

# Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: a systematic review

Rebecca Guy, Judy Gold, Jesus M García Calleja, Andrea A Kim, Bharat Parekh, Michael Busch, Thomas Rehle, John Hargrove, Robert S Remis, John M Kaldor, for the WHO Working Group on HIV Incidence Assays\*

We systematically reviewed the accuracy of serological tests for recent infections with HIV that have become widely used for measuring population patterns incidence of HIV. Across 13 different assays, sensitivity to detect recent infections ranged from 42–100% (median 89%). Specificity for detecting established infections was between 49.5% and 100% (median 86.8%) and was higher for infections of durations longer than 1 year (median 98%, range 31.5–100.0). For four different assays, comparisons were made between assay-derived population incidence estimates and a reference incidence estimate. The median percentage difference between the assay-derived incidence and reference incidence was 26.0%. Serological assays have reasonable sensitivity for the detection of recent infection with HIV, but are vulnerable to misclassifying established infections as recent—potentially leading to biases in incidence estimates. This conclusion is highly qualified by the apparent absence of a standardised approach to assay evaluation. There is an urgent need for an internationally agreed framework for evaluating and comparing these tests.

## Introduction

Programmes for the prevention of HIV are aimed at reducing viral transmission in populations. The monitoring of transmission, indicated by incidence rates of HIV,<sup>1</sup> is therefore essential for establishing the need for prevention programmes and their effectiveness.

The direct measurement of incidence is resource intensive and intrusive, because it needs repeat serological testing of individuals over time.<sup>1</sup> The use of direct measurement has accordingly been limited to particular settings, such as cohort studies. However, longitudinal cohorts cannot be assumed to be representative of the wider populations that are the targets of prevention programmes. To provide a more practical means of estimating the incidence of HIV, several groups have developed specialised testing algorithms that can distinguish recent from established infections with HIV on the basis of single serological specimens.<sup>2</sup>

The underlying principle of these algorithms is the use of an additional test applied to serological specimens from people with newly diagnosed infections with HIV. This test has an extended window period, meaning that it detects a marker that is reliably negative in the early stages of infection, and then becomes positive at a well defined interval after initial infection.<sup>2</sup> This characteristic contrasts with the key requirement for a standard diagnostic test for infection with HIV, which must have as short a window period as possible to ensure that new infections are not missed. The incidence rate of HIV in a population is estimated by applying a formula involving the assay's window period to the number of cases in the population that are detected by the assay as being recent.<sup>2</sup> The first such algorithm, developed by the US Centers for Disease Control and Prevention (CDC), used a modified version of a commercial assay as the test with an extended window period.<sup>3</sup> More recently, the CDC

developed a dedicated assay with an extended window period, known as the BED test,<sup>4</sup> which, internationally, has become the most widely used test of this type. Other assays for recent infection have been developed in France, the USA, Australia, and Italy.<sup>5–8</sup> The assays have been used in multiple countries in a range of surveillance and research settings.<sup>9–16</sup>

In December, 2005, the role of assays for recent infection with HIV was thrown into question when a report by the Joint United Nations Programme on HIV/AIDS claimed that the BED assay substantially overestimated incidence rates in Africa, because it falsely labelled some longstanding infections as being recent.<sup>17</sup> Subsequent efforts were made to correct or adjust the incidence estimates resulting from the use of these assays,<sup>18,19</sup> but the debate has continued about the real accuracy, and hence the value, of assays for recent infection with HIV. To provide an indication of the accuracy of the tests that are available at present, and to inform the development of a more systematic approach to assay development and validation, we systematically reviewed published reports on the accuracy of tests for recent infection with HIV.

## Methods

The biological basis for the various tests developed to detect recent infection with HIV has been reviewed elsewhere.<sup>2</sup> In developing a new assay, investigators first calibrate the assay by establishing its window period—defined by the optical density cut-off (or its equivalent) at which the test should be read as positive. This calibration is done with serial serum specimens from people with very recent infection. Once the assay has been calibrated, its accuracy can be assessed. Our Review addressed assay accuracy as defined by either the measurement of performance characteristics or by validation of assay-derived incidence.

*Lancet Infect Dis* 2009; 9: 747–59

\*Other members listed at the end of the paper

Centre for Population Health, Burnet Institute, Melbourne, VIC, Australia (R Guy PhD, J Gold BBIomedSci); National Centre in HIV Epidemiology and Clinical Research, University of New South Wales, Sydney, NSW, Australia (R Guy, Prof J M Kaldor PhD); HIV/AIDS Department, WHO, Geneva, Switzerland (J M García Calleja MD); Epidemiology and Strategic Information Branch, Global AIDS Program (A A Kim PhD), and Serology/Incidence and Diagnostics Team, GAP International Laboratory Branch (B Parekh PhD), Centers for Disease Control and Prevention, Atlanta, GA, USA; Blood Systems Research Institute, San Francisco, CA, USA (Prof M Busch MD); Human Sciences Research Council, Cape Town, South Africa (Prof T Rehle MD); South African Centre for Epidemiological Modelling and Analysis, Stellenbosch, South Africa (Prof J Hargrove PhD); and Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada (Prof R S Remis PhD)

Correspondence to:

Prof John M Kaldor, National Centre in HIV Epidemiology and Clinical Research, University of New South Wales, Level 2, 376 Victoria St, Darlinghurst, 2010, Sydney, NSW, Australia  
jkaldor@nchecr.unsw.edu.au

For more on the BED test see <http://www.cdc.gov/hiv/topics/surveillance/resources/factsheets/BED.htm>

	Number of sample sets used for assessment of assay performance characteristics (%)	Number of sample sets used for validation of assay-derived HIV incidence (%)
Total number of sample sets	81 (100)	33 (100)
Sample source		
Cohort	53 (65.4)	9 (27.3)
Cross-sectional survey	4 (4.9)	18 (54.5)
Commercial panel	7 (8.6)	1 (3.0)
Other*	7 (8.6)	3 (9.1)
Not specified	10 (12.3)	2 (6.1)
Region of sample origin		
Africa	12 (14.8)	9 (27.3)
Americas	30 (37.0)	15 (45.5)
Asia	5 (6.2)	6 (18.2)
Europe	24 (29.6)	2 (6.1)
Oceania	4 (4.9)	0 (0.0)
Multiple†	0 (0.0)	1 (3.0)
Not specified	6 (7.4)	0 (0.0)
Viral subtype‡		
B	12 (14.8)	3 (9.1)
Combination including subtype B	1 (1.2)	7 (21.2)
C	4 (4.9)	1 (3.0)
Other subtypes (CRF02_AG, E, A1)	6 (6.6)	0 (0.0)
Combination not including subtype B	2 (2.5)	2 (6.1)
Not specified	56 (69.1)	20 (60.6)
Risk group for the acquisition of HIV		
Accidental exposure	0 (0.0)	1 (3.0)
Recipient of blood transfusion	2 (2.5)	3 (9.1)
Heterosexual from country with high prevalence of HIV§	12 (14.8)	8 (24.2)
Heterosexual other	1 (1.2)	0 (0.0)
Use of illicit intravenous drugs	3 (3.7)	7 (21.2)
Men who have sex with men	14 (17.3)	11 (33.3)
Not specified	49 (60.5)	3¶ (9.1)

\*Includes convenience samples, case-based surveillance. †Includes specimens from the USA and Holland. ‡Majority or all of the specimens tested. §From sub-Saharan Africa or Thailand. ¶Included three sample sets from blood donors where the risk group for infection with HIV was not reported.

**Table 1: Characteristics of sample sets used for assessment of assay performance characteristics and validation of assay-derived HIV incidence**

We defined the measurement of performance characteristics of an assay as the process of collecting serum specimens from people infected with HIV for a known duration, applying the assay to these specimens and calculating the sensitivity of the assay to correctly detect infections of short duration as recent, and the specificity of the assay to correctly detect infections of longer duration as not recent. We accepted that individual investigators would define recent in different ways. We defined the validation of assay-derived incidence estimates as the process of collecting serum specimens from the members of a population in which a reference estimate of the incidence of HIV was available, applying the assay for recent infection to these specimens, deriving an estimate of population

incidence of HIV from the results, and comparing it with the reference estimate.

### Search strategy and selection criteria

We searched all published papers and conference abstracts that reported on the calibration, validation, or assessment of a serological test for recent infection with HIV. We then focused on reports of studies that had done either the measurement of assay performance characteristics or a validation of assay-derived incidence estimates resulting from application of the assay.

Reports were excluded if they were based on a review, commentary, or editorial, rather than a primary research report containing original data; described findings that were presented more comprehensively in another report included in our Review; described the application of an assay without any measure of performance characteristics or validation; described a test based on detection of viral nucleic acid or p24 antigen; or where it had been established how recent the infection with HIV in the specimens was using another assay for recent infection.

PubMed was searched to the end of June, 2009, and only English-language papers were reviewed. Reference lists of selected studies were also checked for other potentially relevant studies. Conference presentations were included if the corresponding full report was not available. If the information needed was not available in the conference presentation, authors were contacted for unpublished data. The primary search of published work was done using the terms “HIV” and “incidence”, combined with “immunoassay” and “surveillance”. Variations of the terms were used.

For each paper we cross-checked the author’s name, study location, and key findings with other papers in our Review, ensuring the same data was not duplicated in another study. For abstracts, we specifically searched for a publication or contacted the abstract authors to establish if a paper was in press or recently published.

### Classification and analysis of the published work

For each report, information was extracted on the assay name and type. We also extracted information on the characteristics of the specimens tested with the incidence assay, including the number of people in the study sample, the number of people tested with the assay, the specimen source, demographic characteristics, risk group, HIV subtype, and CD4-T-cell count.

For each report on the measurement of performance characteristics, we abstracted information on the assay window period, as well as the optical density cut off that had been used to read the test result. We then obtained information on the means by which the authors established how recently infection with HIV had been acquired in the samples used to measure performance characteristics. In the absence of standardised terminology in the published work, we adopted a

classification scheme under which a sample was defined as relating to a recent infection if it came from a person documented to have been infected with HIV for a duration that was shorter than the length of the window period of the assay under investigation; the infection was defined as established if it came from a person with infection of longer duration than the assay window period; and the infection was defined as longstanding if the specimen came from a person who was documented to have had been infected with HIV for a year or more. For each population reported in each publication, we then abstracted information on assay sensitivity (proportion of recent infections detected as recent by the assay) and specificity (proportion of non-recent infections detected as non-recent).

For each report on the validation of assay-derived incidence estimates, we abstracted information on the source of the incidence estimate that was used as a reference, and any correction factors that had been applied to the assay-derived estimate of incidence. A Microsoft Access database was created to record the information that was abstracted from the reports.

We did descriptive analyses of the methods that had been used to evaluate assays, both with regard to their performance characteristics and the validation of incidence estimates. These analyses focused on the types of samples used and the comparisons that had been made. We then did quantitative analyses of the findings from the assay evaluations. For publications that measured assay performance characteristics we analysed reported information on sensitivity and specificity, calculating 95% CIs using STATA 9.2. Some reports described findings related to several different sample sets and others described a single finding related to sample sets that had been pooled.

For the reports describing validation of assay-derived estimates of the incidence of HIV, we did a quantitative analysis of reported results on the basis of comparisons between the assay-derived incidence and the reference incidence estimate in the overall study population.

The percentage difference between the estimates was calculated using the following formula:

$$\text{Percentage difference} = \frac{\text{assay-derived incidence} - \text{reference incidence}}{\text{reference incidence}} \times 100$$

The median percentage differences were calculated for subgroups of sample sets defined by assay type, risk group, and whether or not the sample set specimens and the specimens used to calculate the reference incidence estimate were identical or non-identical. Identical referred to studies that tested the same samples in both the assay and reference incidence method. Non-identical referred to studies in

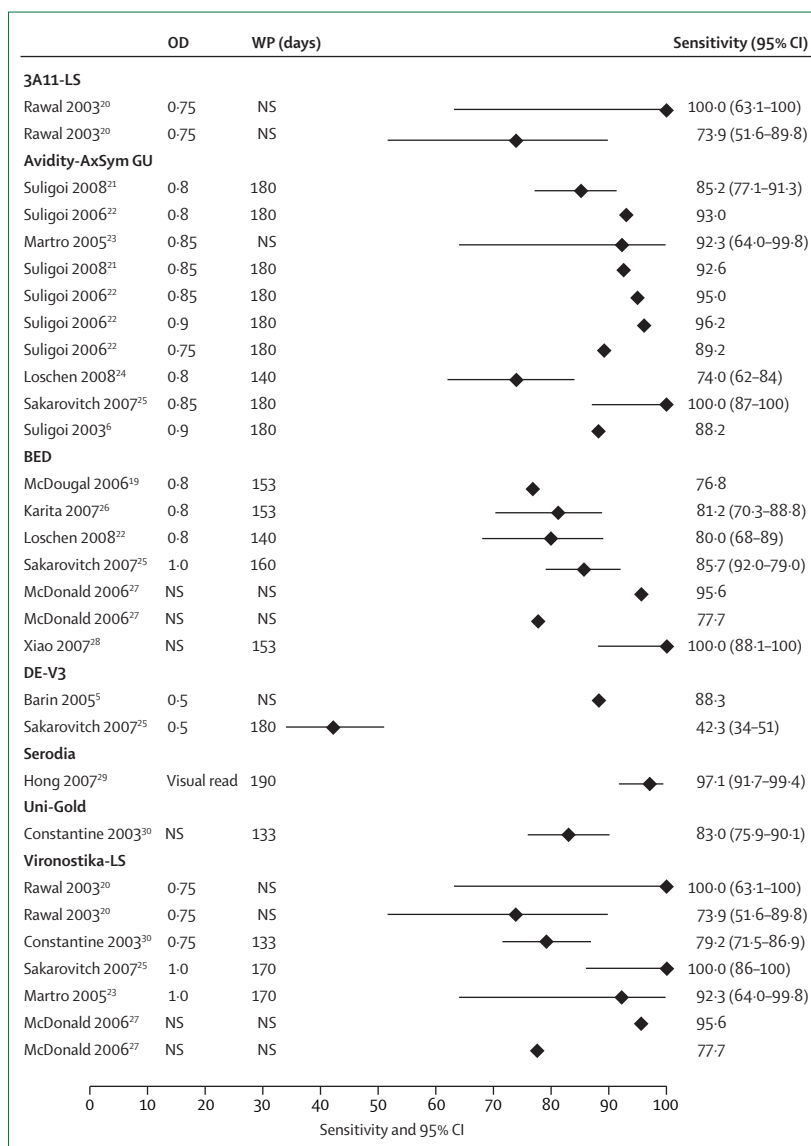
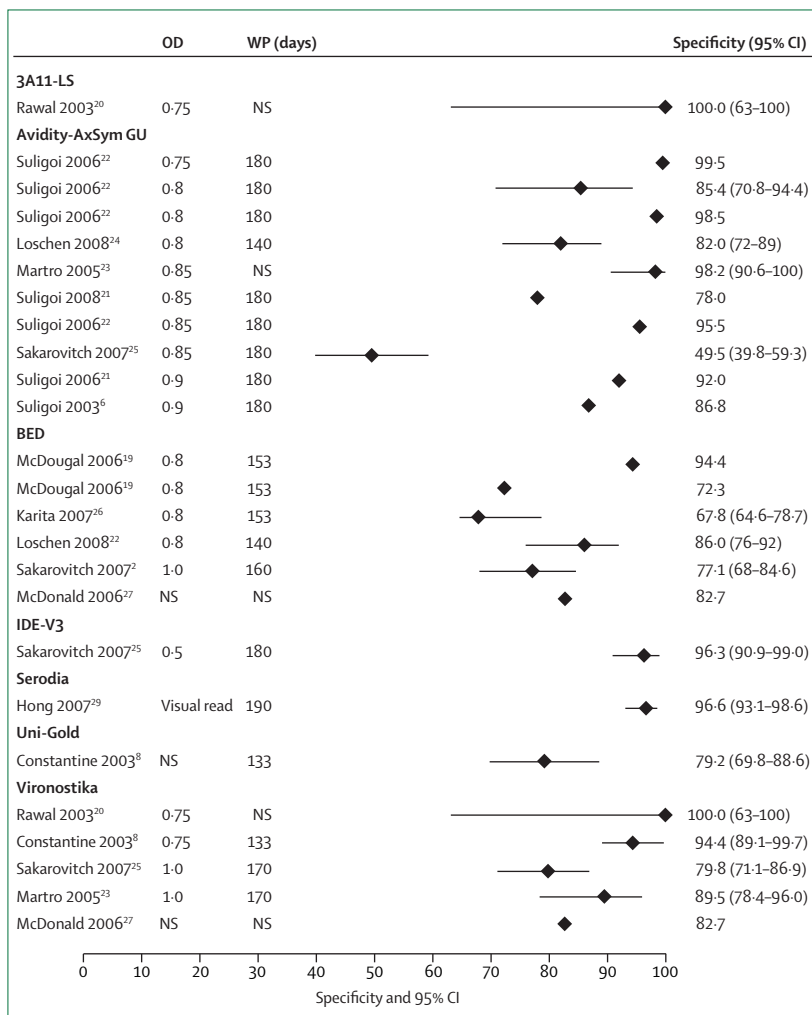


Figure 1: Sensitivity estimates using recent infections  
NS=not specified. OD=optical density. WP=window period.

which different samples were used or a mathematical technique was used to estimate the reference incidence.

Some papers reported both crude assay-derived incidence estimates, and estimates that were corrected by various mathematical techniques. If so, we separately compared the crude and the corrected estimates to the reference estimate.

We assessed the difference between assay-derived and reference incidence estimates by establishing whether or not the 95% CIs for the two estimates overlapped. If the CIs were not reported, we calculated them whenever possible using the binomial approximation method.



**Figure 2: Specificity estimates using established infections**  
NS=not specified. OD=optical density. WP=window period.

## Results

### Overall

A total of 178 reports were identified, including 45 that met the inclusion criteria. The remainder were excluded because they described findings that were presented more comprehensively in another report included in our Review (30 of 178); described a test based on detection of viral nucleic acid or antigen (31); described the application of an HIV-incidence assay without measurement of performance characteristics or validation (36); did not present primary research data (19); or where it had been established how recent the infection with HIV in the specimens was using another assay for recently acquired infection (10). A further seven papers that were excluded described the genetic characterisation of incident HIV subtypes or established the incidence of HIV using another method (cohort, repeat testing, mathematical techniques) and mentioned the potential rather than the actual use of HIV-incidence assays.

The 45 reports that met the inclusion criteria described the measurement of performance characteristics or incidence validation relating to 13 different assays.

### Performance characteristics

26 reports, involving 81 sample sets, described the measurement of performance characteristics of 13 different assays. Individual reports provided information on both recent and established infections (15 of 81); recent and established infections and cases of AIDS (9); recent infections only (4); established infections only (34); and cases of AIDS only (19). Among 59 sample sets that included established infections, 22 were of samples from individuals who had longstanding infections—ie, they had been infected with HIV for 1 year or more.

Most sample sets were derived from cohorts recruited in high-income countries (table 1). Among the 25 sample sets for which the HIV subtype was specified, the majority were of subtype B (table 1). Just over half of the sample sets (48 of 81) provided information on the number of people used to assess performance characteristics, and for 58 sample sets the report provided the total number of specimens tested. The median number of specimens tested per sample set was 100 (range 7–2749).

There were 30 sensitivity estimates reported for seven different assays (figure 1),<sup>5,6,8,19,20-29</sup> with an overall median of 88.8%, and a range 42.3–100%. Among assays for which more than three sensitivity estimates were available, the median estimate was 92.5% for Avidity-AxSym GU (range 74.0–100), 81.2% for BED (range 76.8–100), and 92.3% for Vironostika-LS (range 73.9–100).

There were 25 specificity estimates related to established infections from 12 reports<sup>6,8,19,20-27,29</sup> (excluding longstanding infections and individuals receiving treatment with antiretroviral therapy). The median specificity for established infections was 86.8% (range 49.5–100.0; figure 2). Among assays for which more than three specificity estimates were available, the median was 89.4% for Avidity-AxSym GU (range 49.5–99.5), 79.9% for BED (range 67.8–94.4) and 89.5% for Vironostika-LS (range 79.8–100).

There were 11 specificity estimates (from eight reports)<sup>3,7,18,20,28,31-33</sup> for longstanding infections with a median of 98% (range 31.5–100.0; figure 3) and 23 specificity estimates related to cases of AIDS (from nine reports)<sup>3,5,14,20,23,27,33-35</sup> with a median estimate of 91.6% (range 72.2–100.0; figure 4).

There were also 11 specificity estimates (from five reports)<sup>20,30,32,34,36</sup> that provided information about treatment with antiretroviral therapy. For five specificity assessments (based on two reports), the median specificity was 91.0% (range 85.1–100) for specimens obtained from people who had been treated for 1 year, but it fell to 75.8% (72.7–81.8) for those treated for 2 years (figure 5). In another report the specificity estimate was 100% for

specimens obtained from people who were untreated, falling to 90% after 2 months of antiretroviral therapy (figure 5).<sup>36</sup> The specificity estimates in the two other reports<sup>21,36</sup> did not state the time since starting treatment.

### Validation of assay-derived incidence of HIV

There were 23 published reports that described comparisons between assay-derived estimates of incidence, and corresponding reference estimates. These comparisons involved four different assays: BED, Vironostika-LS, 3A11-LS, and IDE-V3. Those most frequently subjected to validation were the BED (15 of 23) and Vironostika-LS (7) assays.

A total of 33 sample sets were described in the 23 reports. Cross-sectional surveys were the source of 18 of the 34 sample sets (table 1) and were most commonly based on samples obtained through community recruitment (5 of 18), voluntary counselling and testing clinics (4), sexual health clinics (3), and blood donors (3). Cohorts were the next most common source of specimens, providing the basis for nine sample sets, and arising from a vaccine preparedness study (3 of 9), a study at a drug treatment clinic for people who inject drugs (2), a study of antenatal women (1), an HIV-risk-reduction project (1) a study of military personnel (1), and a community cohort (1).

The sample sets used for validation came mainly from high-income countries, particularly the USA (14 of 33). The main risk groups represented in the samples were men who have sex with men (11), heterosexual people from sub-Saharan Africa and Thailand (8), and people that use injected drugs (7) (table 1). HIV-subtype information was only available for 13 sample sets, and most (10 of 13) were described as exclusively or primarily subtype B (table 1). CD4-T-cell count was not reported for any of the sample sets (table 1).

22 of the 33 sample sets reported the number of people in the study sourced for the assay validation, with the median number being 656 people (range 18–32 651).

The 33 sample sets gave rise to 183 separate assay-based incidence estimates that were compared with corresponding reference incidence estimates. Of the comparisons, 34 were for the overall study populations and the remainder were stratified by time (55 of 171), risk group (34), age (26), gender (10), HIV subtype (4), or other characteristics (22).

The most frequently used source of reference incidence data for comparison with overall study populations (table 2) was prospective cohort studies (22 of 34), followed by a database of people repeatedly tested for HIV (6), and mathematical model outputs (6).

Ten of the 34 overall analyses were based on comparing assay-derived incidence with reference incidence from identical samples. Eight of the ten analyses were cohort studies that provided directly measured incidence estimates as reference for the assay-derived incidence estimate. The other 24 of the 34 overall analyses involved

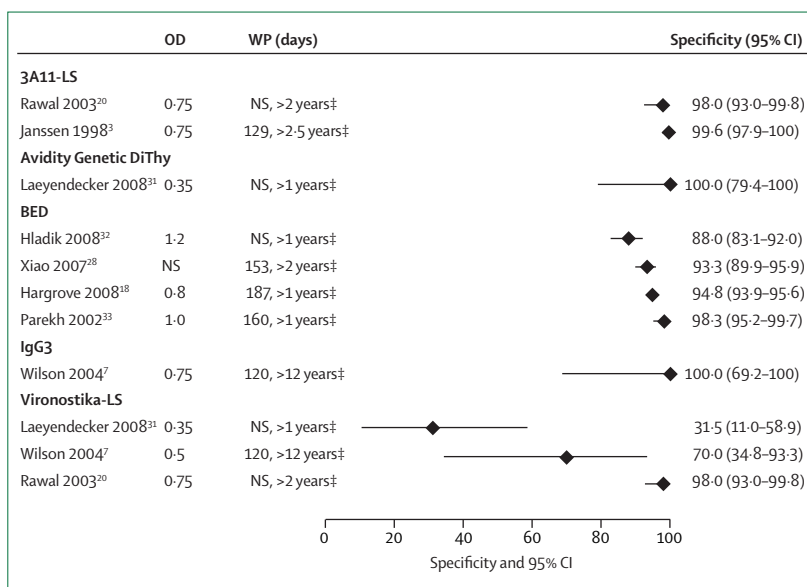


Figure 3: Specificity estimates using longstanding infections\*

NS=not specified. OD=optical density. WP=window period. \*Excluding estimates from samples known to include individuals treated with highly active antiretroviral therapy. †Specimens collected from elite suppressors. ‡Years since infection.

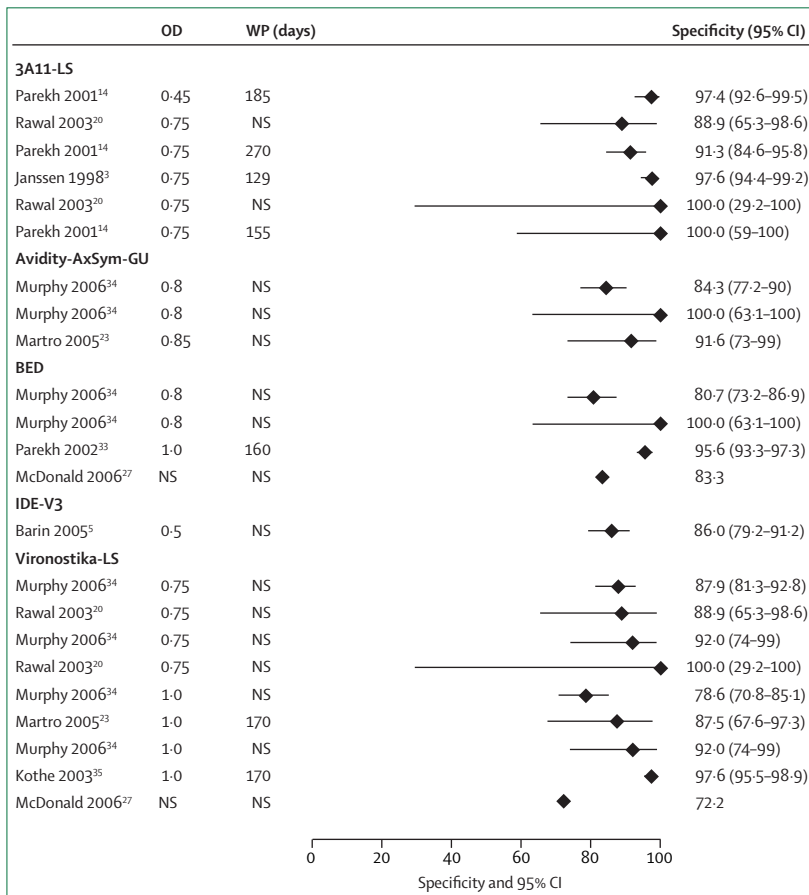
comparison of assay-derived and reference incidence in non-identical samples. For 20 of these comparisons the assay derived incidence estimate was based on specimens from cross-sectional studies, and compared with reference incidence that was either observed in a cohort study (12 of 20), calculated from the results of repeat HIV testing (either recorded in a database or based on self report; 4), or estimated by mathematical techniques (4).

The 34 overall analyses were based on 22 reports (Thai-US BED Validation Working Group, unpublished),<sup>3,12,16,18,26,33,37–49</sup> and the percentage difference between the assay-derived and reference incidence for each of the 34 overall results is presented in figure 6. Aggregating across all 34 overall results, the median percentage difference was 26.0% (range 0.0–483%). Findings from one additional report<sup>50</sup> were excluded as no crude assay-derived incidence estimates were reported.

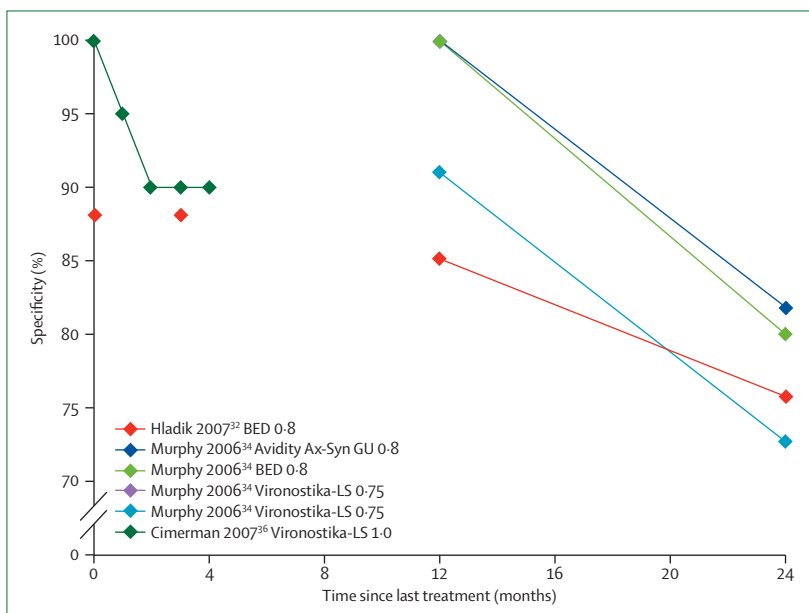
The median percentage difference was 7.1% (range 0–13.5%) for the three validations of the 3A11-LS assay, 27.3% (range 1.7–483%) for 20 validations of the BED assay, and 39.9% (range 8.3–133%) for the 10 validations of the Vironostika-LS assay. Validation results for the IDE-V3 assay were only available in one report, with a percentage difference of 18.8%.

For the 10 validations in which identical samples were used, the median percentage difference was 9.4% (range 2.9–48%) compared with 47.6% (range 0–483) for the 24 validations done using non-identical populations.

Among specific risk groups, the median percentage difference was 39.9% (range 0–121%) for the ten validations done among people who inject drugs, 18.8% (range 13.5–20.5%) for the three validations done among



**Figure 4: Specificity estimates using AIDS cases**  
 NS=not specified. OD=optical density. WP=window period.



**Figure 5: Specificity estimates using established infections with treatment information\***  
 \*The reports of Rawal<sup>20</sup> and Selli<sup>20</sup> (three assessments) were excluded because the timing of treatment was not specified.

blood donors, 7.7% (range 1.7–133%) for the eight validations in men who have sex with men, and 54.1% (range 2.9–483%) for the 13 validations among heterosexual people from sub-Saharan Africa and Thailand, nine of which involved comparisons based on non-identical samples, with assay-derived incidence being estimated from samples obtained from a cross-sectional study.

The statistical difference between the assay-derived incidence and the reference incidence estimate could be assessed by examining the overlap of CIs for 26 of the 34 overall results. In only seven instances was there non-overlap between the two sets of CIs (table 2).

In addition to crude assay-derived incidence estimates, 12 of the 34 overall comparisons also presented corrected estimates. The correction methods varied substantially. Eight comparisons used the McDougal<sup>19</sup> method; one the Hargrove<sup>18</sup> formula; one a simplified version of the McDougal-formula<sup>38</sup> and compared this to the McDougal, Hargrove, and McWalter/Welte<sup>51</sup> formulas; and two used the Satten<sup>32</sup> method (figure 7).

The McDougal and Hargrove formulas adjust for the imperfect specificity of the assays (false-positive rate for longstanding infections). Two comparisons from two reports<sup>26,49</sup> based on these formulas used a locally derived false-positive rate (0.052 and 0.0169, respectively) and eight comparisons (five reports) used false-positive rates (0.052 or 0.0557) that had been established externally in other populations.<sup>26,40,41</sup>

The median percentage difference between the crude and reference estimates was 126% reducing to 36.2% for the difference between the corrected and reference estimates. The results of applying the correction varied substantially across the studies. Four corrected estimates (two reports)<sup>39,40</sup> became on average 20.1% further away from the reference estimate. Another four corrected estimates (three reports)<sup>18,40,41</sup> became on average 41.1% closer to the reference estimate, but on average remained more than 135% away from the reference estimate. The other four corrected estimates (four reports)<sup>38,41,43</sup> became on average 94% closer to the reference estimate, and on average within 2.8% of the reference estimate (figure 7).

One of the ten reports<sup>38</sup> showed that using a locally derived false-positive rate in four equations (McDougal, Hargrove, McWalter/Welte, and a simplified version of McDougal) the median percentage difference between the crude and reference estimates was reduced from 33.7% (4.1% vs 3.1%) to 1.0% (3.1% vs 3.1%; figure 7). However, with an externally-derived false positive rate, the median percentage difference between the crude and reference estimates increased to 100% (0.5% vs 3.1%).

## Discussion

Using a systematic review strategy, we have identified and analysed publications that reported on the accuracy of 13 different assays for recent infection with HIV. We

	Degree of similarity between assay-derived and reference sample sets	Assay-derived sample source	Reference sample source	Crude assay-derived estimate of the incidence of HIV (95% CI)	Reference estimate of the incidence of HIV (95% CI)	Significant difference†
<b>3A11-LS</b>						
Janssen, 1998 <sup>3</sup>	Identical	Cohort, serial samples (San Francisco Men's Health Study)	Cohort	1.5 (NS)	1.4 (NS)	No
Janssen, 1998 <sup>3</sup>	Identical	Cross-sectional (US blood donors 1993–95)	Repeat HIV testing (database)	2.95 (1.1–6.5)	2.6 (1.5–4.2)	No
McFarland, 2000 <sup>37</sup>	Not identical	Convenience sample (San Francisco Drug Treatment Centre)	Repeat HIV testing (database)	0.0 (0.0–1.9)‡	0.0 (0–1.0) ‡	No
<b>BED</b>						
Barnighausen, 2008 <sup>38</sup>	Identical	Cohort, serial samples (KwaZulu-Natal longitudinal HIV surveillance)	Cohort	4.1§ (NS)	3.1§ (2.7–3.5)	..¶
Hall, 2008 <sup>39</sup>	Not identical	Case-based surveillance (US HIV surveillance)	Mathematical techniques	22.8   (19.5–26.1)	22.4   (20.2–24.6)	No
Hargrove, 2008 <sup>38</sup>	Not identical	Cohort, enrolment samples (ZVITAMBO cohort)	Cohort	7.6 (6.7–8.5)	3.5 (2.9–4.2)	Yes
Hargrove, 2008 <sup>38</sup>	Identical	Cohort, serial samples (ZVITAMBO cohort)	Cohort	3.4 (3.0–3.8)	3.5 (2.9–4.2)	No
Hu, 2003 <sup>32</sup>	Identical	Cohort, serial samples (AIDS VAX B/E Vaccine Trial [VAX003])	Cohort	17.3 (12.8–24.2)	9.0 (6.7–11.9)	Yes
Karita, 2007 <sup>26</sup>	Not identical	Cross-sectional (Makaka Uganda Study)	Cohort (incidence from year after)	6.1 (4.2–8.0)	1.3 (0.8–2.0)	Yes
Karita, 2007 <sup>26</sup>	Not identical	Cross-sectional (Makaka Uganda Study)	Cohort (incidence from year before)	6.1 (4.2–8.0)	1.7 (1.3–2.2)	Yes
Kim, 2007 <sup>40</sup>	Not identical	Cross-sectional (Kenya 2003 DHS survey)	Mathematical techniques	3.5 (2.7–4.3)	0.6 (0.5–0.7)	Yes
Kim, 2007 <sup>40</sup>	Not identical	Cross-sectional (Côte d'Ivoire ANC surveys)	Mathematical techniques	3.8 (3.3–4.5)	1.0 (0.6–1.5)	Yes
McDougal, 2006 <sup>39</sup>	Identical	Cohort, serial samples (AIDS VAX B/B Vaccine Trial [VAX004])	Cohort	2.9 (2.3–3.5)	3.1 (2.6–3.6)	No
Mermin, 2008 <sup>41</sup>	Not identical	Cross-sectional (Uganda HIV/AIDS Serobehavioural Survey 2005)	Cohort	2.6 (NS)	1.7 (NS)	..¶
Mermin, 2008 <sup>41</sup>	Not identical	Cross-sectional (Uganda HIV/AIDS Serobehavioural Survey 2005)	Cohort	2.6 (NS)	1.7 (NS)	..¶
Parekh, 2005 <sup>42</sup>	Not identical	Cross-sectional (Survey in Wonji, Ethiopia)	Cohort	0.5 (NS)	0.4 (NS)	..¶
Parekh, 2002 <sup>33</sup>	Identical	Combination (Bangkok Metropolitan Administration [BMA] Study)**	Combination	46.3†† (NS)	50.0†† (NS)	..¶
Parekh, 2002 <sup>33</sup>	Identical	NS (Kenyan seroconverters)	Cohort	57.1†† (NS)	53.1†† (NS)	..¶
Rehle, 2006 <sup>42</sup>	Not identical	Cross-sectional (South Africa National HIV Survey 2005)	Mathematical techniques	2.7 (1.3–4.1)	1.3 (NS)	No
Scheer, 2009 <sup>44</sup>	Not identical	Case-based surveillance (San Francisco HIV Surveillance)	Mathematical techniques (Delphi)	125.7   (88.4–162.9)	127.9   (107.7–145.4)	No
Thai-US BED Validation working group, 2007	Identical	Cohort, serial samples (Bangkok Metropolitan Administration Study)	Cohort	5.2 (3.5–6.9)	6.5 (5.4–7.6)	Yes
Thai-US BED Validation working group, 2007	Identical	Cohort, serial samples (Royal Thai Army Conscripts Study)	Cohort	1.1 (0.3–1.9)	1.2 (0.6–1.8)	No
Xiao, 2007 <sup>28</sup>	Not identical	Cross-sectional (China IDU Study)	Cohort	8.2 (5.9–11.5)	8.8 (6.4–12.0)	No

(Continued on next page)

found that the assays were generally sensitive and specific for identification of recent infection, and were able to provide accurate estimates of population incidence of HIV, but this conclusion is highly qualified by the apparent absence of a standardised approach to assay evaluation.

The process of developing new diagnostic tests generally involves comparisons of test results against those of a designated alternative, if one is available.<sup>53</sup> A test must be rigorously assessed in a range of settings, and shown to have high levels of both sensitivity and specificity, so that pathologists and clinicians can have confidence in findings that will ultimately lead to

decisions about the care of individual patients.<sup>53</sup> Although some authors have pointed out that serological tests for recent infection with HIV can provide some guidance in patient management,<sup>54</sup> they are generally not seen as having a clinical role, so it might be perceived that their assessment can in some sense be less exacting. In fact, these tests are increasingly being called upon to provide the basis for decision making related to major investments in programmes to prevent HIV,<sup>13,15,16,55,56</sup> which will in turn have the potential to affect many people's lives and their health. It is therefore important to have a good understanding of how well these tests have been assessed, and how precise they might be.

	Degree of similarity between assay-derived and reference sample sets	Assay-derived sample source	Reference sample source	Crude assay-derived estimate of the incidence of HIV (95% CI)	Reference estimate of the incidence of HIV (95% CI)	Significant difference†
(Continued from previous page)						
<b>IDE-V3</b>						
Pillonel, 2006 <sup>45</sup>	Not identical	Cross-sectional (French blood donors)	Repeat HIV testing (database)	1.9‡	1.6‡	No
Vironostika-LS						
Busch, 2005 <sup>46</sup>	Not identical	Cross-sectional (US blood donors 2000–01)	Repeat HIV testing (database)	1.8 (1.2–2.3)	2.2 (1.2–3.1)	No
De Jarlais, 2005 <sup>47</sup>	Not identical	Cross-sectional (Beth Israel Detox Study 1990–95)	Cohort (needle syringe programme and methadone users 1992–96)	3.1 (1.9–4.3)	1.4 (0.2–4.6)	No
De Jarlais, 2005 <sup>47</sup>	Not identical	Cross-sectional (Beth Israel Detox Study 1990–95)	Cohort (methadone users [no needle syringe programme] 1992–96)	3.1 (1.9–4.3)	5.3 (2.4–11.5)	No
De Jarlais, 2005 <sup>47</sup>	Not identical	Cross-sectional (Beth Israel Detox Study 1990–95)	Cohort (needle syringe programme users 1992–94)	3.1 (1.9–4.3)	1.6 (0.5–4.7)	No
De Jarlais, 2005 <sup>47</sup>	Not identical	Cross-sectional (Beth Israel Detox Study 1996–2002)	Cohort (users of illicit injected drugs receiving detox 1995–97)	0.9 (0.4–1.3)	1.4 (0.7–2.5)	No
De Jarlais, 2005 <sup>47</sup>	Not identical	Cross-sectional (Beth Israel Detox Study 1996–2002)	Cohort (Young users of illicit injected drugs 1995–97)	0.9 (0.4–1.3)	0.4 (0.0–2.5)	No
Dukers, 2007 <sup>48</sup>	Not identical	Cross-sectional (Amsterdam STI Clinic)	Cohort	2.8‡	1.2‡	No
Kellogg, 2005 <sup>49</sup>	Not identical	Cross sectional (San Francisco anonymous testing)	Repeat HIV testing (database)	1.3 (0.9–1.8)	1.2 (0.8–1.6)	No
Kellogg, 2005 <sup>49</sup>	Not identical	Cross sectional (San Francisco anonymous testing)	Repeat HIV testing (self report)	1.3 (0.93–1.8)	1.0 (1.0–1.1)	No
Vignoles, 2006 <sup>39</sup>	Not identical	Cross-sectional (Survey in MSM, Buenos Aires)	Cohort	6.7 (3.7–9.7)	6 (3.1–11.0)	No

NS=not specified. DBS=dried blood spot, LS=Less sensitive. \*Findings from Truong, 2007, were excluded as no crude assay-derived incidence estimates were reported. †Between assay and reference incidence, based on non-overlapping 95% CIs. ‡One-sided CI. §Per 100 people per year. ¶Not determined as CIs could not be calculated. ||Per 100 000 population. \*\*518 samples from BMA study and 104 samples from US-cases resulting from accidental exposure or use of injected drugs. ††Percent recent, rather than incidence estimate. ‡‡Overall estimate derived from the average of time series results. In each time period the difference between assay-derived and reference incidence was not significant.

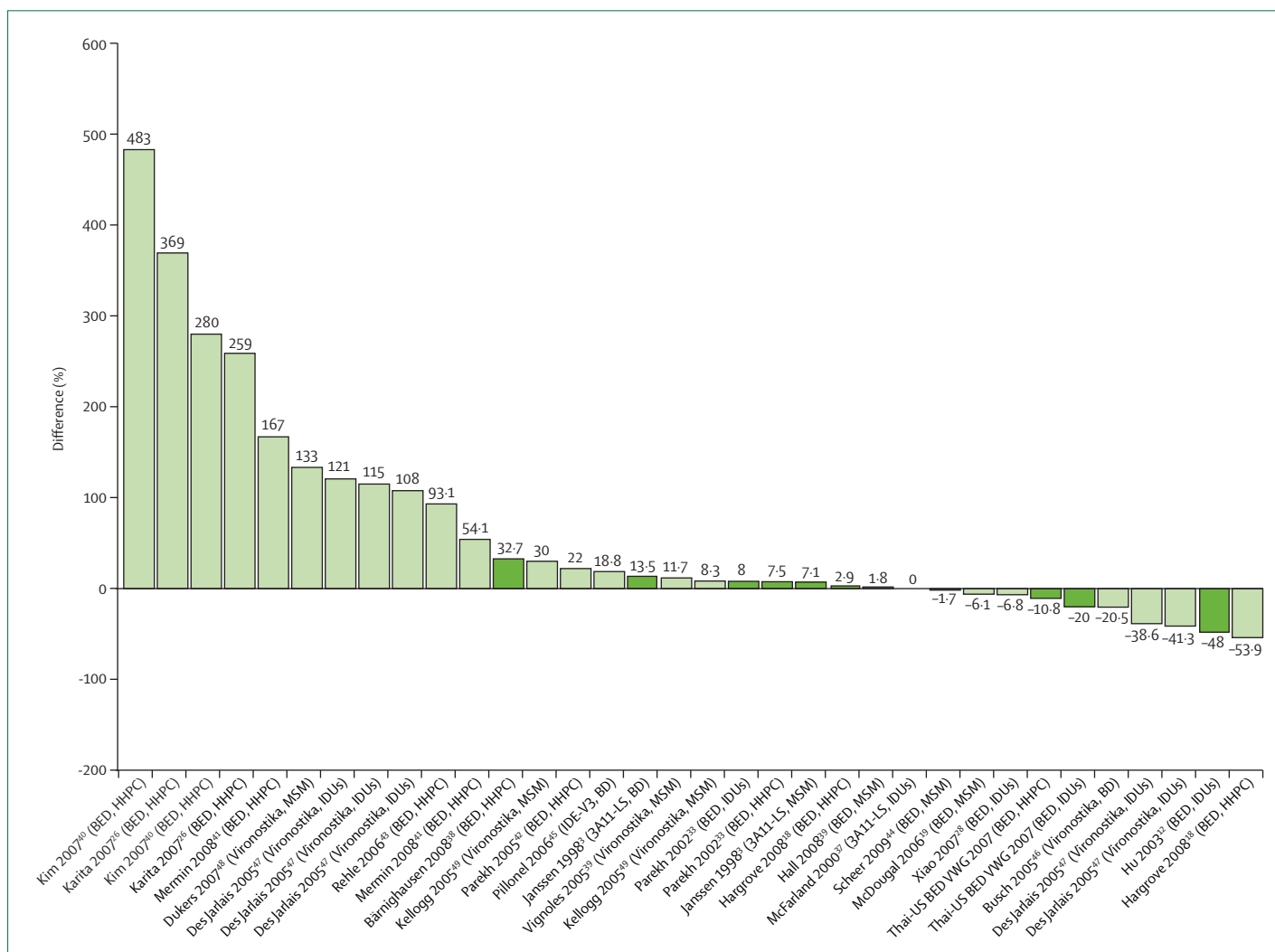
**Table 2: Assay-derived and reference incidence of HIV, overall estimates\***

When tests for recent infection with HIV first became available, they were seen as a major public health breakthrough that could make expensive tracking of cohorts unnecessary for estimating population incidence of HIV—particularly since it is uncertain how representative these cohorts are. The development of assays to establish recent infection with HIV meant identifying an immunological signal that was weak or non-existent very early in infection, and then strengthened over time so that it could eventually be detected in all people infected with HIV. Ideally, the signal should become detectable within an interval, the window period, that was not too close to the time of initial infection and not too far removed from it, and it should evolve in a manner that was quite similar across viral genotypes and population subgroups. A general consensus emerged from the publications on the topic that the ideal evolution time was around 6 months;<sup>2</sup> much shorter and the test would only detect very recent infections, and much longer and the idea of how recent the infection is becomes less relevant from a public health perspective, as well as potentially introducing greater variability in the immunological marker that is the basis for the assay. Furthermore, most immunological markers that were known to evolve in the early phases of infection with HIV had largely stabilised by 6 months, or soon thereafter. By

varying the optical density cut-off at which tests were deemed to be positive or negative, developers had some ability to influence the average window period of an assay.<sup>2</sup>

The accuracy of tests for recent infection with HIV can be compromised in several ways. First, some people with recent infection might have more rapidly evolving immune responses, such that their result on the assay becomes positive before the average window period for the assay has passed.<sup>2</sup> This rapid immune response would result in less than perfect sensitivity, because these cases would not be detected as recent. A second group consists of people with more slowly evolving immune responses who are found to be negative when their infection had been present for longer than the average window period.<sup>2</sup> These cases would be falsely detected as recent, thereby resulting in apparently lowered specificity. However, if the timing of specimen collection is independent of the timing of infection in a population, and the statistical distribution of the immunological marker's evolution times is symmetrical about the average window period, these two types of error should cancel each other out when incidence estimates are derived from the assay results.

Of substantially more concern has been the discovery that in some people the markers of immune response



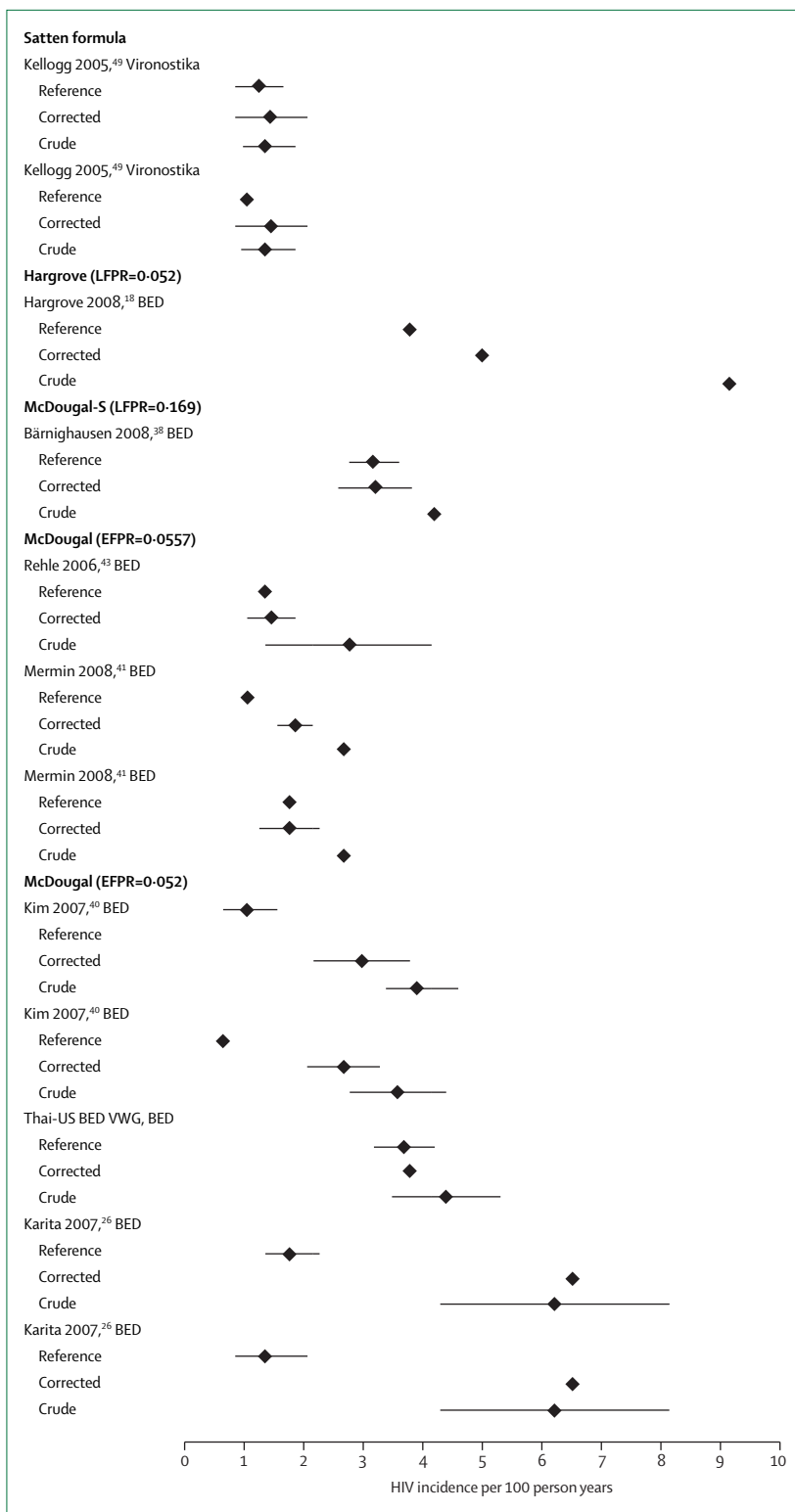
**Figure 6: Difference between crude assay-derived estimates and reference overall estimates of incidence of HIV**

Where sample set specimens used for assay-derived incidence and specimens used to calculate the reference incidence estimate were identical they are shown in dark green. Where the sample set specimens used for assay-derived incidence and specimens used to calculate the reference incidence estimate were non-identical they are shown in light green. Overall estimates for Pillonel, 2006, and Dukers, 2007, derived from the average of time series results. In each time period the difference between assay-derived and reference incidence was not significant. Findings from Truong, 2007, were excluded because no crude assay-derived incidence estimates were reported. BD=blood donors (risk for acquisition of HIV not reported). IDUs=injecting drug users. NS=not specified. MSM=men who have sex with men. HHPC=heterosexuals from high HIV prevalence countries (sub-Saharan Africa and Thailand). VWG=validation working group.

being used as the basis for the assay for recent infection might actually disappear later in the course of infection, either as a result of declining immunological competence or treatment.<sup>32,34</sup> If the tests are then applied in populations with a pool of undiagnosed, longstanding infections, or via unlinked surveys in a setting where treatment is widely used, there is the potential for reduced specificity through the increased number of apparently recent infections.

This issue lies at the heart of the controversy set off by the 2005 report by the Joint United Nations Programme on HIV/AIDS that questioned the accuracy of the BED test on the grounds that it led to overestimation of incidence. In settings of high prevalence, such as southern Africa, where there are substantial numbers of people

with longstanding infection, the incidence estimate derived from an assay for recent infection will be particularly vulnerable to upward bias. Although our Review shows that the assays generally have good characteristics for classifying infections according to how recent they are, even a small loss of specificity can be magnified into a major bias when prevalence of longstanding infection is high. Solutions proposed for this problem have been the exclusion in incidence estimation of people known to have longstanding infection<sup>11</sup> and the application of correction factors to adjust for this bias.<sup>18-20</sup> Such factors should ideally be locally derived as shown by Bärnighausen and colleagues<sup>38</sup> because the proportion of longstanding infections in a population might vary according to the stage of an epidemic.



**Figure 7: Crude assay-derived, corrected, and reference overall estimates of incidence of HIV**  
 The overall estimate of Hargrove, 2008, was derived from the average of time series results. LFPR=locally derived false positive rate. EFPR=externally derived false positive rate. Crude=crude assay-derived incidence. Corrected=corrected assay-derived incidence.

The reports considered in the performance characteristics section of this Review provided some insight into these forms of error. The median sensitivity of the assays to detect infections that were known to be within the window period of the assay was encouragingly high and quite consistent across assays. It was more difficult to assess the specificity in a systematic manner, because the various reports did not all provide comparable information on the duration of infections, and few included cases with documented duration of infection greater than 1 year. For those reports that separated cases of longer duration of infection, there was an average false recent rate of 2% for longstanding infection of 1 year or more duration, increasing to 8.4% for AIDS cases, as reflected by the specificity.

Similarly, antiretroviral therapy also led to reduced specificity of assay results, with two studies showing the false recent rate was more pronounced with increasing duration of treatment—9% among patients on treatment for a year, increasing to 24.2% after 2 years of treatment as reflected by the specificity.<sup>32,34</sup> Barin and colleagues<sup>5</sup> included some additional information about the effect of treatment started at the time of early diagnosis on the performance of the assays. These findings were not included in our Review because they were based on probability calculations but showed that after 180 days of treatment started at early diagnosis of HIV the false recent rate was 76.5%, compared with 0% in patients who did not start treatment early, suggesting that HIV seroconversion is delayed when highly active antiretroviral therapy is started during the acute phase of primary infection.<sup>5</sup>

Although the issue of false recent results from longstanding infections was first raised in the context of the BED assay,<sup>17</sup> it is clearly a broader problem. As shown by similar findings for other assays. Nevertheless, it is also apparent that the problem has the potential to be solved, given reliable information on the false recent rates in populations and the use of correction formulas.

Reports that provided information on the validation of assay-derived incidence estimates also show a high level of accuracy, but caution is needed in interpreting these findings because of the nature of the comparisons being made. It will never be possible to directly assess the accuracy of assay-derived incidence in a true population setting, because there is no practical way to measure the real incidence of infection with HIV from some other source. The published comparisons therefore rely either on simplified situations in which incidence can be directly measured, such as cohorts, which are unrealistic because they do not include the longstanding infections with the potential for false recency; or on estimates of population incidence that have major, inherent methodological weaknesses, such as clinical databases or people repeatedly tested for HIV, or mathematical techniques. There are also likely to be corresponding issues related to the groups of people known to be

recently infected that are used to assess the performance characteristics of the assays. It was often not properly documented how recent infections were established, with some variability between studies.

On this basis, it was perhaps unsurprising that the cohort-derived comparisons showed very close agreement between assay-derived and directly measured incidence. However, it was also impressive that comparisons on the basis of indirect measures of incidence were also very close, with a few exceptions. The studies showing the greatest seeming disagreement between assay-derived and comparison estimates of incidence involved the use of non-identical samples for the two estimates, and were done in populations for which there was likely to be a large pool of longstanding infections, due to high background prevalence of HIV. In interpreting these differences, it is difficult to attribute the extent to which the disagreement is due to one or the other of these two factors.

On the basis of this Review, it would seem that assays for recent infection with HIV can be used to estimate the incidence of HIV, provided attention is paid to longstanding infections in the population. However, the available reports on which this conclusion is based reflect a narrow range of specimen sources, with the majority being done in the USA, and involving subtype B virus. Many publications did not provide detailed information on various parts of their methods or results, so it was difficult to comprehensively assess their conclusions. The comparisons were also based on a small number of samples and incident cases. As a result, the CIs for many of the sensitivity, specificity, and incidence estimates were wide and the probability of detecting statistically meaningful differences was correspondingly reduced.

Further development of these assays remains a public health priority, to ensure that incidence of new infections with HIV can be readily measured for programmatic and research purposes. However, there is at present no comprehensive framework for the development and evaluation of incidence assays. Neither has there been substantial financial backing for work of this kind, whether from public or commercial sources. It is therefore perhaps unsurprising to see such a wide variety of approaches being used for assay validation, and a great unevenness in the way in which they have been reported. The varying approaches used for performance characteristic assessment so far show the need for assay developers to be able to gain access to well-characterised specimen banks from an appropriate range of HIV subtypes and populations groups, so that they can assess performance characteristics before an assay is applied in the field. A standardised approach to assay development would give researchers and funding agencies a clear objective to aim for, whether they are designing new assays, improving existing ones, or combining two or more incidence assays in sequence to improve accuracy.<sup>8,57</sup>

In developing a framework for assay assessment, it will be important to balance the need for accuracy with a

strong dose of pragmatism. The specimens that are available and relevant for the assessment of incidence assays are finite in number, and most are in the hands of investigators who face many demands for this precious material. By setting the bar for assay development too high, there is a risk that assay development will be thwarted rather than encouraged. A starting point for standardisation would be an agreement on guidelines for reporting, ensuring that methods and results can be evaluated in a comprehensive manner.

Reduction in incidence of HIV is one of the world's major public health objectives, and the target of billions of dollars in government and international expenditures. The ability to accurately monitor the incidence of HIV would seem to be an absolute necessity if such programmes are to be delivered and evaluated in an optimum manner.

#### Contributors

RG and JMK were jointly responsible for the design of the study and how it was done, as well as the writing of the paper. RG did the search of published work, data collection, and data analysis. JG contributed to the design, search of published work, data collection, and data analysis. JG also developed the database for the study. The other authors assisted with the search of published work and contributed to the interpretation of the findings.

#### WHO Working Group on HIV Incidence Assays

Chris Archibald, Andre Charlett, Niel Constantine, Elizabeth Dax, Stephane Le Vu, Gary Murphy, John Parry, Josiane Pillonel, Renee Ridzon, Connie Sexton, Barbara Suligoi, Bernard Branson, Eleanor Gows, Anita Sands, Gaby Vercautern, Alex Welte.

#### Conflicts of interest

BP is one of three inventors of the BED assay. BED assay was licensed and commercialised through CDC's technology transfer office. Most of the royalty for BED goes to CDC and the laboratory. A portion is distributed among the three inventors.

#### Acknowledgments

We acknowledge the assistance of Rebecca Jenkinson (Burnet Institute, Melbourne, Australia). We thank all authors who supplied additional unpublished information for inclusion in this Review. The National Centre in HIV Epidemiology and Clinical Research is funded by the Australian Government Department of Health and Ageing and is affiliated with the Faculty of Medicine, University of New South Wales.

#### References

- Rutherford GW, Schwarcz SK, McFarland W. Surveillance for incident HIV infection: new technology and new opportunities. *J Acquir Immune Defic Syndr* 2000; **25** (suppl 2): S115–19.
- Murphy G, Parry JV. Assays for the detection of recent infections with human immunodeficiency virus type 1. *Euro Surveill* 2008; **13**: pii 18966.
- 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.
- Dobbs T, Kennedy S, Pau CP, McDougal JS, Parekh BS. Performance characteristics of the immunoglobulin G-capture BED-enzyme immunoassay, an assay to detect recent human immunodeficiency virus type 1 seroconversion. *J Clin Microbiol* 2004; **42**: 2623–28.
- Barin F, Meyer L, Lancar R, et al. Development and validation of an immunoassay for identification of recent human immunodeficiency virus type 1 infections and its use on dried serum spots. *J Clin Microbiol* 2005; **43**: 4441–47.
- Suligoi B, Massi M, Galli C, et al. Identifying recent HIV infections using the avidity index and an automated enzyme immunoassay. *J Acquir Immune Defic Syndr* 2003; **32**: 424–28.
- Wilson KM, Johnson EI, Croom HA, et al. Incidence immunoassay for distinguishing recent from established HIV-1 infection in therapy-naïve populations. *AIDS* 2004; **18**: 2253–59.

- 8 Constantine NT, Sill AM, Jack N, et al. Improved classification of recent HIV-1 infection by employing a two-stage sensitive/less-sensitive test strategy. *J Acquir Immune Defic Syndr* 2003; **32**: 94–103.
- 9 Center for Epidemiological Studies on HIV/AIDS of Catalonia. SIVES 2004, Integrated HIV/STI Surveillance System of Catalonia (SIVES) Annual Report. Barcelona: Generalitat de Catalunya, Departament de Salut, 2005.
- 10 CDC. HIV incidence among young men who have sex with men—seven US cities, 1994–2000. *MMWR Morb Mortal Wkly Rep* 2001; **50**: 440–44.
- 11 Guy RJ, Breschkin AM, Keenan CM, Catton MG, Enriquez AM, Hellard ME. Improving HIV surveillance in Victoria: the role of the “detuned” enzyme immunoassay. *J Acquir Immune Defic Syndr* 2005; **38**: 495–99.
- 12 Hu DJ, Vanichseni S, Mock PA, et al. 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* 2003; **19**: 727–30.
- 13 Murphy G, Charlett A, Jordan LF, Osner N, Gill ON, Parry JV. HIV incidence appears constant in men who have sex with men despite widespread use of effective antiretroviral therapy. *AIDS* 2004; **18**: 265–72.
- 14 Parekh BS, Hu DJ, Vanichseni S, et al. Evaluation of a sensitive/less-sensitive testing algorithm using the 3A11-LS assay for detecting recent HIV seroconversion among individuals with HIV-1 subtype B or E infection in Thailand. *AIDS Res Hum Retroviruses* 2001; **17**: 453–58.
- 15 Saphonn V, Parekh BS, Dobbs T, et al. Trends of HIV-1 seroincidence among HIV-1 sentinel surveillance groups in Cambodia, 1999–2002. *J Acquir Immune Defic Syndr* 2005; **39**: 587–92.
- 16 Hall HI, Song R, Rhodes P, et al. Estimation of HIV incidence in the United States. *JAMA* 2008; **300**: 520–29.
- 17 UNAIDS reference group on estimates, modelling and projections. Statement on the use of the BED-assay for the estimation of HIV-1 incidence for surveillance or epidemic monitoring. [http://www.unaids.org/en/HIV\\_data/Epidemiology/default.asp](http://www.unaids.org/en/HIV_data/Epidemiology/default.asp) (accessed Oct 11, 2009).
- 18 Hargrove JW, Humphrey JH, Mutasa K, et al. Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS* 2008; **22**: 511–18.
- 19 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**: 945–52.
- 20 Rawal BD, Degula A, Lebedeva L, et al. Development of a new less-sensitive enzyme immunoassay for detection of early HIV-1 infection. *J Acquir Immune Defic Syndr* 2003; **33**: 349–55.
- 21 Suligoi B, Butto S, Galli C, et al. Detection of recent HIV infections in African individuals infected by HIV-1 non-B subtypes using HIV antibody avidity. *J Clin Virol* 2008; **41**: 288–92.
- 22 Suligoi B, Bossi V, Regine V, Rodella A, Manca N, Galli. The accuracy of different cutoffs for the avidity index of HIV antibodies to identify recent infections. *AIDS 2006–16th International AIDS Conference*; Toronto, Canada; Aug 13–18, 2006. Abstract 373.
- 23 Martró E, Suligoi B, Gonzalez V, et al. Comparison of the avidity index method and the serologic testing algorithm for recent human immunodeficiency virus (HIV) seroconversion, two methods using a single serum sample for identification of recent HIV infections. *J Clin Microbiol* 2005; **43**: 6197–99.
- 24 Loschen S, Batzing-Feigenbaum J, Poggensee G, et al. Comparison of the human immunodeficiency virus (HIV) type 1-specific immunoglobulin G capture enzyme-linked immunosorbent assay and the avidity index method for identification of recent HIV infections. *J Clin Microbiol* 2008; **46**: 341–45.
- 25 Sakarovitch C, Rouet F, Murphy G, et al. Do tests devised to detect recent HIV-1 infection provide reliable estimates of incidence in Africa? *J Acquir Immune Defic Syndr* 2007; **45**: 115–22.
- 26 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**: 403–08.
- 27 McDonald A, Cunningham P, Kelleher A, Kaldor J. Comparison of two assays for identifying incident infection among cases of newly diagnosed HIV infection. *STARHS Workshop in association with the 16th International Conference on AIDS 2006*; Toronto, Canada; Aug 13–18, 2006.
- 28 Xiao Y, Jiang Y, Feng J, et al. Seroincidence of recent human immunodeficiency virus type 1 infections in China. *Clin Vaccine Immunol* 2007; **14**: 1384–86.
- 29 Hong L, Ketema F, Sill AM, Kreisel KM, Cleghorn FR, Constantine NT. A simple and inexpensive particle agglutination test to distinguish recent from established HIV-1 infection. *Int J Infect Dis* 2007; **11**: 459–65.
- 30 Selleri M, Orchi N, Zaniratti MS, et al. Effective highly active antiretroviral therapy in patients with primary HIV-1 infection prevents the evolution of the avidity of HIV-1-specific antibodies. *J Acquir Immune Defic Syndr* 2007; **46**: 145–50.
- 31 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**: 211–15.
- 32 Hladik W, Olara D, Were W, Mermin J, Downing R. The effect of antiretroviral treatment on the specificity of the serological BED HIV-1 incidence assay. *HIV implementers meeting*; July 16–19, 2007; Kigali, Rwanda. Abstract 998.
- 33 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**: 295–307.
- 34 Murphy G, Pao D, Fisher M, et al. The effect of confounding factors on the specificity of tests employed in the serological testing algorithm for recent HIV seroconversion (STARHS). *STARHS Workshop in association with the 16th International Conference on AIDS 2006*; Toronto, Canada; Aug 13–18, 2006.
- 35 Kothe D, Byers RH, Caudill SP, et al. Performance characteristics of a new less sensitive HIV-1 enzyme immunoassay for use in estimating HIV seroincidence. *J Acquir Immune Defic Syndr* 2003; **33**: 625–34.
- 36 Cimerman S, Sucupira MC, Lewi DS, Diaz RS. Less sensitive HIV-1 enzyme immunoassay as an adjuvant method for monitoring patients receiving antiretroviral therapy. *AIDS Patient Care STDS* 2007; **21**: 100–05.
- 37 McFarland W, Kellogg TA, Louie B, Murrill C, Katz MH. Low estimates of HIV seroconversions among clients of a drug treatment clinic in San Francisco, 1995 to 1998. *J Acquir Immune Defic Syndr* 2000; **23**: 426–29.
- 38 Bärnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell ML. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. *AIDS* 2008; **22**: 139–44.
- 39 Vignoles M, Avila MM, Osimani ML, et al. HIV seroincidence estimates among at-risk populations in Buenos Aires and Montevideo: use of the serologic testing algorithm for recent HIV seroconversion. *J Acquir Immune Defic Syndr* 2006; **42**: 494–500.
- 40 Kim A, McDougal S, Hargrove J, et al. Toward more plausible estimates of HIV-1 incidence in cross-sectional serologic surveys in Africa: application of a HIV-1 incidence assay with post-assay adjustments. *14th Conference on Retroviruses and Opportunistic Infections*; Feb 25–28, 2007; Los Angeles, California. Abstract 950.
- 41 Mermin J, Musinguzi J, Opio A, et al. Risk factors for recent HIV infection in Uganda. *JAMA* 2008; **300**: 540–49.
- 42 Parekh B. Introduction to the HIV-1 BED incidence assay. *Meeting of the UNAIDS/WHO Reference Group on Estimates, Modelling and Projections*; Athens, Greece; Dec 13–15, 2005.
- 43 Rehle T. HIV-1 incidence estimates: South Africa 2005. *STARHS Workshop in association with the 16th International Conference on AIDS 2006*; Toronto, Canada; Aug 13–18, 2006.
- 44 Scheer S, Chin CS, Buckman A, McFarland W. Estimation of HIV incidence in San Francisco. *AIDS* 2009; **23**: 533–34.
- 45 Pillonel J, Barin F, Laperche S, et al. HIV-1 incidence in blood donors in France between 1992 and 2005: Use of an immunoassay to identify recent infections. *STARHS Workshop in association with the 16th International Conference on AIDS 2006*; Toronto, Canada; Aug 13–18, 2006.

- 46 Busch MP, Glynn SA, Stramer SL, et al. A new strategy for estimating risks of transfusion-transmitted viral infections based on rates of detection of recently infected donors. *Transfusion* 2005; **45**: 254–64.
- 47 Des Jarlais DC, Perlis T, Arasteh K, et al. HIV incidence among injection drug users in New York City, 1990 to 2002: use of serologic test algorithm to assess expansion of HIV prevention services. *Am J Public Health* 2005; **95**: 1439–44.
- 48 Dukers NHTM, Fennema HSA, van der Snoek EM, et al. HIV incidence and HIV testing behavior in men who have sex with men: using three incidence sources, the Netherlands, 1984–2005. *AIDS* 2007; **21**: 491–99.
- 49 Kellogg TA, Loeb L, Dilley J, Adler B, Louie BT, McFarland W. Comparison of three methods to measure HIV incidence among persons seeking voluntary, anonymous counseling and testing. *J Acquir Immune Defic Syndr* 2005; **39**: 112–20.
- 50 Truong H-HM, Kellogg T, Louie B, Klausner J, Dilley J, McFarland W. Comparison of HIV incidence estimates derived from laboratory incidence assays and repeat testing data at HIV testing sites in San Francisco. 15th Conference on Retroviruses and Opportunistic Infections; Boston, MA; Feb 3–7, 2008. Abstract 952.
- 51 Welte A, McWalter TA, Barnighausen T. A simplified formula for inferring HIV incidence from cross-sectional surveys using a test for recent infection. *AIDS Res Hum Retroviruses* 2009; **25**: 125–26.
- 52 Satten GA, Janssen R, Busch MP, Datta S. Validating marker-based incidence estimates in repeatedly screened populations. *Biometrics* 1999; **55**: 1224–27.
- 53 Banoo S, Bell D, Bossuyt P, et al. Evaluation of diagnostic tests for infectious diseases: general principles. *Nat Rev Microbiol* 2006; **4**: S20–32.
- 54 Barin F, Nardone A. Monitoring HIV epidemiology using assays for recent infection: where are we? *Euro Surveill* 2008; **13**: pii 18967.
- 55 Semaille C, Barin F, Cazein F, et al. Monitoring the dynamics of the HIV epidemic using assays for recent infection and serotyping among new HIV diagnoses: experience after 2 years in France. *J Infect Dis* 2007; **196**: 377–83.
- 56 Rehle T, Shisana O, Pillay V, Zuma K, Puren A, Parker W. National HIV incidence measures--new insights into the South African epidemic. *S Afr Med J* 2007; **97**: 194–99.
- 57 Adonsu-Hoyi Y, Calder-Kent B, Malloch L, Archibald C, Sandstrom P, Kim J. Examination of avidity in STARHS testing for HIV incidence. STARHS Workshop in association with the 16th International Conference on AIDS 2006; Toronto, Canada; Aug 13–18, 2006.