Published in final edited form as:

Stat Med. 2014 May 10; 33(10): 1767–1783.

Mixture models for calibrating the BED for HIV incidence testing

Severin Guy Mahiane^{1,2,#}, Agnès Fiamma^{3,4}, and Bertran Auvert^{5,6}

Severin Guy Mahiane: mahianes@yahoo.fr; Agnès Fiamma: afiamma@ucla.edu; Bertran Auvert: bertran.auvert@uvsq.fr

¹Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA

²SACEMA, Stellenbosch University, Stellenbosch, South Africa

³UCLA Program in Global Health, University of California, Los Angeles, USA

⁴CESP INSERM-UVSQ UMRS-1018, Villejuif, France

⁵Université de Versailles-Saint-Quentin, France

⁶Institut Universitaire de France (IUF), Paris, France

Abstract

A number of antibody biomarkers have been developed to distinguish between recent and established Human Immunodeficiency Virus (HIV) infection and used for HIV incidence estimation from cross-sectional specimens. In general, a cut-off value is specified and estimates of the following parameters are needed: a) the mean time interval (w) between seroconversion and reaching that cut-off, b) the probability of correctly identifying individuals who became infected in the last w years (sensitivity) and c) the probability of correctly identifying individuals who have been infected for more than w years (specificity). We develop two statistical methods to study the distribution of a biomarker and derive a formula for estimating HIV incidence from a cross sectional survey. Both methods allow handling interval censored data and basically consist of using a generalised mixture model to model the growth of the biomarker as a function of time since infection. The first uses data from all followed-up individuals and allows incidence estimation in the cohort while the second only uses data from seroconverters. We illustrate our methods using repeated measures of the IgG capture BED enzyme immunoassay. Estimates of calibration parameters i.e. mean window period, mean recency period, sensitivity and specificities obtained from both models are comparable. The formula derived for incidence estimation gives the maximum likelihood estimate of incidence which, for a given window period, depends only on sensitivity and specificity. The optimal choice of the window period is discussed. Numerical simulations suggest that data from seroconverters can provide reasonable estimates of the calibration parameters.

Keywords

Mixture Model; Interval censored data; BED Calibration; HIV; Incidence; Estimation

[#]Corresponding author: Dr. Severin Guy Mahiane, Department of Biostatistics, Room E3137, The Johns Hopkins University Bloomberg School of Public Health, 625 N. Wolfe Street, Baltimore, MD 21205, USA, Tel: 443 287 8761; Fax: 410-955-0958.

1. Introduction

There has been great interest in the application of laboratory methods to measure Human Immunodeficiency Virus (HIV) incidence rates from cross sectional samples, since the first detuned enzyme immunoassay to detect recent HIV infection was described [1–3]. The precision and accuracy of methods based on ‘snapshot’, or cross-sectional sampling, to estimate HIV incidence, rely on the characteristics of the antibody response soon after infection [4–7] and are being continuously developed to identify recent infection [8].

In effect, the original procedure involved testing individuals using Sensitive/Less Sensitive (S/LS) commercial antibody assays (e.g. 3A11-LS, LS EIA), in order to detect differential HIV titre [3, 8]. More recently, a biomarker has been proposed based on the fact that antibodies produced early after infection bind less strongly to the antigen than those produced in established infection [9]. The antibodies binding avidity to the antigen can be measured using the Avidity Index (AI) which is calculated by dividing the sample-to-cut-off (S/CO) ratio from a low-avidity sample treated with guanidine by the S/CO ratio from a control sample [3, 8]. Concurrently, the BED capture enzyme immunoassay (BED) technique, which is based on the increasing proportion of anti-HIV-1 specific IgG in total IgG following seroconversion was developed [10]. In general, depending on the choice of a specific cut-off, commonly 0.8 (for the BED assay and for the AI), individuals with biomarkers measured below the threshold are classified as recently infected and the window period is the time spent below the chosen threshold.

Prior to the usage of biomarkers to estimate incidence, several parameters and assumptions are important. The main assumptions underlying these methods are that the length of time spent in the recent state (window period) is such that there exists a maximum duration, w_m , such that no one stays in the recent state for more than w_m years; and incidence is constant over the past w_m years. Besides these assumptions, the window period, which depends on the antibody response and varies between individuals, should not have a long tail distribution [8].

Two of the current challenges in using assays to characterize HIV incidence are precise knowledge of the mean window period, i.e. the mean time infected individuals spend before crossing a predefined threshold (or cut-off value), and misclassifications [1]. The main source of misclassifications is the number of those falsely identified as recent seroconverters, which increases the number of HIV positive individuals who have been infected for periods markedly longer than the incidence assay’s window period [1].

Attempts have been made to calibrate biomarkers for recent HIV infections. Hargrove et al [11, 12] used the Zimbabwe Vitamin A for Mothers and Babies trial (ZVITAMBO) data set to estimate the mean window period of the BED assay using a linear mixed model [11]; the proportion of individuals misclassified as recent seroconverters was obtained using an empirical estimator. Parekh et al. [13] used a larger sample size to estimate the BED window period in several regions of the world. Fiamma et al. [4] used data from the first male circumcision trial to calibrate the BED assay and the Bio-Rad AI. The methodologies in the above papers assumed a linear growth of the biomarkers. In a more recent approach,

Sweeting et al. [8] carefully modeled the growth of the AI and studied the distribution of the window period in a Bayesian framework.

In general, the main purpose of modelling the growth of biomarkers is to infer the time it would take to reach a given threshold, given that a primary event occurred. In the case of biomarkers for HIV incidence estimation, the mean window period was thought to be a natural parameter. Later, it was argued that, the mean recency duration, i.e. the mean time individuals spend above the cut-off value given that they have not been infected for more than a pre-set duration was to be considered instead [14]. However, in practice, these times are hardly observable. Data often arise from cohort studies where individuals are not monitored in a daily basis, and only the interval where the infection occurred is known, yielding to interval censored data. For simplicity, it is often assumed that the biomarker grows monotonically [8, 11, 13].

The main objective of this paper is to model the growth of the BED Normalized Optical Density (OD-n) as a function of time since seroconversion and estimate the mean window period, the mean recency period together with sensitivity, specificity and the false recent rate, in a frequentist framework, using a generalised mixture model. The second objective is to study the extent to which parameters estimated can be used to provide estimates of HIV incidence using data from a cross-sectional survey.

The remainder of this paper is organized as follows. In Section 2, we present two models. The first model uses data from both HIV-positive and HIV-negative participants to both estimate the incidence rate and describe the growth of the OD-n in the population of HIV-infected individuals. The second model describes the growth of the biomarker using only data from participants who become HIV positive during the study. Section 3 gives formulas and describes how to estimate the mean time spent with an OD-n lower than a cut-off value (mean window period), the mean recency period, sensitivity and specificities as functions of the window period. Section 4 provides the maximum likelihood estimator of incidence, using data from a cross-sectional survey, given a cut-off value, a window period and sensitivity and specificity corresponding to that window. Section 5 is devoted to numerical results. It contains estimates of the characteristics of the test together with HIV incidence rate estimates obtained from the ZVITAMBO study. Finally, Section 6 discusses the results and conclusions.

2 The Models

The framework

Assume that N' individuals were followed up for a certain period and each individual i was observed m_i times, at successive times $t'_{i1} \dots t'_{im_i}$. Let us assume that N ($N < N'$) of these individuals became infected during the follow up period and, for infected individuals, let X_{ij} ($i=1 \dots N, j=1 \dots n_i$) be the random variable describing the observed values of the biomarker at successive visits after infection: ξ_{ij} ($i=1 \dots N, j=1 \dots n_i$). Thus $t_{i1}=t'_{ij}(j'=m_i-n_i+1)$ is the first time the individual i was diagnosed HIV positive. Let $l_i=t'_{ij''}(j''=m_i-n_i)$ be the last time the individual i was diagnosed to be HIV negative. Without loss of generality, we

assume that $t'_{i1}=0$ for each individual i . Let us assume that there are K possible distributions for the random process describing the growth of the biomarker as a function of time since infection among the infected individuals. Define a *class* as the set of individuals whose biomarker measures follow the same distribution over time, so that the infected individuals can be classified into K (non empty) classes. Then, let us further assume that for any individual i in the k^{th} class ($k < K+1$), the biomarker grows as a function of time in the following fashion:

$$\log(X_{t_{ij}}) = a_k + (b_k - a_k) \exp(-c_k(t_{ij} - \tau_i)) + s_k E_{t_{ij}} \quad (1)$$

where K is to be determined, a_k , b_k , c_k and s_k are constants to be estimated ($b_k < a_k$ and $c_k s_k > 0$), τ_i is the (unobserved) seroconversion time and the non-scaled error terms $E_{t_{ij}}$ are independent, identically distributed and follow a standard Gaussian distribution, with probability density function g .

As stated in the introduction, data described in this framework arise from cohort studies and only the interval where the infection occurred is known. When used to estimate the mean time it takes the biomarker measure of individuals to cross a given threshold, such data appear to be interval censored and only parametric models are identifiable [8]. The choice of a function similar to the one used in the right hand side of equation (1) was justified and used to model the growth of the Avidity Index by Sweeting et al. [8]. They used a Bayesian approach in order to obtain the distribution of the mean window period. Hargrove et al. [15] argued that the expectation of the right hand side of equation (1) is biologically appropriate for modelling the growth of the BED assay and provides a good estimate of its mean recency period, using a mixed effect model. However, these studies did not account for the fact that there may be heterogeneity in the growth of the biomarker in the population, which a mixture model can handle.

One important issue in this modelling exercise is that the observed value of the biomarker depends on the seroconversion date and on the individual's class, which are not observed. The only information we have for each individual i is that her infection time, τ_i , belongs to the interval (l_i, t_{i1}) . A naive strategy to deal with it is to fix τ_i at any value in the interval (l_i, t_{i1}) , for each individual i . This approach is subject to inaccuracy due to the increase in uncertainty of seroconversion date when the duration between last negative and first positive date ($t_{i1} - l_i$) is large [8, 13]. Some of the authors who have used this approach had to exclude individuals for whom $t_{i1} - l_i$ is greater than a cut-off value [8, 11], throwing away available information. If data on individuals who did not seroconvert is available, the best strategy would be to determine K and estimate the incidence, the proportion of individuals in

each class ($\pi_k, k=1 \dots K$, with the constraint: $\sum_{k=1}^K \pi_k = 1$), and the vector of parameters $\theta = (\theta_k)_{k=1 \dots K}$ ($\theta_k = (a_k, b_k, c_k, s_k), k=1 \dots K$), simultaneously.

In order to proceed with the estimation procedure, we should first study the distribution of $(X_{t_{ij}})_{j=1 \dots n_i}$ ($i=1 \dots N$), which is equivalent to studying the distribution of $(\log(X_{t_{ij}}))_{j=1 \dots n_i}$. To this end, one should notice that equation (1) and our assumptions imply that for any

infected individual i , if her infection time τ_i and the class k she belongs to are known, $\log(X_{ij})$ follows a Gaussian distribution with mean $a_k + (b_k - a_k)\exp(-c_k(t_{ij} - \tau_i))$ and variance s_k^2 , for $j = 1 \dots n_i$. Now, let us adopt the following notations:

$$v_{ij}(\tau_i; \theta_k) = \frac{\log(\xi_{t_{ij}}) - a_k - (b_k - a_k)\exp(-c_k(t_{ij} - \tau_i))}{s_k}, \text{ for all } i, j \text{ and } k$$
 Thus, for any infected individual i , if her infection time τ_i and her class k are known, the 1 density function of the random variable $\log(X_{ij})$ can be written as $\frac{1}{s_k} g(v_{ij}(\tau_i; \theta_k))$, for $j = 1 \dots n_i$.

Model 1

In the case where data on individuals who did not seroconvert is available, one can estimate the incidence rate at the same time as the other parameters. Let us assume that the incidence rate is constant during the study period. Thus the infection time is distributed according to an exponential: $\tau_i \sim \text{Exp}(\lambda)$, where λ is the constant incidence hazard rate. Then if an individual i became infected during the follow-up, her contribution to the likelihood is

$$Lik_{1,i}(\theta, \pi, \lambda) = \left[\prod_{m=2}^{j''} \exp(-\lambda_{im} t'_{im}) \right] \sum_{k=1}^K \pi_k \int_{l_i}^{t'_{ij'}} \lambda \exp(-\lambda(\tau_i - l_i)) \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k))}{s_k} d\tau_i$$

(where $t'_{im} = t'_{im} - t'_{im-1}$ with $m = m - 1$ and the convention $t'_{i1} = 0$) and, if she remained uninfected, her contribution to the likelihood is:

$$Lik_{1,i}(0, \pi, \lambda) = \prod_{m=2}^{m_i} \exp(-\lambda t'_{im})$$

Hence, the log-likelihood can be expressed as:

$$LLik_1(\theta, \pi, \lambda) = \sum_{i=1}^N \log \left(\sum_{k=1}^K \pi_k \int_{l_i}^{t_{i1}} \exp(-\lambda \tau_i) \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k))}{s_k} d\tau_i \right) - \lambda \sum_{i=N+1}^{N'} l_i + N \log(\lambda)$$

where individuals i ($i = 1 \dots N$) are those who became infected during the study.

The number of parameters increases with the number of classes K so that a significant amount of time may be needed to obtain the maximum likelihood estimates. We suggest using an Expectation Maximisation (EM) like algorithm in order to estimate the parameters. We proceeded as follows:

Initialisation: randomly select $\theta^{(0)} = (\theta_k^{(0)})_{k=1 \dots K}$, $\lambda^{(0)}$ and $\pi^{(0)} = (\pi_k^{(0)})_{k=1 \dots K}$

For $r = 1 \dots$

E-Step—Probability of being in each class for each individual:

$$\pi_{ik}^{(r)} = \frac{\pi_k^{(r-1)} \int_{l_i}^{t_{i1}} \exp(-\lambda^{(r-1)} \tau_i) \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k^{(r-1)}))}{s_k^{(r-1)}} d\tau_i}{\sum_{m=1}^K \pi_m^{(r-1)} \int_{l_i}^{t_{i1}} \exp(-\lambda^{(r-1)} \tau_i) \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_m^{(r-1)}))}{s_m^{(r-1)}} d\tau_i}$$

M-Step—Estimate:

- $\theta^{(r)} = \operatorname{argmax} Q_1(\theta | \theta^{(r-1)})$

where $Q_1(\theta | \theta^{(r-1)}) = \sum_i \sum_{k=1}^K \pi_{ik}^{(r)} \log \left(\int_{l_i}^{t_{i1}} \exp(-\lambda^{(r-1)} \tau_i) \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k))}{s_k} d\tau_i \right);$

- $\lambda^{(r)} = \operatorname{argmax} LLik_1(\theta^{(r-1)}, \pi^{(r-1)}, \lambda)$ and

- $\pi_k^{(r)} = \frac{1}{N} \sum_{i=1}^N \pi_{ik}^{(r)}.$

The E-step consists of estimating the probability that each individual i belongs to the k^{th} class, given the previous approximation of the parameter (θ, π, λ) . This approximates the distribution of the *latent* variable indicating the class each individual belongs to, conditional on the observation. The M-step is then simplified since it amounts to estimating (and thus updating) θ_k ($k=1 \dots K$), λ and π separately instead of estimating the whole vector (θ, π, λ) at once.

Because we have no guarantee that our algorithm will always converge to the global maximum, we randomly selected the starting point (initialisation) more than 100 times, depending of the number of classes. For each draw, the E and M steps were repeated until convergence, i.e. until the relative increase in the likelihood was less than 10^{-10} . Then the “*local*” maximum which provided the highest objective was selected.

Model 2

However, if data on individuals who did not seroconvert are not available, one may use a *uniform prior*, i.e. assume that $\tau_i \sim U(l_i, t_{i1})$. We can also use the EM algorithm to estimate the parameters. We proceeded as follows:

Initialisation: choose $\theta^{(0)} = (\theta_k^{(0)})_{k=1 \dots K}$ and $\pi^{(0)} = (\pi_k^{(0)})_{k=1 \dots K}$

For $r=1 \dots$

E-Step—Probability of being in each class for each individual:

$$\pi_{ik}^{(r)} = \frac{\pi_k^{(r-1)} \int_{l_i}^{t_{i1}} \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k^{(r-1)}))}{s_k^{(r-1)}} d\tau_i}{\sum_{m=1}^K \pi_m^{(r-1)} \int_{l_i}^{t_{i1}} \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_m^{(r-1)}))}{s_m^{(r-1)}} d\tau_i}$$

M-Step—Estimate:

- $\theta^{(r)} = \operatorname{argmax} Q_2(\theta | \theta^{(r-1)})$

$$\text{where } Q_2(\theta | \theta^{(r-1)}) = \sum_i^K \sum_{k=1} \pi_{ik}^{(r)} \log \left(\int_{l_i}^{t_{i1}} \frac{1}{t_{i1} - l_i} \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k))}{s_k} d\tau_i \right)$$

- $\pi_k^{(r)} = \frac{1}{N} \sum_{i=1}^N \pi_{ik}^{(r)}$

Note that Model 2 only deals with individuals who become infected during the follow-up. The contribution to the likelihood for an individual i is:

$$Lik_{2,i}(\theta, \pi) = \sum_{k=1}^K \pi_k \int_{l_i}^{t_{i1}} \frac{1}{t_{i1} - l_i} \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k))}{s_k} d\tau_i$$

Thus the log-likelihood for the model considered here is given by:

$$LLik_2(\theta, \pi) = \sum_{i=1}^N \log \left(\sum_{k=1}^K \pi_k \int_{l_i}^{t_{i1}} \frac{1}{t_{i1} - l_i} \prod_{j=1}^{n_i} \frac{g(v_{ij}(\tau_i; \theta_k))}{s_k} d\tau_i \right)$$

The implementation of the EM algorithm for Model 2 is similar to the one used for Model 1 except that the estimation of λ in the M-step is skipped for Model 2.

Confidence intervals

The bootstrap method consisting of re-sampling individuals with replacement was applied. At each step, a sample of equal size was drawn from the original sample and the estimation procedure was applied. For Model 1, the numbers of seroconverters and non-seroconverters were kept constant, equal to those observed in the original sample. The re-sampling estimation procedure was repeated 1000 times and the confidence intervals were derived using the percentile method.

Model selection

To choose the number of classes, we used the Bayesian Information Criterion (BIC) given by:

$$BIC_u(k) = -2LLik_u + m_u(k) \log(N_u)$$

where $u=1$ for model 1, $u=2$ for model 2, k is the number of classes, $m_1(k)=5k$, $m_2(k)=5k-1$, $N_1=N'$ and $N_2=N$. We used the forward procedure; which was stopped when the BIC reached its minimum.

3 Calibration of the biomarker

In this section, we assume that there were K classes in our population and that the growth of the biomarker in the k^{th} class is described by X_t^k , where t is the time since seroconversion. We further assume that the parameters $\theta = (\theta_k)_{k=1 \dots K}$ and $\pi = (\pi_k)_{k=1 \dots K}$ which describe the growth of the biomarker in each class and the proportion of individuals in each class, respectively, are known.

Sensitivity and long term specificity

It may be worthwhile to know the probability that an individual who is tested as a recent seroconverter is correctly identified by the biomarker. Let w be the window period. Then the probability that an individual in the population does not cross the threshold x_0 before w ($S_e(x_0, w)$), also called sensitivity, can be obtained by:

$$S_e(x_0, w) = \sum_{k=1}^K \pi_k \int_0^w P_r(X_t^k < x_0) h_{0,w}(t) dt \quad (2)$$

where x_0 is the cut-off, and $h_{0,w}(t)$ is the density of individuals infected t years ago among those infected within the interval $(0, w)$.

If $T_{\max}$ is the maximum time an infected individual can survive, the specificity, which is the probability that an individual infected for more than w is correctly identified by the test (having biomarker value above the cut-off), can be obtained using the formula:

$$\rho(x_0, w) = 1 - \sum_{k=1}^K \pi_k \int_w^{T_{\max}} P_r(X_t^k < x_0) h_{w, T_{\max}}(t) dt \quad (3)$$

In the same way, we obtain the short term specificity ($\rho_{w, 2w}(x_0)$) defined by McDougal [5] by replacing $T_{\max}$ by $2w$ in equation (3).

Equation (2) and equation (3) show that the calibration parameters depend on both incidence and survival after infection. However, if we assume that both incidence and the infection related mortality are constant with time, we obtain:

$$S_e(x_0, w) = \frac{1}{w} \sum_{k=1}^K \pi_k \int_0^w P_r(X_t^k < x_0) dt$$

and

$$\varepsilon(x_0, w) = 1 - \rho(x_0, w) = \frac{1}{T_{\max} - w} \sum_{k=1}^K \pi_k \int_w^{T_{\max}} P_r(X_t^k < x_0) dt$$

For fixed T , the quantity $\varepsilon(x_0, T)$ is often noted ε_T and called false recent rates. In the applications, based on the fact that a) the mean duration after HIV infection is approximately 11 years and almost 75% of HIV infected people die within 14 years in absence of Anti-Retroviral Therapy [16, 17], and b) the data set we used was collected before the Anti-Retroviral era, we assumed that $T_{\max} = 14$ years. This assumption was assessed through simulations which will be described later in this section.

Mean window period

The time it takes to cross a given threshold x_0 , for individuals in the k^{th} class, $w_k(x_0)$, can be obtained easily since the biomarker is assumed to be perfect, and grows smoothly. It should be positive and solution of the equation:

$$\log(x_0) = a_k + (b_k - a_k) \exp(-c_k w_k(x_0))$$

which is obtained by setting $X_{tij} = x_0$, $t_{ij} - \tau_i = w_k(x_0)$ and $E_{tij} = 0$ in equation (1).

It comes out that $w_k(x_0)$ is defined if and only if

$$0 < \frac{\log(x_0) - a_k}{b_k - a_k} \leq 1$$

This holds when $\log(x_0)$ belongs to the interval (b_k, a_k) . In this case,

$$w_k(x_0) = \log\left(\frac{\log(x_0) - a_k}{b_k - a_k}\right) \quad (4a)$$

Strictly speaking, the mean window period is well defined if and only if $0 < w_k(x_0) < +\infty$

for all k and, in this case, is given by: $w(x_0) = \sum_{k=1}^K \pi_k w_k(x_0)$. However, when $w_k(x_0)$ is not defined for some k , one can redefine $w(x_0)$ as the mean window period for the sub-set of the population who would have their biomarker measure greater than x_0 .

Mean recency period

The mean recency period $w_T(x_0)$ is defined as the mean time infected individuals spend in the recent state (OD-n above the cut-off value x_0), given that they have been infected for a duration less than T years. This parameter can be approximated in each class k by $w_{k,T}(x_0) = w_k(x_0)$, if $\log(x_0)$ belongs to the interval (b_k, a_k) and T otherwise, i.e.

$$w_{k,T}(x_0) = \begin{cases} 0 & \text{if } \log(x_0) \leq b_k \\ w_k(x_0) & \text{if } \log(x_0) \in (b_k, a_k) \\ T & \text{if } \log(x_0) \geq a_k \end{cases} \quad (4b)$$

Thus the mean recency period is given by: $w_T(x_0) = \sum_{k=1}^K \pi_k w_{k,T}(x_0)$.

Hargrove et al [11] suggested a graphical estimation of the mean recency w_T which is such that the short term specificity is equal to the sensitivity. This can be found by solving the equation:

$$S_e(x_0, \bar{w}_T) = \rho_T(x_0, \bar{w}_T) \quad (4c)$$

As with the other estimators, the mean recency period estimated by equation (4c) depends on the cut-off value: $w_T = w_T^-(x_0)$.

Sensitivity Analysis

In practice, it is not an easy task to get the true h appearing in equations (2) and (3). In order to approximate it, we assumed that both the incidence and the infection related mortality were constant over time and that all the infected individuals lived $T_{\max}$ years.

We performed simulations in order to investigate the effects of our assumptions on our estimates of sensitivity and of the false recent rates. We simulated the first 20 and 30 years of the HIV epidemic in a population with the age distribution similar to that of sexually active individuals (15-to-49 years) in Zimbabwe in the early HIV epidemic [18]. Considering different scenarios for incidence rates over time, we simulated both infection and death of each individual in our population. HIV incidence rates presented different phases over time: increasing, decreasing or constant. HIV negative individuals were exposed to both HIV infection and background mortality which we assumed to be constant, equal to 0.01, while HIV infected individuals were exposed to death with survival varying as a function of their age at infection. We assumed that the hazard of dying as a function of time since infection, for causes related to the infection, followed a Weibull distribution with parameters estimated from studies in Sub-Saharan Africa [19] (see the Supporting Information, Table S1). Our simulated population consisted of 10000 individuals. Individuals were replaced with those aged 15-year-old when they died. The empirical estimation of the sensitivity and of the specificity could be obtained from our population for any cut-off value and at any time t since we have the duration of infection (t_i), for each infected individual i in our simulated population.

In effect, if the growth of the biomarker is perfectly characterized in our population, i.e. if θ and π are known, the sensitivity in the simulated population at time t is:

$$\hat{S}_e(x_0, w) = \frac{1}{n_1(w)} \sum_{t_i \in (0, w)} \sum_{k=1}^K \pi_k P_r(X_{t_i}^k < x_0)$$

the short term specificity is:

$$\hat{\rho}_{w,2w}(x_0) = 1 - \frac{1}{n_2(w)} \sum_{t_i \in (w, 2w)} \sum_{k=1}^K \pi_k P_r(X_{t_i}^k < x_0)$$

and the false recent rate is:

$$\hat{\varepsilon}(x_0, w) = \frac{1}{n_3(w)} \sum_{t_i > w} \sum_{k=1}^K \pi_k P_r(X_{t_i}^k < x_0)$$

where $n_1(w)$, $n_2(w)$ and $n_3(w)$ are the numbers of individuals who became infected in the respective intervals $(t - w, t)$, $(t - 2w, t - w)$ and $(0, t - w)$, and are alive at time t .

The formulas for sensitivity ($\hat{S}_e(x_0, w)$), specificity ($\hat{\rho}_{w,2w}(x_0)$) and false recent rates ($\hat{\varepsilon}(x_0, w)$) given above cannot be used in practice since they express the need for knowledge of the time since infection which is hard to obtain for all the individuals in a real population.

However, these formulas can be useful to obtain estimations of sensitivity, specificity and false recent rates through simulation, if there is enough information on the history of HIV in the population of interest.

4 Practical estimation of the incidence rate from a cross sectional survey

Incidence estimation

The great advance in the development of biomarkers is their ability to provide HIV incidence rates from cross sectional surveys. Many estimators have been provided [5, 11, 14, 20]; McWalter et al. [20] suggested that only the estimates of the mean window period and the proportion of assay non-progressors was necessary to estimate the incidence with good precision. Later, Auvert et al. [1] showed that calibrating parameters such as long and short term specificities were necessary in order to obtain accurate estimates of HIV incidence rate ratios when using the BED assay among young people. However, for a given value of the cutoff, x_0 , any reasonable value of the window period w can be used in order to estimate incidence, provided $S_e = S_e(x_0, w)$ and $\varepsilon = \varepsilon(x_0, w)$ are well calibrated. Indeed, if a cross sectional survey is to be conducted among n individuals in a population, the expectations of the number of HIV positive (N_+), tested recently infected (TR) and tested non-recently infected (TNR) among HIV positive individuals and the expectation of the number of recent infections (R) are related by the formula:

$$R = \frac{1-\varepsilon}{S_e-\varepsilon} TR - \frac{\varepsilon}{S_e-\varepsilon} TNR \quad (5)$$

Now, if the incidence hazard rate λ was constant in the time interval $(t - w, t)$, assuming death and infection are independent events in that interval (conditional on the event of being alive at time t), the probability of becoming infected in that interval can be estimated by 1-

$\exp(-w\lambda)$ and $\frac{R}{N_-+R}$, where N_- is the number of HIV negative among those surveyed. It results that the incidence hazard can be estimated by:

$$\hat{\lambda} = -\frac{1}{w} \log \left(\frac{N_-}{N_-+R} \right) \quad (6)$$

We show in the Supporting Information that the incidence rate estimated by equation (6) is the maximum likelihood estimation of the incidence rate.

We compared the HIV-incidence estimates obtained when using formula (6) to the estimates obtained using: a) McDougal's estimator [5], b) Hargrove's estimator [11], c) Welte's estimator [7], and Kassanjee's estimator [14]. All the estimates were obtained using parameters estimated as suggested in Section 3.

Confidence intervals and uncertainty analysis

The asymptotic variance of $\hat{\lambda}$ can be estimated by the delta method shown in the Supporting Information (Section A). Theoretical calculations of the sensitivity coefficients in the Supporting Information allow for studying the effects of the imprecision of the sensitivity and the specificity on the estimates of the incidence rate.

In this study, we used the bootstrap re-sampling method to estimate all the confidence intervals. The same technique was used to assess the effects of the imprecision of the sensitivity, specificities, the window period, and the mean recency period on the incidence estimates. Both parameters and counts were sampled 1000 times.

The procedures were implemented in R [21], code which is available from the corresponding author on request.

5 Numerical results: Application to the ZVITAMBO data set

The data

We used data from the Zimbabwe Vitamin A for Mothers and Babies trial (ZVITAMBO) which took place in Harare, Zimbabwe, from November 1997 to January 2000. The study was described elsewhere [12]. Briefly, 14,110 women were recruited within 96 hours of giving birth, followed at 6 weeks, 3 months and at 3-monthly intervals thereafter, for at least 1 year. Of these women, 9562, 4495 and 53 mothers tested HIV negative, positive and indeterminate, respectively. Each surviving mother was retested for HIV at each follow-up visit and those who tested HIV positive had their BED OD-n measured. During follow-up, 353 of the initially HIV negative mothers were seen to seroconvert.

For calibration purpose with Model 2, we used data from 186 women who seroconverted during the study and for which the BED OD-n was measured at least two times. The data used consisted of the date at which each woman was observed and the OD-n at that time, for those HIV-positive. Of these women, 89 (47.8%) were observed 2 times, 35 (18.8%) were observed three times, 45 (23%) were observed 4 or 5 times and the others were observed

either 6, 7 or 8 times. The mean (median; IQR) of the duration of the interval to seroconversion was 145 days (110 days; 91.0 days–182 days). The mean (median; IQR) duration of the follow-up since individuals were diagnosed HIV-negative was 428 days (365 days; 308 days–548 days). The mean (median; IQR) value of the OD-n the first time individuals were diagnosed positive and of the overall value of the OD-n were 0.43 (0.21; 0.08–0.52) and 1.0 (0.85; 0.33–1.6), respectively.

For Model 1, we used data from the 7413 women who were HIV-negative at recruitment. This model requires the knowledge of the duration between the first and the last visit for women who stayed HIV negative, and for those who became HIV positive, the model uses the duration between visits, their HIV statuses and OD-n measures once they became infected.

To estimate incidence rates as if they came from a cross-sectional survey, we used data from 9839 women observed during the inclusion visit and twelve months later. The HIV prevalence among these women at month 12 was 3244/9839 (33.0% CI: 32.0%–33.8%) and 279/3244 (8.6% CI: 7.6%–9.6%) of the HIV-positive had an OD-n less than the cut-off 0.8.

Parameter estimation

Model 1 and Model 2 were applied to the data set. For each number of classes (K), the parameters were estimated together with the BIC as described in Section 3. We used the forward procedure, i.e. the number of classes was increased until the BIC reached its minimum. Table 1 presents the BIC as a function of the number of classes; it shows that the appropriate numbers of classes are $K_{est} = 5$ and 7 for Model 1 and Model 2, respectively. Figure 1 illustrates the trajectories of the BED OD-n as functions of time since infection in each of the classes identified by Model 1. It suggests that there is heterogeneity in the growth of the OD-n measures and individuals in some classes never reach some cut-off values. Similar patterns were also observed when using Model 2. The parameters obtained from the above estimation procedure were used to estimate the mean window and mean recency periods, the sensitivity and the specificities for $K = 1 \dots K_{est}$.

Calibration for the conventional cut-off value of 0.8

Mean window and mean recency periods—The mean window and the mean recency periods were calculated using formulas (4a)–(4c). As suggested by Figure 1, our models predicts that approximately 1% of individuals in our population would never have their OD-n greater than 0.8. Such individuals were discarded, by removing their class, in order to calculate the mean window period. In contrast, the mean recency periods always have finite values and use the full set of parameter. When using Model 1, the mean window period was smaller (171 days (95% CI: 155–189)) when all the individuals were pooled in one class ($K=1$), increased to roughly 184 days for Model 1 with 2 and 3 classes and stabilised to 206 days for K greater than 4 but with larger confidence intervals. Figure 2(a) illustrates the estimates of the mean recency period (w_T , for $T = 1$ year) as a function of number of classes ($K=1 \dots 4$) for Model 1, for the cut-off value of 0.8. It shows that the mean recency period is smaller when the number of classes is 1 and stabilises around the same value when the number of classes is greater than 2. Its estimate, for $K=5$, obtained when using Model 1 is

184 days (95% CI: 169–201). In contrast, when the mean recency period is estimated by w_T^- , it tends to be higher, around 209 days when the number of classes is 1 and stabilises around 195 days for K greater than 1. Similar patterns were observed when using Model 2.

Sensitivity and specificities—Independently of whether we used Model 1 or Model 2, we observed the highest sensitivity when the number of classes was 1, with an estimate greater than 0.85, which decreased and stabilised around 0.74 for K greater than 4. The short term specificity was close to 0.74 for K less than 3 and increased to roughly 0.80 for K greater than 4. In contrast, the long-term specificity estimates were similar, approximately equal to 0.95.

False recent rate—The false recent rates, ε_T , which can be seen as one minus the specificity with T fixed, is independent of the mean window period and was also calculated as a function of the number of classes, for $T=1$ year. Figure 2(b) displays these estimates when using Model 1. It indicates that these are smaller for $K=2$ and 3 and present large confidence intervals for K greater than 4. Similar results were observed when using Model 2.

Calibration parameters as functions of the cut-off

We estimated the calibrating parameters as functions of the cut-off, for the optimal number of classes for Models 1 and 2, i.e. for $K=5$ and 7, respectively. Figure 3 presents the mean window period (w), the sensitivity (Se), the short term specificity ($\rho_{w,2w}$) and the proportion of false recent (ε_{2w}) as functions of the cut-off as estimated when using Model 1; results from Model 2 being similar. These mean window periods were calculated after the class of individuals for which the cut-off value never reach 0.1 was removed. The cut-off value varied from 0.3 to 0.9 since results from our models indicate that the OD-n measure of more than 15% of the individuals in our population will not reach any cut-off value greater than 0.1. As shown in Figure 3, the mean window period is an increasing function of time; the sensitivity is likely to be convex while the short term specificity is likely to be concave. The false recent rate is almost constant for cut-off values less than 0.7 and then increases quadratically. The widths of the confidence intervals increase as the cut-off values are increased.

When using formula (4c), estimates of the mean recency period, w_T^- , increase from roughly 63 days to 351 days when the cut-off varies from 0.3 to 1.2. When using equation (4b), estimates of the mean recency period, w_T , increase less rapidly from roughly 88 days to 324 days when the cut-off value varies from 0.3 to 1.9.

Incidence estimates from a cross-sectional survey for the cut-off value of 0.8

We also used the parameters estimated using optimal number of classes for Models 1 and 2 to estimate the incidence rate in the last twelve months. The analysis provided similar results. Sensitivity and specificity are especially required in order to estimate incidence using equation (6). Figure 4 illustrates results from Model 1. Figure 4(a) presents the sensitivity and specificity as functions of the window period. It shows that the specificity stays high, above 0.86, while sensitivity decreases from a value greater than 0.86 for

window periods less than 146 days (0.4 year) to less than a quarter for window periods greater than 547 days (1.5 years). As illustrated in Figure 4(b) the incidence rate estimates obtained using equation (6) are higher than the follow up estimates. These estimates decrease as the window period increases and are very close to the follow up estimate, with a relative difference less than 23% for any window period between 110 days (0.3 year) and 1095 days (3 years). The smallest difference was 5.7%. This difference is comparable to the relative differences obtained when using McDougal's, Hargrove's, Welte's and Kassanjee's estimators, with relative difference approximately equal to 5.6%, -1.4%, 22% and 8.65, respectively. Figure 4(c) gives estimated values and confidence intervals of incidence rate obtained by McDougal's, Hargrove's, Welte's and Kassanjee's estimators. It shows that the confidence intervals for all these estimators contain the follow-up estimate. Finally, when we performed uncertainty analysis, sampling both the counts (number of recently infected, number of non-recently infected and number of HIV-negative), all the estimates varied from 0 to roughly 0.064.

Sensitivity analysis

Many scenarios were investigated to assess the potential effects our hypotheses might have on the estimates of the sensitivity and of the false recent rates. Table 2 contains what we hope is an informative selection of 19 of these scenarios. The parameters values are displayed in columns 2 to 6. Column 2 shows the value of the incidence at the end of the first phase where incidence is increasing, column 3 indicates the incidence at the time when the parameters are estimated, columns 4, 5 and 6 indicate the time at end of the first phase, at the end of the second phase and the time when the parameters are estimated, respectively. Note that the incidence was constant during the second phase and decreasing during the third one. The resulting prevalence is shown in the 7th column while the sensitivity and the false recent rates are presented in the 8th and 9th column, respectively, in the case where the window period is set to one year. It appears that, despite the wide range of possible scenarios for the incidence, the relative differences in the estimates of the sensitivity and of the false recent rate do not exceed 2% and 9%, respectively. Moreover, the values obtained with our model assumptions are in the range of those observed in Table 2.

6 Concluding remarks

We used generalised mixture models to model the growth of the BED OD-n and estimate the *mean window period*, the mean recency period, the sensitivity and specificity as functions of the window period. The first model allows estimation of the incidence rate while the second model assumes that infection occurred following a uniform distribution between the last date the individuals were known HIV-negative and the first date they were diagnosed HIV-positive. The parameters estimated by our models were used to calibrate the BED assay and estimate incidence from a cross-sectional survey, using published estimators together with a newly derived estimator which is the maximum likelihood estimator of the incidence hazard rate.

A threshold of 0.8 was previously suggested as cut-off to classify individuals as recently infected, based on sensitivity and specificity of the biomarker six months after the midpoint of the seroconversion interval [8, 9]. We find such a threshold to be associated with a mean

window period (conditional on the progression of the assay) of about 205 days which is in the range of published values [13] but higher than the value obtained in two studies which used the same data set [11, 15]. In the first study, Hargrove et al. [11] used linear mixed models and assumed that infection occurred at the midpoint of the seroconversion interval and numbers of observation were excluded while they used a non linear mixed model in the second study.

The use of a mixture model to describe the growth of the BED assay with the possibility of estimating the incidence rate using the same model is novel. The incidence hazard rate estimated was 0.039 with a coefficient of variation less than 1%, independently of the number of classes. Model 2, which assumed that infection events are uniformly distributed in the seroconversion interval, appeared to have smaller BIC but a higher parameter dimension compared to Model 1. The latter model uses information about infection rates which is available. Furthermore, the average contribution to the BIC, which we defined as the BIC divided by the number of individuals used in the estimation process, was 6.89 for Model 2, more than 12-fold compared to Model 1. However, the hypothesis of constant incidence hazard rate may become irrelevant for a longer study. In this case, more parameters may be needed in order to adequately capture incidence trends and obtain a better approximation of the infection times of seroconverters. Furthermore, the a posteriori classes predicted by the mixture model can be used to study the factors explaining the differences in the growth of the biomarker.

Obtaining accurate estimates of the sensitivity, specificities and the false rate is a daunting challenge since these parameters depend on the incidence itself. We were able to estimate the sensitivity and the false recent rate comparable to the empirical false recent rate which was estimated using data from individuals who were infected at inclusion. Our sensitivity analysis shows that the approach proposed in this study can provide reasonable approximations of these parameters as functions of the window period. Our results confirm that the *mean window period* is not necessary in order to obtain accurate estimates of the incidence from cross-sectional surveys, using an antibody assay. Indeed, as suggested in this study, the *mean window period* appears not to be well defined, since some individuals can survive after infection and never cross a given threshold (assay non-progressors) [7, 20] and/or as illustrated by Hargrove et al. [22], the OD-n is not a strictly increasing function of time. This concept was recently revisited by Kassanjee et al. [14] who derived a new estimator. In this study, we observed that, for a cut-off value of 0.8, any window period greater than 0.3 years may give estimates relatively close to the follow-up estimate of the incidence; though the width of the confidence intervals appeared to increase as the window period decreases. This suggests that the incidence estimated by equation (6) is in reality a function of the cut-off value which, in turn, suggests that the history of incidence may be reconstructed, using different cut-off values. In fact, one may argue that incidence estimated with a given cut-off value is an estimate of the incidence at time $-w$ years (mean window period).

In this study, we estimated the maximum likelihood estimator of the incidence rate from one cross sectional survey which is at least asymptotically unbiased for any window period, provided the sensitivity and specificity of the test are unbiased. This estimator does not need

more parameters than the previous one; it requires the knowledge of two parameters just as the estimator derived by Kassanjee et al. [14] which, with a perfect knowledge of the mean recency period, is only guaranteed to be unbiased when the false recent rate is zero.

Our analysis suggested that five classes were sufficient to describe the growth of the BED assay in our population. This may be due to the wide variation of individual recency periods and/or differences in subtypes of HIV [13] and implies a large dimensional parameter, which requires a very large sample of seroconverters in order to obtain precise estimates. It would be important to see the extent to which this result may be generalized to other populations. In addition, several assumptions were needed to derive the formulas for specificity and sensitivity. For example, equation (3) shows that incidence and survival after infection as a function of time since infection are necessary for an accurate estimation of the specificity, which may vary as a function of time. The variability of that function around the world, due to differences in HIV epidemiology, may be one of the factors explaining why the estimates of the false recent rates were reported to vary from one part of the world to another [22].

We assumed that survival after infection is piecewise constant and all infected individuals survived 14 years after infection. This simplifying assumption which resulted from lack of data is, of course, one limitation of our study. The alternative way to estimate the specificity is to use the crude estimate. However, this would require recruiting HIV-infected individuals and following them until they die, which is not practical because of obvious ethical considerations and the availability of treatment. Our estimate of the false recent rate was 0.048 when using Model 1 and 0.046 when using Model 2 (for a window period of one year and a cut-off value of 0.8) which are comparable to the crude estimate (5.2%; 95% CI: 4.4–6.1) obtained by Hargrove et al. [11]. This rate increased to 0.049 when using Model 1 and to 0.048 when using Model 2, when we assumed that infected individuals survived 10 years after infection. Moreover, our simulations showed that, since the survival after infection with HIV is now documented [16, 17, 23], if the growth of the assay is adequately modelled, estimates of the calibrating parameters can be approximated through simulations, provided the past incidence rates of HIV can be reconstructed. Furthermore, our formula could be more accurate if used to estimate specificity among young people since among them, those who are infected have been infected for a shorter period and very few of them experience AIDS-related death. Another limitation of this study is that data used to calibrate the biomarker comes from a cohort in whom individuals have been monitored for less than 4 years after infection and before the availability of antiretroviral treatment. It is thus difficult to infer how the biomarker would behave after 4 years of infection or in the presence of treatment.

Our methods allow handling interval censored data and can be generalized for modelling the growth of a wider class of biomarkers increasing (or decreasing) over time. In this study, we observed that the sensitivity of the BED assay decreased considerably while its false recent rate seemed to decrease slowly as the window period increased. This is due to the fact that the BED OD-n reaches its plateau very early, only a few months after infection. The relatively high sensitivity and specificity observed for window periods of approximately six months indicate that this assay can be used to monitor newly HIV infected individuals in

areas with high HIV incidence and low prevalence. However, the BED assay would not be of great relevance for longer window periods if the number of individuals infected within that period is small compared to prevalence [1]. In effect, the poor performances of the BED assay in estimating the false recent rate have pushed efforts to develop biomarkers [24] or algorithms [25] that can perform better in terms of robustness with regard to sub-types of HIV. If such assays or combinations of assays can be characterized by functions which increase (or decrease) as the time since infection increases (or decrease), for a markedly longer period, then methods similar to those presented here can be used to explore the diversity of the response of immune systems, to monitor patients over time and to identify infection times of patients, using these assays.

We used the bootstrap method to estimate all the confidence intervals in this study. For HIV-incidence estimation, the variance estimated using the delta method (See Supporting Information, Section A) can be used to estimate the confidence interval limits, when the sample size is large. Our uncertainty analysis suggested that the precision of incidence estimates obtained by equation (6) can be optimised by choosing the window period such that: a) the sensitivity is equal to one divided by one plus the ratio of the number tested non-recently infected to the number of tested recently infected; b) the false recent rate, epsilon, should be very large or very small compared to sensitivity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Canadian International development Agency (CIDA, Canada) and the Johns Hopkins University Center for AIDS Research (Grant Number 1P30AI094189 from the National Institute of Allergy And Infectious Diseases). The authors would like to thank Dr Jean Humphrey and the ZVITAMBO Study Group who freely provided the data for the numerical experiments. The authors are grateful to Prof. Jean Bretagnolle and Dr. Rachid Ouifki for their valuable comments on some key aspects of the paper. Special thanks to Prof John Hargrove who showed interest to this work and to Prof Lawrence Moulton for comments on the manuscript.

References

1. Auvert B, Mahiane GS, Lissouba P, Moreau T. Statistical power and estimation of incidence rate ratios obtained from BED incidence testing for evaluating HIV interventions among young people. *PLoS One*. 2011; 6:e21149.10.1371/journal.pone.0021149 [PubMed: 21853019]
2. Guy R, Gold J, Calleja JM, Kim AA, Parekh B, Busch M, Rehle T, Hargrove J, Remis RS, Kaldor JM. Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: a systematic review. *Lancet Infect Dis*. 2009; 9:747–759.10.1016/S1473-3099(09)70300-7 [PubMed: 19926035]
3. Janssen RS, Satten GA, Stramer SL, Rawal BD, O'Brien TR, Weiblen BJ, Hecht FM, Jack N, Cleghorn FR, Kahn JO, Chesney MA, Busch MP. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *Jama*. 1998; 280:42–48. [PubMed: 9660362]
4. Fiamma A, Lissouba P, Amy OE, Singh B, Laeyendecker O, Quinn TC, Taljaard D, Auvert B. Can HIV incidence testing be used for evaluating HIV intervention programs? A reanalysis of the Orange Farm male circumcision trial (ANRS-1265). *BMC Infect Dis*. 2010; 10:137.10.1186/1471-2334-10-137 [PubMed: 20507545]

5. McDougal JS, Parekh BS, Peterson ML, Branson BM, Dobbs T, Ackers M, Gurwith M. Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. *AIDS Res Hum Retroviruses*. 2006; 22:945–952. [PubMed: 17067263]
6. Wang R, Lagakos SW. On the use of adjusted cross-sectional estimators of HIV incidence. *J Acquir Immune Defic Syndr*. 2009; 52:538–547.10.1097/QAI.0b013e3181c080a7 [PubMed: 19881357]
7. Welte A, McWalter TA, Barnighausen T. A Simplified Formula for Inferring HIV Incidence from Cross-Sectional Surveys Using a Test for Recent Infection. *AIDS Res Hum Retroviruses*. 2009; 25:125–126.10.1089/aid.2008.0150 [PubMed: 19182926]
8. Sweeting MJ, De Angelis D, Parry J, Suligoi B. Estimating the distribution of the window period for recent HIV infections: a comparison of statistical methods. *Stat Med*. 2010; 29:3194–3202.10.1002/sim.3941 [PubMed: 21170913]
9. Suligoi B, Massi M, Galli C, Sciandra M, Di Sora F, Pezzotti P, Recchia O, Montella F, Sinicco A, Rezza G. Identifying recent HIV infections using the avidity index and an automated enzyme immunoassay. *J Acquir Immune Defic Syndr*. 2003; 32:424–428. [PubMed: 12640201]
10. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, Green T, Hu DJ, Vanichseni S, Young NL, Choopanya K, Mastro TD, McDougal JS. Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. *AIDS Res Hum Retroviruses*. 2002; 18:295–307. [PubMed: 11860677]
11. Hargrove JW, Humphrey JH, Mutasa K, Parekh BS, McDougal JS, Ntozini R, Chidawanyika H, Moulton LH, Ward B, Nathoo K, Iliff PJ, Kopp E. Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *Aids*. 2008; 22:511–518.10.1097/QAD.0b013e3282f2a960 [PubMed: 18301064]
12. Humphrey JH, Hargrove JW, Malaba LC, Iliff PJ, Moulton LH, Mutasa K, Zvandasara P, Nathoo KJ, Mzengeza F, Chidawanyika H, Zijenah LS, Ward BJ. HIV incidence among post-partum women in Zimbabwe: risk factors and the effect of vitamin A supplementation. *Aids*. 2006; 20:1437–1446. [PubMed: 16791019]
13. Parekh BS, Hanson DL, Hargrove J, Branson B, Green T, Dobbs T, Constantine N, Overbaugh J, McDougal JS. Determination of mean recency period for estimation of HIV type 1 Incidence with the BED-capture EIA in persons infected with diverse subtypes. *AIDS Res Hum Retroviruses*. 2011; 27:265–273.10.1089/aid.2010.0159 [PubMed: 20954834]
14. Kassanjee R, McWalter TA, Barnighausen T, Welte A. A new general biomarker-based incidence estimator. *Epidemiology*. 2012; 23:721–728.10.1097/EDE.0b013e3182576c07 [PubMed: 22627902]
15. Hargrove J, Eastwood H, Mahiane G, van Schalkwyk C. How should we best estimate the mean recency duration for the BED method? *PLoS One*. 2012; 7:e49661.10.1371/journal.pone.0049661 [PubMed: 23166743]
16. Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, Mwita W, Suriyanon V, Rangsin R, Nelson KE, Sonnenberg P, Fitzgerald D, Karita E, Zaba B. 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):S55–63.10.1097/01.aids.0000299411.75269.e8 [PubMed: 18032940]
17. Van der Paal L, Shafer LA, Todd J, Mayanja BN, Whitworth JA, Grosskurth H. HIV-1 disease progression and mortality before the introduction of highly active antiretroviral therapy in rural Uganda. *Aids*. 2007; 21 (Suppl 6):S21–29. [PubMed: 18032935]
18. Zimbabwe 1988: results from the Demographic and Health Survey. *Stud Fam Plann*. 1991; 22:395–399. [PubMed: 1792680]
19. Hallett TB, Zaba B, Todd J, Lopman B, Mwita W, Biraro S, Gregson S, Boerma JT. Estimating incidence from prevalence in generalised HIV epidemics: methods and validation. *PLoS Med*. 2008; 5:e80.10.1371/journal.pmed.0050080 [PubMed: 18590346]
20. McWalter TA, Welte A. A comparison of biomarker based incidence estimators. *PLoS One*. 2009; 4:e7368.10.1371/journal.pone.0007368 [PubMed: 19809505]
21. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2012.

22. Hargrove J, van Schalkwyk C, Eastwood H. BED estimates of HIV incidence: resolving the differences, making things simpler. *PLoS One*. 2012; 7:e29736.10.1371/journal.pone.0029736 [PubMed: 22235334]
23. UK Register of HIV Seroconverters Steering Committee. The AIDS incubation period in the UK estimated from a national register of HIV seroconverters. *Aids*. 1998; 12:659–667. [PubMed: 9583607]
24. Duong YT, Qiu M, De AK, Jackson K, Dobbs T, Kim AA, Nkengasong JN, Parekh BS. 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:e33328.10.1371/journal.pone.0033328 [PubMed: 22479384]
25. Brookmeyer R, Konikoff J, Laeyendecker O, Eshleman SH. Estimation of HIV incidence using multiple biomarkers. *Am J Epidemiol*. 2013; 177:264–272.10.1093/aje/kws436 [PubMed: 23302151]

Figure 1.

Trajectories of the BED optical density (OD-n) as functions of time since infection in each class predicted by Model 1. The first component of the vectors $(\pi_k, a_k, b_k, c_k, s_k)$, $k=1 \dots 5$ indicates the proportion of individuals of the k^{th} class in the population and $\theta_k = (a_k, b_k, c_k, s_k)$ characterises the growth of the BED OD-n in that class.

Figure 2.

Mean recency period (w_T) and false recent rate (ϵ_T) estimates as functions of the number of classes in the population when using Model 1 ($T=1$ year). The error bars represent the 95% bootstrapped confidence intervals

Figure 3.

Mean window period sensitivity and specificities as functions of the cut-off value for Model 1 with the number of classes equal to 5. The error bars represent the 95% bootstrapped confidence intervals

Figure 4.

Sensitivity, specificity and incidence rate estimates using different approaches for a cut-off value of 0.8. (a) The sensitivity and specificity were estimated as indicated in Section 3, using parameters estimated using Model 1 with 5 classes. (b) Incidence rates estimated with the estimator presented in Section 4. The error bars show the 95% bootstrapped confidence intervals. The sensitivity and specificities were assumed to be perfectly known. (c) Incidence rates estimated from published estimators, using parameters estimated in this study. The mean recency periods (w_T^- and w_T), and the false recent rate ε_T were estimated for $T=365$. Hargrove's Welte's and Kassanjee's estimators used w_T^- , w and w_T , respectively; and all these three estimators used ε_T . McDougal's estimator requires the mean window

period (w), the short and long term specificities $\rho_{w,2w}$ and ρ_w . All the parameters were fixed. The dotted lines in (b) and (c) indicate the follow-up incidence estimates

Table 1

BIC for Models 1 and 2 as a function of the number of classes

Number of classes	BIC	
	Model 1	Model 2
1	4486	1611
2	4236	1442
3	4183	1365
4	4164	1328
5	4155	1301
6	4156	1283
7	/	1282
8	/	1286

Table 2

Sensitivity analysis: effects of the incidence rate on the sensitivity and on the false recent rate for a cut-off value of 0.8

Scenario #	Input parameters				Estimated parameters* (CoV (%))			
	λ_{12}	λ_S	d_1	d_2	d_S	Prevalence (%)	$\hat{S}_e(w)$	$\hat{e}(w)$
1	0.02	0.02	0	20	20	17 (2.5)	0.52 (6.9)	0.048 (0.7)
2	0.035	0.035	0	20	20	27 (1.8)	0.52 (4.7)	0.047 (0.8)
3	0.02	0.02	0	30	30	18 (2.6)	0.52 (6.3)	0.048 (1.0)
4	0.035	0.035	0	30	30	28 (1.9)	0.52 (4.1)	0.050 (0.8)
5	0.02	0.02	10	20	20	16 (2.6)	0.52 (8.6)	0.046 (0.7)
6	0.035	0.035	10	20	20	25 (2.0)	0.53 (6.0)	0.047 (0.6)
7	0.02	0.02	15	20	20	15 (2.8)	0.52 (1.0)	0.046 (0.7)
8	0.035	0.035	15	20	20	24 (2.2)	0.52 (6.5)	0.046 (0.6)
9	0.02	0.02	20	30	30	17 (3.0)	0.52 (7.7)	0.048 (1.0)
10	0.035	0.035	20	30	30	24 (2.1)	0.53 (5.4)	0.048 (0.9)
11	0.02	0.02	25	30	30	16(3.1)	0.52(8.0)	0.047(1.0)
12	0.035	0.035	25	30	30	24(2.1)	0.53(5.4)	0.048(0.9)
13	0.02	0.02	30	30	30	14 (3.4)	0.52(8.9)	0.047(1.0)
14	0.035	0.035	30	30	30	22 (2.5)	0.52 (5.9)	0.048 (0.9)
15	0.04	0.02	10	15	20	25 (1.8)	0.53 (5.9)	0.047 (0.6)
16	0.04	0.02	10	15	30	22 (2.1)	0.53 (4.6)	0.050 (0.9)
17	0.05	0.02	10	20	30	27 (2.0)	0.53 (3.8)	0.051 (0.9)
18	0.02	0.04	10	15	20	20 (2.4)	0.52 (8.9)	0.046 (0.6)
19	0.02	0.04	10	15	30	25 (2.2)	0.53 (5.2)	0.048 (0.9)

* λ_{12} : incidence at the end of the first phase; λ_S : incidence at the time where the sensitivity and the false recent rates are estimated; d_1 : duration of the phase in which incidence is constant; d_S : duration of the phase in which incidence is increasing; CoV: coefficient of variation.