Tests for recent infection (TRIs), such as the BED assay, provide a convenient way to estimate HIV incidence rates from cross-sectional survey data. Controversy has arisen over how the imperfect performance of a TRI should be characterised and taken into account. Recent theoretical work is providing a unified framework within which to work with a variety of TRI- and epidemic-specific assumptions in order to estimate incidence using imperfect TRIs, but suggests that larger survey sample sizes will be required than previously thought. This paper reviews the framework qualitatively and provides examples of estimator performance, identifying the characteristics required by a TRI to estimate incidence reliably that should guide the future development of TRIs.

Introduction

When monitoring HIV epidemics it is vital to estimate incidence in order to plan and evaluate HIV programmes [1]. Prospective cohort studies are the most direct way to achieve this. They are, however, expensive, prone to recruitment and retention bias, and potentially rendered unrepresentative by ethical obligations. The use of prevalence data in conjunction with mathematical modelling is an alternative approach [2,3], but is indirect and requires accurate knowledge of mortality and migration. The disadvantages of these methods have focused attention on estimating incidence from cross-sectional surveys [4-8], with the result that a number of assays and algorithms that test for recent infection have been developed [9,10]. In the context of HIV, such an assay or algorithm has sometimes been termed a STARHS (Serological Testing Algorithm for Recent HIV Seroconversion) [9,10], but we prefer to use the generic term ‘test for recent infection’ (TRI), because it does not specify a particular disease and method of testing. Recently, the World Health Organization (WHO) Technical Working Group on Statistical Approaches for Development, Validation and Use of HIV Incidence Assays has proposed using the term ‘recent infection testing algorithm’ (RITA). The term has not, however, gained universal acceptance.

TRIs identify HIV-positive individuals who have been infected recently. By using a TRI in a serosurvey, incidence (I) can be estimated by applying the epidemiological relationship (based on ‘Prevalence = Incidence \times \text{Duration}’):

\[ I = \frac{R}{SD} \]

where \( R \) and \( S \) are the counts of ‘recently infected’ and ‘susceptible’ (HIV-uninfected) individuals observed in the cross-sectional survey and \( D \) is the mean duration spent in the ‘recently infected’ state, often called the (mean) window period. This incidence estimate is an average of the instantaneous incidence over a period of approximately \( D \) prior to the survey. The problem of incidence estimation then reduces to measuring the prevalence of ‘recent infection’, given knowledge of its duration.

TRIs usually discriminate recent from established infections by measuring specific aspects of the immune system which evolve during the course of initial infection. For HIV, this is typically the antibody response, with the titre, proportion of HIV-specific IgG, or antibody avidity (or a combination of these) providing quantitative output [10]. Laboratory-defined thresholds are chosen to convert these outputs into categorical results. These results may be augmented with other clinical information, such as CD4 lymphocyte counts and antiretroviral therapy (ART) status, to classify individuals as either TRI-positive (\( P \) i.e. recent) or TRI-negative (\( N \) i.e. non-recent). Positive and negative in this context should not be confused with HIV-positive and HIV-negative.

The interaction between the virus and the immune system is complex, and individuals vary in their response to infection as assessed by a particular TRI. Modest
variation is not intrinsically problematic, but serious complications arise if, in some individuals, the immune response is such that they remain indefinitely classified as TRI-positive or if individuals revert back to a TRI-positive classification as a result of advanced disease or in the presence of antiretroviral therapy. Unfortunately, both these complications arise for TRIs currently in use. This not only limits the applicability of the simple incidence estimator above, but also makes it difficult to define and estimate the mean duration spent in the recently infected state (i.e. to evaluate $D$).

Methods for ‘adjusting’ estimates of incidence have been proposed [7,8] and adopted by the United States Centers for Disease Control and Prevention [11] but are currently under debate [12-15]. Recently, a formally rigorous framework has been developed [16,17]. We provide a summary of the framework and explore its implications for the analysis of surveys and development of new TRIs.

**Theoretical framework**

We now briefly describe the theoretical framework and how it can be generalised. The key results that emerge from the analysis are:

- A TRI is ideal if *all* individuals eventually progress permanently out of the TRI-positive state before there is any disease-related mortality. In this case, the TRI-positive category directly corresponds to a useful definition of ‘recently infected’ [16,18], which means that an estimate for the number of recent infections is:
  \[ R = P \]

- For a non-ideal TRI (i.e. when some individuals never progress out of the TRI-positive state), it is in principle still possible to estimate the number of individuals in a well-defined ‘recently infected’ state, even though this state is not directly observable in all individuals. If $P_{np}$ is the proportion of the HIV-positive individuals who never progress on the TRI under consideration, then an estimate for the number of recent infections is [16]:
  \[ R = P - \frac{P_{np}}{1 - P_{np}} N \]

When the TRI is ideal, then $P_{np} = 0$, and this formula reduces to the previous expression.

- For all applications (including determination of a trend without regard to the absolute level of incidence), an estimate of $P_{np}$ is required.

- To determine the *absolute level* of incidence, it is also necessary to estimate the mean time spent TRI-positive in the subset of individuals who eventually do progress to become TRI-negative. This quantity, which we denote by $\omega$, is analogous to the duration $D$ in the simple estimator, but differs in the requirement that it should be estimated in the subset of individuals that progress on the TRI.

- As $P_{np}$ increases (i.e. a larger fraction of individuals fail to progress on the TRI) and as $\omega$ decreases (i.e. individuals spend less time in the TRI-positive state) statistical power is lost. This means that estimates of incidence will have more uncertainty (i.e. wider confidence intervals), and it is less likely that a true change in incidence will be detected.

Previous work by McDougal *et al.* [7] used terminology usually employed to characterise the performance of diagnostic tests, such as sensitivity and specificity, to characterise TRI performance. ‘Recent infection’ was defined as being infected for less than a particular time (chosen to be the mean window period). A sensitivity and two specificity parameters were introduced to characterise imperfect classification. No procedure incorporating the effect of parameter uncertainty has thus far been proposed to estimate statistical error or power for the McDougal approach. It has recently been shown that use of sensitivity and specificity parameters is a redundant description of the TRI characteristics [17,19]. In contrast, the new framework defines the condition of being ‘recently infected’ directly in terms of the TRI result. This approach is applicable under less restrictive assumptions, is less prone to bias, and admits an equally informative description of TRI performance using only $\omega$ and $P_{np}$ [17].

In deriving the results outlined above, two assumptions were made. Firstly, it was assumed that individuals who do not progress on the TRI have the same survival outcomes as TRI progressors. There is, however, evidence for some TRIs that individuals that fail to progress on the test have a survival advantage. For example, in Baltimore, USA, 60% of elite suppressors (individuals with naturally suppressed virus below 50 copies per ml) failed to progress on the BED assay [20], and elite suppressors have been observed to survive for longer than others [21]. Secondly, it was assumed that TRI progressors never regress back to the TRI-positive state, but there are indications that this is not true for some TRIs. For example, the rate of misclassification by the BED assay is observed to be higher in individuals with advanced infection [22] and individuals on ART [22-24]. When these assumptions are true, $P_{np}$ is always equal to the proportion of non-recently infected individuals who are classified TRI-positive. When the assumptions are violated, this proportion, or false-recent rate, denoted by $\varepsilon$, varies according to the historic trajectory of the epidemic [17,25]. This would be consistent with the apparently higher BED assay false-recent rate in Uganda [26] (an older, declining epidemic) than in South Africa [27] (a younger, growing epidemic) [25]. It is, however, still possible to estimate the number of recent infections by replacing $P_{np}$ in the expression (1) above, with an estimate of $\varepsilon$ applicable to the time and place of an incidence survey [27] (see eAppendix http://www.sun.ac.za/sacema/publications/eAppendix.pdf for justification). The incidence estimator can then be written as:
The inputs to this estimator are of two types: survey counts \((P, N)\) and \((S)\), which need to be estimated in every incidence survey, and parameters that describe the characteristics of the TRI \((\omega, \epsilon)\), which ideally are estimated in a smaller number of parameter estimation studies.

When \(\epsilon\) and \(\omega\) are known with sufficient accuracy, there are no theoretical reasons why an imperfect TRI should not allow the accurate estimation of incidence. However, two distinct types of practical problems arise – counting error and TRI parameter error. An important component of recent developments is the first consistent analysis of incidence uncertainty accounting for both counting and parameter error (see eAppendix http://www0.sun.ac.za/sacema/publications/eAppendix.pdf for a description of the uncertainty expression).

We now illustrate this uncertainty with a somewhat idealised model of the BED assay, which has received much attention and application [28].

**Counting error**

Even in the largest HIV epidemics, infection events are relatively rare (about 2% of the population per year) and ‘recent’ infections (infections in the last 155 days or so, for the BED assay [7,29]) are even less common.

---

**Figure 1**

Uncertainty of incidence point estimates as a result of sample size and background incidence

---

\[
I = \frac{P - \epsilon N}{1 - \epsilon} \frac{1}{\omega S} \tag{2}
\]

CV: coefficient of variation; pyar: person years at risk; TRI: test for recent infection.

The coefficient of variation of estimates of incidence using a TRI depends on the sample size of the survey and the true incidence rate. Note that a sample size of 10,000 approximates to the typical size of household-based surveys in Sub-Saharan Africa, and that incidence in South Africa (where there is one of the largest epidemics) is estimated to be about two per 100 pyar.

Assumptions: \(\omega = 155\) days; \(\epsilon = 0.05\); no TRI parameter uncertainty; steady-state epidemic conditions; mean survival with HIV: 11 years [31-33].
(about 0.85% in a cross-section of the population). Thus, estimates of incidence are associated with substantial uncertainty since there are few recent infections to be counted. Figure 1 shows the coefficient of variation (CV, which is the ratio of the standard deviation to the estimate) for the estimator (2) calculated under various survey sample sizes and steady-state HIV incidence rates (see eAppendix http://www0.sun.ac.za/sacema/publications/eAppendix.pdf for a description of the uncertainty and steady-state calculations). The TRI parameters ($\omega$ and $\epsilon$) are assumed to be known with absolute certainty. Low values of CV are desirable and indicate that estimates of incidence have small confidence bounds, while high values indicate that incidence estimates will be less certain. For example, in a cross-sectional survey of 5,000 individuals from a population with a steady-state incidence of 2.0 per 100 person years at risk (pyar) the CV is 25.8% – i.e. the 95% likelihood interval for an incidence estimate is 1.0 to 3.0 per 100 pyar.

To explore the ability to detect a change in incidence, a substantial reduction (halving) in incidence is simulated (initially in a steady-state epidemic, with prevalence remaining constant between the two surveys), and a two-tailed test of the null hypothesis that incidence is the same in the two surveys is performed. The possible outcomes are: sustaining the null hypothesis, or concluding that incidence has either increased or decreased. Figure 2 shows the probability of correctly inferring a reduction in incidence between two surveys, when incidence has actually been reduced by half, as a function of the sample size of the surveys (both assumed to be the same) and the baseline incidence rate.

**Figure 2**
Probability of correctly inferring a reduction in incidence

pyar: person years at risk; TRI: test for recent infection.

The probability of detecting a reduction in incidence between two surveys, when incidence has actually been reduced by half, as a function of the sample size of the surveys (both assumed to be the same) and the baseline incidence rate.

Assumptions: $\omega=155$ days; $\epsilon=0.05$; no TRI parameter uncertainty; significance $\alpha=5\%$; steady-state epidemic conditions at first survey, with equal prevalence at second survey; mean survival with HIV: 11 years [31,33].
inferring a reduction in incidence, when testing the null hypothesis at a significance level of $\alpha = 5\%$. A probability close to 100% indicates that reductions in incidence will be reliably detected, with a probability of less than 90% indicating that results will be unreliable. The South African National Strategic plan for HIV AIDS [30] has ambitiously set a target of halving incidence between 2007 and 2012. Our calculations suggest that the sample size of each of two surveys (in 2007 and 2012) required to reliably conclude that incidence has decreased, at the 5% significance level, is approximately 25,000.

**TRI parameter error**

In the previous section, it was assumed that the correct TRI parameters were known with certainty. The incidence estimates are very sensitive to changes in the values of $\omega$ and $\epsilon$, however, and small differences between the values used in the calculation and the true values can lead to large errors. These parameters have to be estimated in separate studies, usually using cohorts of individuals whose infection time is known approximately. Such cohorts are rare, however, and the numbers of individuals in them are typically small, resulting in substantial uncertainty for the values of $\omega$ and $\epsilon$. In Figure 3 we explore the uncertainty of the estimator (expressed as a CV), as a function of the uncertainty in the TRI parameters. For example, when

![Figure 3](image-url)

Uncertainty of incidence point estimates as a result of TRI parameter uncertainty

CV: coefficient of variation; pyar: person years at risk; TRI: test for recent infection.

Coefficient of variation of incidence estimator, using a BED-like assay on a sample size of 5,000, in a population exposed to an incidence of two per 100 pyar, as a function of the uncertainty in the TRI parameters, assumed to be normally distributed.

Assumptions: $\omega = 155$ days; $\epsilon = 0.05$; steady-state epidemic conditions; mean survival with HIV: 11 years [31,33].
the BED-like parameters are known with a CV of only 15.0% (reference to a CV for ω relates to the uncertainty of the estimate of ω, not the variation associated with progression times), at a sample size of 5,000 and a steady-state incidence of 2.0 per 100 pyar the CV, as a result of both counting error and parameter uncertainty, is 35.7% – i.e. the 95% likelihood interval for an incidence estimate is 0.6 to 3.4 per 100 pyar.

Since the TRI parameter estimation study may be conducted in a separate population, it is possible to introduce systematic bias if the true values of the TRI parameters vary between populations or over time. The few estimates of ε that have been published vary widely. For example, the false-recent rate is estimated at 1.7% in a South African survey [27] and 26.7% in Rwanda and Zambia [26] presumably due to population differences in the historic courses of the epidemics, viral subtypes, host immune profiles, and uptake of antiretroviral therapy. This undermines confidence in the ability to use an estimate for ε obtained in a different population to the one in which incidence is to be estimated, and could contribute to the apparently inflated estimates of incidence reported recently [34,35]. There is also currently no general theoretically unbiased procedure for estimating ε – work on this problem is in progress [36]. In Figure 4 we explore the systematic error in the incidence estimate, expressed as a percentage of the correct value, introduced by systematic errors in the TRI parameters, also expressed

**Figure 4**
Systematic error in incidence point estimates as a result of systematic error in TRI parameters

Contours of constant error (%)

<table>
<thead>
<tr>
<th>Error in ω (%)</th>
<th>Error in ε (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>−30</td>
<td>−30</td>
</tr>
<tr>
<td>−20</td>
<td>−20</td>
</tr>
<tr>
<td>−10</td>
<td>−10</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

pyar: person years at risk; TRI: test for recent infection.
Systematic error expressed as a percentage of the correct estimate, excluding counting error, observed in the incidence estimator, using a BED-like assay, as a function of a precisely known systematic error in the TRI parameters.
Assumptions: ω=155 days; ε=0.05; steady-state epidemic conditions; mean survival with HIV: 11 years [31,33].
as percentages. There is a region in which bias may be small due to cancellation of systematic errors (see the zero error contour).

**Conclusion**

In the short term, reports from early studies using BED should be interpreted with caution [28], given the substantial uncertainties identified above. Analysis of TRI data should be performed within a more general theoretical framework [16,17], rather than using earlier methods. Most importantly, incidence surveillance should not currently rely on any single methodology, but make use of multiple methods for estimating incidence [37], such as interpretation of prevalence trends and epidemiological and demographic modelling [3,38].

The search for robust means of estimating incidence from cross-sectional surveys is at a crucial juncture. Although an imperfect TRI can be used to estimate HIV incidence reliably, the reliance on having accurate and precise values of two key aspects of TRI performance ($\omega$ and $\varepsilon$) can undermine the use of this technology. The effect of $\omega$ and $\varepsilon$ on statistical power is shown in Figure 5. While larger values of $\omega$ provide sufficient numbers of TRI-positive individuals to ensure statistical power, $\omega$ should not be so large that the estimated incidence is not representative of the recent past. On this basis,

**Figure 5**

Uncertainty of incidence point estimates as a result of TRI performance

![Figure 5](image-url)

CV: coefficient of variation; pyar: person years at risk; TRI: test for recent infection.

Coefficient of variation of incidence estimator, on a sample size of 5,000, in a population exposed to an incidence of two per 100 pyar, as a function of the TRI parameters.

Assumptions: no TRI parameter uncertainty; steady-state epidemic conditions; mean survival with HIV: 11 years [31,33].
a value of approximately six months to a year is desirable.

In the next phase of TRI development, it will be essential to be guided by these insights into the key determinants of test performance, and to focus on characterising the performance of the test within a systematic framework.

Acknowledgements

A Welte and TA McWalter thank the Canadian International Development Agency (CIDA) for funding support. TB Hallett thanks The Wellcome Trust for funding support. This research was partly supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), United States National Institutes for Health (NIH). In addition this research was sponsored in part by The HIV Prevention Trials Network (HPTN) sponsored by NIAID, National Institute on Drug Abuse (NIDA), National Institute of Mental Health (NIMH), and Office of AIDS Research, of the NIH, United States Department of Health and Human Services (U01-AI-068613).

References


