

## Short Communication: Defining Optimality of a Test for Recent Infection for HIV Incidence Surveillance

Reshma Kassanjee,<sup>1,2</sup> Thomas A. McWalter,<sup>1,3</sup> and Alex Welte<sup>1</sup>

### Abstract

The estimation of HIV incidence from cross-sectional surveys using tests for recent infection has attracted much interest. It is increasingly recognized that the lack of high performance recent infection tests is hindering the implementation of this surveillance approach. With growing funding opportunities, test developers are currently trying to fill this gap. However, there is a lack of consensus and clear guidance for developers on the evaluation and optimization of candidate tests. A fundamental shift from conventional thinking about test performance is needed: away from metrics relevant in typical public health settings where the detection of a condition in individuals is of primary interest (sensitivity, specificity, and predictive values) and toward metrics that are appropriate when estimating a population-level parameter such as incidence (accuracy and precision). The inappropriate use of individual-level diagnostics performance measures could lead to spurious assessments and suboptimal designs of tests for incidence estimation. In some contexts, such as population-level application to HIV incidence, bias of estimates is essentially negligible, and all that remains is the maximization of precision. The maximization of the precision of incidence estimates provides a completely general criterion for test developers to assess and optimize test designs. Summarizing the test dynamics into the properties relevant for incidence estimation, high precision estimates are obtained when (1) the mean duration of recent infection is large, and (2) the false-recent rate is small. The optimal trade-off between these two test properties will produce the highest precision, and therefore the most epidemiologically useful incidence estimates.

**T**HE MEASUREMENT OF HIV INCIDENCE, the rate of new infections, is essential in most surveillance and intervention contexts. Recognizing the practical challenges presented by longitudinal studies, the estimation of incidence from cross-sectional surveys using tests for recent infection has attracted considerable interest.<sup>1-7</sup> However, the performance, characterization, and optimization of a test that aims to categorize infections as “recent” or “nonrecent,” specifically for *population-level surveillance*, requires a shift from conventional diagnostic thinking about test performance.

When *individual-level detection* of a condition is of primary interest, sensitivity, specificity, and predictive values are appropriate metrics of performance. These metrics improve as intersubject variability decreases. However, when estimating a population-level summary parameter, such as incidence, the appropriate performance metrics are accuracy and precision of the statistic measured. Biomarker-based cross-sectional incidence estimation utilizes information on the *average* behavior of biomarkers, and is relatively insensitive to the var-

iability underlying this averaging. While the appropriate optimization of tests for recent infection has been noted in passing,<sup>3-7</sup> there is neither consensus nor guidance for developers.

As with any diagnostic, elements of a test for recent infection may be adjusted to alter its performance. In the context of HIV recent infection tests, typically some quantitative host or viral biomarkers are measured, and the infection is categorized as “recent” or “nonrecent” by reference to thresholds.<sup>1-3</sup> For example, the widely used BED assay measures the proportion of HIV-specific immunoglobulin G (IgG) antibodies in total IgG, and a measurement below some threshold classifies the infection as “recent.”<sup>8</sup> While a test may be composed of many elements that can be varied, from the underlying biological processes measured to the assay platforms and specific kits, ultimately the optimization will involve the fine-tuning of thresholds.

It is increasingly recognized that the lack of high performance recent infection tests poses a major obstacle to the

<sup>1</sup>The South African DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), University of Stellenbosch, Stellenbosch, South Africa.

<sup>2</sup>School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa.

<sup>3</sup>Department of Actuarial Science and the African Collaboration for Quantitative Finance and Risk Research, University of Cape Town, Cape Town, South Africa.

widespread implementation of cross-sectional incidence surveillance.<sup>5,7</sup> The World Health Organization (WHO) has maintained a WHO Working Group on HIV Incidence Assays since 2006, the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA) was established in 2010, and both the Bill and Melinda Gates Foundation and the National Institutes of Health have provided substantial funding for the development of better tests.<sup>9–13</sup> Given the current surge in the development of candidate tests for recent infection, it is important to have clarity and consensus on robust metrics of performance, and in particular to avoid the pitfalls of traditional diagnostic thinking.

Prevalence, the fraction of a population with a condition, can at times substantially inform us about incidence. For example, for transient conditions, such as influenza, it is well known that near demographic equilibrium:

$$\text{Incidence} \approx \text{Prevalence} / \text{Mean duration of condition} \quad (1)$$

where incidence is expressed as a rate of cases per person time in the entire population, not just per person time at risk. However, when a condition is enduring, and survival in the state is poorly known and evolving, as is the case with HIV, prevalence becomes uninformative about incidence. In this case, it makes sense to find ways of defining and detecting a robust early phase postinfection, and using a more refined version of the above heuristic to infer incidence from the prevalence of “recent” infection.

Under simplistic assumptions, HIV incidence, expressed as a rate of infections per person time at risk, is then formally estimated, in a cross-sectional setting, by<sup>14</sup>

$$\hat{I} = \frac{p_R}{p_- \Omega} \quad (2)$$

where  $p_R$  and  $p_-$  are the proportions of “recently” infected and HIV-negative subjects in the sample and  $\Omega$  is the mean duration of recent infection. Currently available, and perhaps all conceivable, tests for recent infection present a subtle problem in that some individuals who have been infected for long periods of time may nevertheless yield spurious “recent” results.<sup>15–17</sup> With some simplifying assumptions (which are often not numerically catastrophic) it has been shown how this “false-recent” phenomenon can be intuitively understood as requiring a “subtraction” of the estimated number of “false-recent” results from the observed number of “recent” results.<sup>18–21</sup>

More recently, a very general analysis has been obtained by introducing a convenience recency time cut-off,  $T$  (presumed to be 1 year for the purposes of model scenarios throughout this article), which represents the time, postinfection, after which a “recent” test result is a “false-recent” result.<sup>21</sup> The test properties then are (1) a false-recent rate,  $\beta_T$ , which is the (population-dependent) proportion of those individuals infected for more than time  $T$  who produce “recent” test results, and (2) a somewhat subtly defined mean duration of recent infection,  $\Omega_T$ , which is the average time spent “recently” infected while infected for less than  $T$ .<sup>21</sup> Note that  $1 - \beta_T$  is the (population-dependent) specificity of the test if it aimed to identify infections that have occurred within the preceding period  $T$ . This leads to the following incidence estimator<sup>21</sup>:

$$\hat{I}_T = \frac{p_R - \beta_T p_+}{p_- (\Omega_T - \beta_T T)} \quad (3)$$

which depends on the proportions of subjects in the sample who are classified as “recently” infected ( $p_R$ ), HIV-positive ( $p_+$ ), and HIV-negative ( $p_- = 1 - p_+$ ), and the test properties ( $\Omega_T$  and  $\beta_T$ ) for a chosen recency time cut-off  $T$ . When there are no “false-recent” results ( $\beta_T = 0$ ), Eq. (3) reduces to Eq. (2). In terms of epidemiological and demographic context, the applicability of Eq. (3) requires only that the susceptible population not vary substantially over a period of duration  $T$ .<sup>21</sup> In terms of the biomarkers underlying the recency test, these should mainly capture stable biological, as opposed to environmentally dependent, factors over the period  $T$  postinfection ( $\Omega_T$  should not vary significantly by context, but be a true property of the test).<sup>21</sup> It is understood that  $\beta_T$  will have contextual variability.<sup>21</sup>

Uncertainty in the incidence estimate arises from statistical fluctuations of the proportions of subjects in the sample in the various classes, as well as uncertain test properties. The uncertainty in the incidence estimate, described here by its coefficient of variation (ratio of standard deviation to mean),  $c$ , can be approximated using the delta method<sup>21</sup>:

$$c^2 \approx \frac{1}{n P_+} \left( \frac{1}{P_-} + \frac{P_R P_{NR}}{(P_R - \beta_T P_+)^2} \right) + \sigma_{\hat{\Omega}_T}^2 \left( \frac{1}{\Omega_T - \beta_T T} \right)^2 + \sigma_{\hat{\beta}_T}^2 \left( \frac{P_{NR} \Omega_T - P_R (T - \Omega_T)}{(P_R - \beta_T P_+) (\Omega_T - \beta_T T)} \right)^2 \quad (4)$$

where  $P_R$  and  $P_{NR}$  are the proportions of “recently” and “nonrecently” infected individuals in the study population;  $P_+ = P_R + P_{NR}$  and  $P_- = 1 - P_+$  are the proportions of HIV-positive and HIV-negative individuals in the population; and  $\sigma_{\hat{\Omega}_T}$  and  $\sigma_{\hat{\beta}_T}$  are the uncertainties (standard deviations) with which the test properties  $\Omega_T$  and  $\beta_T$  are measured. It is certainly possible that the normality assumptions intrinsic to deriving Eq. (4) could be violated in practice, in which case the same underlying theory can be used as a basis for a numerically more complex calculation of the variance of incidence estimates, such as by bootstrap resampling methods.<sup>22</sup> Whether or not Eq. (4) is used to estimate the coefficient of variation is not fundamental to the present discussion about optimization.

The familiar statistical objective when estimating any parameter is to find a sample statistic that estimates the parameter with the greatest accuracy (least bias) and greatest precision (smallest variance). As shown in the general derivation of Eq. (3),<sup>21</sup> the bias is negligible compared to variance in the epidemiologically/demographically relevant regime, using any reasonable test for recent infection. The remaining goal is therefore the minimization of variance. The apparent bias reported by other researchers<sup>23</sup> is a result of an alternative summary parameterization of biomarker dynamics, which declares the false-recent rate to be zero. This leads to a mean duration of recent infection that is complex, context-dependent, and difficult to estimate, and hence produces a similarly context-dependent implicit weighting over historical incidence.

To minimize variance, the state of “recent” infection should not be too transient, so that a realistically sized cross-sectional survey can capture a sufficient number of “recent” cases for

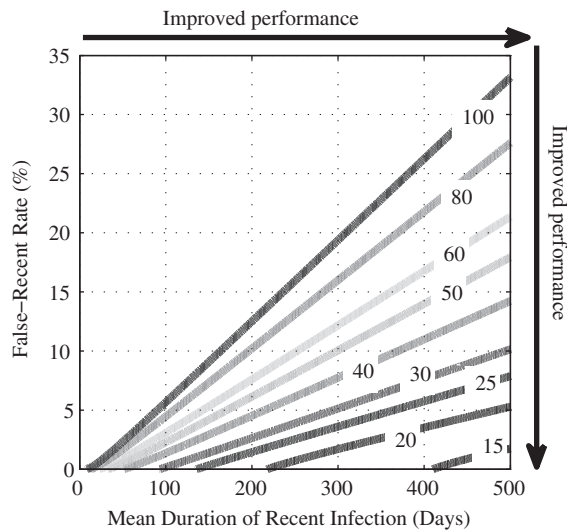
the estimation of the “recent” proportion to be statistically robust. The larger the adjustment for “false-recent” results, the greater the overall uncertainty arising from fluctuations of the sample proportions. The two test properties cannot be independently adjusted, as increases in the mean duration of recent infection typically also increase the false-recent rate. Hence, the central goal of test design is an optimal balance between these two properties. An ideal test would have a near-zero false-recent rate and a mean duration of recent infection of the order of a year (considering a setting where we are interested in the average incidence over approximately the last year).

Figure 1 illustrates this trade-off between the mean duration of recent infection,  $\Omega_T$ , and false-recent rate,  $\beta_T$ . The contours show the coefficient of variation (CoV) of the incidence estimate, in an example context (epidemiological and demographic equilibrium of 1% per annum HIV incidence and 10% HIV prevalence;  $\Omega_T$  and  $\beta_T$  estimated with a 5% and 30% CoV, respectively; and incidence measured in a cross-sectional survey of 5,000 subjects). The CoV is shown as a function of the mean duration of recent infection and false-recent rate. Moving to the right (to a large mean duration of recent infection) and down (to a low false-recent rate) in the contour plot, the CoV of the incidence estimate decreases (i.e., precision increases). For a test to begin to move into a regime of usefulness, the mean duration of recent infection should be at least 6 months and the false-recent rate below 2%.<sup>4,5,12</sup> In the context captured in Fig. 1, this implies a CoV of the incidence estimate of 30%, which implies that we can be 95% confident of estimating the true incidence of 1% per annum as a point estimate between 0.4% and 1.6% per annum.

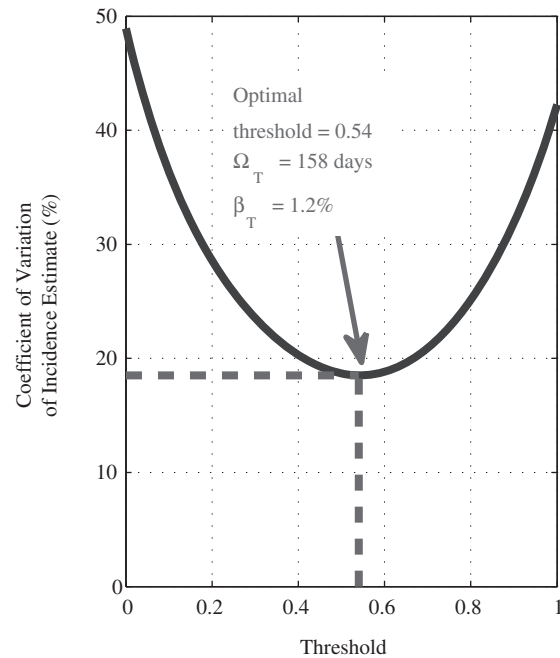
To illustrate the optimization of test design in a very simplistic scenario, consider a biomarker where the reading at a time  $t$  after infection (years) is  $y(t) = 1 - \exp(-2t) + 0.2\varepsilon$ , where  $\varepsilon$  is standard normal noise and all individuals survive for 10

years after infection. Readings below some chosen recency threshold indicate “recent” infection. Here, the only source of variability is noise in the biomarker, while in reality there is often substantial intersubject variability and effects due to immune system decline and treatment. Nevertheless, this simple model captures the same trade-off, that increasing the recency threshold increases both the mean duration of recent infection and false-recent rate, that would be observed with more complex biomarker dynamics. In Fig. 2, we demonstrate the variation of the precision of incidence estimates, with changes in the recency threshold. Again, the precision is quantified for a chosen context (epidemiological and demographic equilibrium of 1% per annum HIV incidence and 10% HIV prevalence;  $\Omega_T$  and  $\beta_T$  exactly known; and a cross-sectional survey size of 10,000 subjects). Note that  $\beta_T$  depends on the survival dynamics and the epidemiological and demographic history of the population (up until the maximum postinfection survival time). A tool to calculate the CoV of incidence estimates is available ([www.incidence-estimation.com/page/tools](http://www.incidence-estimation.com/page/tools)).

Unfortunately, there is no single test design that will be optimal in all settings. This is because uncertainty is determined by both the dynamics of the recent infection test as well as the context-specific epidemiological and demographic dynamics (captured by incidence and prevalence). Therefore, a range of anticipated contexts should be considered in evaluating a candidate test or in fine-tuning test properties. This context-specific performance may be discouraging and regrettably complicated, but it is not unique to this surveillance application: even in a conventional simplistic diagnostics setting, the sensitivity and specificity of a test, if these can be assumed to be context-independent, must be combined with a



**FIG. 1.** Test performance as a function of test properties. The coefficient of variation (CoV) of the incidence estimate (%), as a function of the mean duration of recent infection,  $\Omega_T$  (days) and false-recent rate,  $\beta_T$  (%):  $n = 5,000$ ,  $\sigma_{\hat{\Omega}_T} = 5\% \times \Omega_T$  and  $\sigma_{\hat{\beta}_T} = 30\% \times \beta_T$ , HIV incidence = 1% per annum, and HIV prevalence = 10% [see Eq. (4)].



**FIG. 2.** Optimal threshold for a test for recent infection based on a single biomarker. The coefficient of variation (CoV) of the incidence estimate (%), as a function of the recency threshold:  $n = 10,000$ ,  $\sigma_{\hat{\Omega}_T} = 0\%$  and  $\sigma_{\hat{\beta}_T} = 0\%$ , HIV incidence = 1% per annum, and HIV prevalence = 10% [see Eq. (4)].

contextual prevalence to determine the predictive value performance of the test.

The minimization of the variance of the incidence estimate, or maximization of precision, by trading the mean duration of recent infection off against the false-recent rate, provides a completely general criterion for optimizing test design, regardless of the complexity of the test. For example, there is a trend toward devising a dichotomous test for recent infection based on multiple biomarkers, where various approaches for combining the individual biomarker results could be employed (for example, the final classification could be based on whether the sum of the biomarkers readings is below a single threshold or on the number of individual readings below biomarker-specific thresholds). The optimal test design and thresholds are those that provide the lowest variance in incidence estimates across intended contexts.

Obtaining the most precise incidence estimates also consistently captures the optimization that would be appropriate in studies that aim to test for differences in incidence or identify risk factors for HIV acquisition. Statistical tests for differences between groups (for example, capturing different ages, genders, or social and sexual behaviors) are more highly powered when incidence is more precisely estimated in each group.

Much of the literature introducing new tests for recent infection has attempted to assess their utility in terms of sensitivity and specificity.<sup>24–30</sup> As would be appropriate in the more familiar diagnostic applications, values close to 100% have been regarded as realistic targets, with these two measures summarized into, for example, receiver operating characteristic (ROC) curves and the overall classification accuracy. This line of thinking runs into three major obstacles. (1) Any workable definition of sensitivity and specificity requires a notion of “recent” infection defined by a strict threshold on time since infection and a fully specified distribution of times since infection in a population. (2) Even if thus defined, sensitivity and specificity cannot be accurately estimated from interval censored seroconverter data sets. (3) Intersubject variability of infection-related biomarkers naturally increases with time postinfection, and therefore diagnostic optimization will tend to motivate for a category of “recent” restricted to the lower variability period close to infection, whereas incidence estimation requires the most enduring notion of “recent” infection that does not bring a substantial false-recent rate.

In a clinical setting there may be substantial value in having some evidence of time since infection, at the time of HIV diagnosis. This opens up a multitude of new questions beyond the scope of this discussion. Most importantly, there needs to be further work to support reporting and interpretation of individual biomarker values beyond a “recent” or “nonrecent” categorical result, and so in this context the optimization of a test may not be the fine-tuning of a recency threshold.

Any evaluation and optimization of a test for recent infection should be based on the specific purpose for which the test is to be used, with the current work focusing on incidence estimation. The goal of HIV incidence estimation from cross-sectional surveys, using tests for recent infection, has attracted the interest of test developers. However, the assessment and optimization of these tests, for purposes of estimating a population-level average, require a fundamental shift from traditional criteria for measuring performance. In this article we have laid out the relevant performance metric of such tests, namely the precision of incidence estimates produced in an

intended context. The central goal of the test developer, then, is the minimization of the variance of incidence estimates through a trade-off between the mean duration of recent infection and false-recent rate. A test performance calculator is available at [www.incidence-estimation.com/page/tools](http://www.incidence-estimation.com/page/tools).

## Acknowledgments

This work was supported in part by grants from the Canadian International Development Agency and the Bill and Melinda Gates Foundation. Authors are grateful to their colleagues at the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA) for stimulating discussions on test development, evaluation and optimization.

## Author Disclosure Statement

No competing financial interests exist.

## 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(36):11–16.
2. Murphy G and Parry JV: Assays for the detection of recent infections with human immunodeficiency virus type 1. *Euro Surveill* 2008;13(36):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(18):2763–2771.
4. Welte A, McWalter TA, Laeyendecker O, and Hallett TB: Using tests for recent infection to estimate incidence: Problems and prospects for HIV. *Euro Surveill* 2010;15(24): pii = 19589.
5. Incidence Assay Critical Path Working Group: More and better information to tackle HIV epidemics: Towards improved HIV incidence assays. *PLoS Med* 2011;8(6):e1001045.
6. Hallett TB: Estimating the HIV incidence rate: Recent and future developments. *Curr Opin HIV AIDS* 2011;6(2):102–107.
7. Sharma UK, Schito M, Welte A, *et al.*: Workshop summary: Novel biomarkers for HIV incidence assay development. *AIDS Res Hum Retroviruses* 2012;28(6):532–539.
8. Parekh BS and McDougal JS: New approaches for detecting recent HIV-1 infection. *AIDS Rev* 2001;3:183–193.
9. UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance: When and how to use assays for recent infection to estimate HIV incidence at a population level. 2011. [www.who.int/diagnostics\\_laboratory/110906\\_guidance\\_hiv\\_incidence.pdf](http://www.who.int/diagnostics_laboratory/110906_guidance_hiv_incidence.pdf).
10. WHO Technical Working Group on HIV Incidence Assays: [www.who.int/diagnostics\\_laboratory/links/hiv\\_incidence\\_assay/en/](http://www.who.int/diagnostics_laboratory/links/hiv_incidence_assay/en/). Accessed March 14, 2013.
11. The Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA): [www.incidence-estimation.com/page/cephia](http://www.incidence-estimation.com/page/cephia). Accessed March 14, 2013.
12. Bill and Melinda Gates Foundation Letters of Inquiry (LOI): New Biomarkers for HIV Incidence Measurement. [www.nuhs.edu.sg/wbn/slot/u3394/HIV\\_Incidence\\_LOI\\_Rules\\_and\\_Guidelines.pdf](http://www.nuhs.edu.sg/wbn/slot/u3394/HIV_Incidence_LOI_Rules_and_Guidelines.pdf). Accessed March 14, 2013.
13. National Institutes of Health Funding Opportunity Announcements: HIV Incidence Assays with Improved Specificity. <http://grants.nih.gov/grants/guide/pa-files/PA-10-212.html>. Accessed March 14, 2013.

14. Brookmeyer R and Quinn TC: Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests. *Am J Epidemiol* 1995;141(2):166–172.
15. Chaillon A, Le Vu S, Brunet S, *et al.*: Decreased specificity of an assay for recent infection in HIV-1-infected patients on highly active antiretroviral treatment: Implications for incidence estimates. *Clin Vaccine Immunol* 2012;19(8):1248–1253.
16. Laeyendecker O, Brookmeyer R, Oliver AE, *et al.*: Factors associated with incorrect identification of recent HIV infection using the BED capture immunoassay. *AIDS Res Hum Retroviruses* 2012;28(8):816–822.
17. Wendel SK, Mullis CE, Eshleman SH, *et al.*: Effect of natural and ARV-induced viral suppression and viral breakthrough on anti-HIV antibody proportion and avidity in patients with HIV-1 subtype B infection. *PLoS One* 2013;8(2):e55525.
18. 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(10):945–952.
19. Hargrove JW, Humphrey JH, Mutasa K, *et al.*: Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS* 2008;22(4):511–518.
20. McWalter TA and Welte A: Relating recent infection prevalence to incidence with a sub-population of assay non-progressors. *J Math Biol* 2010; 60(5):687–710.
21. Kassanjee R, McWalter TA, Barnighausen T, and Welte A: A new general biomarker-based incidence estimator. *Epidemiology* 2012;23(5):721–728.
22. Efron B and Tibshirani RJ: *An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability 57)*. Chapman & Hall/CRC, Boca Raton, FL, 1993.
23. Brookmeyer R: On the statistical accuracy of biomarker assays for HIV incidence. *JAIDS* 2010;54(4):406–414.
24. Parekh BS, Kennedy MS, Dobbs T, *et al.*: Quantitative detection of increasing HIV type 1 antibodies after seroconversion: A simple assay for detecting recent HIV infection and estimating incidence. *AIDS Res Hum Retroviruses* 2002;18(4):295–307.
25. Karita E, Price M, Hunter E, *et al.*: Investigating the utility of the HIV-1 BED capture enzyme immunoassay using cross-sectional and longitudinal seroconverter specimens from Africa. *AIDS* 2007;21(4):403–408.
26. Sakarovich C, Rouet F, Murphy G, *et al.*: Do tests devised to detect recent HIV-1 infection provide reliable estimates of incidence in Africa? *JAIDS* 2007;45(1):115–122.
27. Guy R, Gold J, Calleja JM, *et al.*: Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: A systematic review. *Lancet Infect Dis* 2009;9(12):747–759.
28. Masciotra S, Dobbs T, Candal D, *et al.*: Antibody avidity-based assay for identifying recent HIV-1 infections based on Genetic Systems™ 1/2 Plus O EIA. Abstract 937 at the 17th Conference on Retroviruses and Opportunistic Infection, 2010.
29. Braunstein S, Nash D, Ingabire C, Mwamarangwe L, and Wijgert Jvd: Performance of BED-CEIA and Avidity Index assays in a sample of ART-naïve, female sex workers in Kigali, Rwanda. Abstract 939 at the 17th Conference on Retroviruses and Opportunistic Infection, 2010.
30. Park SY, Love TM, Nelson J, Thurston SW, Perelson AS, and Lee HY: Designing a genome-based HIV incidence assay with high sensitivity and specificity. *AIDS* 2011;25(16):F13–F19.

Address correspondence to:  
Reshma Kassanjee  
SACEMA  
c/o SIIAS  
Private Bag X1  
Matieland  
Stellenbosch 7602  
South Africa

E-mail: rkassanjee@sun.ac.za