood for these data furnished by the incidence assay is therefore:

$$\begin{aligned}\log L\_inc = {} & N_s\log(1-\rho) + N_{NR}\log[\rho - (1-\rho)I \\ & (\Omega_T - \beta_T T)/365 - \beta_T\rho] \\ & + N_R\log[(1-\rho)I(\Omega_T - \beta_T T)/365 + \beta_T\rho]\end{aligned} \tag{9}$$

And, the log-likelihood of the prevalence data and the incidence assays combined is the sum of the two log-likelihoods.

To demonstrate the use of this approach, we generate synthetic data sets that do and do not include data from incidence assays and compare the estimates of incidence so derived. We base our synthetic data construction on the data from urban areas of the following countries,

(1) Eastern Africa: Burundi, Ethiopia, Eritrea, Kenya, Malawi, Rwanda, Tanzania, Uganda, Zambia;
(2) Central Africa: Cameroun, RCA, Chad, Congo, RDC, Equatorial Guinea, Gabon;
(3) Southern Africa: Botswana, Lesotho, Namibia, Zimbabwe;
(4) Western Africa: Benin, Burkina Faso, RCI, Gambia, Ghana, Guinee, Liberia, Mali, Sierra Leone, Togo.

ANC data and NPBS data are available in these countries starting from 1990–2000 and up to 2010. We assume that the incidence assays become available in the most recent year, or the year before, that prevalence data were also available.

Our approach to comparison of model results is as follows:

(1) Fit the EPP model using only prevalence data. For each posterior sample of EPP output, read off the incidence and prevalence at 2010, and use them for simulation.
(2) Set $\beta_{\mathrm{T}} = 2.5\%$ and $\Omega_{\mathrm{T}} = 150$. Simulate incidence assay data from the trinomial distribution following equations (4–6) with an assumed sample size of 5000.
(3) Refit the EPP model with incidence data.
(4) Compare the results with/without incidence data.

The above simulation offers a scenario wherein the incidence assay data 'agree' with surveillance data and will be referred as 'Scenario I' thereafter. We also simulate two more scenarios to mimic the cases that the incidence assay data imply a different incidence rate: either double the inferred incidence (Scenario II) or half the inferred incidence (Scenario III). All the scenarios are then repeated with sample sizes of 10 000 and 20 000 to investigate the influence of having larger surveys.

Finally, we conduct a sensitivity analysis to understand how incorrect estimates β_{T} and Ω_{T} would impact our results by fixing the values of β_{T} and Ω_{T} at different values to those used for generating the synthetic incidence assay data. Instead of using the EPP model, the simulated data are produced using values of prevalence = 20%, incidence = 2%, $\beta_{\mathrm{T}} = 2.5\%$ and $\Omega_{\mathrm{T}} = 150$. In the estimation process, we specify the following five scenarios:

(1) Set $\beta_{\mathrm{T}} = 2.5\%$ and $\Omega_{\mathrm{T}} = 150$, (all parameters are correct)
(2) Set $\beta_{\mathrm{T}} = 1.5\%$ and $\Omega_{\mathrm{T}} = 150$, ($\beta_{\mathrm{T}}$ too low)

(3) Set $\beta_T = 3.5\%$ and $\Omega_T = 150$, (β_T too high)
(4) Set $\beta_T = 2.5\%$ and $\Omega_T = 130$, (Ω_T too low)
(5) Set $\beta_T = 2.5\%$ and $\Omega_T = 170$. (Ω_T too high).

For fixed β_T and Ω_T, we can approximate the distributions of prevalence and incidence at the time of the incidence assays by using the maximum likelihood. Note that the above analysis uses neither the EPP model nor any real prevalence data because that information will complicate the relationships between parameters involved in incidence assays: prevalence, incidence, β_T and Ω_T, and make the effect of using wrong values harder to summarize.

We emphasize that the use of data from countries is for illustration only and results should therefore not be seen as replacing official estimates published by those countries or by UNAIDS.

Results

We examine the proposed method of incorporating incidence assay data using the antenatal clinic (ANC), first assuming that the calibrating parameters of incidence assays are perfectly known. Table 1 summarizes the posterior median and standard deviation of prevalence and incidence estimates based on the combination of ANC data and other possibly available data sources such as NPBS and incidence assay data. If we assume that the incidence assays estimate the same HIV incidence level as implied by the EPP model with prevalence data (Scenario I), then introducing incidence assay data narrows the uncertainty bounds, especially when the incidence rate at 2010 is high and NPBS are not available, although the posterior median of prevalence and incidence in 2010 remain the same (Fig. 1 for six selected countries, Table 1 left two columns, and Appendix, http://links.lww.com/QAD/A578). If the estimates from a country already benefit from additional prevalence data from NPBS, then the overall uncertainty of prevalence and incidence is reduced compared with when only ANC data were used. In this case, the incidence assay data can still make a contribution by reducing uncertainty overall, albeit by a lesser amount (Fig. 1 for six selected countries, Table 1, right two columns, and Appendix, http://links.lww. com/QAD/A578). This effect is relatively powerful, reducing the standard deviation of the posterior samples on incidence in 2010 from an average of 0.15–0.07% when NPBS data are not available. When NPBS data are available, the effect is more modest, reducing the standard deviation from an average of 0.08–0.06%.

In Scenarios II and III, the incidence assay data imply that incidence in 2010 is double or half that which would be suggested by the prevalence data alone, respectively. Figure 2 compares the estimated prevalence and incidence trajectories for Lesotho with and without using incidence assay data. The model fits find a balance between the

Table 1. Scenario I: the prevalence and incidence estimates at 2010 with or w/o National Population Based Surveys data and incidence assays in form of posterior median (standard deviation).

	w/o NPBS				With NPBS			
	w/o incidence assays		With incidence assays		w/o incidence assays		With incidence assays	
	Prevalence	Incidence	Prevalence	Incidence	Prevalence	Incidence	Prevalence	Incidence
Benin	2.8 (0.3)	0.2 (0)	2.9 (0.2)	0.2 (0)	2 (0.2)	0.1 (0)	2 (0.1)	0.1 (0)
Botswana	34.1 (2.4)	2.2 (0.4)	34.3 (0.7)	2.2 (0.2)	22.6 (0.5)	0.9 (0.1)	22.6 (0.4)	0.9 (0.1)
BurkinaFaso	3.1 (0.4)	0.2 (0)	3.1 (0.2)	0.2 (0)	2.8 (0.3)	0.2 (0)	2.8 (0.2)	0.1 (0)
Burundi	5.4 (1.2)	0.3 (0.1)	5.4 (0.3)	0.3 (0)	5.6 (0.4)	0.3 (0.1)	5.6 (0.3)	0.3 (0)
Cameroun	9.7 (1)	0.8 (0.2)	9.7 (0.4)	0.8 (0.1)	5.9 (0.4)	0.5 (0.1)	5.9 (0.3)	0.5 (0.1)
Chad	4.7 (0.7)	0.4 (0.1)	4.8 (0.3)	0.4 (0.1)	5.5 (0.5)	0.4 (0.1)	5.5 (0.3)	0.4 (0.1)
Congo	5.9 (0.9)	0.5 (0.1)	5.9 (0.3)	0.5 (0.1)	3.3 (0.2)	0.2 (0)	3.3 (0.2)	0.2 (0)
Ethiopia	6.3 (0.6)	0.2 (0)	6.3 (0.3)	0.2 (0)	4.4 (0.3)	0.1 (0)	4.5 (0.2)	0.1 (0)
Ghana	3.7 (0.4)	0.3 (0)	3.7 (0.2)	0.3 (0)	2.5 (0.2)	0.2 (0)	2.5 (0.2)	0.2 (0)
Guinea Ecuatorial	8.9 (1.3)	0.7 (0.2)	9 (0.4)	0.7 (0.1)	7.3 (1.1)	0.6 (0.2)	7.3 (0.3)	0.6 (0.1)
Kenya	7.5 (0.9)	0.4 (0.1)	7.4 (0.3)	0.3 (0)	6.8 (0.3)	0.3 (0.1)	6.8 (0.2)	0.3 (0.1)
Lesotho	35.3 (2.7)	4.7 (0.7)	35.4 (0.7)	4.7 (0.3)	28 (0.9)	3.3 (0.3)	28 (0.6)	3.1 (0.3)
Liberia	4.5 (0.6)	0.2 (0.1)	4.5 (0.3)	0.2 (0.1)	2.4 (0.2)	0.1 (0)	2.4 (0.2)	0.1 (0)
Malawi	15.6 (1.3)	0.9 (0.2)	15.7 (0.5)	0.9 (0.1)	15.5 (0.5)	1 (0.1)	15.5 (0.4)	1 (0.1)
Mali	3.8 (0.5)	0.2 (0.1)	3.8 (0.2)	0.2 (0)	2.1 (0.2)	0.1 (0)	2.1 (0.1)	0.1 (0)
RCA	10 (1.1)	0.8 (0.2)	10.1 (0.4)	0.8 (0.1)	6.5 (0.4)	0.4 (0.1)	6.5 (0.3)	0.4 (0.1)
RCI	5.3 (0.6)	0.2 (0.1)	5.4 (0.3)	0.2 (0)	4.1 (0.3)	0.2 (0)	4.1 (0.2)	0.2 (0)
RDC	4.7 (0.4)	0.4 (0.1)	4.7 (0.3)	0.4 (0)	2.9 (0.3)	0.3 (0)	2.9 (0.2)	0.3 (0)
Rwanda	7.4 (0.8)	0.3 (0.1)	7.4 (0.4)	0.3 (0)	6.8 (0.5)	0.2 (0.1)	6.8 (0.3)	0.2 (0)
SierraLeone	5 (0.6)	0.3 (0.1)	5 (0.3)	0.3 (0.1)	2.9 (0.3)	0.2 (0.1)	3 (0.2)	0.2 (0)
Tanzania	9.9 (0.8)	0.9 (0.1)	9.8 (0.4)	0.9 (0.1)	8.7 (0.3)	0.8 (0.1)	8.7 (0.3)	0.8 (0.1)
Uganda	9.8 (0.9)	0.8 (0.1)	9.8 (0.4)	0.8 (0.1)	9.7 (0.6)	0.8 (0.1)	9.6 (0.3)	0.8 (0.1)
Zambia	25.3 (1.7)	2.1 (0.3)	25.2 (0.6)	2.1 (0.1)	19.9 (0.6)	1.4 (0.1)	19.8 (0.4)	1.4 (0.1)
Zimbabwe	15.1 (1.6)	0.6 (0.2)	15.3 (0.5)	0.6 (0.1)	12.7 (0.5)	0.4 (0.1)	12.6 (0.4)	0.4 (0.1)

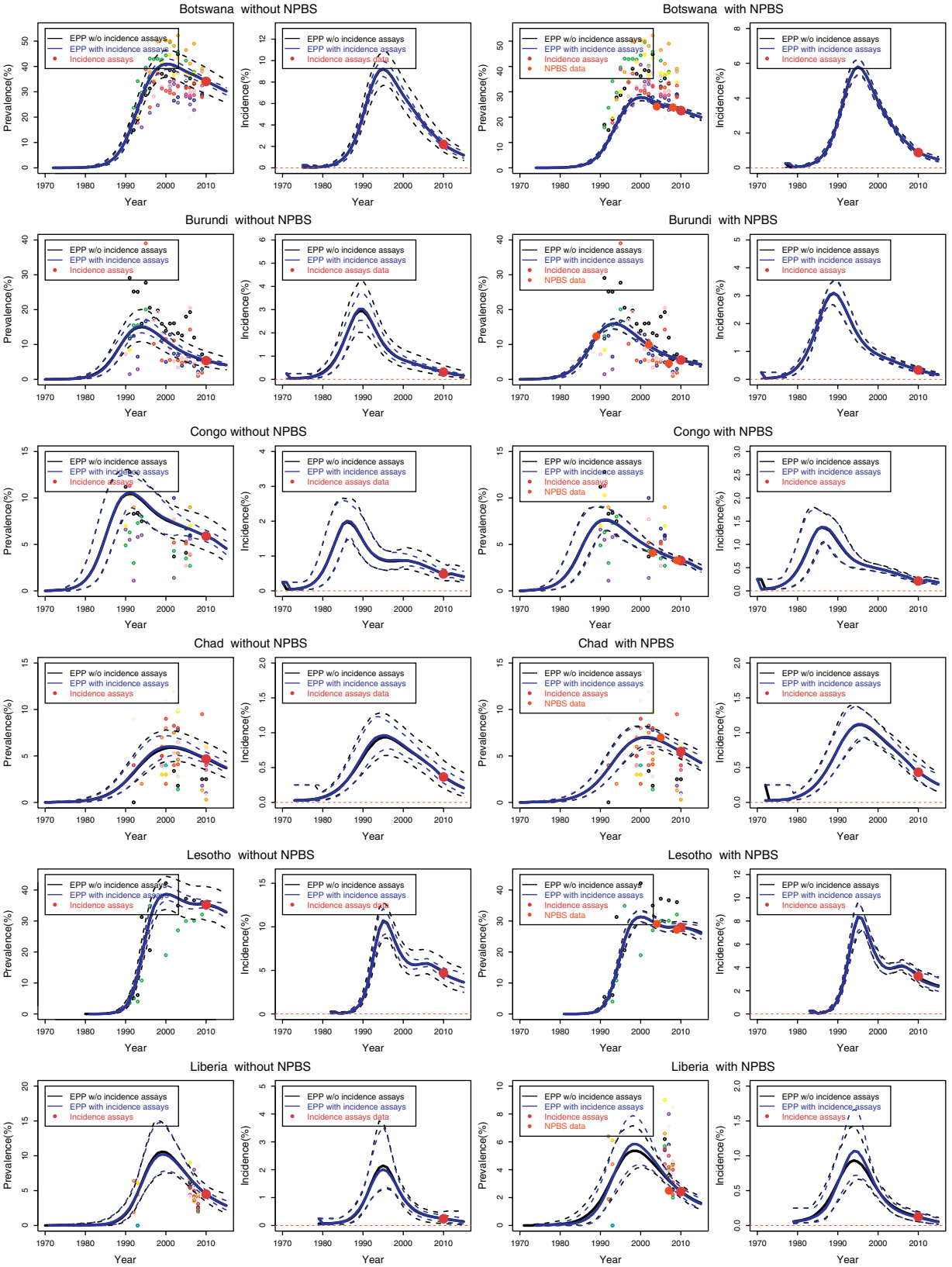

Fig. 1. Scenario I: the prevalence and incidence trajectories using incidence assays (blue) and not using incidence assays (black) when incidence assay data are consistent with surveillance data. The solid lines are posterior median, the dashed lines are 95% credible intervals, the coloured dots are surveillance data collected from different clinics, the big red dots are the prevalence data collected from NPBS and the big brown dots are the prevalence and incidence in 2010 used for generating incidence assay data.

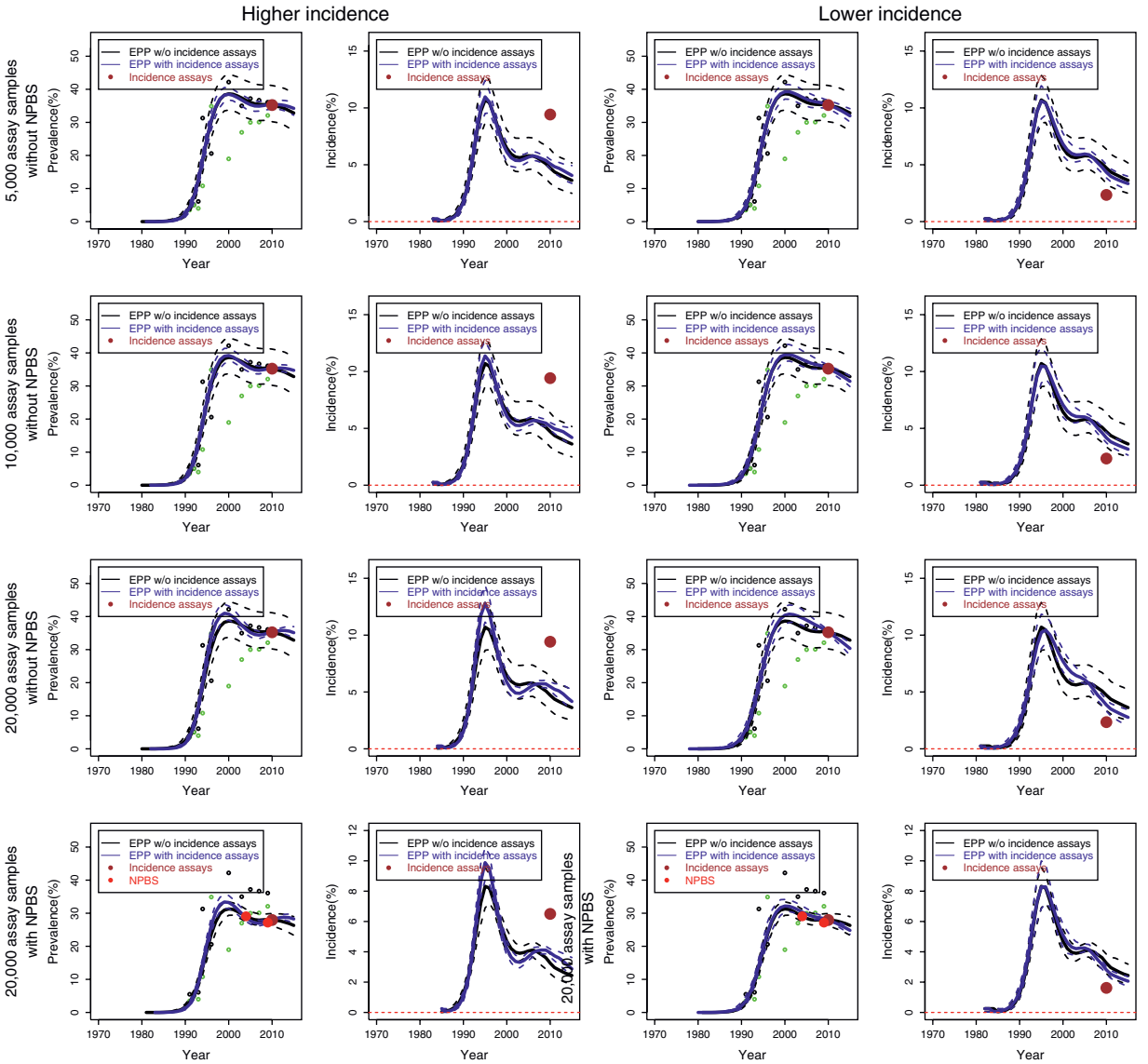

Fig. 2. Scenarios II and III: The prevalence and incidence trajectories using incidence assays (blue) and not using incidence assays (black). The left two columns show prevalence and incidence when incidence assay data are simulated from a higher incidence. The right two columns show prevalence and incidence when incidence assay data are simulated from a lower incidence. The sample sizes of incidence assays vary across rows from 5000 to 20 000. The solid lines are posterior median, the dashed lines are 95% credible intervals, the coloured dots are surveillance data collected from different clinics, the big red dots are the prevalence data collected from NPBS and the big brown dots are the prevalence and incidence in 2010 used for generating incidence assay data.

incidence levels implied by the prevalence data and that implied by incidence assays, with that balance being towards incidence assays when the sample size of incidence assays is increased. With the NPBS present, the model tries to find the best way of representing the joint information from ANC, NPBS and incidence assays, and the incidence assay data will still contribute to determining the trajectories of incidence and prevalence as shown in Fig. 2, bottom row.

Finally, we evaluate the effects of using incorrect calibrating parameters for incidence assays (Table 2). When β_T and Ω_T are underestimated (B and D), the incidence assay data lead to an overestimate of incidence; when β_T and Ω_T are overestimated (C and E), we underestimate incidence.

Discussion

We have developed an approach by which data from incidence assays can be incorporated into the calibration of mathematical models of HIV, and this will be a useful

Table 2. The prevalence and incidence distributions derived from incidence assays, summarized by mean (standard deviation), under prevalence, 20%; incidence, 2%; and different fixed values of β_T and Ω_T.

	5000 assay samples		10 000 assay samples		20 000 assay samples	
Scenario	Prevalence	Incidence	Prevalence	Incidence	Prevalence	Incidence
A. True value	20.0 (0.57)	1.98 (0.48)	20.0 (0.40)	2.01 (0.34)	20.0 (0.28)	2.01 (0.24)
B. β too low	20.0 (0.57)	2.56 (0.47)	20.0 (0.40)	2.58 (0.33)	20.0 (0.28)	2.58 (0.24)
C. β too high	20.0 (0.57)	1.36 (0.49)	20.0 (0.40)	1.39 (0.35)	20.0 (0.28)	1.40 (0.25)
D. Ω too low	20.0 (0.57)	2.30 (0.56)	20.0 (0.40)	2.34 (0.40)	20.0 (0.28)	2.34 (0.28)
E. Ω too high	20.0 (0.57)	1.73 (0.42)	20.0 (0.40)	1.75 (0.30)	20.0 (0.28)	1.75 (0.21)

addition to the EPP model in those settings wherein incidence assays have been collected. The formulation allows EPP estimates of incidence, as well as other epidemiological quantities, to benefit from this additional source of information. We have also demonstrated the use of this approach on simulated datasets. The impact of incidence assay data depends on both sample size and reliable estimates of the false recent rate, β_T, and the mean duration of recent infection, Ω_T. The approach will also be suitable for application in other models of HIV.

The incidence assay data are assumed to be generated from a trinomial distribution. We also considered modelling the joint distribution of incidence and prevalence by a bivariate normal distribution, with the mean and standard error estimated from assays. We found that the posterior distribution of incidence or prevalence can be highly skewed if its point estimate is close to 0%, and the normal approximation could be problematic because the normal distribution is symmetric and may take values less than zero or greater than one. Therefore, we selected the trinomial distribution for incidence assays. However, our approach is generic in the sense that it incorporates the information from incidence assays as an additional part of likelihood function. Other ways of characterizing the performance of the incidence assays can be used if they are properly defined.

As shown in Fig. 2, the estimated incidence curve from EPP is nearly identical whether the assay incidence estimate doubled or halved for some settings and when the sample size for the incidence assay is small. Again, the model tries to find a balance between the information carried by the surveillance data, survey data and incidence assays. The identical estimates of incidence with or without using incidence assays suggest that the incidence assays are not informative enough to suggest a different incidence estimate. However, developments of incidence assays and measurement of appropriate parameters will continue, so that the impact of using data from incidence may become greater. In addition, although the effects of the incidence assays are reasonably modest under assumptions of reflecting expected survey size and calibration parameters for the incidence assays – particularly when household surveys already furnish prevalence data – it is nevertheless useful to incorporate all available information into estimates. The impact of incidence assays is also weakened because with a short recency period for the incidence assay, few individuals will return a positive biomarker. We define the recency duration as 150 days as indicative because work measuring the recency duration for LAg is ongoing and the approach is also generic to an assay with any recency duration. It is indeed the case that the recency duration has a profound impact on how informative the incidence assay is on the HIV incidence rate, so that surveys with much greater sample sizes are required when using incidence assays with shorter recency duration to achieve the same precision in estimates [7].

On one hand, we have accentuated the impact of incidence assays uses by assuming that we have complete knowledge about the calibrating parameters β_T and Ω_T. In fact, we do not have such good knowledge (http://www.incidence-estimation.com/page/cephia) and the false recent rate, for instance, might be expected to vary by time and by place [20]. And as our final analysis shows, uncertainty in those parameters substantially reduces the information content of incidence assays.

On the other hand, several aspects of our analysis count against the usefulness of incidence assay data. We did not look at how the addition of incidence assay data affects the ability of this modelling method to detect changes in HIV incidence over time. That advantage may be highly meaningful because methods based on inferring incidence only from prevalence cannot reliably detect changes in incidence until many years after the fact. Also, the uncertainty in the estimates of incidence without the use of incidence assays is understated because uncertainty in the natural history of HIV is not represented, and nor is the substantial uncertainty about the level of ART use and its effects on incidence and survival. This is the uncertainty that the data from the incidence assay, being an independent source of information, may partly help resolve. Thus, our analysis could have underestimated the ability of data from the incidence assay to reduce uncertainty in incidence estimates. Each of these issues will be investigated further.

Our approach does not mix the estimation of β_T and Ω_T with the estimation of EPP parameters. If one wants to implement uncertainty analysis of β_T and Ω_T, informative priors for β_T and Ω_T should be used to avoid a failure

to converge, as those four parameters are not identifiable given the trinomial incidence assay data: prevalence, incidence, β_T and Ω_T.

The performance of this new approach should be further tested with actual incidence data. It would be also interesting to explore the minimum sample sizes and accuracy of β_T and Ω_T that allow incidence assays to successfully detect recent changes in incidence.

Acknowledgements

This research was supported by the Joint United Nations Programme on HIV/AIDS and The Bill & Melinda Gates Foundation. T.B.H. thanks UNAIDS, The World Bank and The Bill & Melinda Gates Foundation for funding support. The authors are grateful to Alex Welte, Hilmarie Brand, Runze Li, Xiaoyue Niu, Peter Ghys, Mary Mahy, Kelsey Case, Tim Brown, John Stover and the reviewers and editor of this supplement for helpful comments and discussions.

Conflicts of interest

There are no conflicts of interest.

References

1. Joint United Nations Programme on HIV/AIDS (UNAIDS). Global report: UNAIDS report on the global AIDS epidemic 2013. UNAIDS, Geneva, Switzerland.
2. Ghys PD, Brown T, Grassly NC, Garnett G, Stanecki KA, Stover J, Walker N. **The UNAIDS estimation and projection package: a software package to estimate and project national HIV epidemics.** *Sex Trans Infect* 2004; **80**:i5–i9.
3. Brown T, Grassly NC, Garnett G, Stanecki K. **Improving projections at the country level: the UNAIDS Estimation and Projection Package 2005.** *Sex Trans Infect* 2006; **82 (Suppl 3)**:iii34–iii40.
4. Brown T, Bao L, Raftery AE, Salomon JA, Baggaley RF, Stover J, Gerland P. **EPP 2009: bringing the UNAIDS Estimation and Projection Package into the ART era.** *Sex Trans Infect* 2010; **86 (Suppl 2)**:ii3–ii10.
5. Brown T, Bao L, Eaton JW, Hogan DR, Mahy M, March K, *et al*. **Improvements in prevalence trend fitting and incidence estimation in EPP 2013.** *AIDS* 2014; **28 (Suppl 4)**:S415–S425.
6. Incidence Assay Critical Path Working Group.. **More and better information to tackle HIV epidemics: towards improved HIV incidence assays.** *PLoS Med* 2011; **8**:e1001045.
7. Welte A, McWalter TA, Laeyendecker O, Hallet TB. **Using tests for recent infection to estimate incidence: problems and prospects for HIV.** *Euro Surveill* 2010; **15**:19589.
8. Kassanjee R, McWalter TA, Bärnighausen T, Welte A. **A new general biomarker-based incidence estimator.** *Epidemiology* 2012; **23**:721–728.
9. Parekh BS, McDougal JS. **New approaches for detecting recent HIV-1 infection.** *AIDS Rev* 2001; **3**:183–193.
10. Bärnighausen T, McWalter TA, Rosner Z, Newell ML, Welte A. **HIV incidence estimation using the BED capture enzyme immunoassay: systematic review and sensitivity analysis.** *Epidemiology* 2010; **21**:685–697.
11. Duong YT, Qiu M, De AK, Jackson K, Dobbs T, Kim AA, *et al*. **Detection of recent HIV-1 infection using a new limiting-antigen avidity assay: potential for HIV-1 incidence estimates and avidity maturation studies.** *PloS One* 2012; **7**:e33328.
12. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, *et al*. *South African national HIV prevalence, incidence and behaviour survey, 2012*. Cape Town: HSRC Press; 2014.
13. Bao L, Raftery AE. **A stochastic infection rate model for estimating and projecting national HIV prevalence rates.** *Sex Trans Infect* 2010; **86 (Suppl 2)**:ii93–ii99.
14. Bao L, Salomon J, Brown T, Hogan DR. **Modeling HIV/AIDS epidemics in the UNAIDS Estimation and Projection Package 2011.** *Sex Trans Infect* 2012; **88 (Suppl 2)**:i3–i10.
15. Hogan D, Zaslavsky A, Hammitt J, Salomon JA. **Flexible epidemiological model for estimates and short-term projections in generalised HIV/AIDS epidemics.** *Sex Trans Infect* 2010; **86 (Suppl 2)**:ii84–ii92.
16. Hogan DR, Salomon JA. **Spline-based modelling of trends in the force of HIV infection, with application to the UNAIDS Estimation and Projection Package.** *Sex Trans Infect* 2012; **88 (Suppl 2)**:i52–i57.
17. Bao L. **A new infectious disease model for estimating and projecting HIV/AIDS epidemics.** *Sex Trans Infect* 2012; **88 (Suppl 1)**:i58–i64.
18. Alkema L, Raftery AE, Clark SJ. **Probabilistic projections of HIV prevalence using Bayesian melding.** *Ann Appl Stat* 2007; **1**:229–248.
19. Marsh K, Mahy M, Salomon JA, Hogan DR. **Assessing and adjusting for differences between HIV prevalence estimates derived from national population-based surveys and antenatal care surveillance, with applications for Spectrum 2013.** *AIDS* 2014; **28 (Suppl 4)**:S497–S505.
20. Hallett TB, Zaba B, Todd J, Lopman B, Mwita W, Biraro S, *et al*. **Estimating incidence from prevalence in generalised HIV epidemics: methods and validation.** *PLoS Med* 2008; **5**:611–622.