

A Comparison of Biomarker Based Incidence Estimators

Thomas A. McWalter^{1*}, Alex Welte^{1,2}

1 School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa, **2** DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa

Abstract

Background: Cross-sectional surveys utilizing biomarkers that test for recent infection provide a convenient and cost effective way to estimate HIV incidence. In particular, the BED assay has been developed for this purpose. Controversy surrounding the way in which false positive results from the biomarker should be handled has led to a number of different estimators that account for imperfect specificity. We compare the estimators proposed by McDougal et al., Hargrove et al. and McWalter & Welte.

Methodology/Principal Findings: The three estimators are analyzed and compared. An identity showing a relationship between the calibration parameters in the McDougal methodology is shown. When the three estimators are tested under a steady state epidemic, which includes individuals who fail to progress on the biomarker, only the McWalter/Welte method recovers an unbiased result.

Conclusions/Significance: Our analysis shows that the McDougal estimator can be reduced to a formula that only requires calibration of a mean window period and a long-term specificity. This allows simpler calibration techniques to be used and shows that all three estimators can be expressed using the same set of parameters. The McWalter/Welte method is applicable under the least restrictive assumptions and is the least prone to bias of the methods reviewed.

Citation: McWalter TA, Welte A (2009) A Comparison of Biomarker Based Incidence Estimators. PLoS ONE 4(10): e7368. doi:10.1371/journal.pone.0007368

Editor: Landon Myer, University of Cape Town, South Africa

Received: June 11, 2009; **Accepted:** August 22, 2009; **Published:** October 7, 2009

Copyright: © 2009 McWalter, Welte. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mcwalter@cam.wits.ac.za

Introduction

Although prospective follow-up of an initially HIV-negative cohort is widely regarded as the “gold-standard” for estimating incidence, the idea of utilizing a biomarker to define a suitable class of “recently infected” individuals, and then to use the prevalence of this class as the basis for estimating HIV incidence, is attractive for a number of reasons. Since this can be implemented using a cross-sectional survey, it is logistically simpler, cheaper and less prone to the biases that result from intervention and loss to follow-up.

The BED capture enzyme immunoassay (BED assay) has been developed for this purpose [1,2] and widely used [3]. It measures the proportion of IgG that is HIV-1 specific as a normalized optical density (OD_n). Since this proportion increases over time after the infection event, specifying an OD_n threshold allows seropositive individuals to be classified as recently infected, if they are below threshold, and as non-recently infected, if they are above threshold. Initially, an incidence formula was proposed [1] that did not explicitly account for the possibility of assay non-progressors (i.e. individuals who never develop enough of an immunological response to cross the threshold). This method was similar to the earlier approaches of Brookmeyer and Quinn [4], and Janssen *et al.* [5]. Later, the methodology proposed by McDougal *et al.* [6] was the first to deal with assay non-progressors. They derived an incidence formula which can be expressed in terms of the prevalence of below-threshold seropositive, above-threshold seropositive and seronegative individuals, and four assay

calibration parameters, being the mean window period (ω), sensitivity (σ), short-term specificity (ρ_1) and long-term specificity (ρ_2). Introducing the long-term specificity parameter provided a way to quantify assay non-progression.

Two other incidence paradigms that explicitly account for assay non-progressors have since been formulated. Hargrove *et al.* [7] proposed a simpler incidence estimator which is equivalent to the McDougal estimator when one sets $\sigma = \rho_1$. Recently, we have also proposed a formally rigorous incidence paradigm [8], which accounts for assay non-progression using fewer assumptions than are made by McDougal *et al.* The parameters that emerge naturally in our estimator are a mean window period and a probability of not progressing on the assay (which can also be expressed as a long-term specificity).

A large portion of this paper is dedicated to an analysis of the assay parameters of the McDougal methodology, showing how they are related. By using a survival analysis formulation of the problem, we are able to write down precise expressions for the parameters. This allows us to derive a relationship between three of the parameters, which simplifies the McDougal estimator by showing that only ω and ρ_2 , which are considerably easier to calibrate than σ and ρ_1 , are required in the final formula. The reduction of the McDougal approach is important in that it shows that all three incidence estimators are, in effect, based on the same underlying parameters characterising the performance of the assay, and are therefore amenable to direct comparison.

We then compare the performance of the three incidence estimators by substituting analytic expressions for population

counts, derived from a model steady state epidemic, into the various formulae. This analysis shows that only our formula [8] produces a bias-free result. Although the biases are typically small, we demonstrate, using numerical examples, that there are regimes where bias may be significant.

The paper is structured as follows: We start by describing the McDougal methodology and, in doing so, write down mathematical expressions for the assay calibration parameters. In the next section we restate the assumptions made by McDougal *et al.* in a mathematically precise manner. This allows us to derive the identity that shows the relationship between the parameters. We then present the three incidence formulae and compare them by inserting the population counts from a model steady state epidemic. We conclude the paper with a discussion of the implications of the identity and the steady state analysis.

Analysis

The McDougal Methodology

Denote the number of individuals in a cross-sectional sample who are respectively *under-threshold*, *over-threshold* and *healthy* (susceptible) by N_U , N_O and N_H . Then the McDougal estimator [6] can be written as

$$\bar{I} = \frac{fN_U}{fN_U + \omega N_H}, \tag{1}$$

where ω is specified in years and the “correction factor”,

$$f := \frac{P_t}{P_o} = \frac{P_o + \rho_2 - 1}{P_o(\sigma - \rho_1 + 2\rho_2 - 1)}, \tag{2}$$

is the ratio of the “true” proportion P_t of recent infections and the proportion $P_o = N_U / (N_U + N_O)$ of the HIV positive individuals that are under the threshold. This correction factor, which depends on subtle definitions for the sensitivity and specificity parameters, explicitly accounts for the fact that the BED assay imperfectly classifies individuals as “recently infected”.

McDougal *et al.* calibrate these parameters using seroconversion panels which show BED optical density as a function of time since infection (some of these are published [1,2]). The calibration occurs in two stages. A window period is estimated, and then estimates of the sensitivity, short-term specificity and long-term specificity are determined with respect to the window period.

The window period is estimated as “the mean period of time from initial seroconversion to reaching an ODn of 0.8” [6]. Although it is not explicitly stated, we presume that those individuals that never reach the threshold, either because they do not progress above the threshold or because they die before reaching the threshold, are not included in the calculation of the mean. More specifically this implies that the window period is the mean observable threshold crossing time, conditional on assay progression (i.e. actually reaching the threshold).

In order to calibrate the sensitivity, short-term specificity and long-term specificity, “a plot of the proportion of specimens positive in the assay versus time since seroconversion” is generated (also later referred to as “the curve”). This is the sampled survival function (essentially a Kaplan-Meier curve) in the state of being *under* the threshold, conditional on being *alive*, which we denote $S_{U|A}(t)$.

The sensitivity of the test is estimated for an interval corresponding to the window period by “integrating the curve within the window”. Short-term specificity is calculated for “the interval immediately after, and equal in duration to, the window period”. Long-term specificity is for “the period thereafter (where

the curve is flat)”. McDougal *et al.* explicitly make the following assumptions, with the justification that they “are reasonable as very little attrition (from death) during the first two time intervals after infection would be expected”:

1. “Recent infections are randomly distributed within the first window period”.
2. “The number of persons in the interval of equal duration immediately after the mean window period equals the number in the first window period”.
3. “The remainder of the population is more than two window periods since seroconversion”.

While it may be true in the situation being explored here, we note that it is not *a priori* obvious that the choice of equal window periods ensures that $S_{U|A}(t)$ is flat after twice the window period. With this in mind, we propose a generalization in which there are two window periods with arbitrary values ω_1 and ω_2 , chosen so that all individuals that progress do so in a time less than $\omega_1 + \omega_2$ after seroconversion (i.e. $S_{U|A}(t)$ is flat for $t > \omega_1 + \omega_2$, see the bottom graph of Figure 1). It should be noted that this is a special survival curve in that it never reaches a zero value, capturing the fact that a certain proportion of individuals will never progress above the threshold. This is what differentiates this approach from other approaches that do not account for assay non-progression (Such as Brookmeyer and Quinn [4], Janssen *et al.* [5], and Parekh *et al.* [1]).

For analytical convenience, we introduce $S_{PU|A}(t)$, the survival of assay *progressors* in the state of being *under-threshold*. We also introduce P_{NP} , the probability of individuals *not progressing* on the assay. Then $S_{U|A}(t)$, $S_{PU|A}(t)$ and P_{NP} are related by

$$S_{U|A}(t) = (1 - P_{NP})S_{PU|A}(t) + P_{NP}.$$

The introduction of $S_{PU|A}(t)$ allows us to provide a precise definition of the window period used by McDougal *et al.* It is the mean time between seroconversion and reaching threshold, for individuals who progress:

$$\omega = \int_0^\infty S_{PU|A}(t) dt. \tag{3}$$

Assumption 1 above can only mean that infection times in the first window period are *uniformly distributed*. Although assumption 2 merely states that the number of infections in the second window period is equal to the number in the first, we shall see later that for ρ_1 to be a property of the assay, independent of the epidemic state, we require the stronger assumption that the infection events in the second window period are also *uniformly distributed* with the same intensity as in the first window period. We see below that this assumption is implicit in the work of McDougal *et al.* To make this more explicit, we define $f(t)$ to be the density of times since infection realized in the sample. The number of *seropositive* individuals is then given by

$$N_{sp} = \sum_{i=1}^6 n_i = \int_0^\infty f(t) dt,$$

where n_i are the counts of individuals in the various categories depicted in the top graph in Figure 1.

Setting $f(t) = f_0$ over the first two window periods means that the ratio of the number of infected individuals in the second window

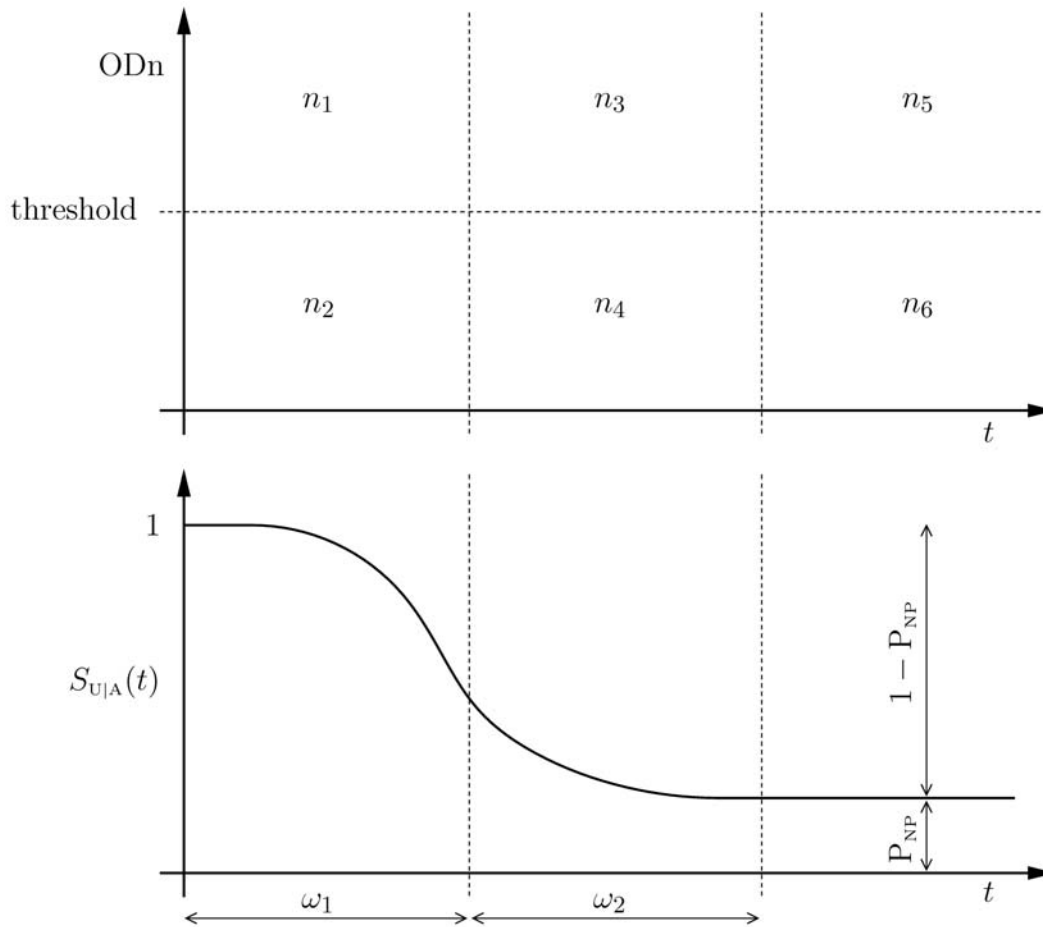


Figure 1. The six sector model of McDougal et al. The top graph shows counts n_i and the bottom graph shows the survival function $S_{U|A}(t)$ versus time since infection. doi:10.1371/journal.pone.0007368.g001

period to those in the first period is ω_2/ω_1 . Assumption 2 is recovered when the length of the window periods is equal. It should be noted that $f(t)$ depends on incidence, susceptible population and life expectancies over the history of the epidemic. With reference to Figure 1, we are now in a position to write expressions for the number of seropositive individuals in each sector:

$$n_1 = \int_0^{\omega_1} f(t)(1 - S_{U|A}(t)) dt = f_0(1 - P_{NP}) \int_0^{\omega_1} (1 - S_{PU|A}(t)) dt$$

$$n_2 = \int_0^{\omega_1} f(t)S_{U|A}(t) dt = f_0\omega_1 P_{NP} + f_0(1 - P_{NP}) \int_0^{\omega_1} S_{PU|A}(t) dt$$

$$n_3 = \int_{\omega_1}^{\omega_1 + \omega_2} f(t)(1 - S_{U|A}(t)) dt = f_0(1 - P_{NP}) \int_{\omega_1}^{\omega_1 + \omega_2} (1 - S_{PU|A}(t)) dt$$

$$n_4 = \int_{\omega_1}^{\omega_1 + \omega_2} f(t)S_{U|A}(t) dt = f_0\omega_2 P_{NP} + f_0(1 - P_{NP}) \int_{\omega_1}^{\omega_1 + \omega_2} S_{PU|A}(t) dt$$

$$n_5 = \int_{\omega_1 + \omega_2}^{\infty} f(t)(1 - S_{U|A}(t)) dt = (1 - P_{NP}) \int_{\omega_1 + \omega_2}^{\infty} f(t) dt$$

$$n_6 = \int_{\omega_1 + \omega_2}^{\infty} f(t)S_{U|A}(t) dt = P_{NP} \int_{\omega_1 + \omega_2}^{\infty} f(t) dt.$$

Using the above expressions, the sensitivity, the short-term specificity and the long-term specificity are given by

$$\sigma = \frac{n_2}{n_1 + n_2} = \frac{(1 - P_{NP}) \int_0^{\omega_1} S_{PU|A}(t) dt + \omega_1 P_{NP}}{\omega_1}$$

$$\rho_1 = \frac{n_3}{n_3 + n_4} = \frac{(1 - P_{NP}) \int_{\omega_1}^{\omega_1 + \omega_2} (1 - S_{PU|A}(t)) dt}{\omega_2}$$

$$\rho_2 = \frac{n_5}{n_5 + n_6} = 1 - P_{NP}.$$

We can now see why the assumption of uniformly distributed infection events for the first and second window periods is required – it is the only way in which a cancellation of $f(t)$ in the expressions for σ and ρ_1 is possible. Note that under bias-free

recruitment into a survey, at time $t=0$, we have

$$f(t) = \frac{N_{sp}}{T_{sp}} I(-t)H(-t)S_A(t), \tag{4}$$

where $I(t)$ is the instantaneous incidence, $H(t)$ is the number of healthy (susceptible) individuals, $S_A(t)$ is the life-expectancy survival function measured from the time since infection and

$$T_{sp} = \int_0^\infty I(-t)H(-t)S_A(t) dt$$

is the total number of seropositive individuals alive in the population at the time of the survey. The ratio N_{sp}/T_{sp} is just the fraction of the total population that has been recruited. Thus, the only sensible way to ensure that $f(t)=f_0$ for $t \leq \omega_1 + \omega_2$, is to assume that the incidence and the susceptible population are constant, and the survival function $S_A(t) = S_A(0) = 1$.

We also see why $S_{U|A}(t)$ must be flat after both window periods – this ensures that $S_{U|A}(t)$ is constant and can be pulled out of the integrals in the expressions for n_5 and n_6 as the factor P_{NP} . This is necessary for ρ_2 to be independent of $f(t)$.

Furthermore, in order to specify ρ_2 so that it is independent of the state of the epidemic, an implicit assumption is being made that survival is the same for assay progressors and assay non-progressors. Note that $f(t)$ appears in the expressions for both n_5 and n_6 . If different life expectancies were used in these formulae, reflecting a difference in survival for assay progressors and assay non-progressors, the f 's in these formulae would need to be different, and would not cancel in the expression for ρ_2 . This assumption is not explicitly stated by McDougal *et al.* but is implicit in their assumption that ρ_2 is independent of epidemic state.

With the calibration parameters specified in the more general setting of unequal window periods ω_1 and ω_2 , we now generalize the expression for the correction factor

$$f = \frac{P_t}{P_o},$$

where $P_t = (n_1 + n_2)/N_{sp}$ is the proportion of seropositive individuals who are *truly* infected at a time less than ω_1 . Recalling that $P_o = (n_2 + n_4 + n_6)/N_{sp}$ and using the definitions of the parameters, it is easy to verify that

$$P_o = P_t \sigma + P_t \frac{\omega_2}{\omega_1} (1 - \rho_1) + \left(1 - P_t - P_t \frac{\omega_2}{\omega_1}\right) (1 - \rho_2).$$

This means that the correction factor can be expressed as

$$f = \frac{P_o + \rho_2 - 1}{P_o \left[\sigma - \frac{\omega_2}{\omega_1} \rho_1 + \left(1 + \frac{\omega_2}{\omega_1}\right) \rho_2 - 1 \right]}. \tag{5}$$

Note that this equation simplifies to the previous expression (2) when one sets $\omega_1 = \omega_2$.

Elimination of Parameters

For completeness, we now provide a precise specification of the assumptions that are required in order to facilitate the analysis in

the rest of this paper. We note that with the exception of arbitrary sized window periods, these assumptions are equivalent to the assumptions – either explicit or implicit – that are being made by McDougal *et al.* [6].

Model Assumptions. Specify window periods ω_1 and ω_2 . We assume that:

1. The window periods are chosen so that the survival function $S_{U|A}(t)$ is flat (and equal to P_{NP}) for $t > \omega_1 + \omega_2$. This means that $S_{PU|A}(t)$ only has support on the time interval $t \in [0, \omega_1 + \omega_2]$.
2. Arrival times of infection events are uniformly distributed on the interval $[0, \omega_1 + \omega_2]$. An equivalent way of stating this assumption is that over the interval $t \in [0, \omega_1 + \omega_2]$, $H(t)$ and $I(t)$ are constant and $S_A(t) = 1$.
3. Survival is the same for assay progressors and assay non-progressors.

We are now able to provide the identity relating the parameters in the McDougal approach.

Proposition 1. Under the model assumptions stated above, the following identity holds:

$$\sigma - \frac{\omega_2}{\omega_1} \rho_1 + \left(1 + \frac{\omega_2}{\omega_1} - \frac{\omega}{\omega_1}\right) \rho_2 = 1. \tag{6}$$

Proof. Since we assume that $S_{PU|A}(t)$ only has support on $t \in [0, \omega_1 + \omega_2]$, we have

$$\int_0^{\omega_1 + \omega_2} S_{PU|A}(t) dt = \int_0^\infty S_{PU|A}(t) dt = \omega$$

Then, simply evaluating

$$\begin{aligned} \sigma - \frac{\omega_2}{\omega_1} \rho_1 &= \frac{(1 - P_{NP}) \int_0^{\omega_1} S_{PU|A}(t) dt + \omega_1 P_{NP}}{\omega_1} \\ &\quad - \frac{\omega_2 (1 - P_{NP}) \int_{\omega_1}^{\omega_1 + \omega_2} (1 - S_{PU|A}(t)) dt}{\omega_2} \\ &= \frac{(1 - P_{NP}) \int_0^{\omega_1 + \omega_2} S_{PU|A}(t) dt - \int_{\omega_1}^{\omega_1 + \omega_2} (1 - P_{NP}) dt + \omega_1 P_{NP}}{\omega_1} \\ &= \frac{(1 - P_{NP})(\omega - \omega_2 - \omega_1) + \omega_1}{\omega_1} \\ &= 1 - \left(1 + \frac{\omega_2}{\omega_1} - \frac{\omega}{\omega_1}\right) \rho_2, \end{aligned}$$

yields the result directly.

Using the proposition, the correction factor (5) simplifies to

$$f = \frac{\omega_1 P_o + \rho_2 - 1}{\omega P_o \rho_2}.$$

This expression no longer relies on estimates for σ and ρ_1 . It is also interesting to note that it does not depend explicitly on ω_2 . Calibrating ρ_2 , however, requires identifying individuals who have been infected for *at least* $\omega_1 + \omega_2$. Thus, ω_2 need not be precisely known, but a safe upper bound for $\omega_1 + \omega_2$ is required.

Furthermore, if we set $\omega_1 = \omega$ as in McDougal *et al.* then we recover

$$f = \frac{P_o + \rho_2 - 1}{P_o \rho_2}. \tag{7}$$

Note that (2) as stated in McDougal *et al.* contains three calibration parameters (σ , ρ_1 and ρ_2), while (7) contains only one calibration parameter (ρ_2). Incidence estimates using (1) and (7), however, still require the estimation of ω . The method of McDougal *et al.* can in principle be applied to an arbitrarily declared (as opposed to measured) window period, as long as σ , ρ_1 and ρ_2 are calibrated for that value. We have therefore reduced the number of calibration parameters by one.

Estimation of extra parameters may unnecessarily dilute the statistical power of the calibration data at hand. Moreover, estimates of the uncertainty due to calibration, based on the assumption of the independence of σ , ρ_1 and ρ_2 , will be incorrect. Note that when one sets $\omega_1 = \omega_2 = \omega$, the identity is reduced to

$$\sigma - \rho_1 + \rho_2 = 1.$$

Substituting the estimates of the parameters found by McDougal *et al.*, namely $\sigma = 0.768$, $\rho_1 = 0.723$ and $\rho_2 = 0.944$, into this equation gives a value of $0.989 \approx 1$ for the left hand side. The slight discrepancy is a manifestation of the combined fluctuations in the estimates of σ , ρ_1 , ρ_2 and ω . Although ω is superficially absent in the identity, it enters as the period over which the other parameters are defined.

When one assumes that $\rho_2 = 1$ (corresponding to the situation where there are no assay non-progressors) and $\omega_1 = \omega$, the identity reduces to

$$\omega(1 - \sigma) = \omega_2(1 - \rho_1) \tag{8}$$

and the ratio of counts over this period is given by

$$\frac{n_1 + n_2}{n_3 + n_4} = \frac{\omega}{\omega_2}.$$

Using this ratio and substituting the definitions for σ and ρ_1 into (8) yields $n_1 = n_4$. Therefore, for tests with perfect long-term specificity, the observed count of individuals who are under-threshold is an unbiased estimate of the number of infections in the last period ω . This was noted in a less general analysis of Brookmeyer [9] where assay non-progressors were *a priori* excluded.

It should be noted that there is a subtlety in the definition of the window period that emerges in the above analysis. If, instead of (3), the window period is defined by

$$\omega := \int_0^\infty S_{PU|A}(t) S_A(t) dt. \tag{9}$$

then the two definitions are equivalent under the model assumptions leading to the proposition. This follows from the fact that $S_{PU|A}(t)$ only has support on $t \in [0, \omega_1 + \omega_2]$ and that $S_A(t) = 1$ over that interval. We have suggested an alternative incidence estimation paradigm [8] which requires fewer assumptions than the method of McDougal *et al.* In this approach P_{NP} and ω , as defined in (9), emerge as the natural calibration parameters.

Comparison of Estimators Under Steady State Conditions

We now provide a simplified form for the McDougal incidence estimator based on the proposition. Substituting the new correction factor (7) into their estimator (1) and expressing the result in terms N_U , N_O and N_H gives

$$\bar{I}_a = \frac{N_U - P_{NP}(N_U + N_O)}{N_U - P_{NP}(N_U + N_O) + \omega(1 - P_{NP})N_H}, \tag{10}$$

where ω is specified in years. Here the subscript a indicates that the estimator is quoted as an ‘‘annualized incidence’’. Note that in writing down this expression, we have chosen to use P_{NP} rather than the long-term specificity as this is a biologically more intuitive parameter. In addition, the other two estimators to which this estimator will be compared were originally specified in terms of P_{NP} .

In a previous attempt to simplify the McDougal formula, Hargrove *et al.* [7] proposed the following incidence formula

$$\tilde{I}_a = \frac{N_U - P_{NP}(N_U + N_O)}{N_U - P_{NP}(N_U + N_O + N_H) + \omega N_H}, \tag{11}$$

where ω is specified in years. Note that they use the symbol ε where we use P_{NP} .

We have recently rigorously derived a weighted incidence estimator under less restrictive assumptions than those that are required for the McDougal or Hargrove approach [8]. Unlike the other two estimators, our estimator is expressed as a rate (indicated by a subscript r) and is given by

$$\begin{aligned} \hat{I}_r &= \frac{N_U - \frac{P_{NP}}{1 - P_{NP}} N_O}{\omega N_H} \\ &= \frac{N_U - P_{NP}(N_U + N_O)}{\omega(1 - P_{NP})N_H}. \end{aligned} \tag{12}$$

To convert between an annualized incidence and an incidence expressed as a rate, one can use the standard conversion formula

$$I_a = 1 - e^{-I_r T} \quad \Leftrightarrow \quad I_r = \frac{-\ln(1 - I_a)}{T},$$

where $T = 1$ year.

In Appendix S1 we show that, under steady state conditions, N_U and N_O are specified in terms of N_H and an incidence rate I as

$$N_O = I N_H (1 - P_{NP})(\alpha - \omega) \tag{13}$$

and

$$N_U = I N_H (1 - P_{NP})\omega + I N_H P_{NP}\alpha \tag{14}$$

where α is the post-infection life expectancy. Using these population counts, it is now possible to compare the performance of the incidence estimators. Substituting (13) and (14) into the McDougal formula (10) yields

$$\bar{I}_a = \frac{I}{I + 1}.$$

Converting this to a rate, we have

$$\bar{I}_r = \ln(I+1) = I + O(I)^2,$$

where the last step results from a Taylor series expansion. Thus the estimator is accurate for small values of I , but yields a discrepancy at $O(I)^2$. The reason for this discrepancy is subtle. In deriving the correction factor, McDougal *et al.* assume uniform infection events over the window periods. We have shown that this is consistent with assuming that the incidence and susceptible population are constant. In using this factor to estimate an incidence with (1) they have, however, inconsistently assumed that these infection events are generated in a susceptible population which is being depleted by the infection events over a period of a year. This is implied by their choice of denominator in that formula, which adds back an annualized number of recent infections into the susceptible population. This is at odds with the assumption of a constant susceptible population, and leads to dimensionally inconsistent incidence estimators, (1) and (10).

To illustrate the magnitude of the bias, Figure 2 shows the difference between the McDougal incidence estimate and the equilibrium incidence, expressed as a percentage. Note that the range of incidence values used is large (up to 50% per annum). Although incidence for HIV is not likely to be larger than about 15% in the highest risk groups (e.g. injection drug users [10]), if this methodology were used to monitor other rapidly spreading epidemics, where incidence is large when stated in units of years, it would certainly produce unacceptable bias.

Substituting the counts into the Hargrove formula (11) yields

$$\tilde{I}_a = \frac{I}{I + \frac{\omega - P_{NP}}{(1 - P_{NP})\omega}},$$

which, when converted to a rate, gives

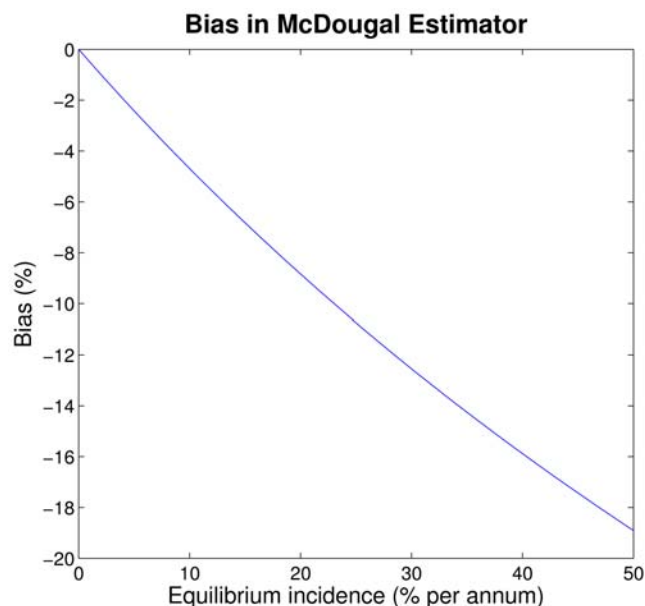


Figure 2. Bias in the McDougal estimator. Relative difference between the McDougal estimate and the equilibrium incidence plotted as a function of equilibrium incidence. doi:10.1371/journal.pone.0007368.g002

$$\tilde{I}_r = \ln\left(\frac{I(1 - P_{NP})\omega}{\omega - P_{NP}} + 1\right) = I + \frac{P_{NP}(1 - \omega)}{\omega - P_{NP}}I + O(I^2).$$

The Hargrove estimator incorporates the same form of denominator which leads to the second order discrepancy and dimensional inconsistency in the McDougal formula, and, in addition, it includes a linear bias term. Figure 3 demonstrates the bias introduced as a function of ω and P_{NP} for an equilibrium incidence of 5% per annum. Although the bias is worst in the regimes where all the estimators have little statistical power and are unlikely to be used, there are nevertheless intermediate regimes where the bias is significant. Note that the estimator produces the same result (and bias) as the McDougal estimator when $P_{NP} = 0$ or $\omega = 1$.

Finally, substituting the counts into our formula (12), which is already specified as a rate, yields

$$\hat{I}_r = I.$$

Thus, under the assumption of a steady state epidemic, our weighted incidence estimator recovers the steady state incidence exactly. It is also the maximum likelihood estimator. This can be seen by writing the estimator in terms of the population proportions

$$\hat{I}_r = \frac{P_U - \frac{P_{NP}}{1 - P_{NP}}P_O}{\omega P_H}, \quad \text{where } P_X = \frac{N_X}{N_{sp}},$$

and noting that, since the counts are trinomially distributed, the sample proportions are the maximum likelihood estimates of the population proportions. We have already seen that the estimator solves for the equilibrium incidence. Thus, by the invariance property of maximum likelihood estimators (see e.g. p. 105 of van den Bos [11]), it is the maximum likelihood estimator for the incidence. This has also recently been demonstrated by Wang and Lagakos [12] by explicit maximization of the log likelihood function.

A weighted incidence will in general not be equal to the instantaneous incidence under non-steady state conditions. We should, however, demand that any incidence formula exactly recover the incidence under this rather idealized situation.

Discussion

We have shown that under a precise restatement of the McDougal *et al.* assumptions, there exists a redundancy in the parameters they chose to characterise the assay. This allows the elimination of σ and ρ_1 from their estimator, with the important advantage that the remaining parameters are easier to calibrate. The calibration of σ and ρ_1 requires obtaining specimens from individuals with confidence about their time since infection (i.e. using frequent follow-up). On the other hand both ω and P_{NP} (or equivalently ρ_2) can be estimated through long follow-up intervals. The estimate for P_{NP} is the proportion of under-threshold samples known to be obtained more than $\omega_1 + \omega_2$ post-infection. Given an estimate for P_{NP} , an estimate of ω can be obtained from data with follow-up intervals greater than $\omega_1 + \omega_2$ using an extended version [13] of the Bayesian approach previously described by Welte [14].

We have also shown that under steady state conditions the only estimator that is dimensionally consistent and produces an unbiased result is the one we have previously derived [8]. It is

Bias in Hargrove Estimator (Equilibrium incidence=5% per annum)

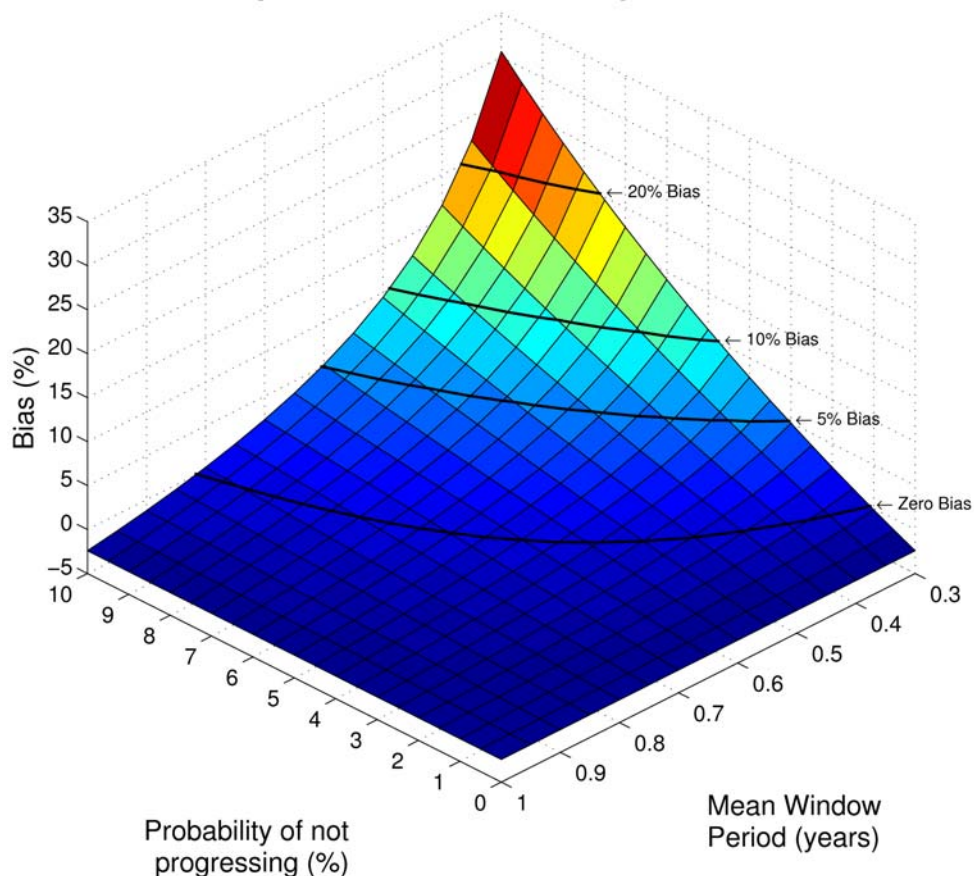


Figure 3. Bias in the Hargrove estimator. Relative difference between the Hargrove estimate and the equilibrium incidence plotted as a function of ω and P_{NP} for an equilibrium incidence of 5% per annum. Black lines indicate contours of equal bias.
doi:10.1371/journal.pone.0007368.g003

also the maximum likelihood estimator. The new approach makes fewer assumptions than the other methods. In particular, it consistently accounts for a dynamic epidemic by adopting a weighted definition of incidence. This overcomes a drawback of the other two methods which assume epidemic equilibrium for at least a period equal to the maximum progression time ($\omega_1 + \omega_2$). It should be noted that this methodology is applicable to any biomarker, not only the BED assay – all that is needed is a suitable calibration of the assay parameters. It also follows that cross-sectional incidence estimates using this approach are applicable to infections other than HIV, as long as suitably calibrated assays that test for recent infection are available.

A shortcoming of all the methods explored here is that they make the assumption, either implicitly or explicitly, that survival

for assay non-progressors and assay progressors is the same. As we have shown, relaxing this assumption means that the long-term specificity becomes epidemic state dependent and hence is time dependent. We are involved in ongoing work to address this issue.

Supporting Information

Appendix S1

Found at: doi:10.1371/journal.pone.0007368.s001 (0.04 MB DOC)

Author Contributions

Wrote the paper: TAM AW. Conceived and preformed the analysis: TAM AW.

References

1. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, et al. (2002) Quantitative detection of increasing HIV type 1 antibodies after seroconversion: A simple assay for detecting recent HIV infection and estimating incidence. *AIDS Res Hum Retroviruses* 18(4): 295–307.
2. Dobbs T, Kennedy MS, Pau CP, McDougal JS, Parekh B (2004) Performance characteristics of the Immunoglobulin G capture BED-enzyme immunoassay, an assay to detect recent human immunodeficiency virus type 1 seroconversion. *J Clinical Microbiology* 42(6): 2626–2628.
3. Bärnighausen T, McWalter TA, Rosner Z, Newell M-L, Welte A (2009) Use of the BED capture enzyme immunoassay to estimate HIV incidence: systematic review and sensitivity analysis. *Int J Epidemiol*, (submitted).
4. Brookmeyer R, Quinn TC (1995) Estimation of current human immunodeficiency virus incidence rates from a cross sectional survey using early diagnostic tests. *Am J Epidemiol* 141: 166–172.
5. Janssen RS, Satten GA, Stramer SL, Rawal BD, O'Brien TR, et al. (1998) New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *JAMA* 280: 42–8.
6. McDougal JS, Parekh BS, Peterson ML, Branson BM, Dobbs T, et al. (2006) 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* 22(10): 945–952.

7. Hargrove JW, Humphrey JH, Mutasa K, Parekh BS, McDougal JS, et al. (2008) Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS* 22(4): 511–518.
8. McWalter TA, Welte A (2009) Relating recent infection prevalence to incidence with a sub-population of assay non-progressors *J Mathematical Biology*. In press.
9. Brookmeyer R (2009) Should biomarker estimates of HIV incidence be adjusted? *AIDS* 23(4): 485–491.
10. Hu DJ, Vanichseni S, Mock PA, Young NL, Dobbs T, et al. (2003) HIV type 1 incidence estimates by detection of recent infection from a cross-sectional sampling of injection drug users in Bangkok: use of the IgG capture BED enzyme immunoassay. *AIDS Res Hum Retroviruses* 19(9): 727–30.
11. van den Bos A (2007) *Parameter estimation for scientists and engineers*. John Wiley & Sons.
12. Wang R, Lagakos SW (2009) On the use of adjusted cross-sectional estimators of HIV incidence, (working paper).
13. Kassanjee R, Welte A, McWalter TA, Viljoen J, Bärnighausen T, et al. (2009) Calibration of BED assay for use in incidence estimation. Fifth IAS Conference on HIV Pathogenesis, Treatment and Prevention, Cape Town, South Africa CDB018.
14. Welte A (2008) Relating incidence to 'recent infection' prevalence: Application to HIV. *South African Journal of Science* 104: 199–202.