

On the Statistical Accuracy of Biomarker Assays for HIV Incidence

Ron Brookmeyer, PhD

Objective: To evaluate the statistical accuracy of estimates of current HIV incidence rates from cross-sectional surveys, and to identify characteristics of assays that improve accuracy.

Methods: Performed mathematical and statistical analysis of the cross-sectional estimator of HIV incidence to evaluate bias and variance. Developed probability models to evaluate impact of long tails of the window period distribution on accuracy.

Results: The standard cross-sectional estimate of HIV incidence rate is estimating a time-lagged incidence where the lag time, called the shadow, depends on the mean and the coefficient of variation of window periods. Equations show how the shadow increases with the mean and the coefficient of variation. We find with an assay such as BED capture enzyme immunoassay, if only 0.5% are elite controllers who remain in the window until death, then the shadow is over 2.3 years, implying that estimates reflect HIV incidence more than 2 years in the past rather than current levels. If even 5% of AIDS cases are unrecognized and not excluded from the numbers in the window, then the shadow is more than 2.2 years.

Conclusions: Small perturbations to the tail of the window period distribution can have large effects on the accuracy of current HIV incidence estimates. The shadow and mean window period are useful for comparing the accuracy of assays. The results help explain differences reported between cohort and cross-sectional HIV incidence estimates. Screening out elite or viremic controllers by RNA polymerase chain reaction testing, and persons with advanced HIV disease (with AIDS or on antiretrovirals) may considerably improve the accuracy of HIV incidence estimates based on BED or similar assays.

Key Words: BED, biomarker, epidemiology, incidence, statistics

(*J Acquir Immune Defic Syndr* 2010;54:406–414)

INTRODUCTION

HIV incidence is the rate at which new infections occur in populations and tracks the leading edge of the epidemic. In recent years, considerable progress has been made in the development of biomarkers for determining current HIV incidence rates.¹ The basic idea is to use biomarkers in cross-

sectional samples to identify persons who were recently infected. The approach was first suggested using an assay for P24 antigenemia.² However, that biomarker necessitates large sample sizes because the durations that persons are in that window (P24 antigen positive and HIV antibody negative) are relatively short. Subsequently, the Serological Testing Algorithm for Recent HIV Seroconversion) was developed which consists of a dual antibody testing system in which persons who are positive on the standard HIV antibody assay are tested with a second assay to distinguish recent infections from long-standing infection.³

Several classes of assays have been proposed within the Serological Testing Algorithm for Recent HIV Seroconversion framework for HIV incidence estimation.⁴ One class is based on the quantity of HIV antibodies. The “detuned assay”, for example, is purposefully made less sensitive than the standard enzyme immunoassay by diluting and altering the incubation time of each tested specimen. Persons are said to be in the “window period” if they are confirmed positive on the first assay (standard enzyme immunoassay) and negative on the second detuned (or less sensitive) assay. A second class of assays is based on the proportion of HIV antibodies. The BED capture enzyme immunoassay, for example, is based on the proportion of HIV-1–specific IgG to total IgG.⁵ A third class of assays is based on the avidity of HIV antibodies that relies on the principle that antibodies produced shortly after incident HIV infection bind more weakly to the antigen than those produced later in infection.

Associated with each of these assays is a window period. Persons are in the “window period” if they are confirmed positive on the first assay (standard enzyme immunoassay) and identified as a recent infection by the second assay (eg, negative on the detuned or BED assays). Assay cutoff values are specified; for example, the BED and detuned assays specify an optical density cutoff, below which specimens are considered negative and thus recently infected. Higher cutoffs lead to larger mean durations of the window period.

United Nations AIDS has issued a statement that the BED assay should be used for neither estimating nor monitoring trends in HIV incidence.⁶ Their concern about the accuracy of the approach arises from reports primarily from Africa that the cross-sectional estimate overestimates HIV incidence.^{7–8} For example, a study of new mothers in Zimbabwe, the ZVITAMBO study, found that the unadjusted BED HIV incidence estimate was over 2 times greater than that based on the longitudinal follow-up of the cohort.⁸ One theory that has been advanced to explain that discrepancy is that the BED has a high “false recent” misclassification rate, whereby long-standing infections

Received for publication November 22, 2009; accepted March 5, 2010.
 From the Department of Biostatistics, School of Public Health, University of California, Los Angeles, CA.
 Correspondence to: Ron Brookmeyer, PhD, Department of Biostatistics, School of Public Health, University of California, Los Angeles, CA (e-mail: rbrookmeyer@ucla.edu).
 Copyright © 2010 by Lippincott Williams & Wilkins

are incorrectly labeled as new incident infections. This concern has led to several proposals to statistically adjust the cross-sectional estimates⁸⁻¹¹ and subsequent discussion about if and when such statistical adjustments are appropriate.¹²⁻¹⁷

The concerns that some persons remain in the window period for prolonged periods especially arise with antibody-based assays such as the BED which depend on antibody quantification. Elite or viremic controllers who have low or undetectable viral loads may remain in the window period until AIDS or death. Furthermore, persons with AIDS or on antiretroviral therapy may experience declines in HIV antibody levels causing re-entry into the window period. The Centers for Disease Control and Prevention recommended excluding persons with AIDS or persons on antiretrovirals from being counted in the window period regardless of their optical density values.¹⁸

The objective of this article is to set forth a framework for evaluating the statistical accuracy of estimates of current HIV incidence rates from cross-sectional surveys using assays for biomarkers and to identify characteristics of assays for biomarkers that improve accuracy. The framework permits evaluation of the impact of the window period distribution on accuracy. We use the framework to investigate the effects on accuracy of 3 phenomena that can contribute to long right tails of window period distributions: persons who remain in the window period for years, perhaps until AIDS onset or initiation of antiretrovirals (eg, viremic controllers or long-term nonprogressors); persons with persistently undetectable viral loads who never develop AIDS and remain in the window period until death (eg, elite controllers); and persons who re-enter the window period after onset of advanced HIV disease (eg, AIDS or beginning antiretrovirals), but they are not recognized as having advanced HIV disease, and are thus not excluded from counts in the window period.

METHODS

Statistical Framework

The cross-sectional estimator of HIV incidence is

$$\hat{I} = \frac{w}{n\mu} \tag{1}$$

where w is the number of persons in the window period, n is the total number of persons who are uninfected (negative on the standard enzyme immunoassay) and μ is the mean window period duration. The durations of window periods are variable and have a probability distribution. The survival function of window periods, $S(t)$, is the probability that the window period is greater than t days. Equation (1) is justified by assuming that the incidence rate of new infections is approximately constant over the support of $S(t)$. Kaplan and Brookmeyer studied the behavior of the incidence estimator (1) when the assumption of constant incidence is violated¹⁹ and showed that the estimator \hat{I} converges to a time weighted average of past HIV incidence rates:

$$I^* = \int_0^{\infty} I(t)f(t) dt \tag{2}$$

where $I(t)$ is the HIV incidence rate t years before the specimen collection, and the weighting function $f(t)$

is the backward recurrence density of the window period given by

$$f(t) = \frac{S(t)}{\mu}$$

Kaplan and Brookmeyer¹⁹ show that I^* (equation 2) is approximately equal to the HIV incidence at the time point ψ years before specimen collection where ψ is the mean backward recurrence time (Appendix for discussion of assumptions to justify the approximation such as linearity of HIV incidence trends). That is,

$$I^* \approx I(\psi)$$

We refer to ψ as the “shadow” because the cross-sectional sample is casting a shadow back in time. The shadow depends on the both the mean and the coefficient of variation of window periods through the equation:

$$shadow = \psi = \frac{\mu(1 + c^2)}{2}$$

where c , the coefficient of variation of the window periods, is the ratio of the standard deviation (σ) to the mean (μ) of window periods, that is, $c = \sigma/\mu$.²⁰ In summary, the estimator in equation (1) is estimating a time lagged incidence rather than the “current” HIV incidence, where the lag time is the “shadow.”

Framework for Evaluating the Statistical Accuracy of Assays

Evaluation of the statistical accuracy of the biomarker estimator of incidence requires consideration of both the estimator’s bias and variance. The statistical bias of an estimator is the difference between what it is actually estimating and what we hope it is estimating (ie, the current HIV incidence rate). The absolute statistical bias of the biomarker estimator is approximately $I(\psi)-I(0)$, which is the change in HIV incidence rates over the duration of the shadow. To compare 2 estimators, we define the relative bias as the ratio of their biases. The relative bias is approximately equal to the ratio of the shadows (Appendix). For example, if assay 1 has a shadow that is twice as large as assay 2, then assay 1 has a relative bias that is also approximately twice as large as that of assay 2. Although the (absolute) statistical bias depends on the magnitude of the change in HIV incidence rates over the duration of the shadow, the relative bias does not and depends only on the ratio of their shadows. As such, the shadow is a useful tool for comparing the statistical biases of assays.

The statistical stability of an estimator can be measured by its variance. Assuming that the numbers of persons in the window period (w) follows a Poisson distribution, then the variance of the estimator in equation (1) is approximately $\frac{I(\psi)}{n\mu}$, where $I(\psi)$ is the HIV incidence rate at the shadow. To compare the statistical stability of estimators with 2 assays, we consider the relative sample size requirements of the 2 assays, which we define as the ratio of the sample sizes required by the 2 assays to achieve the same variances. As discussed in the Appendix, the relative sample size requirement is approximately the ratio of mean window periods. For example,

if assay 2 has a mean window period twice as large as assay 1, then assay 1 would require a sample size twice as large as assay 2 to have the same level of statistical stability (ie, the same variances).

The preceding discussion suggests that 2 quantities, the shadow and the mean window period, are useful tools for comparing the statistical accuracy of assays in determining current HIV incidence from cross-sectional samples. The shadow measures how “current” is the estimate of incidence produced by the biomarker from a cross-sectional sample. The shadow determines the bias of the estimator. The mean window period measures the “statistical stability” of the incidence estimator and is a key determinant of the variance of the estimate. The shadow depends not only on the mean of the window period distribution but also its coefficient of variation. The coefficient of variation is sensitive to the tails of the distribution.

The Effect of the Tails of the Window Period Distribution on Accuracy

We investigated the effects on accuracy of several phenomena, which cause the window period distribution to be skewed with a long right tail. First, some persons may remain in the window period until the onset of AIDS.^{21–23} These may include persons who have been called “viremic controllers”, or long-term nonprogressors, and we assume that once these persons progress to advanced HIV disease (AIDS onset or initiation of antiretrovirals) that their advanced disease state is recognized and they are no longer included in the counts of persons in the window period. Second, some persons with persistently undetectable viral loads may never develop AIDS and they remain in the window period until death (from causes other than AIDS). We will refer to this group as elite controllers.^{22–24} Third, some persons may revisit the window period a second time after onset of AIDS or after starting antiretrovirals, and if their advanced HIV disease is not recognized, then they may be improperly counted in the window period.¹⁸ To evaluate the effects of these 3 phenomena on statistical accuracy, we developed 2 models. A mixture model was used to describe the first 2 phenomena. In the mixture model, a proportion (*p*) of newly infected persons have long window periods (eg, until onset of AIDS or death) with window period distribution $S_2(t)$, mean μ_2 , and coefficient of variation c_2 ; the remaining proportion $1-p$ have more “typical” window periods characterized by the distribution $S_1(t)$, mean μ_1 , and coefficient of variation c_1 . The overall mean window period duration of the mixture population is $\mu = \mu_1(1 - p) + \mu_2p$. The shadow ψ of the mixture population is (Appendix)

$$\psi = \frac{\mu_1^2(1 + c_1^2)(1 - p) + \mu_2^2(1 + c_2^2)p}{2[\mu_1(1 - p) + \mu_2p]} \quad (3)$$

We numerically evaluated μ and ψ (from equation 3) using the mixture models to determine the sensitivity of the accuracy of the estimator to the right tails of the window period distribution.

A model for window period re-entry was used to describe the third phenomena, which concerns persons with

AIDS (or persons on antiretrovirals) who are improperly counted in the window. Our model for window period re-entry (Appendix) allows persons to initially enter the window period and then re-enter the window period upon the onset of AIDS (or upon starting antiretrovirals). In this model, a proportion α are not recognized as AIDS cases and are thus (incorrectly) included in the window period count, whereas the remaining proportion $(1-\alpha)$ are correctly recognized as AIDS cases and properly excluded from the window count tally. We vary the parameter α to evaluate the quantitative effects of unrecognized AIDS cases on accuracy.

RESULTS

Table 1 tabulates the shadow as a function of the mean and the coefficient of variation of the window periods. For example, if an assay had a mean window period of 0.5 years and a coefficient of variation of 2.0, then the shadow would be 1.25 years, which implies that the assay is estimating HIV incidence approximately 1.25 years before the time point of collection of specimens. Table 1 illustrates the fact that if the coefficient of variation is greater than 1.0, then the shadow is greater than the mean window period. The shadow takes a minimum value of one half of the mean window period when the coefficient of variation is 0.

Table 2 displays the mean and coefficient of variation of window periods required to guarantee that the shadow is less than 1 year. For example, to guarantee a shadow less than 1 year, an assay with a mean window period of 1.0 years would require a coefficient of variation no larger than 1.0, whereas an assay with a mean window period of 0.25 years could afford a coefficient of variation as large as 2.6. Assays with smaller mean window periods can afford larger coefficients of variation, however, that comes at the price of a larger sample size requirements. Assays with smaller mean window periods require larger sample sizes to achieve the same level of statistical stability (variance). Table 2 also shows the relative sample size requirements (relative to a mean window of 1 year). For example, an assay with a mean of 0.25 years and coefficient of variation of 2.6 would require 4 times the sample

TABLE 1. Shadow (years) as a Function of the Mean (Years) and the Coefficient of Variation of Window Periods*

Mean Window (yrs)	Coefficient of Variation						
	0	0.25	0.50	1.0	1.5	2.0	3.0
0.25	0.13	0.13	0.16	0.25	0.41	0.63	1.25
0.50	0.25	0.27	0.31	0.50	0.81	1.25	2.50
0.75	0.38	0.40	0.47	0.75	1.22	1.88	3.75
1.00	0.50	0.53	0.63	1.00	1.63	2.50	5.00
1.50	0.75	0.80	0.94	1.50	2.44	3.75	7.50
2.00	1.00	1.06	1.25	2.00	3.25	5.00	10.0
3.00	1.50	1.60	1.88	3.00	4.88	7.50	15.0

*Based on the equation that the mean of the backward recurrence distribution is the shadow which is equal to $\psi = \frac{\mu(1+c^2)}{2}$, where μ and c are the mean and coefficient of variation of window periods, respectively.

TABLE 2. Maximum Tolerable Coefficient of Variation* of Window Periods to Insure the Shadow Is 1 Year or Less, and Relative Sample Size Requirements†

Mean Window (yrs)	Maximum Coeff Var (c)*	Relative Sample Size†
0.25	2.6	4.00
0.50	1.7	2.00
0.75	1.3	1.33
1.00	1.0	1.00
1.50	0.6	0.67
2.00	0.0	0.50

*The maximum coefficient of variation (c) that is tolerable to insure the shadow is $<M$ is $\sqrt{\frac{2M-\mu}{\mu}}$; table gives the largest values of the c to guarantee a shadow ≤ 1 year ($M=1$).

†The relative sample sizes refer to the ratio of sample sizes of the cross-sectional surveys (relative to when the mean window period is 1 year) required to produce equal variances of the estimates. Here it is assumed that the HIV incidence rates at the shadows are equal; alternatively, different assumptions about HIV incidence could be used and incorporated into the variance formula to refine the sample size calculations.

size as an assay with a mean of 1 year and a coefficient of variation of 1 year, yet both assays have shadows of 1 year.

Tables 3 and 4 evaluate the effects of viremic controllers and elite controllers on accuracy. Suppose a proportion $(1-p)$ have window periods with a mean of $\mu_1 = 0.5$ years and a proportion p have mean window period of $\mu_2 = 10$ years (Table 3) or $\mu_2 = 20$ years (Table 4). The coefficients of variation c_1 and c_2 were held fixed in these tables (eg, $c_1 = 0.363$ based on published reports on the BED window distribution⁵ and $c_2 = 0.50$ as suggested by time to AIDS and survival distributions.²⁴ Table 3 with $\mu_2 = 10$ years is intended to investigate the scenario in which a small proportion of persons (eg, viremic controllers) remain in the window until the onset of AIDS or initiation of antiretroviral therapy; and at that point, their advanced HIV disease is recognized and they are excluded from the count of the numbers in the window period. We found, for example, even if there are only 2% of such persons, that it is sufficient to cause the shadow to increase to over 2.02 years from only 0.28 years if $p = 0.0$ (Table 3). Table 4 with $\mu_2 = 20$ years is intended to investigate the scenario in which a very small proportion of persons (eg, elite controllers) never develop AIDS and remain in the window period until death. We found, if 1.0% are elite controllers (ie, $p = 0.01$), then the shadow is 3.81 years. Even if

TABLE 3. Impact on Accuracy of “Viremic Controllers”

p	μ (yrs)	Coefficient of Variation (CV)	Shadow ψ (yrs)
0.00	0.50	0.363	0.28
0.01	0.60	1.83	1.29
0.02	0.69	2.20	2.02
0.05	0.98	2.42	3.35
0.10	1.45	2.25	4.41
0.20	2.40	1.84	5.27

Based on a mixture model where a proportion $1-p$ have a windows with mean $\mu_1 = 0.5$ years (coefficient of variation $c_1 = 0.363$) and a proportion p have a window with mean $\mu_2 = 10$ years (coefficient of variation $c_2 = 0.50$). The overall mean window μ , coefficient of variation (CV), and shadow ψ in the mixture population are given.

TABLE 4. Impact on Accuracy of Elite Controllers

p	μ (yrs)	Coefficient of Variation (CV)	Shadow ψ (yrs)
0.000	0.50	0.36	0.28
0.001	0.52	1.37	0.75
0.005	0.60	2.61	2.33
0.010	0.70	3.15	3.81
0.015	0.79	3.37	4.91
0.020	0.89	3.46	5.77

Based on a mixture model where a proportion $1-p$ have a windows with mean $\mu_1 = 0.5$ years (coefficient of variation $c_1 = 0.363$) and a proportion p have a window with mean $\mu_2 = 20$ years (coefficient of variation $c_2 = 0.50$). The overall mean window μ , coefficient of variation (CV), and shadow ψ in the mixture population are given.

only 0.5% of persons are elite controllers (ie, $p = 0.005$), the shadow is 2.33 (Table 4). These results imply that even a very small proportion of persons who are elite controllers can cause a large increase in the shadow, which potentially decreases the statistical accuracy of biomarker estimates of current HIV incidence.

Figure 1 illustrates the window period distributions for 3 of the mixture models considered in Table 4 (for elite controllers comprising 0.1%, 1.0%, and 2% of the population). The 3 curves in Figure 1 are visually indistinguishable from each other, however, their shadows are very different (0.75 years, 3.81 years, and 5.77 years). The cautionary warning

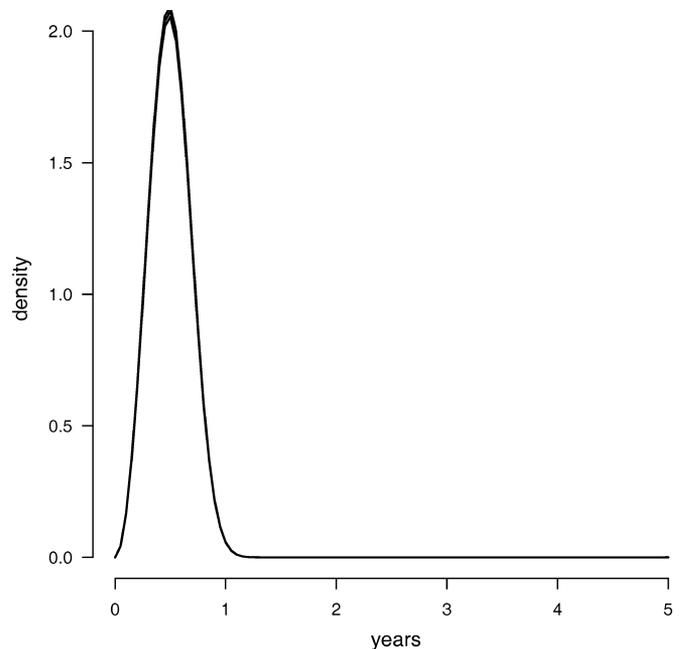


FIGURE 1. Window period densities if the proportions that are elite controllers are $p = 0.001$, $p = 0.01$, and $p = 0.02$ using the mixture models in Table 4 (and additionally assuming Weibull models for S_1 and S_2). The 3 curves are visually indistinguishable from each other (the curve for $p = 0.001$ has the highest peak, $p = 0.01$ the middle peak, and $p = 0.02$ the lowest peak), however, their shadows are very different (0.75, 3.81, and 5.77 years, respectively).

here is that the shadow depends on very subtle characteristics of the tails of the window periods, which may not be readily apparent from visual inspection of the curves.

Table 5 evaluates the effect of unrecognized AIDS case who revisit the window period, but because they are not recognized as AIDS cases they are not excluded from the count of the number in the window period. We find that even a very small proportion of unrecognized AIDS cases are sufficient to greatly increase the shadow. For example, if 0%, 1%, and 5% of AIDS cases are unrecognized, the shadows are respectively 0.28, 0.73, and 2.23 years. To keep the shadow under 1 year, no more than 1.6% of AIDS cases may go unrecognized.

A practical question is whether assays could be designed more optimally to increase accuracy for HIV incidence estimation. Here we consider 2 design parameters that could potentially be manipulated, the mean window (μ_1) and the coefficient of variation (c_1) (among the first subpopulation with typically short window periods in the mixture population). For example, the mean and coefficient of variation of the window periods change if the optical density cutoff of an assay is changed. We investigated whether there was an optimal choice for μ_1 under the mixture model. We varied the mean window period μ_1 in equation (3) and calculated the shadow. Figure 2 shows the relationship of the shadow to the mean window period μ_1 (μ_2 was fixed at 10 years and the coefficients of variation c_1 and c_2 were also fixed). Figure 3 is a similar figure except μ_2 was fixed at 20 years. The figures illustrate that there are optimal choices of the mean window (μ_1) that minimize the shadow for a given mixing proportion p . For example, in Figure 3, if $p = 0.005$, the shadow is minimized at 1.63 years by an assay with mean of $\mu_1 = 1.44$ years. It is a remarkable and surprising fact that Figures 2 and 3 are approximately U shaped: the explanation for the U shape is that if μ_1 gets too small, then persons who are “caught” in the window period become increasingly dominated by the subpopulation with long window periods as opposed to the subpopulation with shorter window periods. It can be shown analytically that the shadow is approximately minimized if the mean window period μ_1 of the assay is (Appendix):

$$\mu_1 = \mu_2 \sqrt{\frac{(1 + c_2^2)p}{(1 + c_1^2)}} \tag{4}$$

which yields a minimum shadow of:

$$\psi = \mu_2 \sqrt{(1 + c_1^2)(1 + c_2^2)p} \tag{5}$$

Equation (5) shows that the shadow decreases as the coefficient of variation c_1 decreases. By setting $c_1 = 0$ in equation (5), we can determine the minimum value of the shadow (ψ_{\min}) that is ever achievable with the mixture distribution:

$$\psi_{\min} = \mu_2 \sqrt{(1 + c_2^2)p} \tag{6}$$

For example, suppose elite controllers have a mean window period of 20 years with coefficient of variation of $c_2 = 0.5$. With optimal choices for the assay’s design parameters [ie, set $c_1 = 0$ and μ_1 equal to equation (4)], the minimum

TABLE 5. Impact on Accuracy of Unrecognized AIDS Cases Who Re-Enter the Window

α	μ^* (yrs)	Shadow ψ (yrs)
0.00	0.50	0.28
0.01	0.52	0.73
0.02	0.54	1.15
0.05	0.60	2.23
0.10	0.70	3.63
0.25	1.00	6.14
0.50	1.50	8.09
0.75	2.00	9.07
1.00	2.50	9.66

The model for window period re-entry assumes the initial window period has mean $\mu_1 = 0.5$ years (coefficient of variation $c_1 = 0.363$), mean time to AIDS is 10 years at which point cases re-enter the window and the mean survival time with AIDS is 2 years (exponential distributions). Model allows for a proportion α of AIDS cases to be counted in the window period, whereas a proportion $(1-\alpha)$ of AIDS cases are excluded from being counted in the window period.

*The total mean time (μ) spent in the window period is the sum of the mean time in the window during the initial visit plus the mean time spent in the window after onset of AIDS.

values of the shadow corresponding to $p = 0.001, 0.005, 0.010, 0.015,$ and 0.020 are respectively 0.71, 1.58, 2.23, 2.74, and 3.16 years. The implication of these calculations is that if elite controllers compose greater than 0.5% of the population, then it is not possible to “tune” an assay (eg, the BED) to have a shadow of less than 1.58 years even if the coefficient of variation (c_1) was successfully reduced to 0.

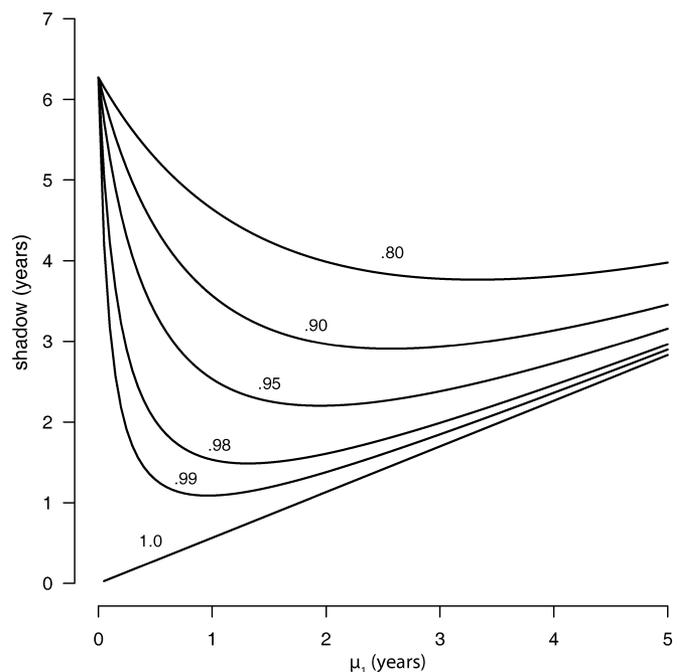


FIGURE 2. Shadow versus the mean window period (μ_1 in years) in the mixture model (where $\mu_2 = 10$ years, $c_2 = 0.5$ and $c_1 = 0.363$ are fixed). Curves are shown for $(1-p) = 0.80, 0.90, 0.95, 0.98, 0.99,$ and 1.0 .

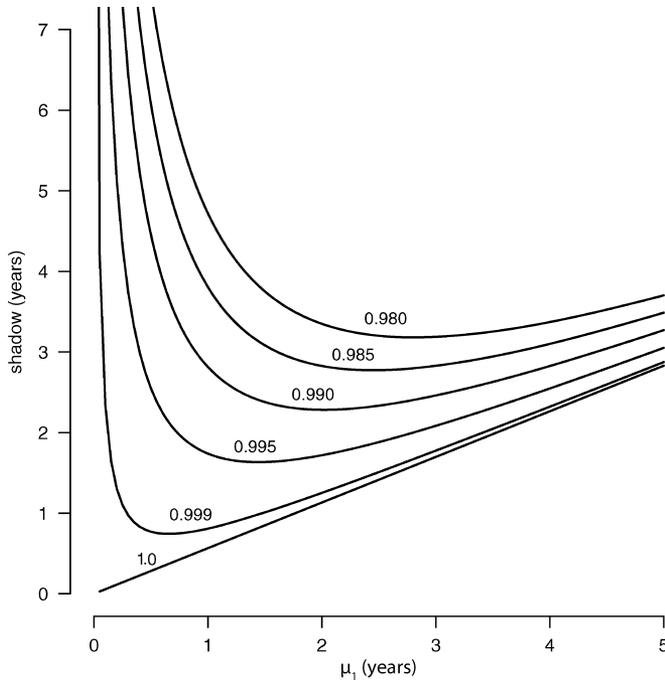


FIGURE 3. Shadow versus the mean window period (μ_1 in years) in the mixture model (where $\mu_2 = 20$ years, $c_2 = 0.5$ and $c_1 = 0.363$ are fixed). Curves are shown for $(1-p) = 0.980, 0.985, 0.990, 0.995, 0.999,$ and 1.0 .

DISCUSSION

Evaluation of the statistical accuracy of the biomarker approach to HIV incidence requires consideration of both the bias and variance of the estimators. Both the shadow and the mean window period are useful in this regard. The shadow is useful for comparing statistical bias. Assays with smaller shadows have smaller relative bias. The mean window period is useful for comparing the statistical stability (variance) of assays. Assays with larger variances have greater sample size requirements.

The biomarker approach is estimating a time-lagged incidence back into the past where the lag time is the shadow. The shadow measures how “current” is the biomarker based estimate of HIV incidence. The shadow depends on both the mean and the coefficient of variation of window periods. The shadow increases as the mean window period increases and as the coefficient of variation increases.

The validity of the shadow interpretation of the cross-sectional incidence estimate depends to some extent on the linearity assumption of HIV incidence trends (Appendix, and section 1.4 of¹⁹). Although linearity is not an absolutely necessary condition for validity, the degree of nonlinearity that can be tolerated should be assessed and depends on a number of factors including the window period distribution.¹⁹

We have evaluated the impact of 3 phenomena on accuracy: viremic controllers who are counted in the window period until onset of advanced HIV disease; elite controllers who remain in the window until death; and unrecognized AIDS cases who re-enter the window period. We have shown that even a small proportion of persons with long window

periods (eg, elite controllers) can significantly increase the shadow, which may substantially bias the biomarker approach for estimating current HIV incidence levels. Furthermore, a very small proportion of AIDS cases, who are unrecognized and not excluded from the counts in the window period, can significantly increase the shadow.

Our results help explain discrepancies between the cohort and biomarker approaches that have been reported in the literature. One study in Zimbabwe, the ZVITAMBO study, recruited and followed new mothers within 96 hours of delivery and compared the BED cross-sectional incidence estimate with the cohort estimate.⁸ The BED estimate was over 2 times larger than the cohort estimate. Our results help explain that discrepancy. We have shown that if as little as 0.5% of persons are elite controllers, the shadow is over 2.5 years. In that case, the BED approach is estimating HIV incidence into the prepartum period, whereas the cohort approach is estimating incidence exclusively during the postpartum period.¹⁶ It is very plausible that HIV incidence rates are higher during the prepartum than the postpartum period because the women were sexually active during the prepartum period, and in the ZVITAMBO study, the new mothers were receiving counseling and other behavioral interventions during the postpartum period. The cautionary warning here is that if HIV incidence is changing sharply over time (such as between the prepartum and postpartum periods among new mothers), shadows of over 2 years duration can create discrepancies between cohort and biomarker estimates. Neither the biomarker approach nor the cohort approach are gold standards in which to validate the other because they are estimating HIV incidence rates at different points in time.

Our results provide guidance for the design of new assays to improve accuracy. Judicious choice of the mean window period (by changing the optical density cutoff for example) can help minimize the shadow as illustrated in Figures 2 and 3, and afford some protection from long tails of the window period distribution. Precisely how changes in the optical density cutoff would influence the shadow is an open research question and is complicated because changes to the cutoff may affect not only μ_1 but also other parameters in equation 3 including the coefficient of variations and p .

Furthermore, our results establish the limits to the statistical accuracy of some assays. We find, for example, that if as little as 0.5% of persons are elite controllers who remain in the window until death, it is not possible to reduce the shadow to less than 1.58 years even if the assay’s coefficient of variation (c_1) is successfully reduced to 0. Whether or not a shadow of 1.58 years is adequate depends on how much HIV incidence changes over the duration of the shadow and the ultimate purposes of the incidence estimate. We illustrated a number of our results with a specific value for μ_2 (20 years.) The interpretation of μ_2 is the life expectancy of an elite controller, which may depend on overall health conditions of the population, and should be assessed at the local level.

The limits to the statistical accuracy of some assays discussed above begs the question if there are alternative strategies for improving the accuracy of current HIV incidence estimates? One promising strategy is based on algorithms involving multiple assays such as CD4, BED, avidity assays,

antiretroviral, P24 antigen, and HIV RNA screening.^{21,25} For example, an algorithm could consist of HIV-positive persons who are tested for CD4 to exclude persons with AIDS, those without advanced HIV disease are then tested with the biomarker assays (eg, BED), and those in the window period are assayed for HIV RNA to exclude elite controllers. Tables 3–5 suggest that screening with additional assays would increase accuracy considerably; the tables show by how much the shadow would decrease if p or α are decreased by screening. Two complications deserve attention with algorithms of multiple assays. First, with each specific algorithm, the window period distribution would need to be recalibrated. Second, if elite controllers with nondetectable RNA are automatically excluded from the window count tally, it will be necessary to account for the fact that some of those exclusions may have been in the (counterfactual) window period; here we are considering the “window period” of elite controllers as a counterfactual, that is, an unobservable duration which follows the same probability distribution as the window period of noncontrollers. That is, if w is the observed number in the window period, then the corrected number in the window would be $w/(1-p)$ where p is the probability an individual is an elite controller. In practice, however, because p is likely small, that suggested correction may have a minor numerical effect. It is worth reiterating that p refers to the proportion of elite controllers in a cohort of newly infected persons. The proportion of elite controllers in a group of prevalent infected individuals is likely to be larger because of the survival advantage of elite controllers.

Other measures of statistical accuracy have been discussed in the literature. For example, the “false-recent rate” refers to the proportion of long-standing infections that are in the window period; however, that definition is ambiguous without further qualification about the meaning of long-standing infection.²⁵ We could qualify the term, for example, and define the 3-year false recent rate as the proportion with window periods greater than 3 years. However, that rate by itself does not determine the bias. To illustrate that fact, consider 2 assays, where 1 assay has a mean window of 0.50 years and coefficient of variation of 1.7, and a second assay has a mean of 1 year and coefficient of variation of 1; Although both assays have a shadow of 1 year (Table 2), the 3-year false recent rates are quite different (.017 versus .041 under a lognormal model for the window period distribution). The key point is that false recent rates at one time point do not fully characterize the statistical accuracy of assays for current HIV incidence either with respect to bias or variance.

The purpose of this article was to provide a framework for understanding how the window period distribution affects accuracy. The framework considers 2 dimensions to statistical accuracy, bias, and variance. Alternatively, one could combine the bias and variance measures into a single overall summary measure of accuracy such as the mean squared error; however, we prefer to describe the 2 dimensions of accuracy separately to clarify the tradeoff between bias and variance and because their relative importance may depend on the circumstances of the setting (eg, if sample size limitations are of primary concern, controlling variance may be paramount; if monitoring rapidly changing HIV incidence is of primary concern,

controlling bias may be paramount). The discussion in this article assumed the window period distribution was known. Of course, accuracy will be reduced if that is not the case, and there is error in the mean window period (μ) used in equation 1. There are confidence interval procedures that account for uncertainty in μ including both analytic²⁶ and Monte-Carlo approaches.²⁷ The framework set forth in this article provides guidance for the designing improved assays and also identifies the limits to accuracy achievable by some assays. Algorithms that include testing for RNA, CD4, avidity, P24 antigen, and antiretrovirals offer a very promising avenue for improving the accuracy of current HIV incidence estimation from cross-sectional studies.

REFERENCES

1. Le Vu S, Pillonel J, Semaille C, et al. Principles and uses of HIV incidence estimation from recent infection testing—A review. *Eurosurveillance*. 2008;13:11–16.
2. Brookmeyer R, Quinn TC. Estimation of current human immunodeficiency virus incidence rates from a cross sectional survey using early diagnostic tests. *Am J Epidemiol*. 1995;141:166–172.
3. Janssen RS, Satten GA, Stramer S, et al. 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.
4. Murphy G, Parry JV. Assays for the detection of recent infections with Human Immunodeficiency Virus Type 1. *Eurosurveillance*. 2008;13:4–10.
5. Parekh BS, Kennedy MS, Dobbs T, et al. 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.
6. UNAIDS Reference Group for Estimates, Modeling and Projections. *Statement on the Use of the BED-assay for the Estimation of HIV-1 Incidence for Surveillance or Epidemic Monitoring. Report of a meeting of the UNAIDS Reference Group for Estimates, Modeling and Projections. Athens, Greece*. Geneva, Switzerland: UNAIDS; Available at: <http://www.epidem.org/Publications/BED%20statement.pdf>. Accessed December 13–15, 2005.
7. Karita E, Price M, Hunter E, et al. Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from Africa. *AIDS*. 2007;21:403–408.
8. Hargrove JW, Humphrey JH, Mutasa K, et al. Improved HIV-1 incidence estimates using BED capture enzyme immunoassay. *AIDS*. 2008;22:511–518.
9. McDougal JS, Parekh BS, Peterson ML, et al. Comparison of HIV-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.
10. McWalter TA, Welte A. Relating recent infection prevalence to incidence with a subpopulation of nonprogressors. *J Math Bio*. 2009;DOI10.1007/s00285-009-0282-7.
11. McWalter TA, Welte A. A comparison of biomarker based incidence estimators. *PLoS ONE*. 2009;4(10):e7368.
12. Brookmeyer R. Should biomarker estimates of HIV incidence be adjusted? *AIDS*. 2009;23:485–491.
13. Hargrove JW. BED estimates of HIV incidence must be adjusted. *AIDS*. 2009;23:2064–2065.
14. McDougal JS. BED estimates of HIV incidence must be adjusted. *AIDS*. 2009;23:2064–2065.
15. Welte A, McWalter TA, Barnighausen T. Reply to “Should biomarker estimates of HIV incidence be adjusted?” *AIDS*. 2009;23:2062–2063.
16. Brookmeyer R. Response to correspondence on “Should biomarker estimates of HIV incidence be adjusted?” *AIDS*. 2009;23:2066–2068.
17. Wang R, Lagakos SW. On the use of adjusted cross-sectional estimators of HIV incidence. *J Acquir Immune Defic Syndr*. 2009;52:538–547.
18. Using the BED HIV-1 Capture EIA Assay to estimate incidence using STARHS in the context of surveillance in the United States, October

2007. Centers for Disease Control and Prevention website. Available at: <http://www.cdc.gov/hiv/topics/surveillance/resources/factsheets/BED.htm>. Accessed March 5, 2010.

19. Kaplan E, Brookmeyer R. "Snapshot estimators of recent HIV incidence rates." *Oper Res.* 1999;47:29–37.
20. Cox DR. *Renewal Theory*. London, United Kingdom: Methuen and Co. Ltd; 1962.
21. Laeyendecker O, Rothman RE, Henson C, et al. The effect of viral suppression on cross-sectional incidence testing in the Johns Hopkins Hospital Emergency Department. *J Acquir Immune Defic Syndr.* 2008;48: 211–215.
22. Laeyendecker O, Redd AD, Lutalo T, et al. Frequency of long-term nonprogressors in HIV-1 seroconverters from Rakai Uganda. *J Acquir Immune Defic Syndr.* 2009;52:316–319.
23. Okulicz O, Marconi VC, Landrum ML, et al. Clinical outcomes of elite controllers, viremic controllers and long term non-progressors in the US Department of defense HIV natural history study. *J Infect Dis.* 2009; 1714–1723.
24. AAlen O, Farewell VT, DeAngelis D, et al. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment; application to AIDS prediction in England and Wales. *Stat Med.* 1997;16: 2191–2210.
25. Mastro TD, Kim A, Hallett T, et al. Estimating HIV incidence in populations: issues, challenges and the way forward. *J HIV/AIDS Surveill Epidemiol.* 2010; (in press).
26. Brookmeyer R. Accounting for follow-up bias in estimation of Human Immunodeficiency virus incidence rates. *J R Stat Soc A.* 1997;160: 127–140.
27. Cole SR, Chu R, Brookmeyer R. Confidence intervals for biomarker-based Human Immunodeficiency Virus incidence estimates and differences using prevalent data. *Am J Epidemiol.* 2007;165:94–100.

APPENDIX: THE SHADOW

It has been shown¹⁹ that the cross-sectional estimator is in fact not estimating the current HIV incidence but is instead an approximation of the HIV incidence ψ years before the collection of the specimens in the cross-sectional sample, where, ψ , the shadow, is the mean of the backward recurrence time distribution, and ψ is given by:

$$\psi = \frac{\int_0^{\infty} tS(t) dt}{\mu}$$

The validity of the approximation that the cross-sectional estimate is the HIV incidence rate ψ years earlier can be formally assessed by a second order Taylor series expansion. The error in the approximation is equal to $I''(\psi)V/2$, where $I''(\psi)$ is the second derivative of the incidence function and V is the variance of the backward recurrence times, (equation 16 in¹⁹). Therefore, it follows that the error in the approximation becomes small under either or combination of 2 conditions: the time trend in HIV incidence rates is approximately linear (ie, $I(t) \approx I(0) + bt$ in which case $I''(\psi)$ is zero), or the variance (V) is small.¹⁹

Bias

The (absolute) bias in the cross-sectional estimator is $I(\psi)-I(0)$, where $I(t)$ is the incidence t years before the collection of the specimens. The relative bias of 2 assays with shadows ψ_1 and ψ_2 is the ratio of their biases, which under the assumption HIV incidence trends are linear in time is the ratio of their shadows because

$$\frac{I(\psi_1) - I(0)}{I(\psi_2) - I(0)} = \frac{I(0) + b\psi_1 - I(0)}{I(0) + b\psi_2 - I(0)} = \frac{\psi_1}{\psi_2}$$

Variance (Statistical Stability)

Suppose we wish to compare the statistical stability of 2 assays where assay 1 has the mean window period and shadow given respectively by μ_1 and ψ_1 , and assay 2 has the mean window and shadow given respectively by μ_2 and

ψ_2 . To compare the statistical stability of the estimates from the 2 assays, we consider the relative sample size requirements of the 2 assays (sizes of the cross-sectional surveys) which is defined as the ratio of the sample sizes required by two assays (say N_1 and N_2 which refer to the total numbers in the cross-sectional samples) in order that they each achieve the same level of statistical precision (ie, the same variances). The variance of the estimator in equation (1) is $\frac{I(\psi)}{\mu}$. Therefore, the relative sample size is

$$R = \frac{N_2}{N_1} = k \frac{\mu_1}{\mu_2}$$

where $k = \frac{I(\psi_2)}{I(\psi_1)}$ is the ratio of the HIV incidence rates at the shadows of the two assays. The relative sample sizes depend both on the ratio of the mean window periods and the ratio k of HIV incidence rates at the shadows. In the absence of prior knowledge about the magnitude of k , consider $k = 1$ as the leading case for assessing statistical stability. In that case, the relative sample size requirement is just the ratio of mean window periods, $R = \frac{\mu_1}{\mu_2}$. For example, if assay 2 has a mean window twice as large as assay 1, then assay 1 would require a sample size twice as large as assay 2 to have the same level of statistical stability (eg, variances).

Mixture Model

The window period distribution for the mixture population is

$$S(t) = (1 - p)S_1(t) + pS_2(t)$$

To derive equation (3), we note that the mean of a mixture distribution is

$$\mu = \mu_1(1 - p) + \mu_2p$$

and its variance is

$$\sigma^2 = (\sigma_1^2 + (\mu_1 - \mu)^2)(1 - p) + (\sigma_2^2 + (\mu_2 - \mu)^2)p$$

where σ_i and μ_i are respectively the standard deviation and mean for subpopulation i with survival distribution $S_i(t)$. The mean of the backward recurrence distribution is

$$\psi = \frac{\mu^2 + \sigma^2}{2\mu}$$

Substituting the mean μ and variance σ^2 of the mixture distribution into the above equation gives equation (3).

To derive equations (4) to (6), we make the approximation that the mean of the mixture distribution $\mu_1(1 - p) + \mu_2p \approx \mu_1$ which is an adequate approximation when p is small. Then the mean of the backward recurrence time of the mixture is

$$\psi \approx \frac{\mu_1^2(1 + c_1^2)(1 - p) + \mu_2^2(1 + c_2^2)p}{2\mu_1}$$

Differentiating ψ with respect to μ_1 and setting $\frac{\partial \psi}{\partial \mu_1} = 0$, we obtain equations (4). Substituting equation (4) into the above equation gives equation (5). The results from these analytic equations agree well with the direct numerical results of the optimum (minimum) values of ψ in Figures 2 and 3. Setting $c_1 = 0$ in equation (5) yields equation (6).

Model for Window Period Re-Entry

The model structure is as follows. Persons spend an initial duration of time in the window period according to the survival distribution $S_1(t)$ which has mean μ_1 . The time to AIDS after infection follows a survival distribution $S_2(t)$ which has mean μ_2 . Persons re-enter the window period at the onset of AIDS and remain in the window until death. The duration from AIDS onset until death follows the survival distribution $S_3(t)$, which has mean μ_3 . A proportion α of AIDS cases are not recognized as AIDS and thus not excluded from the window count tally, whereas the remaining proportion $(1-\alpha)$ are correctly identified as AIDS cases and properly excluded from being counted in the window. As shown in¹⁹ if there is re-entry into window periods, the shadow is the mean (expectation) of a random variable with probability density:

$$f(t) = \frac{\phi(t)}{\int_0^{\infty} \phi(s) ds}$$

where, $\phi(t)$ is the probability of being in the window period t years after infection (that includes either the events of being in the window during the initial period or the second period after AIDS onset). Then,

$$f(t) = \frac{S_1(t) + \alpha \int_0^t f_2(u) S_3(t-u)}{\mu_1 + \alpha \mu_3}$$

where, $f_2(u)$ is the probability density of the time to AIDS. The shadow is the expectation associated with the probability density f , that is, $\psi = \int_0^{\infty} t f(t) dt$