Published in final edited form as:

Biometrics. 2012 March ; 68(1): 62–74. doi:10.1111/j.1541-0420.2011.01632.x.

Augmented Cross-Sectional Studies with Abbreviated Follow-up for Estimating HIV Incidence

B. Claggett^{*}, S.W. Lagakos, and R. Wang^{**}

Department of Biostatistics, Harvard University School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A

Summary

Cross-sectional HIV incidence estimation based on a sensitive and less-sensitive test offers great advantages over the traditional cohort study. However, its use has been limited due to concerns about the false negative rate of the less-sensitive test, reflecting the phenomenon that some subjects may remain negative permanently on the less-sensitive test. Wang and Lagakos (2010) propose an augmented cross-sectional design which provides one way to estimate the size of the infected population who remain negative permanently and subsequently incorporate this information in the cross-sectional incidence estimator. In an augmented cross-sectional study, subjects who test negative on the less-sensitive test in the cross-sectional survey are followed forward for transition into the nonrecent state, at which time they would test positive on the less-sensitive test. However, considerable uncertainty exists regarding the appropriate length of follow-up and the size of the infected population who remain nonreactive permanently to the less-sensitive test. In this paper, we assess the impact of varying follow-up time on the resulting incidence estimators from an augmented cross-sectional study, evaluate the robustness of cross-sectional estimators to assumptions about the existence and the size of the subpopulation who will remain negative permanently, and propose a new estimator based on abbreviated follow-up time (AF). Compared to the original estimator from an augmented cross-sectional study, the AF Estimator allows shorter follow-up time and does not require estimation of the mean window period, defined as the average time between detectability of HIV infection with the sensitive and less-sensitive tests. It is shown to perform well in a wide range of settings. We discuss when the AF Estimator would be expected to perform well and offer design considerations for an augmented cross-sectional study with abbreviated follow-up.

Keywords

Augmented; cross-sectional studies; false negative; incidence estimators

1. Introduction

Cross-sectional HIV incidence estimation based on a sensitive and less-sensitive test offers great advantages over the traditional cohort study for incidence estimation (Brookmeyer and Quinn, 1995; Janssen et al., 1998). Based on features of the evolving HIV-1 antibody response during the months after primary HIV-1 infection, several enzyme-immunoassays

^{*}bclagget@hsph.harvard.edu

^{**}rwang@hsph.harvard.edu

Supplementary Material

Web Appendices and Tables referenced in Sections 2.1.3, 3, and 4 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

(EIAs) have been developed to differentiate early and longstanding HIV infections. Subjects who test positive on an EIA sensitive assay but negative on a less-sensitive assay are identified as early infections while subjects who are reactive on both are considered to be long-term HIV infections. An incidence estimate can be obtained by equating incidence with prevalence of persons with early infection divided by the time between seroconversion on the 2 tests (commonly termed “window period”). This approach has attracted much attention. One popular assay is the IgG capture BED EIA (BED) assay (Parekh et al., 2002). Balasubramanian and Lagakos (2010) postulate a longitudinal history statistical model of HIV seroconversion and subsequent reactivity to a less sensitive assay, and show that the cross-sectional estimator arises as a maximum likelihood estimator under this model. They further extend this method to allow the assessment of covariate effects. However, concerns have been raised in recent years regarding the reliability of this method (Sakarovich et al., 2007). Some less-sensitive tests, for example, the BED assay, may yield false negative results; that is, subjects may remain negative to the less-sensitive assay long after they have been infected (Hargrove et al., 2008). It has been noted that if such a subpopulation exists, then the original cross-sectional incidence estimator must be adjusted to account for this fact (Brookmeyer, 2009; Welte and McWalter, 2009; Wang and Lagakos, 2009).

Recently, Wang and Lagakos (2010) propose an augmented cross-sectional study design which adds a longitudinal component to the traditional cross-sectional study by following forward those subjects who test positive on the sensitive assay and test negative on the less-sensitive assay until either they become reactive to the less-sensitive assay or for a predetermined length of follow-up time τ , whichever occurs first. The augmented cross-sectional study provides one way to estimate the size of the subpopulation who will remain negative permanently on the less-sensitive test. An estimate of the false negative rate can be obtained by calculating the proportion of subjects who remain negative at the end of the follow-up period. An underlying assumption for an unbiased estimate of the false negative rate is that τ is chosen to be long enough so that those who have not become reactive to the less-sensitive assay by the end of the follow-up period will remain negative indefinitely. Thus, the choice of τ requires knowledge of an upper limit for the window period (W^*) among those who will become reactive. Two issues arise. First, does there exist a subpopulation where infected subjects would remain nonreactive to the less-sensitive assay permanently? If so, what is the upper limit of the window period among those who will become reactive? Although there is some belief that those who will become reactive to the BED assay will do so within a year, considerable uncertainty regarding this upper limit exists.

In this paper, we assess the impact of varying follow-up time on estimating HIV incidence within the context of an augmented cross-sectional design, evaluate the robustness of incidence estimators to the assumptions about the existence and size of the subpopulation where infected subjects will remain negative permanently on the less-sensitive test, and propose a new estimator based on abbreviated follow-up time (AF). In Section 2.1, we generalize the likelihood function in Wang and Lagakos (2010) by incorporating length of follow-up and analyze the behavior of the estimators when different underlying models are assumed. In Section 2.2, we introduce the AF Estimator, which allows shorter follow-up time than the original estimator based on the augmented design and does not require estimation of the window period. In Section 3, we use simulations to compare the performance of various estimators. In Section 4, we discuss areas of future research. All proofs are relegated to Appendices.

2. Methods

2.1 Notation and Model

We consider the four-state progressive-disease model as in Section 3 of Wang and Lagakos (2010). Here, State 1 represents the pre-seroconversion state (uninfected or infected but not detectable by the sensitive test). State 2 represents the “recent infection” state, in which an infected subject is detectable by the sensitive diagnostic test, but not yet by the less-sensitive test (though he/she will eventually test positive). State 3 represents the “nonrecent infection” state in which an infected individual is detectable by both the sensitive and less-sensitive diagnostic tests. State 4 represents subjects who will remain nonreactive to the less-sensitive assay permanently.

Using the same notation as in Wang and Lagakos (2010), let time 0 denote birth of an individual and T denote the calendar time of HIV seroconversion. Let $f(u)$, $\lambda(u)$, and $F(u)$ denote the density, incidence rate, and cumulative distribution functions of T at time $u \geq 0$. Let W denote the individual's time spent in State 2, with cumulative distribution function $G(\cdot)$. Denote the upper limit of support for W by W^* . Let $\mu = E(W)$ denote the mean time from infection to testing positive on less-sensitive test, amongst the population who will test positive eventually. Let p denote the proportion of such patients among all infected subjects. Let t denote the calendar time of the cross-sectional sample. As in Wang and Lagakos (2010), we assume that T is independent of W and $f(u) = f$, for $u \in (t - W^*, t)$. Our goal is to make inferences about $\lambda(t) = f / \{1 - F(t)\}$, the HIV incidence rate at the time, t , of the cross-sectional sample. Hereafter we denote $\lambda(t)$ as λ for simplicity because the remainder of the paper focuses on a fixed calendar time t , the time at which the cross-sectional sample is obtained.

Suppose that a random sample of size N is drawn from a population of asymptomatic individuals at calendar time t , and tested using both a sensitive and less-sensitive diagnostic test. Let N_1 , N_2 , and N_3 denote the numbers of subjects who test negative on both tests (State 1), positive only on the sensitive test (State 2 or 4), and positive on both tests (State 3), respectively. In an augmented cross-sectional study, we will follow and periodically retest the N_2 subjects who test negative on the less-sensitive test for some time period τ , at which time we will have observed $n_{1\tau}$ transitions to the “non-recent infection” state. Let $n_{0\tau} = N_2 - n_{1\tau}$ denote the number of remaining subjects, including subjects who may become positive on the less-sensitive test at a later time as well as those who will remain negative permanently. Let X denote the forward recurrence time, which represents the time elapsed from t until entrance into State 3. Let $h(x | t)$ and $H(x | t)$ denote the conditional density and cumulative function of X . Each of the $n_{1\tau}$ subjects gives rise to an interval censored observation $[a_i, b_i]$ of the forward recurrence time X_i , where a_i denotes the elapsed time between t and the last negative test result and b_i denotes the elapsed time between t and the first positive test result. Let n_1 and n_0 denote the true number of subjects who are in State 2 and 4, respectively, at time t .

Let $\pi_1(t)$, $\pi_2(t)$, $\pi_3(t)$, and $\pi_4(t)$ denote the prevalence probabilities in State 1, 2, 3, and 4 at time t , respectively. The likelihood function corresponding to an arbitrary τ is given by (see Appendix 1):

$$\begin{aligned} L = & \{\pi_1(t)\}^{N_1} \{\pi_2(t)H(\tau|t)\}^{n_{1\tau}} \{\pi_3(t)\}^{N_3} \times [\pi_4(t) + \pi_2(t)\{1 - H(\tau|t)\}]^{n_{0\tau}} \times \prod_{i=1}^{n_{1\tau}} \frac{H(b_i|t) - H(a_i|t)}{H(\tau|t)} \\ = & \varphi^{N_1} \{p(1 - \varphi - \varphi\lambda\mu)\}^{N_3} (\varphi\lambda p\mu)^{n_{1\tau}} \times [(1 - p)(1 - \varphi) + \varphi\lambda\mu p\{1 - H(\tau|t)\}]^{n_{0\tau}} \times \prod_{i=1}^{n_{1\tau}} \{H(b_i|t) - H(a_i|t)\} \end{aligned} \quad (1)$$

where $\varphi = 1 - F(t)$ and

$$H(x|t) = \int_0^x \frac{1 - G(v)}{\mu} dv, \quad (2)$$

as shown in Wang and Lagakos (2010). This relationship implies that $H(\cdot)$ is uniquely determined once $G(\cdot)$ is specified. Because the right hand side of (2) is free of t , and the remainder of the paper focuses on a fixed time t , we hereafter use $H(x)$ to represent $H(x / t)$ for simplicity. We note that without parametric assumptions about the distribution $G(\cdot)$, the parameters (μ, p) are non-identifiable, as the true state of subjects who have yet to transition to State 3 by the end of follow-up cannot be known. Even with a parametric assumption, the parameters (μ, p) are weakly identifiable, in the sense that the likelihood function is relatively flat near the true parameter values, resulting in a near-singular information matrix. Below we attempt to address this issue by making additional assumptions on either p or $G(\tau)$.

2.1.1 Assume $p = 1$ —If all infected subjects will eventually test positive on the less-sensitive test, then $p = 1$ and the above likelihood reduces to

$$\begin{aligned} L &= \varphi^{N_1} (1 - \varphi - \varphi\lambda\mu)^{N_3} (\varphi\lambda\mu)^{n_{1\tau}} \times \prod_{i=1}^{n_{1\tau}} \{H(b_i) - H(a_i)\} [\varphi\lambda\mu\{1 - H(\tau)\}]^{n_{0\tau}} \\ &= \varphi^{N_1} (1 - \varphi - \varphi\lambda\mu)^{N_3} (\varphi\lambda\mu)^{N_2} \prod_{i=1}^{N_2} \{H(b_i) - H(a_i)\}. \end{aligned}$$

For those subjects who have not become reactive to the less-sensitive assay at time τ , we let $a_i = \tau$ and $b_i = \infty$. The above likelihood is given in equation (5) of Wang and Lagakos (2010). The assumption of $p = 1$ implies that all recently infected individuals will eventually test positive on the less-sensitive test. If this assumption is true, then the resulting estimator

$$\widehat{\lambda}_A = \frac{N_2}{N_1 \widehat{\mu}}$$

(hereafter referred to as Estimator A) is consistent regardless of follow-up time τ , provided that the parametric assumption about the underlying distribution for W is correct. Here we use $\widehat{\mu}$ to denote the maximum likelihood estimate of μ .

However, if this assumption of $p = 1$ is violated, then it was shown in Wang and Lagakos (2009) that this estimator $\widehat{\lambda}_A$ converges in probability to

$$\frac{p\varphi\lambda\mu + (1-p)(1-\varphi)}{\varphi\mu_{\tau A}} \quad (3)$$

where $\mu_{\tau A}$ is the probability limit of the (biased) estimate of μ resulting from τ units of follow-up and the (improper) assumption that $p = 1$.

Under mild violations of this assumption ($p = .99$), we find that the asymptotic bias is modest for $\tau \leq \mu$, but then increases quickly, with $E(\widehat{\lambda}_A)/\lambda > 1.25$ when $\tau \geq 2\mu$ under many plausible scenarios, with the degree of bias increasing with the population prevalence, leading to a significant overestimation of the true incidence rate.

2.1.2 Assume that follow-up time is longer than W^* —When the follow-up time τ is longer than W^* , we have $H(\tau) = 1$, and the likelihood function (1) reduces to

$$L = \varphi^{N_1} \{p(1 - \varphi - \varphi\lambda\mu)\}^{N_3} (\varphi\lambda p\mu)^{n_{1\tau}} \{(1 - p)(1 - \varphi)\}^{n_{0\tau}} \times \prod_{i=1}^{n_{1\tau}} \{H(b_i) - H(a_i)\}.$$

This corresponds to the scenario described in Section 3 of Wang and Lagakos (2010) when all individuals in State 2 are followed until they have entered State 3. If we know W^* up to an upper limit, then we can choose $\tau > W^*$, so that $H(\tau) = 1$.

The resulting estimator,

$$\widehat{\lambda}_B = \frac{n_{1\tau}}{N_1 \widehat{p}_\tau \widehat{\mu}_\tau}$$

(hereafter referred to as Estimator B) is consistent, where $\widehat{p}_\tau = (N_3 + n_{1\tau}) / (N_3 + N_2)$.

If this assumption $\tau > W^*$ is violated, then some of the $n_1 - n_{1\tau}$ truly recently infected individuals who have not transitioned during the follow-up period may be incorrectly classified as “long-term non-progressors”. It is shown in Appendix 2 that this estimator converges in probability to

$$\lambda \frac{\mu}{\mu_{\tau B}} \frac{(1 - \varphi)H(\tau)}{1 - \varphi - \varphi\lambda\mu\{1 - H(\tau)\}} \quad (4)$$

where $\mu_{\tau B}$ is the probability limit of the (biased) estimate of μ resulting from τ units of follow-up and the (improper) assumption that $H(\tau) = 1$.

In Figure 1, we show an example of the probability limits of Estimators A and B as a function of follow-up time τ , for $p = 1$ (left panel) and $p = .95$ (right panel). For this example, we assumed prevalence and incidence values of 15% and 2%, respectively, with mean window period of 6 months and W following a Weibull distribution with shape parameter $k = 2$. We used (3) and (4) to calculate the probability limits at each time τ . Asymptotic values of $\widehat{\mu}_{\tau A}$ and $\widehat{\mu}_{\tau B}$ were estimated by simulating data representing approximately one million subjects with biweekly follow-up. The asymptotic bias of Estimator A is seen to be quite sensitive to departures from the assumption of $p = 1$, and increases as τ increases. The limit of Estimator A rises to 2.9% for $\tau = 52$ and 3.4% when $\tau = 78$ weeks. The asymptotic bias of Estimator B is comparatively small even when τ is much smaller than W^* .

2.1.3 Assume $p = p_0 < 1$ —If the assumption of $p = 1$ is likely to be violated, and we are unable to posit a value for W^* , we can assume a pre-specified value of p (denoted p_0), possibly on the basis of prior studies in a similar population, in which case the likelihood becomes

$$L = \varphi^{N_1} \{p_0(1 - \varphi - \varphi\lambda\mu)\}^{N_3} \times [(1 - p_0)(1 - \varphi) + \varphi\lambda p_0\mu\{1 - H(\tau)\}]^{n_{0\tau}} \times (\varphi\lambda p_0\mu)^{n_{1\tau}} \times \prod_{i=1}^{n_{1\tau}} \{H(b_i) - H(a_i)\} \quad (5)$$

To obtain the MLEs for (λ, ϕ, μ) , we can again assume some parametric distribution for $G(\cdot)$ and numerically maximize L . We refer to this maximum likelihood estimate of λ , $\hat{\lambda}_C$, as Estimator C. An estimate of the covariance matrix of (λ, ϕ, μ) is provided by the sample Fisher information corresponding to L in (5). In Web Appendix A, we derive the estimated covariance matrix assuming that $G(\cdot)$ follows a Weibull distribution, a flexible distribution with support on positive numbers which is commonly-used in parametric survival modeling. This estimator is consistent when p_0 and the parametric assumption about $G(\cdot)$ are correctly specified, regardless of length of follow-up τ . In Section 3, we examine the robustness of this estimator to violation of assumptions about p_0 and parametric distributions through simulations. For estimators A, B, and C, a $(1 - \alpha)100\%$ confidence interval is given by

$$(\hat{\lambda}e^{-Z_{1-\alpha/2}\hat{s}/\hat{\lambda}}, \hat{\lambda}e^{Z_{1-\alpha/2}\hat{s}/\hat{\lambda}}), \quad (6)$$

where $\hat{s}$ is the estimated standard error for $\hat{\lambda}$. The confidence intervals in (6) are additive on the log scale. As suggested by the associate editor, we also consider the use of profile likelihood confidence intervals in Section 3.

2.2 AF Estimator

2.2.1 Motivation—If we were to conduct a cohort study to estimate HIV incidence, i.e., the rate of transition from State 1 to State 2 in our disease model, we could estimate the incidence rate by # of events/person-years, where “number of events” refers to the number of transitions from State 1 to State 2, and “person-years” refers to the cumulative time contributed by the uninfected individuals in the cohort. Let T_2 denote the time of entering State 3 and $f_2(\cdot)$ denotes its corresponding density function. Under the assumption of a constant infection density, $f(u) = f$ for $u \in (t - W^*, t)$, the probability of transition from State 1 to State 2 at time t , $f(t)$, is related to the probability of transition from State 2 to State 3, $f_2(t)$. In Appendix 3 we show that $f(t) = f_2(t)/p$, which implies that

$$\lambda = \frac{f}{1 - F(t)} = \frac{f_2(t)}{p(1 - F(t))}. \quad (7)$$

Assume $f_2(u) = f_2$, for $u \in (t, t + \tau)$. This is reasonable for a relatively small τ (e.g., 12 weeks). From (7), we have $\lambda = f_2/[p\{1 - F(t)\}]$. Suppose there is some minimum time, say, τ^* , spent in the recent infection state. That is, an infected individual will test negative on the less-sensitive assay for at least τ^* from seroconversion. If the follow-up time τ is chosen such that τ is less than this minimum time τ^* , then $G(\tau) = 0$. This implies that $P(\text{Subject in State 1 at } t / \text{Subject in State 3 at } t + \tau) = 0$. Hence, if we are interested in observing the number of transitions into State 3 that occur in our sample between the time of the cross-sectional testing, t , and the end of follow-up, $t + \tau$, we only need to follow the N_2 patients who test positive on the sensitive test and negative on the less-sensitive test at time t . The total number of transitions into State 3 during interval $(t, t + \tau)$ is given by $n_{1\tau}$ and f_2 can be estimated by $n_{1\tau}/(N\tau)$. This, combined with the fact that $F(t)$ can be estimated by $(N - N_1)/N$ (Wang and Lagakos, 2009), motivates us to consider the following estimator:

$$\hat{\lambda} = \frac{n_{1\tau}}{N_1 \hat{p} \tau}, \quad (8)$$

where

$$\widehat{p} = \frac{N_3 + n_{1\tau}}{N_3 + N_2}, \quad (9)$$

as used in Estimator B. We note that $\widehat{p}$ will usually underestimate the true value of p because the use of a shorter follow-up time means that $n_{1\tau}$ will underestimate the true number of transitions, n_1 , to State 3, had we been able to follow all N_2 subjects found to be in the recent state for W^* . However, this will help to compensate for the fact that $G(\tau)$ may not be strictly equal to 0 and hence $n_{1\tau}$ could also underestimate the true number of transitions to State 3 had we been able to follow all $N_1 + N_2$ subjects “at risk” for transition into State 3 during the interval $(t, t + \tau)$. We refer to this design as an augmented cross-sectional design with abbreviated follow-up and the corresponding estimator as the AF Estimator.

We note that this estimator has the same form as Estimator B, with the value $\hat{\mu} = \tau$. In fact, if τ is small, but we continue to use observed forward recurrence times for the $n_{1\tau}$ subjects to estimate the mean and the shape parameter (k) of a Weibull distribution $G(\cdot)$ as described in Estimator B, we find through simulations that $\hat{k}$ tends to be quite large and $\hat{\mu} \approx \tau$.

2.2.2 Properties—In Appendix 4 we show that

$$E(\widehat{\lambda}_{AF}) \approx \frac{\lambda \mu H(\tau)}{\tau} \frac{1}{1 - p_r \{1 - H(\tau)\}},$$

where

$$p_r = \frac{\lambda \varphi \mu}{1 - \varphi} = P(\text{recent infection} | \text{currently infected}). \quad (10)$$

In a region with high prevalence and incidence, for example, $(\lambda, \varphi, \mu) = (.04, .7, .5)$, we have $p_r < .05$. As another example, a region with low prevalence and incidence with $(\lambda, \varphi, \mu) = (.01, .92, .5)$ would yield $p_r < .06$. Because p_r is small and for small τ , $H(\tau) \approx \tau h(0) = \tau/\mu$, we have

$$E(\widehat{\lambda}_{AF}) \approx \frac{\lambda \mu H(\tau)}{\tau} \frac{1}{1 - p_r \{1 - H(\tau)\}} \approx \frac{\lambda \mu \tau h(0)}{\tau} = \frac{\lambda \mu \tau}{\tau \mu} = \lambda.$$

It is shown in Appendix 5 that

$$\widehat{\text{Var}}\{\log(\widehat{\lambda}_{AF})\} \approx \frac{1}{n_{1\tau}}.$$

Thus, for shorter follow-up times τ , a simple estimator and 95% CI could be calculated by counting the number of transitions occurring during the τ follow-up time, with the point estimate $\widehat{\lambda}_{AF}$ given in (8) and a $(1 - \alpha)\%$ confidence interval

$$(\widehat{\lambda}_{AF} e^{-Z_{1-\alpha/2} \sqrt{\frac{1}{n_{1\tau}}}}, \widehat{\lambda}_{AF} e^{Z_{1-\alpha/2} \sqrt{\frac{1}{n_{1\tau}}}}).$$

For the AF Estimator to be expected to be close to the true incidence λ by a factor of B , where $B < (1 - p_r)$, i.e. $B\lambda \leq E(\hat{\lambda}_{AF}) \leq \lambda/B$, we can choose τ such that (see Appendix 6)

$$\tau \leq \frac{\int_0^\tau 1 - G(x) dx}{B[1 - p_r\{1 - H(\tau)\}]} \quad (11)$$

Because $[1 - p_r\{1 - H(\tau)\}] \leq 1$ and close to 1, a slightly more strict condition that will guarantee the boundedness of the bias of the AF Estimator is

$$\tau \leq \frac{\int_0^\tau 1 - G(x) dx}{B} \quad (12)$$

The above inequality highlights the dependence of the AF Estimator on the assumption of a relatively “flat” survival function $1 - G(x)$: $x \in (0, \tau)$, i.e. some minimum amount of time spent in State 2. A “perfect” test ($W = \mu$, with probability 1) would result in $g(x) = 0$ on the range $(0, \tau)$, and thus $\int_0^\tau 1 - G(x) dx \equiv \tau$ for any $\tau < \mu$. In practice, the less-sensitive test does not have this “perfect” property. We find that the bias of the AF Estimator generally increases as the choice of follow-up time τ increases. Because the width of the confidence interval based on the AF Estimator is purely “event-driven”, it becomes smaller with each additional observed transition. Therefore, a larger value of τ is associated with increasing bias and decreasing variance. We have found through extensive simulations that $\tau \approx \mu/2$ is generally a good choice. Unless prior information is known about the value of μ in the population of interest, we recommend the choice of $\tau = 12$ weeks for use with the AF Estimator. We show in Section 3 that this leads to good performance in a variety of settings.

In practice we do not know the exact distribution of window period (W) in a given population. Rather, we may have some idea of mean window period μ and possibly coefficient of variation $c = \sigma/\mu$. With a parametric assumption for the window period distribution and (μ, c) , we can calculate $H(\tau)$ and verify whether a 12-week AF Estimator satisfies $B\lambda \leq E(\hat{\lambda}_{AF}) \leq \lambda/B$ for a fixed B and a hypothetical value for p_r , which is determined by (λ, ϕ, μ) as in (10). In Figure 2(a), we provide the acceptable region, in terms of μ and c , of the 12-week AF Estimator for $B = 0.9$ and $p_r = .02, .04, .06$, and $.1$. The boundaries presented in this plot are generated by finding $\mu_{min;c}$ that satisfies (11) for each fixed c . We note that for $\mu_1 < \mu_2$, if (μ_1, c) is in the acceptable region, then (μ_2, c) must also be in the acceptable region. Therefore, $\mu_{min;c}$ provides a lower bound for all μ that satisfies (11) for each c . We also find that the region which guarantees $B\lambda \leq E(\hat{\lambda}_{AF}) \leq \lambda/B$ when the true distribution is Weibull is somewhat smaller than the corresponding region when the true distribution is Lognormal. Because of this, we recommend using the reference region based on the Weibull assumption when deciding whether to use the AF Estimator, in order to be conservative. An even more conservative boundary can be generated using (12), which does not depend on p_r and hence does not require assumptions about (ϕ, λ) . Figure 2(b) shows the actual coverage levels of the nominal 95% confidence intervals of the 12-week AF Estimator for different values of (c, μ) , in a setting where $(N, \lambda, \phi, p) = (6000, .03, .85, .99)$. We find that the actual coverage levels are close to the nominal level for all parameter pairs that fall within the acceptable region corresponding to $B = .9$, as shown in Figure 2(a).

2.2.3 Design Considerations—Suppose that one wanted to conduct a study to estimate the incidence of HIV in a region with some hypothesized parameter values (ϕ, λ, p) , and had decided that the usage of the AF Estimator is both desirable and appropriate. Under this proposed design plan, biweekly follow-up is not necessary, as only the total number of

transitions occurring during the 12 weeks of follow-up is needed to produce this estimate of incidence, and could be obtained via a single follow-up visit 12 weeks ($\tau=12/52$ years) after the initial testing. This designed study would require the administration of N (typically thousands) sensitive tests and $N - N_1$ less-sensitive tests at time t , and a further N_2 (typically < 100) less sensitive tests at time $t + \tau$. The precision of the AF Estimator is closely tied with the number of transitions into State 3 within 12 weeks ($n_{1\tau}$). Below we show how to calculate the total sample needed so that the ratio of the true incidence rate and the estimated rate, $e^{Z_{1-\alpha/2} \sqrt{1/n_{1\tau}}}$, is smaller than δ with $(1 - \alpha)100\%$ confidence. To achieve this, we require that $n_{1\tau} \geq \{Z_{1-\alpha/2}/\log(\delta)\}^2$. For example, for $\delta = 2$ and $\alpha = .05$, we would need the number of transitions $n_{1\tau}$ to be at least 8. If we would like to achieve better precision, and let $\delta = 1.5$, then we would require $n_{1\tau} \geq 24$.

Earlier we showed that the probability that a recently infected subject will transition to State 3 within τ units of follow-up is $\phi\lambda p\mu H(\tau) \approx \phi\lambda p\tau$ for small τ . The sample size N has the negative binomial distribution with number of failures $r = n_{1\tau}$ and success rate $P_s = 1 - \phi\lambda p\tau$. Therefore, the sample size N required to obtain the observation of at least $n_{1\tau}$ failures with probability β is $Q_{NB}(\beta, r = n_{1\tau}, P_s = 1 - \phi\lambda p\tau)$, where Q refers to the quantile function of negative binomial distribution. A close approximation to this value can be obtained by the following formula

$$\tilde{N} = \frac{(n_{1\tau} + Z_\beta \cdot \sqrt{n_{1\tau}})(1 - \phi\lambda p\tau)}{\phi\lambda p\tau}. \quad (13)$$

For example, in a high prevalence ($\phi = .75$) and high incidence ($\lambda = .04$) population, if we assume that $p = .94$, and would like to obtain at least 8 transitions into State 3 within 12 weeks with probability .9, then we would need to include at least $N = 1799$ subjects in the cross-sectional sample. Suppose, instead, that we are in a setting with lower prevalence ($\phi = .9$) and incidence ($\lambda = .02$) and wish to observe at least 24 transitions with probability .7. In this case, we would need $N = 6713$. Using the formula in (13) yields $\tilde{N} = 1775$ and 6778, respectively, for the two settings discussed.

For a fixed sample size N and given parametric distribution $G(\cdot)$ and parameter values (ϕ, λ, p), we can find the optimal follow-up time τ that minimizes the mean squared error (MSE) of the AF Estimator numerically. For example, for the scenario where $\phi = .75$, $\lambda = .04$, and $N = 1799$, assuming that $G(\cdot)$ is Weibull with $\mu = .3$ and $c = .6$, we find that the MSE is minimized near $\tau = 13$ weeks, with relatively little change in the MSE over the range $\tau \in (10, 15)$ weeks.

2.3 Relationship Among Estimators

All of the estimators discussed in Sections 2.1 and 2.2 can be viewed as special cases of the most general maximum likelihood estimator of incidence given τ units of follow-up time,

$$\hat{\lambda} = \frac{n_{1\tau}}{N_1 \hat{p} \hat{\mu} \hat{H}(\tau)}$$

In general, it is not possible to estimate all of these quantities $\{p, \mu, H(\tau)\}$ well without external information or assumptions.

If we are in a setting where all apparent recent infections will eventually test positive on the less-sensitive test, and that the concept of “long-term non-progressor” does not apply, then $p = 1$, $n_1 = N_2$ and $\hat{H}(\tau) = n_{1\tau}/N_2$. Therefore,

$$\widehat{\lambda} = \frac{n_{1\tau}}{N_1 \widehat{p} \widehat{\mu} \widehat{H}(\tau)} = \frac{N_2}{N_1 \widehat{\mu}},$$

which represents Estimator A. Alternatively, we could assume that the value of p is known to be p_0 , and not equal to 1. In this case we need to estimate $\{\mu, H(\tau)\}$, and results in Estimator C,

$$\widehat{\lambda} = \frac{n_{1\tau}}{N_1 p_0 \widehat{\mu} \widehat{H}(\tau)}.$$

Rather than making assumptions about the value of p , one could instead make assumptions about $G(\cdot)$, and thus $h(\cdot)$, $H(\cdot)$, and μ . As discussed earlier in Wang and Lagakos (2010), if we have some knowledge of W^* , and choose $\tau \geq W^*$, then $\hat{H}(\tau) = 1$ and $n_1 = n_{1\tau}$, and

$$\widehat{\lambda} = \frac{n_{1\tau}}{N_1 \widehat{p} \widehat{\mu} \widehat{H}(\tau)} = \frac{n_1}{N_1 \widehat{p} \widehat{\mu}},$$

which represents Estimator B. If we do not have any knowledge of W^* , or we are unable to follow subjects for $\tau > W^*$ for logistical reasons, the AF Estimator may offer an attractive alternative if (μ, c) falls within the acceptable region. The AF Estimator works well when $G(\tau) \approx 0$. This results in $\hat{H}(\tau) = \tau/\hat{\mu}$ for small τ , and hence

$$\widehat{\lambda} = \frac{n_{1\tau}}{N_1 \widehat{p} \widehat{\mu} \widehat{H}(\tau)} = \frac{n_{1\tau}}{N_1 \widehat{p} \tau}.$$

3. Simulation Studies

We have discussed four estimators for estimating HIV incidence in an augmented cross-sectional study with follow-up time τ . We performed an extensive simulation study to compare their performance under a variety of situations. We assumed $(N, \lambda, \varphi) = (6000, .02, .85)$ throughout. We let $p = 0.95$ or 1 . We considered Weibull and Lognormal for the underlying distribution $G(\cdot)$, with mean $\mu = .3, .4$, or $.5$, and the coefficient of variation $c = .4$, or $.6$. The length of follow-up τ was taken to be 12, 24, or 52 weeks respectively. The frequency of follow-up visits is assumed to be biweekly, except for the AF Estimator, which only requires a single follow-up visit at τ weeks. In the case of Estimators A and C, we recorded right-censored observations using $b = 999$ weeks as the upper bound for the transition time of any patient whose transition is not observed during follow-up. For Estimator C, which calls for a pre-specified value for p , p_0 was taken to be $.91, .95$, or $.99$. For Estimators A, B, and C, we assumed a Weibull distribution in our maximization procedure throughout the paper because we found that using a Weibull distribution yields more robust results than using a Lognormal distribution, as in Wang and Lagakos (2010). Results corresponding to the usage of a Lognormal distribution in the maximization

procedure are included in the Web Tables 1 and 2. We also considered confidence intervals based on profile likelihood methods. The width and coverage associated with profile likelihood confidence intervals are comparable to those obtained based on (6) (see Web Table 3), although we noticed that some computational difficulties arose when attempting to profile out the shape parameter of the Weibull distribution, forcing us to impose a lower bound on this value.

When $p = 1$, although the actual coverage of the confidence intervals based on Estimator A is close to the nominal level, we noticed that it tends to overestimate the true incidence rate in the settings we examined. For smaller value of τ , the variance associated with Estimator A is quite large in relative to the true incidence rate. Because λ is bounded below at 0, large variation in $\hat{\lambda}$ makes it more likely for us to obtain larger values of $\hat{\lambda}$ than smaller values of $\hat{\lambda}$, resulting in the average of $\hat{\lambda}$ to be biased high. We believe that this is a finite sample problem. We also performed additional simulations with larger sample size, and noticed that this overestimating phenomenon goes away. The precision of Estimator A improves for larger values of τ . When $p=.95$, the bias associated with Estimator A increases with larger τ and becomes substantial when $\tau = 52$ weeks. Moreover, its precision also worsens substantially as τ gets larger. These results suggest that Estimator A is very sensitive to assumptions about p . Unless we are certain about $p = 1$, the use of Estimator A would best be avoided.

When $H(\tau) \approx 1$, Estimator B performs well as expected, for Weibull or Lognormal $G(\cdot)$, as in Wang and Lagakos (2010). Its actual coverage level is generally near the nominal level. When $H(\tau) < 1$, we observe modest bias for Estimator B and the actual coverage level can be substantially lower than the nominal level when $\tau = 24$ weeks. Interestingly, we find that Estimator B provides adequate coverage when $\tau = 12$ weeks, and showed greater precision than Estimator A in such scenarios. This phenomenon is closely related to the good performance of the AF Estimator and was also part of our motivation for considering the AF Estimator.

When p_0 is correctly specified, Estimator C performs well in general, showing coverage near the nominal level and precision improving with increasing values of τ . When the hypothesized p_0 is smaller than the true p , Estimator C behaves like Estimator B, maintaining actual coverage greater than 92% for $\tau = 12$ or $\tau = 52$ weeks and can have the under-coverage problem for $\tau = 24$ weeks in some settings. When the hypothesized p_0 is larger than the true p , Estimator C behaves like Estimator A, and can lead to substantial bias and greater variances.

The 12-week AF Estimator performs well in scenarios (see Table 3 and Web Table 4) within the acceptable region outlined in Figure 2(a). For these situations, it shows greater precision than all other estimators with $\tau = 12$ weeks. For settings in the non-acceptable region, the use of the AF Estimator should be avoided. As the follow-up time increases, the AF Estimator becomes less variable yet more biased. We observe severe bias when the follow-up is longer than 6 months, highlighting the fact that its performance depends greatly on the assumption that the follow-up time τ is sufficiently small. Therefore, we are primarily interested in the 12-week AF Estimator throughout the paper.

Illustrative Examples

Because data following such an augmented cross-sectional design do not yet exist, we chose 2 of our simulated data sets from the scenario with Weibull $G(\cdot)$ and $(\mu, c) = (.5, .4)$ for illustration. The first data set was generated with $p=1$, indicating that the conditions for Estimator A were met for any choice of τ , while the assumptions underlying Estimator B were approximately satisfied ($H(\tau) > .99$) for $\tau \geq 44$ weeks. In Figure 3 (left panel), we see

that both Estimators A and B have point estimates near the true incidence value and confidence intervals containing the true value, for any choice of τ . Furthermore, we see that the width of the confidence interval of Estimator B is comparable to that of Estimator A for most values of τ . The second data set was generated with $p=.95$, indicating that the conditions for Estimator A were never met in this setting. In this example (right panel), we see that Estimator A suffers greatly in its precision as τ increases. The AF Estimator exhibits severe bias with longer follow-up times. The 12-week AF Estimator performs well in both cases.

4. Discussion

In this paper, we assess how the performance of various estimators are affected by length of follow-up τ and/or assumptions about the size of the “long-term non-progressor” population. We found that Estimator A, which assumes $p = 1$, can result in substantial bias and great loss of precision if the true p is less than 1. This pattern also holds for Estimator C when the assumed p_0 is higher than the true p . When p is correctly specified, both Estimator A and Estimator C are consistent regardless of length of follow-up. However, p is often not known in practice and is expected to vary according to the type of less-sensitive test and population of interest. Estimator B does not require prior knowledge of p , but instead requires length of follow-up to be long enough so that subjects who have not become reactive to the less-sensitive test at the end of the follow-up will remain non-reactive permanently. We found that Estimator B is robust to modest violation of this assumption. Interestingly, when we apply Estimator B with very short follow-up, say, 12 weeks, it works reasonably well and has led us to consider a new HIV incidence estimator based on an abbreviated follow-up time, the AF Estimator.

The AF Estimator has the benefit of not requiring any specific external estimate of p or μ , as well as substantially shortening the amount of follow-up compared to earlier augmented cross-sectional designs without substantial loss of precision. The AF Estimator requires rather general assumptions about the distribution of the time spent in the “recent infection” state, and its use will become more feasible as more information becomes available characterizing these distributions associated with particular less-sensitive tests in different populations. We note that the length of follow-up τ may be chosen to be somewhat smaller than 12 weeks, which will further relax the distributional assumptions necessary for the AF Estimator. However, the sample size will need to be increased accordingly, as shown in (13), to maintain desired precision. Since the precision of the AF Estimator is dependent solely on the observed number of transitions into State 3, it is recommended that all N_2 patients be retested at the 12-week follow-up visit. However, if it is only feasible to follow a random subsample of size $n_2 < N_2$ of these patients, the AF Estimator can be easily adapted by replacing the value $n_{1\tau}$ with $n_{1\tau} \cdot N_2/n_2$ in (8) and (9). The procedure for calculating the confidence intervals remains unchanged.

The AF Estimator, despite its attractive properties, is biased and therefore currently can only serve as a crude estimator when resources are limited. When possible, we recommend the use of Estimator B proposed in Wang and Lagakos (2010) and to follow subjects for a reasonably long period of time. To facilitate the implementation of the augmented cross-sectional design and the use of Estimator B, it would be helpful to provide guidance on several design issues such as sample size of the cross-sectional sample, number of subjects needed to be followed and visit frequency. Because an explicit variance formula for $\hat{\mu}$ is not readily attainable, we do not have a formula for sample size yet. Using simulations and a grid search, we can calculate sample sizes needed to estimate incidence in a population of interest. For example, for one of the settings we examined in 2.2.3, when $(\phi, \lambda, p) = (.75, .04, .94)$, if we assume the underlying distribution for the window period is Weibull with μ

$= .5$ and $c = .3$, we would need about 2200 subjects in the cross-sectional sample and to follow all subjects found to be in ‘recent state’ for about 1 year, in order for $\delta = 2$ with probability 80%. Interestingly, we found that the changes in the required sample size were negligible when we relaxed a 2-week visit schedule to a 6-week visit schedule. Whether this holds true in general warrants further investigation.

We emphasize the importance of quantifying p . With correct knowledge of p , Estimator A or Estimator C would work well with shorter follow-up than that required by Estimator B. Furthermore, if p is known, and it is possible to devise a less-sensitive test which has the property that subjects will not become reactive for a minimum period of time τ^* , then the AF Estimator would be unbiased with $\tau < \tau^*$.

We investigated the implementation of a Bayesian model to address the problem at hand with what we consider to be fairly vague priors (Web Appendix B). The results are included in Web Table 5. When the true parameters lie well within the interior of the prior distributions, the Bayesian method produces credible intervals with width narrower than those obtained from Estimators A, B, and C, and comparable to those from the AF Estimator. However, the Bayesian method is seen to have under-coverage problems when the true parameters are close to the boundary of their prior distributions. For example, when true $\mu = .3$ years and the prior distribution was chosen as Uniform on the interval from .25 years to 1.5 years, even though the prior distribution has non-negligible mass near the true value, the posterior distribution is not symmetric around the true value, due to the truncation at .25 years. The strong negative correlation between $\hat{\lambda}$ and $\hat{\mu}$ can result in $\hat{\lambda}$ being biased low. Thus the corresponding 95% credible interval may not have the nominal coverage. It would also be useful to investigate whether the EM algorithm can be used effectively in the calculation of Estimator C, by iteratively assigning the $n_{0\tau}$ patients to either State 2 or State 4, based on the estimated parameter values at each step, then re-estimating the parameter values by maximizing the updated likelihood function, based on the state assignments.

Recent evidence suggests that use of antiretroviral treatments (ART) can affect the results of less-sensitive diagnostic tests (Marinda et al., 2010). For this reason, it is critical that patients’ ART use be documented. One suggestion is to exclude patients with known ART use from the initial cross-sectional sample (McDougal et al., 2006), though this could potentially impact the validity of the resulting prevalence estimate if ART use is common in the study region. Suppose that some number of the N_2 infected patients, say, n^* , in the cross-sectional survey who test negative on the less-sensitive test are known to be receiving ART. Assuming that these patients’ test results do not change during the follow-up period, all n^* patients could be reassigned to State 3 or “non-recent infection”. An alternative would be to reassign $\hat{p}n^*$ of these patients to State 3, while keeping the remaining $(1 - \hat{p})n^*$ in State 4. Both methods could be employed and the resulting estimates compared as a type of sensitivity analysis. The assumption being made here is that the population of patients currently receiving ART at a given time is likely to be comprised of a disproportionately smaller percentage of “long-term non-progressors” than the overall infected population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We dedicate this paper to our mentor, Dr. Stephen Lagakos, whose guidance and support will not be forgotten. We thank Professor Zucker, an associate editor and two referees for their comments, which improved the paper. This research was supported by grants R37 AI24643 and T32 AI007358 from the National Institutes of Health.

References

- Balasubramanian R, Lagakos SW. Estimating HIV incidence based on combined prevalence testing. *Biometrics*. 2010; 66:1–10. [PubMed: 19397583]
- Brookmeyer R, Quinn TC. Estimation of current Human Immunodeficiency Virus incidence rates from a cross-sectional survey using early diagnostic tests. *American Journal of Epidemiology*. 1995; 141:166–172. [PubMed: 7817972]
- Brookmeyer R. Should Biomarker Estimates of HIV Incidence Be Adjusted? *AIDS* 2009. 2009; 23:485–491.
- Hargrove JW, Humphrey JH, Mutasa K, Parekh BS, McDougal JS, Ntozinie 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. [PubMed: 18301064]
- Janssen RS, Satten GA, Stramer SL, Rawal BD, O'Brien TR, Weiblen BJ, Hecht FM, Jack N, Cleghorn J, 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. *Journal of the American Medical Association*. 1998; 280:42–48. [PubMed: 9660362]
- Marinda ET, Hargrove J, Preiser W, Slabbert H, van Zyl G, Levin J, Moulton L, Welte A, Humphrey J. Significantly Diminished Long-Term Specificity of the BED Capture Enzyme Immunoassay Among Patients With HIV-1 With Very Low CD4 Counts and Those on Antiretroviral Therapy. *JAIDS*. 2010; 53:496–499. [PubMed: 20306555]
- McDougal JS, Parekh BS, Peterson ML, Branson BM, Dobbs T, Ackers M, Gurwith M. Comparison of HIV type I incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. *AIDS Research and Human Retroviruses*. 2006; 22:945–952. [PubMed: 17067263]
- Parekh BS, Kennedy MS, Dobbs T, Pau C, Byers R, Green T, Hu DJ, Vanichseni S, Young NL, Choopanya K, Mastro TD, McDougal S. Quantitative detection of increasing HIV Type 1 antibodies after seroconversion: A simple assay for detecting recent HIV infection and estimating incidence. *AIDS Research and Human Retroviruses*. 2002; 18:295–307. [PubMed: 11860677]
- Sakarovitch C, Rouet F, Murphy G, Minga AK, Alioum A, Dabis F, Costagliola D, Salamon R, Parry JV, Barin F. Do Tests Devised to Detect Recent HIV-1 Infection Provide Reliable Estimates of Incidence in Africa? *Journal of Acquired Immune Deficiency Syndrome*. 2007; 45:115–122.
- Wang R, Lagakos SW. On the Use of Adjusted Cross-Sectional Estimators of HIV Incidence. *The Journal of Acquired Immune Deficiency Syndromes (JAIDS)*. 2009; 52:538–47.
- Wang R, Lagakos SW. Augmented Cross-Sectional Prevalence Testing for Estimating HIV Incidence. *Biometrics*. 2010; 66:864–874. [PubMed: 19912174]
- Welte A, McWalter TA, Bärnighausen 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. [PubMed: 19182926]
- Woolf B. On estimating the relation between blood group and disease. *Annals of Human Genetics*. 1955; 19:251–253. [PubMed: 14388528]

APPENDICES

Appendix 1: Likelihood Function for Model in Section 2.1

The state prevalence probabilities at time t are given by (Balasubramanian and Lagakos 2010, Wang and Lagakos 2009)

$$\begin{aligned}
\pi_1(t) &\stackrel{\text{def}}{=} P(\text{in State 1 at time } t) = 1 - F(t) = \varphi, \\
\pi_2(t) &\stackrel{\text{def}}{=} P(\text{in State 2 at time } t) = p\varphi\lambda\mu, \\
\pi_3(t) &\stackrel{\text{def}}{=} P(\text{in State 3 at time } t) = p(1 - \varphi - \varphi\lambda\mu), \\
\pi_4(t) &\stackrel{\text{def}}{=} P(\text{in State 4 at time } t) = (1 - p)(1 - \varphi),
\end{aligned}$$

The cumulative distribution function of X , for $x > 0$, conditional on t , is (Wang and Lagakos 2010):

$$H(x|t) = \int_0^x \frac{1 - G(v)}{\mu} dv.$$

In our setting, only those who are observed to test positive on the less-sensitive test within the τ units of follow-up (i.e. those with forward recurrence time $X < \tau$), will be determined to definitely be in State 2 at time t . Thus the probability of actually being in State 2 and being observed to transition to State 3 during the follow-up time is $\pi_2 H(\tau | t) = \varphi\lambda p\mu H(\tau | t)$. The probability of actually being in State 2 at the time of the cross-sectional sample, but not being observed to transition into State 3 during the follow-up period is then

$$\pi_2\{1 - H(\tau|t)\} = \varphi\lambda p\mu\{1 - H(\tau|t)\}.$$

The additional information from the interval-censored forward recurrence times of those patients determined to be in State 2 is

$$P(a < x < b | x < \tau) = \frac{H(b|t) - H(a|t)}{H(\tau|t)}.$$

Appendix 2: Probability Limit of $\hat{\Lambda}_B$

$$\hat{\lambda}_B = \frac{n_{1\tau}}{N_1 \hat{p}_\tau \mu_\tau}, \text{ where } \hat{p}_\tau = \frac{N_3 + n_{1\tau}}{N_3 + n_{1\tau} + n_{0\tau}} = \frac{N_3 + n_{1\tau}}{N_3 + N_2}.$$

Let μ_τ be the probability limit of $\hat{\mu}_\tau$. From

$$\frac{n_{1\tau}}{N} \xrightarrow{p} p\varphi\lambda\mu H(\tau), \quad \frac{N_1}{N} \xrightarrow{p} \varphi, \quad \frac{N_2 + N_3}{N} \xrightarrow{p} (1 - \varphi),$$

and

$$\frac{N_3}{N} \xrightarrow{p} p(1 - \varphi - \varphi\lambda\mu),$$

we have

$$\widehat{p}_\tau = \frac{(N_3 + n_{1\tau})/N}{(N_3 + N_2)/N} \xrightarrow{P} \frac{p(1-\varphi-\varphi\lambda\mu) + p\varphi\lambda\mu H(\tau)}{1-\varphi} \\ = \frac{p[1-\varphi-\varphi\lambda\mu\{1-H(\tau)\}]}{1-\varphi},$$

and

$$\widehat{\lambda}_B = \frac{n_{1\tau}/N}{N_1/N\widehat{p}_\tau\widehat{\mu}_\tau} \xrightarrow{P} \frac{p\varphi\lambda\mu H(\tau)}{\varphi \frac{p[1-\varphi-\varphi\lambda\mu\{1-H(\tau)\}]}{1-\varphi} \mu_\tau} \\ = \lambda \frac{\mu}{\mu_\tau} \frac{(1-\varphi)H(\tau)}{[1-\varphi-\varphi\lambda\mu\{1-H(\tau)\}]},$$

Appendix 3: Relationship of f(t) and f2(t)

$$f_2(t) = P(\text{transition from State 2 to State 3 at time } t) \\ = \int P(\text{infected at time } t-u) \times P(\text{in subgroup that will transition to State 3}) \times P(\text{sojourn time in State 2} = u | \text{in subgroup that will transition to})$$

Under the assumption that $f(u)$ is constant over $u \in (t-W^*, t)$,

$$f_2(t) = \int_0^{W^*} f(t-u) \cdot p \cdot g(u) du = f \cdot p \int_0^{W^*} g(u) du = f \cdot p \cdot G(W^*) = pf$$

Appendix 4: Expectation of AF Estimator

$$\text{Recall } \widehat{\lambda}_{AF} = \frac{n_{1\tau}}{N_1 \widehat{p}_\tau \tau} = \frac{n_{1\tau}(N_3 + N_2)}{N_1 \tau (N_3 + n_{1\tau})} = \frac{N_3 + N_2}{N_1} \frac{n_{1\tau}}{N_3 + n_{1\tau}} \frac{1}{\tau},$$

therefore,

$$E(\widehat{\lambda}_{AF}) = E\left(\frac{N_3 + N_2}{N_1} \frac{n_{1\tau}}{N_3 + n_{1\tau}} \frac{1}{\tau}\right) = \frac{1}{\tau} E\left(\frac{N_3 + N_2}{N_1} \frac{n_{1\tau}}{N_3 + n_{1\tau}}\right).$$

Because

$$(N_1, n_{1\tau}, n_{0\tau}, N_3) \sim \text{Multinomial}(N; p_1, p_2, p_3, p_4),$$

we have

$$n_{1\tau} | (N_1, n_{0\tau}) \sim \text{Binomial}(N - N_1 - n_{0\tau}, \frac{p_2}{1 - p_1 - p_3}).$$

Now, we want to show $E(XY) = E(X) E(Y)$, where

$$X = \frac{N_3 + N_2}{N_1}, Y = \frac{n_{1\tau}}{N_3 + n_{1\tau}}$$

$$\begin{aligned} E(Y) &= E\{E(Y|N_1, n_{0\tau})\} = E\{E(\frac{n_{1\tau}}{n_{1\tau} + N_3} | N_1, n_{0\tau})\} \\ &= E\{\frac{1}{n_{1\tau} + N_3} E(n_{1\tau} | N_1, n_{0\tau})\} \\ &= E\{\frac{1}{n_{1\tau} + N_3} \frac{(n_{1\tau} + N_3)p_2}{p_2 + p_4}\} = E(\frac{p_2}{p_2 + p_4}) = \frac{p_2}{p_2 + p_4} \end{aligned}$$

and

$$\begin{aligned} E(Y|X) &= E\{E(Y|N_1, n_{0\tau})|X\} = E(\frac{p_2}{p_2 + p_4} | X) = \frac{p_2}{p_2 + p_4} \\ E(Y|X) &= E(Y) \Rightarrow E(XY) = E(X)E(Y) \end{aligned}$$

so

$$\begin{aligned} E(\widehat{\lambda}_{AF}) &= \frac{1}{\tau} E(\frac{N_3 + N_2}{N_1}) E(\frac{n_{1\tau}}{N_3 + n_{1\tau}}) \\ &\approx \frac{1}{\tau} \frac{1 - \varphi}{\varphi} \frac{p\lambda\varphi\mu H(\tau)}{1 - \varphi - p\varphi\lambda\mu\{1 - H(\tau)\}} \\ &= \frac{\lambda\mu H(\tau)}{\tau} \frac{1}{1 - p_r\{1 - H(\tau)\}} \end{aligned} \quad \begin{matrix} \text{(A.} \\ \text{1)} \end{matrix}$$

where

$$p_r = \frac{\lambda\varphi\mu}{1 - \varphi} = P(\text{recent infection} | \text{currently infected}). \quad \begin{matrix} \text{(A.} \\ \text{2)} \end{matrix}$$

Appendix 5: Approximate Variance of AF Estimator

Using (8) and (9), we have

$$\log(\widehat{\lambda}_{AF}) = \log(\frac{N_2 + N_3}{N_1}) + \log(\frac{n_{1\tau}}{N_3 + n_{1\tau}}) - \log(\tau)$$

Because the two terms are shown to be independent in Appendix 4, this leads to

$$\text{Var}\{\log(\widehat{\lambda}_{AF})\} = \text{Var}\{\log(\frac{N_2 + N_3}{N_1})\} + \text{Var}\{\log(\frac{n_{1\tau}}{N_3 + n_{1\tau}})\}$$

Note that $\log\{(N_2 + N_3)/N_1\}$ is the log odds of HIV prevalence, its variance can be consistently estimated by $1/N_1 + 1/(N_2 + N_3)$ (Woolf 1955). Similarly, by application of the δ -method and common variance formula for sample proportion, the variance of $\log\{n_{1\tau}/(N_3 + n_{1\tau})\}$ can be approximated by $N_3/\{n_{1\tau}(N_3 + n_{1\tau})\}$. Therefore,

$$\widehat{\text{Var}}\{\log(\widehat{\lambda}_{AF})\} \approx \frac{1}{N_1} + \frac{1}{N_2 + N_3} + \frac{N_3}{n_{1\tau}(N_3 + n_{1\tau})} \approx \frac{1}{n_{1\tau}},$$

where the last approximation follows from that N_1 and N_3 are usually much larger than $n_{1\tau}$, and therefore $1/N_1$ and $1/(N_2 + N_3)$ are negligible compared to $1/n_{1\tau}$, and $N_3/(N_3 + n_{1\tau}) \approx 1$.

Appendix 6: Bounding the Bias of AF Estimator

Suppose we wish to bound the bias of the AF Estimator, such that the expected value of the AF Estimator is close to the true incidence rate, to within a factor of B . That is, $B\lambda \leq E(\widehat{\lambda}_{AF}) \leq \lambda/B$. Using (A.1), after some simplification we have

$$B\tau[1 - p_r\{1 - H(\tau)\}] \leq \int_0^\tau 1 - G(x) dx \leq \frac{\tau}{B}[1 - p_r\{1 - H(\tau)\}]$$

where p_r is defined in (A.2).

If the desired bound B is chosen such that $B \leq (1 - p_r)$, then the second inequality is always satisfied, because $\tau \leq \tau[1 - p_r\{1 - H(\tau)\}]/B$ and $\int_0^\tau 1 - G(x) dx \leq \tau$ since $G(\cdot) \geq 0$ always. In this case, we only need to consider the first inequality, which can be expressed as

$$\tau \leq \frac{\int_0^\tau 1 - G(x) dx}{B[1 - p_r\{1 - H(\tau)\}]}.$$

If a tighter bound B is desired, such that $B > (1 - p_r)$, we must also satisfy the second inequality, which can be expressed as

$$\tau \geq B \frac{\int_0^\tau 1 - G(x) dx}{[1 - p_r\{1 - H(\tau)\}]}.$$

Figure 1. Probability Limits of Estimator A (open circles) and Estimator B (solid circles) for Various τ and p , with $\phi = .85$, $\lambda = .02$, and Assuming that $G(\cdot)$ is Weibull with Mean .5 and Shape Parameter $k = 2$.

Figure 2.

Bias and Coverage of 12-week AF Estimator

Left: The 12-week AF Estimator will have acceptably small bias when (c, μ) lie above the line associated with desired bound $B=0.90$

Right: Actual Coverage of Nominal 95% CI, Coverage values obtained via simulation using $(\phi, \lambda, p) = (0.85, 0.03, 0.99)$ and $G(\cdot) \sim \text{Weibull}$

Figure 3. Point estimates and 95% confidence intervals for the incidence rate at different follow-up times τ . The data sets were chosen from the simulation study described in Section 3. Black Dashed: Estimator A, Black Solid: Estimator B, Gray Dotted: AF Estimator $G(\cdot) \sim \text{Weibull}, (\mu, c) = (.5, .4)$

Table 1

Simulation results for Estimators A and B, with $N=6000$, $\phi = .85$ and $\lambda = .02$, and assuming Weibull $G(\cdot)$. $E(\hat{\lambda})$ and $sd(\hat{\lambda})$ denote average and standard deviation of estimates from 1000 simulated studies. $E(\hat{s})$ denotes average of likelihood-based estimates of the standard deviation from the 1000 experiments. Coverage denotes the proportion of experiments in which the true λ is contained in the nominal 95% confidence interval, and Width refers to the median width of the nominal 95% CI. Values in boldface represent scenarios where conditions for this estimator are met.

Estimator A	p=1				p=0.95			
	$E(\hat{\lambda})$	$sd(\hat{\lambda})$	$E(\hat{s})$	Cov.	Width	$E(\hat{s})$	Cov.	Width
Weibull($\mu=0.5, c=0.4$)								
$\tau=12$ w, $H(\tau)=0.45$	0.026	0.0137	0.0137	0.981	0.028	0.0078	0.0110	0.977 0.034
$\tau=24$ w, $H(\tau)=0.80$	0.022	0.0060	0.0059	0.962	0.021	0.0082	0.0099	0.959 0.035
$\tau=52$ w, $H(\tau)=0.999$	0.020	0.0047	0.0044	0.934	0.017	0.0128	0.0232	0.996 0.108
Weibull($\mu=0.3, c=0.6$)								
$\tau=12$ w, $H(\tau)=0.64$	0.026	0.0191	0.0175	0.971	0.038	0.0108	0.0169	0.962 0.074
$\tau=24$ w, $H(\tau)=0.92$	0.022	0.0101	0.0090	0.946	0.030	0.026 0.0127	0.0208	0.991 0.097
$\tau=52$ w, $H(\tau) > 0.999$	0.021	0.0077	0.0073	0.931	0.026	0.029 0.0190	0.0295	0.991 0.150
Lognorm($\mu=0.5, c=0.4$)								
$\tau=12$ w, $H(\tau)=0.46$	0.026	0.0140	0.0126	0.977	0.026	0.022 0.0075	0.0107	0.978 0.031
$\tau=24$ w, $H(\tau)=0.81$	0.022	0.0058	0.0057	0.949	0.021	0.023 0.0082	0.0099	0.968 0.034
$\tau=52$ w, $H(\tau)=0.992$	0.022	0.0052	0.0053	0.940	0.020	0.035 0.0133	0.0247	0.997 0.117
Lognorm($\mu=0.3, c=0.6$)								
$\tau=12$ w, $H(\tau)=0.67$	0.026	0.0199	0.0158	0.982	0.037	0.023 0.0105	0.0165	0.968 0.072
$\tau=24$ w, $H(\tau)=0.92$	0.025	0.0108	0.0107	0.971	0.036	0.027 0.0129	0.0229	0.996 0.115
$\tau=52$ w, $H(\tau)=0.995$	0.024	0.0097	0.0099	0.957	0.036	0.030 0.0187	0.0308	0.992 0.173
Estimator B								
Weibull($\mu=0.5, c=0.4$)								
$\tau=12$ w, $H(\tau)=0.45$	0.022	0.0051	0.0059	0.960	0.023	0.022 0.0053	0.0061	0.963 0.024
$\tau=24$ w, $H(\tau)=0.80$	0.020	0.0037	0.0036	0.963	0.014	0.020 0.0038	0.0037	0.965 0.014
$\tau=52$ w, $H(\tau)=0.999$	0.020	0.0045	0.0043	0.933	0.016	0.020 0.0045	0.0044	0.933 0.017
Weibull($\mu=0.3, c=0.6$)								
$\tau=12$ w, $H(\tau)=0.64$	0.020	0.0060	0.0060	0.980	0.022	0.020 0.0064	0.0065	0.980 0.022
$\tau=24$ w, $H(\tau)=0.92$	0.019	0.0065	0.0054	0.861	0.018	0.019 0.0066	0.0057	0.842 0.018
$\tau=52$ w, $H(\tau) > 0.999$	0.021	0.0076	0.0072	0.932	0.026	0.021 0.0079	0.0075	0.910 0.026

Estimator A	p=1				p=0.95			
Lognorm($\mu=0.5, \sigma=0.4$)								
$\tau=12$ w, $H(\tau)=0.46$	0.022	0.0051	0.0059	0.941	0.023	0.022	0.0054	0.954
							0.0060	0.024
$\tau=24$ w, $H(\tau)=0.81$	0.020	0.0037	0.0037	0.952	0.014	0.020	0.0039	0.962
							0.0038	0.014
$\tau=52$ w, $H(\tau)=0.992$	0.021	0.0048	0.0048	0.943	0.019	0.021	0.0048	0.951
							0.0049	0.019
Lognorm($\mu=0.3, \sigma=0.6$)								
$\tau=12$ w, $H(\tau)=0.67$	0.021	0.0060	0.0061	0.975	0.022	0.021	0.0061	0.985
							0.0064	0.023
$\tau=24$ w, $H(\tau)=0.92$	0.020	0.0068	0.0061	0.897	0.021	0.020	0.0064	0.911
							0.0060	0.021
$\tau=52$ w, $H(\tau)=0.995$	0.023	0.0085	0.0089	0.958	0.033	0.023	0.0081	0.968
							0.0088	0.032

Table 2

Simulation results for Estimator C, with $N=6,000$, $\phi = .85$ and $\lambda = 0.02$, and assuming Weibull $G(\cdot)$, $E(\hat{\lambda})$ and $sd(\hat{\lambda})$ denote average and standard deviation of estimates from 1000 simulated studies. $E(\hat{s})$ denotes average of likelihood-based estimates of the standard deviation from the 1000 simulations. Coverage denotes the proportion of simulations in which the true λ is contained in the nominal 95% confidence interval, and Width refers to the median width of the nominal 95% CI. Values in boldface represent scenarios where conditions for this estimator are met.

Estimator C	p=1					p=0.95				
	$E(\hat{\lambda})$	$sd(\hat{\lambda})$	$E(\hat{s})$	Cov.	Width	$E(\hat{\lambda})$	$sd(\hat{\lambda})$	$E(\hat{s})$	Cov.	Width
$p_0=0.95$										
Weibull ($\mu=0.5, c=0.4$)										
$\tau=12$ w, $H(\tau)=0.45$	0.022	0.0056	0.0060	0.971	0.022	0.025	0.0157	0.0138	0.992	0.028
$\tau=24$ w, $H(\tau)=0.80$	0.019	0.0037	0.0036	0.960	0.013	0.022	0.0075	0.0062	0.963	0.020
$\tau=52$ w, $H(\tau)=0.999$	0.020	0.0043	0.0042	0.936	0.016	0.021	0.0060	0.0051	0.927	0.018
$p_0=0.99$										
$\tau=12$ w, $H(\tau)=0.45$	0.025	0.0149	0.0124	0.994	0.027	0.023	0.0090	0.0121	0.998	0.031
$\tau=24$ w, $H(\tau)=0.80$	0.020	0.0043	0.0042	0.964	0.014	0.023	0.0085	0.0100	0.975	0.034
$\tau=52$ w, $H(\tau)=0.999$	0.020	0.0043	0.0042	0.936	0.016	0.038	0.0152	0.0263	0.989	0.121
Weibull($\mu=0.3, c=0.6$)										
$p_0=0.95$										
$\tau=12$ w, $H(\tau)=0.64$	0.020	0.0071	0.0064	0.984	0.021	0.027	0.0228	0.0210	0.989	0.036
$\tau=24$ w, $H(\tau)=0.92$	0.019	0.0066	0.0054	0.843	0.018	0.024	0.0161	0.0132	0.932	0.030
$\tau=52$ w, $H(\tau) > 0.999$	0.021	0.0077	0.0073	0.924	0.027	0.022	0.0150	0.0099	0.926	0.027
$p_0=0.99$										
$\tau=12$ w, $H(\tau)=0.64$	0.023	0.0165	0.0107	0.981	0.023	0.025	0.0131	0.0198	0.999	0.078
$\tau=24$ w, $H(\tau)=0.92$	0.019	0.0067	0.0056	0.850	0.019	0.029	0.0166	0.0251	0.990	0.116
$\tau=52$ w, $H(\tau) > 0.999$	0.021	0.0077	0.0073	0.924	0.027	0.042	0.0302	0.0505	0.999	0.296
Lognorm($\mu=0.5, c=0.4$)										
$p_0=0.95$										
$\tau=12$ w, $H(\tau)=0.46$	0.023	0.0053	0.0059	0.950	0.022	0.026	0.0133	0.0136	0.985	0.027
$\tau=24$ w, $H(\tau)=0.81$	0.020	0.0038	0.0037	0.955	0.014	0.022	0.0059	0.0058	0.969	0.019
$\tau=52$ w, $H(\tau)=0.992$	0.021	0.0046	0.0048	0.954	0.019	0.022	0.0055	0.0055	0.948	0.020
$p_0=0.99$										
$\tau=12$ w, $H(\tau)=0.46$	0.025	0.0120	0.0110	0.978	0.025	0.023	0.0088	0.0120	0.996	0.031

Estimator C	p=1					p=0.95				
$\tau=24$ w, $H(\tau)=0.81$	0.021	0.0044	0.0042	0.950	0.015	0.024	0.0087	0.0099	0.994	0.033
$\tau=52$ w, $H(\tau)=0.992$	0.021	0.0046	0.0048	0.954	0.019	0.038	0.0157	0.0279	0.988	0.132
Lognorm($\mu=0.3, \sigma=0.6$)										
$p_0=0.95$										
$\tau=12$ w, $H(\tau)=0.67$	0.021	0.0059	0.0059	0.972	0.022	0.026	0.0192	0.0170	0.992	0.034
$\tau=24$ w, $H(\tau)=0.92$	0.020	0.0064	0.0061	0.901	0.022	0.025	0.0139	0.0123	0.962	0.033
$\tau=52$ w, $H(\tau)=0.995$	0.023	0.0080	0.0088	0.965	0.033	0.025	0.0143	0.0117	0.960	0.035
$p_0=0.99$										
$\tau=12$ w, $H(\tau)=0.67$	0.022	0.0094	0.0081	0.971	0.023	0.025	0.0137	0.0194	1.000	0.071
$\tau=24$ w, $H(\tau)=0.92$	0.020	0.0066	0.0063	0.910	0.023	0.031	0.0167	0.0284	1.000	0.144
$\tau=52$ w, $H(\tau)=0.995$	0.023	0.0080	0.0088	0.965	0.033	0.046	0.0357	0.0573	0.999	0.351

Table 3

Simulation results for the AF Estimator, with $N=6000$, $\phi = .85$ and $\lambda = 0.02$. $E(\hat{\lambda})$ and $sd(\hat{\lambda})$ denote average and standard deviation of estimates from 1,000 simulated studies. $E(\hat{s})$ denotes average of likelihood-based estimates of the standard deviation from the 1,000 simulations. Coverage denotes the proportion of simulations in which the true λ is contained in the nominal 95% confidence interval, and Width refers to the median width of the nominal 95% CI.

AF Estimator	p=1					p=0.95				
	$E(\hat{\lambda})$	$sd(\hat{\lambda})$	$E(\hat{s})$	Cov.	Width	$E(\hat{\lambda})$	$sd(\hat{\lambda})$	$E(\hat{s})$	Cov.	Width
Weibull($\mu=0.5, c=0.4$)										
$\tau = 12$ w, $H(\tau)=0.45$	0.020	0.0043	0.0042	0.945	0.017	0.021	0.0042	0.0043	0.951	0.018
$\tau = 24$ w, $H(\tau)=0.80$	0.018	0.0028	0.0027	0.886	0.011	0.018	0.0028	0.0028	0.887	0.011
$\tau = 52$ w, $H(\tau)=0.999$	0.010	0.0015	0.0014	0.000	0.006	0.010	0.0015	0.0014	0.000	0.006
Weibull($\mu=0.3, c=0.6$)										
$\tau = 12$ w, $H(\tau)=0.64$	0.017	0.0039	0.0038	0.915	0.015	0.017	0.0039	0.0039	0.923	0.016
$\tau = 24$ w, $H(\tau)=0.92$	0.012	0.0022	0.0023	0.158	0.009	0.012	0.0024	0.0023	0.161	0.009
$\tau = 52$ w, $H(\tau) > 0.999$	0.006	0.0011	0.0011	0.000	0.004	0.006	0.0012	0.0011	0.000	0.004
Lognorm($\mu=0.5, c=0.4$)										
$\tau = 12$ w, $H(\tau)=0.46$	0.021	0.0041	0.0042	0.952	0.017	0.021	0.0045	0.0043	0.935	0.017
$\tau = 24$ w, $H(\tau)=0.81$	0.018	0.0028	0.0027	0.887	0.011	0.018	0.0028	0.0028	0.899	0.011
$\tau = 52$ w, $H(\tau)=0.992$	0.010	0.0014	0.0014	0.000	0.006	0.010	0.0014	0.0014	0.000	0.006
Lognorm($\mu=0.3, c=0.6$)										
$\tau = 12$ w, $H(\tau)=0.67$	0.018	0.0039	0.0039	0.930	0.016	0.018	0.0040	0.0040	0.944	0.016
$\tau = 24$ w, $H(\tau)=0.92$	0.012	0.0023	0.0022	0.170	0.009	0.012	0.0023	0.0023	0.180	0.009
$\tau = 52$ w, $H(\tau)=0.995$	0.006	0.0011	0.0011	0.000	0.004	0.006	0.0011	0.0011	0.000	0.004