Europe PMC Funders Group

Author Manuscript

Epidemiology. Author manuscript; available in PMC 2013 March 01.

Published in final edited form as:

Epidemiology. 2012 September; 23(5): 721–728. doi:10.1097/EDE.0b013e3182576c07.

A New General Biomarker-Based Incidence Estimator

Reshma Kassanjee^{1,2}, Thomas A. McWalter^{1,2}, Till Bärnighausen^{3,4}, and Alex Welte¹
¹DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), University of Stellenbosch, South Africa

²School of Computational and Applied Mathematics, University of the Witwatersrand, South Africa

³Department of Global Health and Population, Harvard School of Public Health, Boston, United States

⁴Africa Centre for Health and Population Studies, University of KwaZulu-Natal, South Africa

Abstract

Background—It is attractive to estimate disease incidence from cross-sectional surveys, using biomarkers for "recent" infection. Despite considerable interest in applications to HIV, there is currently no consensus on the correct handling of "recent" biomarkers appearing in persons long after infection.

Methods—We derive a general expression for a weighted average of recent incidence that – unlike previous estimators – requires no particular assumption about recent infection biomarker dynamics, or about the demographic and epidemiologic context. This is possible through the introduction of an explicit timescale *T* that truncates the period of averaging implied by the estimator.

Results—The recent infection test dynamics can be summarized into two parameters, similar to those appearing in previous estimators: a mean duration of recent infection and a false-recent rate. We identify a number of dimensionless parameters that capture the bias that arises from working with tractable forms for the resulting estimator, and elucidate the utility of the incidence estimator in terms of the performance of the recency test and the population state. Estimation of test characteristics and incidence is demonstrated using simulated data. The observed confidence interval coverage of the test characteristics and incidence is within 1% of intended coverage.

Conclusions—Biomarker-based incidence estimation can be consistently adapted to a general context without the strong assumptions of previous work about biomarker dynamics and epidemiologic and demographic history.

The measurement of disease incidence — the rate of new cases in a population — is essential for effectively monitoring the spread of disease, and for targeting and assessing interventions. Longitudinal studies to directly count new infections are costly, time-consuming, and prone to capturing unrepresentative behavior. Estimating incidence by modeling multiple prevalence values requires knowledge of the survival of those affected and unaffected by the condition. For incurable conditions such as HIV, prevalence emerges as a slow convolution (averaging) of historic incidence with survival and the variation in the

Corresponding author: Reshma Kassanjee Address: SACEMA, c/o StIAS, Private Bag X1, Matieland, Stellenbosch, South Africa, 7602 Telephone: +27 21 808 2589 Fax: +27 21 808 2586.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

size of the susceptible population. Thus, changes in prevalence over time are poor proxies for recent incidence. On the other hand, it has long been noted that prevalence of recent infection can be a very good proxy for recent incidence. Deriving an incidence estimate from a single cross-sectional survey has enormous practical advantages. There has consequently been considerable interest in developing recent infection tests based on host or viral biomarkers. This approach has been explored particularly in the context of HIV incidence. 1-4 A number of methodologies have been proposed, 5-16 and these have been reviewed and critiqued elsewhere. 1-4,17-22

The limitations of current methodologies (using biomarkers of recent infection to estimate incidence) are hindering consensus, the development of test technology, and field implementation. There is no simple solution to the problem of estimating incidence, a rate, from a single cross-sectional survey, because there is an unavoidable loss of information when a population history is summarized into an instantaneous population state. Previously proposed estimators have been derived under very specific assumptions (known to be substantially violated) concerning both the epidemiologic and demographic context, as well as the behavior of the recent infection tests ⁵⁻¹⁷ (as will be described below).

Methodological background

Most tests for recent HIV infection classify persons as "recently" infected based on a "below-threshold" immune response such as antibody titre, avidity, or HIV-specific IgG proportion. There is some evidence, for all tests proposed to date, that a small minority of persons remain classified as "recently" infected long after infection. Additionally, late-stage HIV disease or treatment leading to viral suppression may diminish the host immune response, returning long-infected persons to the "recent" infection state. This is the physiologic basis for the introduction of the notion of "false-recent" results, the effects of which are encoded into a population-level parameter widely called the "false-recent rate". In it is not a "rate" in the conventional sense, but the proportion of persons not "truly" recently infected, who nevertheless produce a "recent" result with the biomarker. In earlier analyses, dynamics were summarized into only a mean duration of recent infection. An are recent infection.

Much of the analytic complexity and methodological contention arises from the difficulty of formally defining "true-recent" and "false-recent" results. Initially, attempts to account for "false-recent" results were inspired not just by the biologic variability noted above, but by a pattern of cross-sectional incidence estimates that were higher than prospectively obtained estimates in the same populations. However, as pointed out by Brookmeyer, ¹⁸ subtracting "false-recent" results is not the only way to obtain consistency – one can simply account for all times spent in the "recent" state when defining the mean duration of recent infection. In practice, however, this creates other problems. For one, development of a new test, a process that includes the estimation of a mean duration of recent infection, cannot feasibly wait for a decade or two of follow-up of a seroconverter cohort. Also, a long duration of recent infection (as defined by existing tests) can cause problematic temporal bias or blurring of incidence estimates – the extreme case being the use of prevalence as a proxy for incidence. Indeed, the notion of recent infection does not provide a totally unbiased estimate of the instantaneous incidence, but, at best, a weighted average of recent incidence, which in principle can be very close to a uniformly weighted average over recent times. This statistical weighting may be understood by noting that, as people can persist in the "recent" state for some time, a range of past values of incidence contributes to the current population count of "recently" infected individuals.

Kaplan and Brookmeyer⁷ and, more recently, Brookmeyer²⁷ explored this in the special case of incidence varying linearly with time, in which case the temporal statistical weighting can be summarized as a time lag of the incidence estimator, which they refer to as the "shadow". The longer this "shadow," the less informative the estimate, with less power to detect changes in incidence over short periods of time. A key benefit of a rigorous notion of "false-recent" results (in addition to a complementary notion of mean duration of recent infection that is more practical to measure) is the reduction of this temporal bias.

In order to implement a formally consistent definition of both a "false-recent rate" and mean duration of recent infection, we introduce an explicit time cut-off, T, between "true-recent" and "false-recent" results. To lead to an informative estimator, this cut-off, though theoretically arbitrary, must be chosen to reflect the temporal dynamic range of the test for recent infection; i.e. at a time T post infection, the overwhelming majority of infected people should no longer be testing "recent", and furthermore, T should not be larger than necessary to achieve this criterion. While this is reminiscent of a time cut-off in previous analyses, 9,10 the present work dispenses with problematic assumptions of past analyses that have prevented the widespread estimation of incidence using cross-sectional data on recent infection. The eAppendix (http://links.lww.com) provides supplementary material for this work.

Analysis

Our exposition proceeds in four key steps:

- (1) The derivation of a simple, general expression for a weighted mean recent incidence, which can be constructed without any particular assumptions about the demographic or epidemiologic history or the dynamics of the biomarker used to classify persons with a disease (such as HIV infection) as "recently" or "non-recently" infected.
- (2) The derivation of an incidence estimator by expressing the general weighted incidence in terms of (a) quantities that can be known by an experimenter, and (b) a bias term whose size can be approximately estimated in terms of a number of dimensionless parameters that characterize the failure of the test and context to conform to certain idealizations.
- (3) The estimation of the test characteristics.
- (4) The application of the methodology to estimate incidence in simulated scenarios, to confirm the consistency of confidence intervals.

A general expression for weighted incidence

A test for recent infection may employ an arbitrarily complex combination of criteria to classify infected persons as "recently" or "non-recently" infected.³ It is understood that there will be natural inter-subject variability in progression through these categories after becoming infected. This range of responses may be captured in a function $P_R(\tau)$, which is the probability of still being alive and "recently" infected at a time τ post infection. Let $P_A(\tau)$ be the probability of being alive at a time τ post infection.

Throughout this text, "infection" refers to the detectable infection of HIV, which depends on the diagnostic test being used. Practically, the delay between actual infection and detectable infection merely implies an epidemiologically inconsequential delay in entering the operationally HIV-positive state. This point is explored in more detail in eAppendix 1 (http://links.lww.com).

Assuming a continuous population dynamic, and using the reference time t = 0 as the time at which a survey is conducted, consider the following explicit weighted averaging of incidence over a period of duration T:

$$I_{T} = \frac{\int_{-T}^{0} I(t) N_{S}(t) P_{R}(-t) dt}{\int_{-T}^{0} N_{S}(t) P_{R}(-t) dt}, \quad (1)$$

where the, possibly time-dependent, incidence is given by I(t) and the susceptible population by $N_S(t)$. The incidence at time t contributes to I_T with weight $N_S(t)P_R(-t)$, i.e. with a weight proportional to (i) the susceptible population vulnerable to being infected at time t, $N_S(t)$, and (ii) the probability of a person infected at time t still being alive and "recent" at the time of the survey, $P_R(-t)$. It will emerge that, for practical purposes, this is very close to a uniformly weighted average of incidence over a period preceding the survey. There is no exact way to extract either a uniformly weighted average or an estimate of incidence at one point in time, given the substantial compression of information from a population history into a cross-sectional survey.

If there is a critical value, T_{crit} with the property that $N_S(-t)P_R(t) = 0$ for all $t > T_{crit}$ then all choices of $T > T_{crit}$ yield the same result – the one obtained by Brookmeyer and Quinn. ⁵ For any finite value of T, the "shadow", or temporal bias, is strictly less than T, and it is less than T/2 if $N_S(-t)P_R(t)$ is a strictly decreasing function of time in the interval [0, T].

In the interpretation of Equation 1, it is useful to consider the labeled areas in the Figure:

- Areas A and D (the areas under the curve I(t)N_S(t)P_R(-t)) represent the "recently" infected population at t = 0, infected for times less than and greater than T, respectively. A is the numerator of Equation 1.
- Areas B and C (the areas between the curves $I(t)N_S(t)P_A(-t)$ and $I(t)N_S(t)P_R(-t)$, i.e. the infected population excluding the "recently" infected population) represent the "non-recently" infected population at t = 0, infected for times less than and greater than T, respectively.
- Area E (the area under the curve $N_S(t)P_R(-t)$) is the denominator of Equation 1. Hence, Equation 1 can be rewritten as

$$I_T = \frac{A}{E}$$
. (2)

Previous derivations of incidence estimators have relied on assumptions of (at least recent) epidemiologic and demographic equilibrium, some simplifications of the post-infection recency test dynamics, or both. To elucidate the impact of general non-equilibrium conditions, it is useful to express the crucial time-dependent quantities in terms of time-dependent relative deviations from conveniently chosen constants:

$$I(t) = I_T (1 + f_I(t))$$
 (3)

$$N_{S}(t) = N_{S}(0) \left(1 + f_{N_{S}}(t)\right)$$
 (4)

$$P_{A}(\tau) = 1 + f_{P_{A}}(\tau)$$
. (5)

These equations do not represent any approximations or truncated power series, but rather formally arbitrary exact decompositions of I(t), $N_S(t)$ and $P_A(\tau)$, the point of which is to characterize, rather than assume away, the non-ideal aspects of the population and test dynamics. More specifically: $f_I(t)$ is defined to capture the time dependence of the fractional deviation of incidence relative to the weighted incidence as defined in the period preceding t=0; t0, t1 is defined to capture the time dependence of the fractional deviation of the susceptible population from its instantaneous value at t=0; and t2 is defined as the probability of not surviving for at least a time t3 after infection. The averages, over the period of duration t3 preceding the survey, of products of t4 in t5 is defined as the shown to summarize the effect of these deviations on incidence estimates.

Particular forms for an incidence estimator

Incidence estimation involves expressing Equation 2 as a function of test characteristic parameters, and population states. The key link between the numerator and the population states is

$$A=N_R-D$$
, (6)

where N_R is the size of the "recently" infected population at t=0 (Figure). Various authors have used different terms for the situation that D=0 (Figure). These terms have included "misclassification", "false-positivity," "false-recency," "imperfect long-term specificity," the "long tail," or "non-progression" of the test for recent infection. 9,10,14,15,17-21 Formally, the increasingly used term "false-recent rate," β_T , given a cut-off T, is defined in this work as the probability that a randomly chosen person infected for more than time T will be classified as "recently" infected by the recency test.

A variety of approaches may be used for estimating the area D. For example,

$$D = \frac{D}{C+D} \times (C+D)$$

= $\beta_T \times (N_+ - (A+B)),$ (7)

where

$$\beta_T = \frac{D}{C + D}, \quad (8)$$

and the areas A + B and C + D represent the current population infected for times less than and greater than T, respectively, so that $A + B + C + D = N_+$ is the size of the HIV-positive population.

By inspection of the Figure,

$$A + B = \int_{-T}^{0} I(t) N_{S}(t) P_{A}(-t) dt.$$
 (9)

Using the parameterization in terms of the dimensionless $f_I(.)$, $fN_S(.)$ and $f_{P_A}(.)$ introduced earlier, this gives:

$$A+B = \int_{-T}^{0} I_{T} N_{S}(0) (1+f_{I}(t)) (1+f_{N_{S}}(t)) (1+f_{P_{A}}(-t)) dt$$

= $I_{T} N_{S}(0) T \times (1+\gamma_{1}+\gamma_{2}+\gamma_{3}+\gamma_{4}+\gamma_{5}+\gamma_{6}+\gamma_{7}),$ (10)

where the (also dimensionless) γ terms capture the consequences of the time-dependence of I(t), $N_S(t)$ and $P_A(-t)$:

$$\gamma_1 = \frac{1}{T} \int_{-T}^0 f_I(t) dt \quad (11)$$

$$\gamma_2 = \frac{1}{T} \int_{-T}^0 f_{N_S}(t) dt$$
 (12)

$$\gamma_3 = \frac{1}{T} \int_{-T}^0 f_{P_A} (-t) dt$$
 (13)

$$\gamma_4 = \frac{1}{T} \int_{-T}^{0} f_I(t) f_{N_S}(t) dt$$
 (14)

$$\gamma_5 = \frac{1}{T} \int_{-T}^{0} f_I(t) f_{P_A}(-t) dt$$
 (15)

$$\gamma_6 = \frac{1}{T} \int_{-T}^0 f_{N_S}(t) f_{P_A}(-t) dt$$
 (16)

$$\gamma_7 = \frac{1}{T} \int_{-T}^{0} f_I(t) f_{N_S}(t) f_{P_A}(-t) dt.$$
 (17)

These γ corrections may be positive or negative, but γ_3 is always non-positive.

Substituting the expression for area D (Equation 7), and then A + B (Equation 10), into Equation 6, the numerator becomes:

$$A = N_R - \beta_T N_+ + \beta_T I_T N_S (0) T \times \left(1 + \sum_{k=1}^{7} \gamma_k \right).$$
 (18)

In the denominator of Equation 2,

$$E = \int_{-T}^{0} N_{s}(t) P_{R}(-t) dt$$

$$= \int_{-T}^{0} N_{s}(0) (1 + f_{N_{s}}(t)) P_{R}(-t) dt \quad (19)$$

$$= N_{s}(0) \Omega_{T} \times (1 + \gamma_{8}),$$

where

$$\Omega_{T} = \int_{0}^{0} P_{R}(-t) dt$$

$$= \int_{0}^{\overline{T}} P_{R}(t) dt,$$
(20)

and

$$\gamma_8 = \frac{1}{\Omega_T} \int_{-T}^0 f_{N_S}(t) P_R(-t) dt.$$
 (21)

The mean duration of recent infection, Ω_T , thus defined, given a cut-off T, is the average time spent both alive and "recently" infected, within a time T post infection.

Substituting the expressions for the numerator (Equation 18) and the denominator (Equation 19) into Equation 2 gives:

$$I_{T} = \frac{N_{R} - \beta_{T} N_{+} + \beta_{T} I_{T} N_{S} T \times \left(1 + \sum_{k=1}^{7} \gamma_{k}\right)}{N_{S} \Omega_{T} \times (1 + \gamma_{8})}.$$
 (22)

The right-hand side of this expression contains not only the mean duration of recent infection, Ω_T ; the false-recent rate, β_T ; and the uninfected, "recently" infected and "non-recently" infected populations at t = 0, $N_S = N_S(0)$, N_R and N_{NR} , where $N_+ = N_R + N_{NR}$; but also the weighted incidence, I_T , itself. Rearranging and solving for I_T yields

$$I_{T} = \frac{N_{R} - \beta_{T} N_{+}}{N_{S} (\Omega_{T} - \beta_{T} T)} (1 + e)^{-1}, \quad (23)$$

where

$$e = \left(\frac{\Omega_T}{\Omega_T - \beta_T T}\right) \gamma_8 - \beta_T \left(\frac{T}{\Omega_T - \beta_T T}\right) \sum_{k=1}^7 \gamma_k \quad (24)$$

contains all the details that cannot be directly evaluated from an experimenter's point of view.

Using the sample counts of uninfected, "recently" infected and "non-recently" infected subjects at t = 0, n_S , n_R and n_{NR} , with $n_+ = n_R + n_{NR}$, a simple estimator of weighted incidence, with relative (fractional) error e, is obtained:

$$\widehat{I}_{T} = \frac{n_{R} - \beta_{T} n_{+}}{n_{S} \left(\Omega_{T} - \beta_{T} T\right)}.$$
 (25)

By using definitions of the test characteristics (Ω_T and β_T), which are subtly different from those used previously,^{4,9,10,13-15} an incidence estimator is thus obtained in which multiple transitions between "recently" and "non-recently" infected states are allowed, and no assumption is required about the independence of progression through the "recent" / "non-recent" states and post-infection survival. This estimator caters to completely general recent infection test dynamics. Bias arising from a non-constant incidence or susceptible population (in the period T before the incidence study) or imperfect survival (for T after infection) is fully described by e, and further discussed below.

The functional form for the estimator in Equation 25 can be obtained directly by assuming the system is in demographic and epidemiologic equilibrium.¹⁵ The present analysis shows that, when the system is away from equilibrium, in particular when incidence is not close to constant, this functional form provides an estimate of a particular weighted average of recent incidence, with a fractional bias *e*. In eAppendix 1 (http://links.lww.com), the structure and meaning of the terms in *e* are discussed, and the bias is computed in model scenarios.

The γ_8 term, closely related to a bias implicit in all previously proposed estimators, ^{14,17} is zero when the susceptible population is constant for T preceding the survey, but a time-

dependent susceptible population imposes a fundamental limitation to cross-sectional incidence estimation. This highlights a key motivation for introducing T– namely to decouple the short-term dynamics of the recency test from any long-term dynamics (which become convolved with the epidemiology and demography).

The remaining γ terms appear only in conjunction with two further multiplicative factors: (i) the fraction, dominated by T and Ω_T , which is of order unity, or perhaps typically closer in value to two, and (ii) a factor of β_T . Therefore, if β_T is sufficiently small, the estimator can yield an arbitrarily accurate weighted incidence, even when the incidence and survival are varying substantially over the timescale set by T. It is already well known that informative incidence estimation requires that β_T be small, 3,4,14,29 and developers of recency tests are seeking new technologies and algorithms to achieve this. 30,31

Ultimately, the utility of a test for recent infection lies in its ability to produce accurate and precise incidence estimates. The expectation value and variance of the incidence estimator are approximated in eAppendix 2 (http://links.lww.com). The uncertainty in the estimator, and its dependence on the test characteristics, is context-specific, depending on the historic HIV incidence and prevalence in the study population. The precision of the estimator improves with increasing mean duration of recent infection and with decreasing false-recent rate. This trade-off has been previously noted,⁴ with the additional subtlety in the present analysis that the choice of *T* and the test characteristics are intrinsically related.

As noted in the introduction, the choice of *T* is theoretically arbitrary, but given the dynamics of available and foreseeable recency tests, it will need to be of the order of one year. This implies a "shadow"^{7,27} of the order of (but probably comfortably less than) half a year. This is considerably smaller than the "shadows" of well over one year obtained for a number of realistic scenarios considered by Brookmeyer²⁷ when using the original one-parameter incidence estimator, ⁵ and indeed implies less temporal bias or blurring than incurred in a cohort followed up for one year.

Estimation of the test characteristics

As with any method aiming to infer incidence from the cross-sectional application of a recent infection test, use of the newly derived estimator requires measuring some characteristics of the test ahead of its application in the surveillance context. This test characterization should be performed as locally as feasible, because test performance may be context-specific. The false-recent rate, β_T , and the mean duration of recent infection, Ω_T , are intuitively close to previously proposed definitions.^{4,9,10,13-15} However, the definitions of the test characteristics emerging from this work allow, for the first time, arbitrary and complex test dynamics to be exactly captured. The estimation of each of the characteristics is briefly discussed below, with a slightly more technical discussion provided in eAppendix 3 (http://links.lww.com).

The false-recent rate, β_T , would ideally be estimated by the proportion of "recently" infected subjects in a representative sample of people infected for longer than T. It is also conceivable that β_T could be estimated from a combination of convenience samples, knowledge of the dynamics of anomalous subpopulations (who persist in, or return to, the "recent" state despite being infected for a time greater than T) and knowledge of the embedding demography and epidemiology.

The mean duration of recent infection emerges as naturally in longitudinal surveillance settings (where well-pedigreed biologic specimens may be obtained repeatedly over time) as it does in the context of cross-sectional incidence estimation analyzed above. An idealized experiment, which revisits initially HIV-negative persons after a time T, and counts the

frequency of "recent" results in those who have become HIV-positive, provides a direct estimate for Ω_T , assuming a uniform distribution of infection times over the inter-test interval. Specifically, the ratio Ω_T/T is the probability that a seroconverter is "recently" infected. This idea can be expanded to account for varying inter-test intervals, depending on available data and knowledge of the recency test dynamics, with an example of such an extension provided in eAppendix 3 (http://link.lww.com).

More traditionally, measurement of the mean duration of recent infection has been based on the frequent follow-up and recent infection testing of seroconverters. A form of survival analysis or regression can then be used to characterize the time taken to exit the "recent" state or the evolution of the biomarker with time after infection, respectively, thereby estimating the mean duration of recent infection.

Incidence estimation from the cross-sectional application of tests for recent infection

Having derived this new, general incidence estimator, and having outlined potential approaches for estimating the required test characteristics, we now demonstrate the implementation of the full set of analyses to infer incidence using simulated data.

Assuming a particular epidemiologic and demographic history, post-infection survival, and recency test dynamics, we performed one thousand simulations, each producing (independent) datasets to (i) estimate the false-recent rate, β_T ; (ii) estimate the mean duration of recent infection, Ω_T ; and (iii) provide sample counts to infer incidence, I_T , using the incidence estimator in Equation 25.

The generation of the datasets and the maximum likelihood estimation of the test characteristics are described in eAppendix 4 (http://links.lww.com). We used asymptotic normality of maximum likelihood estimators (using estimated characteristics as proxies for true values) to approximate distributions for the estimated parameters and to obtain confidence intervals. Confidence intervals for incidence were then based on these results, the approximate normality of the trinomial sample counts (with sample statistics approximating population parameters), and the approximate normality of the incidence estimator and its estimated variance as provided in eAppendix 2 (http://links.lww.com). Almost exactly 95% of the one thousand thus generated 95% confidence intervals (one for each of the datasets) contained the relevant population parameter used in the simulation (Table), demonstrating the numerical consistency of the full set of analyses.

Discussion

The use of tests for recent infection to infer incidence is of considerable and increasing interest, especially for HIV surveillance. It is a fundamental limitation that all currently available (and perhaps all conceivable) tests with a long enough mean duration of recent infection for statistical robustness also have some people with "recent" results at arbitrarily long times after infection. If there were no such "false-recent" results, the use of recency tests for incidence estimation would be straightforward, as shown, for example, by Brookmeyer and Quinn. Various methodological advances to accommodate a non-zero "false-recent rate" have attracted attention, but consensus has not emerged on the best approach. Previous derivations of incidence estimators have relied on strong assumptions: perhaps most crucially that the "false-recent rate" is an innate property of the test, rather than a convolution of test properties with the demographic and epidemiologic context. This assumption is known to be substantially violated.

We suggest here a formal approach to summarize an arbitrarily complicated recent infection test dynamic into two parameters, namely a mean duration of recent infection and a false-

recent rate. A crucial construct is the introduction of a timescale *T*, describing the dynamic range of recency. The consequence of relaxing the assumptions made by the incidence estimators developed previously is that demography and epidemiology are no longer perfectly separated from test characteristics, reflecting fundamental limitations to the inference of rates from instantaneous population states. If the false-recent rate is very close to zero, the limitations imposed by a non-zero false-recent rate become minor and its variation over time and place is restricted.

The present analysis offers the opportunity to consistently account for imperfect accuracy and precision of the incidence estimator. The utility of the estimator may be assessed in terms of changes in incidence and the susceptible population over the preceding period of duration T, the probability of survival for T post infection, and the recency test performance. The cross-sectional incidence estimator will be informative at feasible sample sizes, in a given context, for a suitably well-behaved recency test, i.e., a test with a suitably long mean duration of recent infection and low false-recent rate.

The approach presented here is broad enough to recover previously-proposed estimators, with minor modification. It also clarifies the use of estimators that do not account for falserecency at all. Setting T to a very large value forces the false-recent rate arbitrarily close to 0, and the one-parameter estimator of Brookmeyer and Quinn⁵ is obtained. The properties of the test are then summarized by the mean duration of recent infection. However, this mean duration, which is now the average time spent alive and "recently" infected, is considerably more difficult to measure and more likely to change over time than one based on individual durations that are each explicitly limited to T. Also, a large T (effectively infinite, if T is not explicitly introduced) leads to a weighting scheme for averaging incidence that extends far back into the past. For heuristic purposes, the weighted incidence that emerges from the use of a realistically available test and a judicious choice of T can be viewed as a good proxy for the uniformly weighted mean incidence in the period of duration Ω_T preceding the survey. One may consider whether there is any benefit in using additional parameters to characterize the dynamics of the recency test. Incidence inference would then be based on a more complex distribution of test results than counts of "recent" and "non-recent" cases. eAppendix 5 (http://links.lww.com) presents a brief argument that suggests this approach has limited prospects.

Relaxing all formal assumptions about the dynamics of a putative test for recent infection and the demographic and epidemiologic context leads to an estimator that substantially increases the robustness of incidence estimation based on cross-sectional surveys using tests for recent infection. The general analysis leads to a clearer characterization of the utility of the estimator than previously possible. While the analysis is fundamentally novel, the resulting estimator has similarities to some previously published estimators. ^{13,14} These similarities imply that the intuition about the crucial concepts of a false-recent rate and mean duration of recent infection are substantially retained. Numerically, the improvement in incidence estimates implied by the new estimator will vary with context.

While the motivation for this work has been to improve our capacity to estimate HIV incidence, the methodology is general, and the approach could be applied to estimate incidence of other incurable conditions. Examples of such conditions include viral infections such as herpes simplex virus and human papilloma virus. Future studies should examine the application of the methodology to a wider range of diseases, with the practical challenge being the development of suitable tests for recent infection.

Conclusion

For incurable conditions such as HIV, where prevalence emerges as a slow convolution of historic incidence with survival and the dynamics of the susceptible population, changes in prevalence are a poor proxy for recent incidence. Estimating incidence from cross-sectional surveys has many potential advantages over using longitudinal studies, and has attracted much interest in recent years, particular in the HIV context. However, previously proposed HIV incidence estimators have been derived under conditions of epidemiologic and demographic equilibrium, or specific assumptions about the biomarker-based recent infection test dynamics, or both. These assumptions are known to be violated in many settings, and this has diminished the practical utility of previous methodologies in many settings.

In this work, biomarker-based incidence estimation, which uses data obtained in cross-sectional surveys, is consistently adapted to a general context. The generalization implies that the strong assumptions about epidemiologic and demographic history and biomarker dynamics required by previous estimators are no longer necessary for valid incidence estimation. Our new estimator thus substantially improves and clarifies the utility of tests for recent infection for estimating disease incidence. The familiar remaining practical challenge is making available better and better characterized recency tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial Support: This work was supported in part by a grant from the Canadian International Development Agency.

TB was supported by grant 1R01-HD058482-01 from the National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH).

References

- 1. Le Vu S, Pillonel J, Semaille C, et al. Principles and uses of HIV incidence estimation from recent infection testing a review. Euro Surveill. 2008; 13:11–16.
- 2. Murphy G, Parry JV. Assays for the detection of recent infections with Human Immunodeficiency Virus type 1. Euro Surveill. 2008; 13:4–10.
- 3. Busch MP, Pilcher CD, Mastro TD, et al. Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation. AIDS. 2010; 24:2763–2771. [PubMed: 20975514]
- 4. Welte A, McWalter TA, Laeyendecker O, et al. Using tests for recent infection to estimate incidence: problems and prospects for HIV. Euro Surveill. 2010; 15 pii=19589.
- 5. 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. [PubMed: 7817972]
- 6. Janssen RS, Satten GA, Stramer SL, 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.
 [PubMed: 9660362]
- Kaplan EH, Brookmeyer R. Snapshot estimators of recent HIV incidence rates. Oper Res. 1999; 47:29–37.
- Parekh BS, McDougal JS. New approaches for detecting recent HIV-1 infection. AIDS Rev. 2001; 3:183–193.

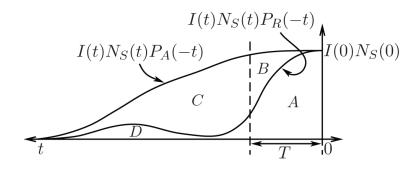
9. McDougal JS, Parekh BS, Peterson ML, et al. Comparison of HIV type 1 incidence observed during longitudinal follow-up with incidence estimated by cross-sectional analysis using the BED capture enzyme immunoassay. AIDS Res Hum Retroviruses. 2006; 22:945–952. [PubMed: 17067263]

- 10. Hargrove JW, Humphrey JH, Mutasa K, et al. Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. AIDS. 2008; 22:511–518. [PubMed: 18301064]
- 11. Welte A. Relating incidence to 'recent infection' prevalence: Application to HIV. S Afr J Sci. 2008; 104:199–202.
- 12. Balasubramanian R, Lagakos SW. Estimating HIV incidence based on combined prevalence testing. Biometrics. 2010; 66:1–10. [PubMed: 19397583]
- Wang R, Lagakos SW. On the use of adjusted cross-sectional estimators of HIV incidence. J Acquir Immune Defic Syndr. 2009; 52:538–547. [PubMed: 19881357]
- 14. McWalter TA, Welte A. Relating recent infection prevalence to incidence with a sub-population of assay non-progressors. J Math Biol. 2010; 60:687–710. (epub 2009). [PubMed: 19633851] Also see poster MOPDB105 at: 5th IAS Conference on HIV Pathogenesis, Treatment and PreventionSouth Africa19-22 July 2009
- McWalter, TA.; Kassanjee, R.; Welte, A. Incidence from cross-sectional surveys: Improved characterization of tests for recent infection. E-poster CDC0473 at: XVIII International AIDS Conference; Austria. 18-23 July 2010;
- McKeown K, Jewel NP. Current status observation of a three-state counting process with application to simultaneous accurate and diluted HIV test data. Can J Stat. 2011; 39:475

 –487.
- 17. McWalter TA, Welte A. A comparison of biomarker based incidence estimators. PLoS One. 2009; 4(10):e7368. doi: 10.1371/journal.pone.0020027. [PubMed: 19809505]
- Brookmeyer R. Should biomarker estimates of HIV incidence be adjusted? AIDS. 2009; 23:485–491. [PubMed: 19165087]
- 19. Hargrove JW. BED estimates of HIV incidence must be adjusted (correspondence). AIDS. 2009; 23:2061–2062. [PubMed: 19755864]
- McDougal JS. BED estimates of HIV incidence must be adjusted (correspondence). AIDS. 2009;
 23:2064–2065. [PubMed: 19755866]
- 21. Welte A, McWalter TA, Bärnighausen T. Reply to 'Should biomarker estimates of HIV incidence be adjusted?' (correspondence). AIDS. 2009; 23:2062–2063. [PubMed: 19755865]
- 22. Bärnighausen T, McWalter TA, Rosner Z, et al. HIV incidence estimation using the BED capture enzyme immunoassay: systematic review and sensitivity analysis. Epidemiology. 2010; 21:685–697. [PubMed: 20699682]
- 23. Hayashida T, Gatanaga H, Tanuma J, et al. Effects of low HIV type 1 load and antiretroviral treatment on IgG-capture BED-enzyme immunoassay. AIDS Res Hum Retroviruses. 2008; 24:495–498. [PubMed: 18327979]
- 24. Marinda ET, Hargrove J, Preiser W, et al. 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. J Acquir Immune Defic Syndr. 2010; 53:496–499. [PubMed: 20306555]
- Hladik W, Olara D, Mermin J, et al. Effect of CD4(+) T cell count and antiretroviral treatment on two serological HIV incidence assays. AIDS Res Hum Retroviruses. 2012; 28:95–99. (epub 2011). [PubMed: 21314476]
- 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. [PubMed: 18520680]
- 27. Brookmeyer R. On the statistical accuracy of biomarker assays for HIV incidence. J Acquir Immune Defic Syndr. 2010; 54:406–14. [PubMed: 20436353]
- 28. Sommen C, Commenges D, Le Vu S, et al. Estimation of the distribution of infection times using longitudinal serological markers of HIV: implications for the estimation of HIV incidence. Biometrics. 2011; 67:467–475. (epub 2010). [PubMed: 20731647]
- 29. Hallett TB. Estimating the HIV incidence rate: recent and future developments. Curr Opin HIV AIDS. 2011; 6:102–107. [PubMed: 21505383]

30. Braunstein SL, Nash D, Kim AA, et al. Dual testing algorithm of BED-CEIA and AxSYM avidity index assays performs best in identifying recent HIV infection in a sample of Rwandan sex workers. PLoS One. 2011; 6(4):e18402. doi:10.1371/journal.pone.0018402. [PubMed: 21532753]

31. Curtis K, Kennedy S, Charurat ME, et al. Development and characterization of a bead-based, multiplex assay for estimation of recent HIV-1 infection. AIDS Res Hum Retroviruses. 2012; 28:188–197. (epub 2011). [PubMed: 21585287]



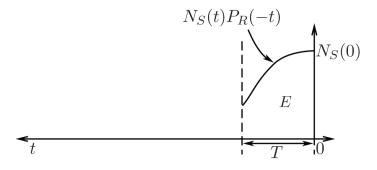


Figure. Epidemiologic, demographic and recent infection test dynamics

Table

Observed 95% confidence interval (CI) coverage of parameters using simulated data^a

Parameter ^b	Input value ^c	Average point estimate	Average CI width ^d	CI Coverage ^e
$oldsymbol{eta}_T$	2.5%	2.52%	1.93%	95.5%
$\Omega_T^{\ f}$	160 days	160.56 days	17.96 days	93.6%
I_T	2%	1.98%	1.38%	94.6%

^aBased on 1000 simulations (each producing a point estimate and confidence interval for each parameter), for the specific modelled scenario and using estimation methods as described in the main text and eAppendix 4 (http://links.lww.com).

 $b_{T=450 \text{ days.}}$

 $^{^{\}it C}$ The true parameter values, as input into the simulated model scenarios.

dAverage width of realized 95% confidence intervals.

 $[\]ensuremath{^{e}}\xspace$ Percentage of realized 95% confidence intervals containing the true parameter value.

^f Uncertainty in the input θ for estimation of Ω T neglected (true parameter value $\theta = 1.7\%$, while the estimated β T was used as the input θ).