

# Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository

Reshma Kassanjee<sup>a,b</sup>, Christopher D. Pilcher<sup>c</sup>, Sheila M. Keating<sup>d</sup>, Shelley N. Facente<sup>c</sup>, Elaine Mckinney<sup>e</sup>, Matthew A. Price<sup>c,f</sup>, Jeffrey N. Martin<sup>c</sup>, Susan Little<sup>g</sup>, Frederick M. Hecht<sup>c</sup>, Esper G. Kallas<sup>h</sup>, Alex Welte<sup>a</sup>, Michael P. Busch<sup>d</sup>, Gary Murphy<sup>e</sup>, on behalf of the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA)

**Objective:** Cross-sectional HIV incidence surveillance, using assays that distinguish 'recent' from 'nonrecent' infections, has been hampered by inadequate performance and characterization of incidence assays. In this study, the Consortium for the Evaluation and Performance of HIV Incidence Assays presents results of the first independent evaluation of five incidence assays (BED, limiting antigen avidity, less-sensitive Vitros, Vitros avidity and BioRad avidity).

**Design:** A large repository of diverse specimens from HIV-positive patients was established, multiple assays were run on 2500 selected specimens, and data were analyzed to estimate assay characteristics relevant for incidence surveillance.

**Methods:** The mean duration of recent infection (average time 'recent' while infected for less than some time cut-off  $T$ ) was estimated from longitudinal data on seroconverters by regression. The false-recent rate (FRR, probability of testing 'recent' when infected for longer than  $T$ ) was explored by measuring the proportions of 'recent' results in various subsets of patients.

**Results:** Assays continue to fail to attain the simultaneously large mean duration of recent infection and small FRR demanded by existing performance guidelines. All assays produce high FRRs amongst virally suppressed patients (>40%), including elite controllers and treated patients.

**Conclusions:** Results from this first independent evaluation provide valuable information about the current performance of assays, and suggest the need for further optimization. Variation of recent/nonrecent thresholds and the use of multiple antibody-maturation assays, as well as other biomarkers, can now be explored, using the rich data generated by the Consortium for the Evaluation and Performance of HIV Incidence Assays. Consistently high FRRs amongst those virally suppressed suggest that viral load will be a particularly valuable supplementary marker.

---

<sup>a</sup>The South African DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, <sup>b</sup>School of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa, <sup>c</sup>University of California, San Francisco, <sup>d</sup>Blood Systems Research Institute, San Francisco, California, USA, <sup>e</sup>Public Health England, London, UK, <sup>f</sup>International AIDS Vaccine Initiative (IAVI), Department of Medical Affairs, New York City, New York, <sup>g</sup>University of California San Diego, Department of Medicine, San Diego, California, USA, and <sup>h</sup>Division of Clinical Immunology and Allergy, School of Medicine, University of São Paulo, São Paulo, Brazil.

Correspondence to Reshma Kassanjee, The South African DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa.

E-mail: rkassanjee@sun.ac.za

Received: 11 June 2014; revised: 23 July 2014; accepted: 24 July 2014.

DOI:10.1097/QAD.0000000000000429

**Video abstract:** <http://links.lww.com/QAD/A569>

© 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins

*AIDS* 2014, **28**:000–000

**Keywords:** biomarkers, HIV, incidence assays, incidence estimation, recent infection

## Introduction

Reliable measurement of HIV incidence (the rate of new infections) is essential for monitoring the epidemic, assessing interventions and planning studies. Traditionally, incidence is measured by counting the number of new infections acquired in a cohort of patients followed up over time. However, such longitudinal studies are often costly, time-consuming, and unrepresentative. Therefore, the estimation of incidence from cross-sectional surveys, using ‘incidence assays’ that distinguish ‘recent’ from ‘nonrecent’ infection, has attracted wide interest [1–4].

Cross-sectional surveillance is founded on the heuristic that a high prevalence of ‘recent’ infection indicates a high incidence. However, current incidence assays that provide a reasonably enduring state of ‘recent’ infection also tend to produce substantial ‘false-recent’ results at large times after infection. As the methodology matured, a general theoretical framework was developed that supports the consistent analysis of ‘false-recent’ results [5]. However, there have not been independent assessments of candidate assays, or consensus metrics of an assay’s utility for incidence estimation.

In 2010, the ‘Bill and Melinda Gates Foundation’ supported the establishment of the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA) [6]. Over the past 3 years, CEPHIA has entered into collaborations and material transfer agreements to establish a large repository of precious plasma specimens with sufficient clinical background data. Test developers can apply for access to a small ‘qualification panel’ of specimens, and, if the assay is suitably promising, the assay can be independently applied (by a CEPHIA laboratory) to a much larger ‘evaluation panel’.

In this study, results are presented for the first five assays that have successfully passed through the full evaluation: limiting antigen avidity (LA<sub>g</sub>) [7], BED [8], less-sensitive/detuned Vitros [9], Vitros avidity [9] and BioRad avidity [10]. In principle, a test for recent infection can be arbitrarily complex in design, and can be optimized by tuning numerous parameters. The present evaluation is of tests for recent infection which are based on ‘single-incidence assays’, applied according to the developers’ test conditions and interpretive guidelines. Test optimization, by the application of alternative

thresholds in the interpretation of results, and using the assays in combination with one another or with supplemental markers (such as viral load), is ongoing.

Translating survey counts (of HIV-negative, ‘recently’ HIV-positive and ‘nonrecently’ HIV-positive patients) into incidence estimates [5] requires knowledge of two test properties:

- (1) The mean duration of recent infection (MDRI) – the average time spent alive and ‘recently’ infected, while infected for less than some time cut-off, denoted by  $T$ .
- (2) The false-recent rate (FRR) – the probability that a randomly chosen patient, infected for longer than  $T$ , will produce a ‘recent’ result.

A ‘target product profile’ (TPP) for tests for recent infection has been developed and attracted some attention [3,4,11], and it provides a number of objectives that incidence assays should meet to be of utility for incidence estimation. To achieve usefully precise incidence estimates, in real-world household surveys in high-incidence settings, an incidence assay should have a sufficiently enduring MDRI (on the order of 1 year) and small FRR (clearly less than 2%, and ideally zero). Furthermore, for feasible widespread use of the assay, results should be highly reproducible, and the training, equipment and sample type requirements should be modest.

In this study, we evaluate each assay’s MDRI and FRR. As the performance of incidence assays may vary across subpopulations, the characteristics of the incidence assays in various specimen sets are also explored.

## Methods

### The CEPHIA specimen repository and the evaluation panel

The CEPHIA repository is housed at the ‘Blood System Research Institute’ (San Francisco, California, USA) and currently consists of more than 5000 plasma specimens obtained from over 1200 patients. The specimens were obtained through collaborations with blood banks, and clinical research studies enrolling and following patients over time: American Red Cross [12]; Blood Centers of the Pacific [13]; South African National Blood Service [14];

Hemocentro do São Paulo [15]; the University of California, San Francisco, Options study [16]; San Francisco Men's Health Study [17]; the San Diego Primary Infection Cohort [18]; the multicenter 'AMPLIAR' cohort [19]; the multicenter International AIDS Vaccine Initiative (IAVI) African Early Infection Cohort (Protocol C) [20]; and the University of California, San Francisco, SCOPE study [21].

Two 'panels' of specimens were created for the present purpose: a 250-member 'qualification panel' for preliminary assessments (see [22] for results), and a 2500-member 'evaluation panel' for the full assessments of assays showing suitable promise, which forms the basis of this study.

The evaluation panel specimens were drawn from 928 patients, and 60% of the patients contributed multiple specimens over time (2–13 specimens, median of 3 specimens per patient). Follow-up after infection ranged from 1 week to more than 10 years, with a median follow-up of 3 years (for patients with estimable infection dates, discussed below).

### Laboratory procedures and interpretation of assay results

Each of the five assays measures an aspect of an individual's immune response, with measurements below some threshold interpreted as indicative of 'recent' infection.

BED [8,23] and LAg [7,24] (Sedia Biosciences Corporation, Portland, Oregon, USA) were developed specifically as incidence assays by the Centers for Disease Control and Prevention (CDC). The IgG capture BED enzyme immunoassay (EIA) measures the proportion of immunoglobulin G (IgG) that is specific to HIV, with a normalized optical density (ODn) below 0.8, indicating 'recent' infection. The single-well LAg EIA is responsive to the avidity of HIV-1-specific IgG, as it presents marginally low concentrations of a multisubtype recombinant HIV-1 antigen, typically affording just a single binding site to the multivalent IgG or IgM antibodies. Whereas a recent/nonrecent threshold of 1.0 ODn was initially proposed, this was recently revised to 1.5 [24,25], following a review of the assay in which CEPHIA participated.

Both less-sensitive Vitros (LS-Vitros) and Vitros avidity [9] are based on the VITROS ECi/ECiQ Immunodiagnostic System – a chemiluminescence assay that gives a quantitative measure of HIV antibodies (Ortho-Clinical Diagnostics, Inc., Rochester, New York, USA). For LS-Vitros, a reported signal-to-cut-off (S/C) below 20, for a diluted specimen, is interpreted as a 'recent' result. For Vitros avidity, the ratio of the S/C in an aliquot treated with a chaotropic agent (guanidine) to that in an aliquot not thus treated yields an avidity index, and a recent/

nonrecent threshold of 60% is used to classify the infection.

The BioRad avidity test [10] is based on a modification of the Genetic Systems HIV-1/HIV-2 + O EIA (Bio-Rad Laboratories, Inc., Hercules, California, USA), which involves the testing of each specimen in the presence and absence of a chaotropic agent (diethylamine). The ratio of the reactivity of the treated to untreated aliquot produces an avidity index, with values below 40%, indicating 'recent' infection.

All assays were applied according to developers' standard operating procedures and package inserts [7–9,23,24], and protocols are available on the CEPHIA project website [6]. Testing was performed independently in CEPHIA laboratories, by technicians trained by the test developers and blinded to specimen background information. Three large-volume 'control' specimens (obtained from blood donations, and chosen to represent a range of serological responses) were supplied to laboratory technicians with each panel, for regular confirmation of reproducibility and stability of assays.

### Data analysis

All data captured within CEPHIA are stored in a (MySQL) relational database. Database queries linked assay results to the background information on patients and specimens for data analysis (performed in Matlab R2013b, the MathWorks Inc.).

Test properties were evaluated in specimen sets defined by stratifying on treatment history, viral load, CD4<sup>+</sup> cell count, time from infection to specimen draw, and HIV subtype (based on country, for the 48% of specimens which lack explicit laboratory subtype confirmation). The performance of assays in 'elite controllers', broadly defined as patients who maintain undetectable or very low HIV viral loads without antiretroviral therapy (ART), is of particular interest. As the SCOPE study purposefully recruited elite controllers, these data were analyzed separately. These patients were ART-naïve (or without ART for at least 6 months), with all off-treatment viral load measurements (HIV-1 RNA) below 200 copies/ml and at least 50% of these measurements below 75 copies/ml.

The definitions of the MDRI and FRR rely on the previously mentioned construct of a postinfection time cut-off  $T$  [5]. If  $T$  is chosen to be too short, this limits the possible MDRI and typically raises the FRR. If  $T$  is chosen to be too long, it becomes difficult to obtain sufficient data to characterize the test with sufficient precision over this time after infection, and the MDRI will also develop variation by time and place (properties inevitable for the FRR) rather than capture stable biological properties of the test. A cut-off value of  $T$  equal to 2 years is used throughout this study.

In practice, the notion of ‘infection’ implicit in the test property definitions refers to ‘detectable infection’, which depends on the particular HIV diagnostic test used in the incidence study. In this analysis, ‘detectable infection’ was defined as the time of seroconversion on an HIV viral lysate-based Western blot assay. On the basis of a methodology described by the authors elsewhere (manuscript in preparation, by CEPHIA), infection dates were estimated for the 56% of patients who had recorded dates of last HIV-negative and first HIV-positive tests (not more than 120 days apart) and descriptions of the diagnostic assays used. Average durations of Fiebig stages [26,27] were used to estimate times at which patients seroconverted (corresponding to entering Fiebig stage 5). Patients with unambiguous acute retroviral syndrome (ARS) symptom onset dates [28–31] between their last HIV-negative and first HIV-positive test dates were estimated to seroconvert 17 days after ARS onset (on the basis of the observation that the incubation period of ARS symptoms is about 14 days [32–35], and that the time from exposure to Western blot seroconversion averages 31 days [26,27]).

A number of methods can reasonably be used to estimate the MDRI, each with its own accuracy, precision and complexity – as explored in a separate, detailed benchmarking exercise (manuscript in preparation, by a working group operating on behalf of the ‘HIV Modelling Consortium’ [36]). In this analysis, binomial regression, an approach found to be robust across a number of scenarios in this benchmarking project, and previously used for this purpose [37], has been applied. The model form is  $g[P_R(t)] = f(t)$ , where  $P_R(t)$  is the probability of testing ‘recent’ at time  $t$  after infection,  $g$  is the chosen link function and  $f(t)$  contains the model parameters, which are estimated by a maximum likelihood approach. Results from a 4-parameter model form are presented, where  $g$  is the logit link, and  $f(t)$  is a cubic polynomial in  $t$  (model A). Data points more than  $1.1 \times T$  after infection were discarded before model fitting (data exclusion rule I), with the aim of achieving the best fit of the model over  $[0, T]$  after infection, while avoiding diluting the data around the boundary at  $T$ . Sensitivity of results when increasing the data exclusion cut-off to  $2 \times T$  (data exclusion rule II) was also considered. Variation in results was explored when fitting two other model forms, namely a more restrictive 2-parameter model, where  $g$  is the log-log link and  $f(t)$  is a linear function of  $\ln(t)$  (model B); and a flexible 7-parameter model, where  $g$  is the logit link and  $f(t)$  is a linear function of the natural cubic spline basis functions with interior knots occurring every 3 months after infection, between 0 and  $T$  after infection (model C). In all cases, the MDRI, expressed mathematically as  $\int_0^T TPR(T) dt$ , was estimated using the fitted  $P_R(t) = g^{-1}[f(t)]$ .

To correctly account for the structure of the data, in the absence of explicit patient-level clustering in the fitted

models, bootstrapping was performed by the sampling patients (not observations) with replacement. The 2.5th and 97.5th percentiles of 10 000 MDRI estimate replicates provided 95% confidence interval (CI) limits [38].

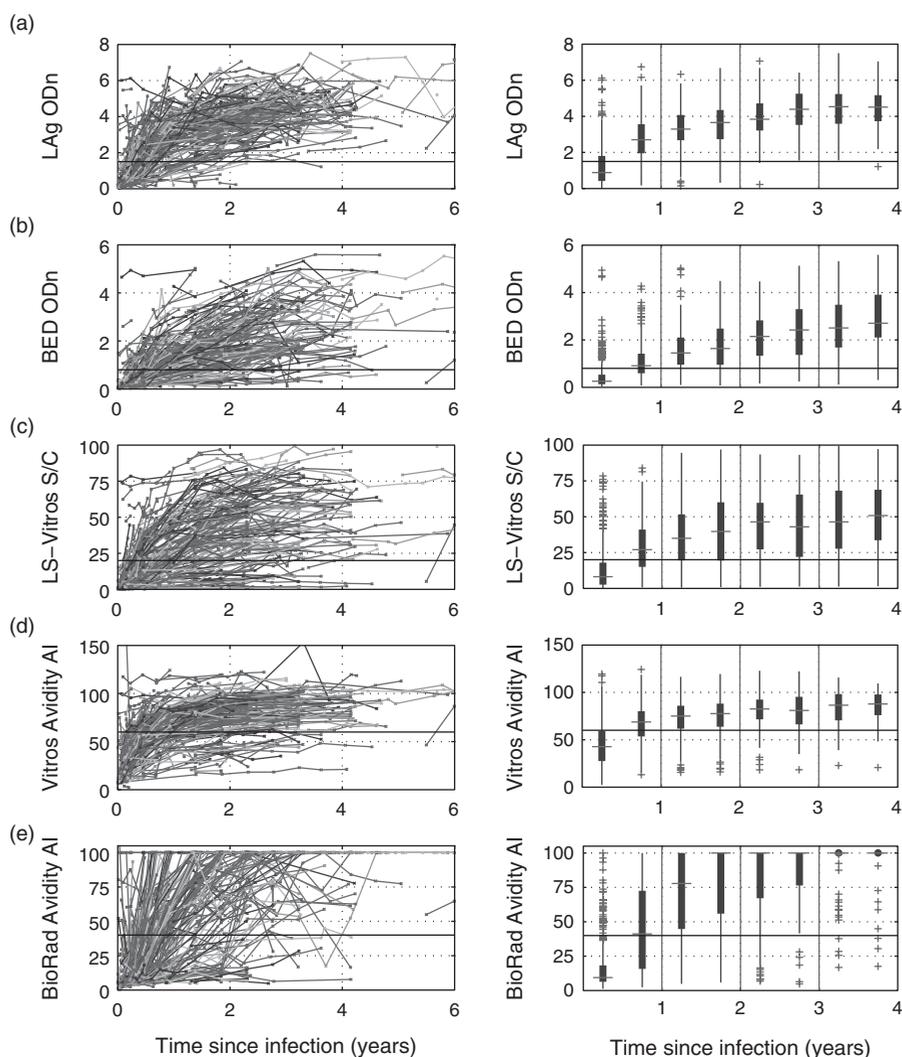
A population-level FRR is inherently dependent on the epidemiological and demographic history of a study population [5], and so a set of specimens, such as in the CEPHIA repository, can only be used to estimate the FRR in well defined subpopulations. Therefore, specimens from long-infected patients were identified (specimens drawn at least  $T$  after the patient’s first recorded HIV-positive visit), and the proportion of ‘recently’ infected patients estimated in each of the specimen sets described above. To capture patient-level clustering, when a patient provided more than one result to any FRR estimate, the most frequent classification was used. Exact Clopper–Pearson 95% CIs [39] are provided.

## Results

The incidence assay dynamics, excluding specimens from treated patients and SCOPE elite controllers, are shown in Figs 1–3. The evolution of assay readings by time since infection is shown in Fig. 1. The distribution of results for specimens drawn more than  $T = 2$  years after infection is shown in Fig. 2. In Fig. 3, the proportion of ‘recent’ results (assay measurements below the recent/nonrecent threshold) is plotted by time since infection, also stratified by HIV subtype (A1, B, C and D). Note that there is natural variability in biomarker maturation, leading to a significant number of patients reaching the standard recent/nonrecent threshold more than 1, but less than 2 years postinfection; and there is significant delay or failure to achieve maturation to ‘nonrecent’ status among specimens of subtypes A1 and D.

Table 1 provides estimated test properties for the various specimen sets. LAg has an estimated MDRI of 188 days (95% CI 165–211), whereas the remaining assays have MDRI estimates of 285–333 days (CIs spanning 254–363 days). Results were insensitive (less than a 2% change in results) to whether ARS symptoms onset dates were used to adjust estimated infection dates, a change to data exclusion rule II, and the use of alternative model C. MDRI estimates increased by 2–4% when changing to model B, which was the most sensitive to the data exclusion rules (4–10% increase in estimates when changing to data exclusion rule II).

Excluding treated patients and SCOPE elite controllers, and analyzing all remaining specimens drawn more than  $T = 2$  years after infection, the measured FRR ranges from 1% (95% CI 0.3–3%) for LAg, to 6–10% (95% CIs spanning 3–14%) for the remaining assays.



**Fig. 1. Incidence assay measurements as a function of time since infection, excluding treated patients and elite controllers (1376 data points from 418 patients), for (a) LAg, (b) BED, (c) LS-Vitros, (d) Vitros avidity and (e) BioRad avidity.** A spaghetti plot (left) shows patients' trajectories, and box-and-whisker plots (right) show percentiles of measurements in 6-monthly intervals after infection (central 50% and median of measurements captured by box and dividing line, whiskers and '+' symbols capture remaining measurements and outliers, respectively; 40–450 data points per group). Recent/nonrecent thresholds are shown by horizontal solid lines. LAg, limiting antigen avidity.

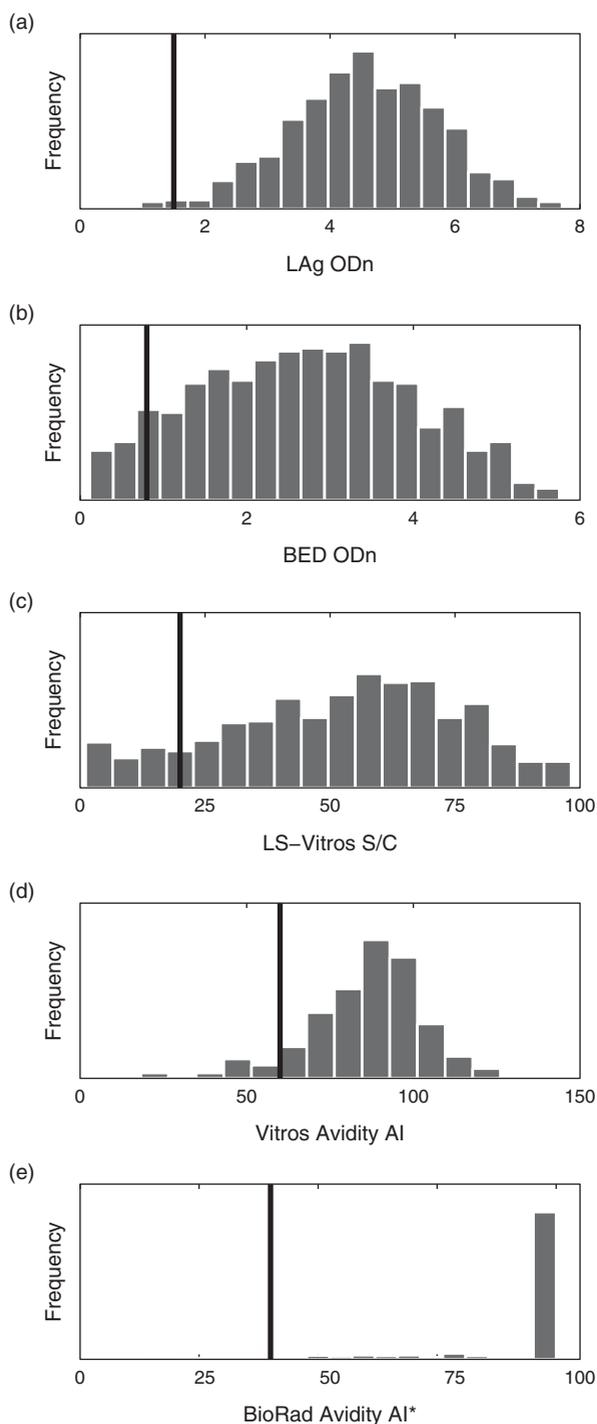
When stratifying by time since infection, the varying persistence of 'recent' classifications across assays is evident, with LAg exhibiting the leanest tail of persistence of 'recent' infection.

The FRR amongst elite controller specimens is high for all assays, and averages 25% (minimum of 13% to a maximum of 48% across assays). The FRR amongst treated patients is even higher, averaging 65% (minimum of 50% to a maximum of 76% across assays). Further stratifying treated patients by time from infection to treatment initiation, the FRR decreases as the time to treatment initiation increases: for early treatment initiation (within 6 months of infection) the average

FRR is 84% (64–93%), whereas for later treatment initiation (more than 6 months after infection) it is 41% (27–57%).

The FRR for patients with low viral loads – here defined as below 75 copies/ml – is high, averaging 55% (41–69%). This is consistent with the results above, as 92% of this specimen set is made up of specimens from the identified elite controllers and treated patients (and 94% of specimens from SCOPE elite controllers and treated patients have a low viral load).

Lastly, the FRR amongst patients with low CD4<sup>+</sup> cell counts, namely less than 200 cells/ $\mu$ l and acting as a proxy



**Fig. 2. Empirical distribution of incidence assay measurements for specimens drawn greater than  $T=2$  years after infection, excluding treated patients and elite controllers (665 data points from 316 patients), for (a) LAg, (b) BED, (c) LS-Vitros, (d) Vitros avidity and (e) BioRad avidity. Recent/nonrecent thresholds are shown by vertical solid lines. \*The peak of BioRad avidity results at 100% is due to a large proportion of (treated and untreated) aliquots returning the maximum possible S/C on the equipment used. LAg, limiting antigen avidity; S/C, signal-to-cut-off.**

for AIDS identification, was relatively low, averaging 2% (0–4%). Further stratifying this group by CD4<sup>+</sup> cell count (not shown) did not reveal any patterns.

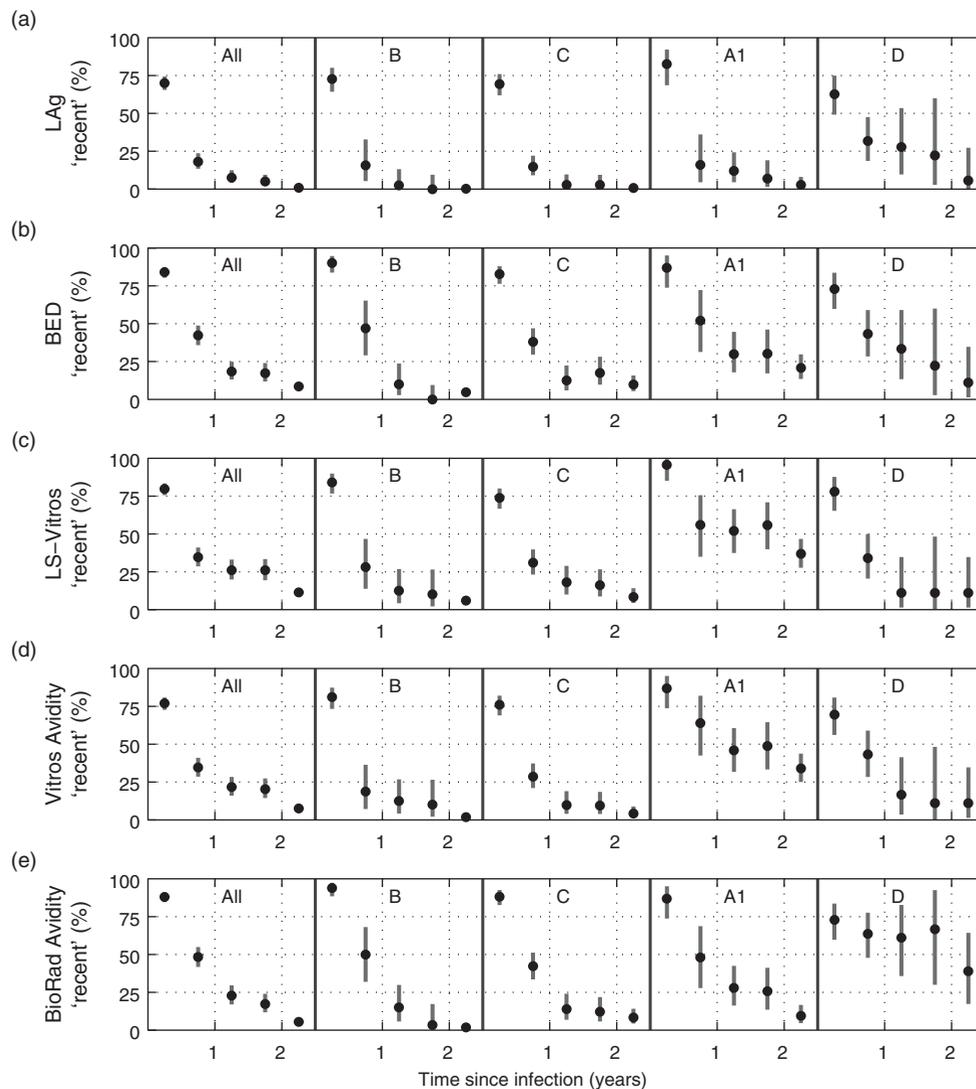
Table 2 lists MDRI and FRR by subtype. The most significant pair-wise differences in the MDRI were between subtype A1 and any other, on the Vitros platform. With one exception, notably small  $P$  values for subtype, pair-wise differences in the FRRs involve A1 or D and a non-A1, non-D subtype, dominated by LS-Vitros, Vitros avidity and BioRad avidity results. Whereas these initial results highlight potential subtype differences, a more definitive analysis (beyond the present scope) should be based on a large number of subtype D and A1 specimens, and estimation procedures specifically adapted to this stratification.

## Discussion

The application of cross-sectional HIV incidence surveillance, utilizing tests for recent infection, has been hampered by the lack of high-performance incidence assays and the lack of independent, rigorous and consistent evaluations of candidate assays [2–4]. Over the past 3 years, CEPHIA [6] has developed a substantial repository of precious specimens, and begun using these specimens to evaluate the most promising incidence assays. Results for LAg, BED, LS-Vitros, Vitros avidity and BioRad avidity are presented above.

Assays can be evaluated against a TPP [3,4,11]: not only should the technology be affordable, practical and transferable to other laboratories, but the MDRI should be sufficiently long (of order 1 year) and FRR small (ideally zero, and less than 2%). Results suggest that incidence assays continue to struggle to simultaneously achieve these two test property goals, with no single assay unequivocally meeting the criteria set out in the TPP. Compared to the increasingly used LAg assay, the other assays provide larger MDRI, but also higher FRRs.

While a stable, high-performance incidence assay should ideally produce a consistently small FRR, regardless of the study population, data from this work help to understand some of the reasons why an assay's performance could be unstable and FRRs may be large. All assays produce particularly high FRRs amongst elite controllers (>10%) and treated patients (>50%), and the size of these subpopulations will vary by region and time. In a surveillance study, identifying these patients is problematic, as there is no universal definition of or test for elite controllers, and self-reported treatment status may be unreliable. Furthermore, earlier initiation of treatment is associated with a higher FRR, in line with varying impacts of treatment on immune responses by treatment timing [41,42]. Context strongly affects when patients begin treatment: for example, in some states in the USA, patients are offered treatment immediately following HIV



**Fig. 3. The proportion of 'recent' results (%) as a function of time since infection, excluding treated patients and elite controllers and stratifying by HIV subtype (A1, B, C and D) for (a) LAg, (b) BED, (c) LS-Vitros, (d) Vitros avidity and (e) BioRad avidity.** Circles show observed proportions and lines capture 95% confidence intervals. Specimens are grouped into 6-monthly intervals of time since infection until 2 years, after which all specimens are grouped together (25 to 665 data points per group, other than for subtype D which has fewer than 20 points 1–2 years after infection). LAg, limiting antigen avidity.

diagnosis [43], whereas in South Africa, most HIV-positive patients are unable to access treatment until  $CD4^+$  cell counts drop below 350 copies/ $\mu$ l [44]. In this study, 94% of specimens from elite controllers and treated patients also had a low viral load ( $<75$  copies/ml), and so viral load testing provides a potential tool to screen for these high-FRR patients – specimens with viral loads below an optimized threshold would be classified as 'nonrecent'. Note that such a change in the 'recent' infection classification rule will also impact (reduce) the MDRI. Surveys could also directly test for the presence of antiretroviral drugs to identify treated patients [45].

Properties for each assay have been estimated here on the standardized basis of a Western blot being used to identify HIV-positive patients. However, other diagnostic

screening tests are likely to be used in incidence studies, and the time between HIV exposure and reactivity on these tests can differ by several weeks [26,27,46]. Therefore, for application to incidence studies, the base case MDRI reported here would need to be increased or decreased – depending on the particular screening test or algorithm used in the study to classify a specimen as HIV-positive, and hence eligible for 'recent' infection testing.

The results presented here should not be viewed as discouraging, as they provide a consistent, independent characterization of these candidate incidence assays. Large FRRs continue to limit the utility of single incidence assays, and subtype-specific test behavior should be further explored. This study provides the basis for exploring optimization through such adjustments as

**Table 1. Estimated test properties (and 95% confidence intervals) for each assay, for various specimen sets.**

Recent/nonrecent threshold (unit)	Number of patients (data points)	LAG 1.5 (ODn)	BED 0.8 (ODn)	LS-Vitros 20 (S/C)	Vitros avidity 60 (AI as %)	BioRad avidity 40 (AI as %)
MDRI (days) <sup>a</sup>						
All specimens <sup>b</sup>	400 (1032)	188 (165–211)	302 (274–331)	306 (274–338)	285 (254–316)	333 (302–363)
FRR (%) <sup>a</sup>						
All specimens <sup>b</sup>	316 (665)	1.3 (0.3–3.2)	7.4 (4.8–10.9)	9.7 (6.6–13.5)	6.5 (4.0–9.8)	6.2 (3.8–9.4)
By time since infection (years) <sup>b</sup>						
(2, 3)	140 (208)	2.5 (0.6–6.6)	12.5 (7.5–19.1)	17.5 (11.6–24.8)	12.5 (7.5–19.1)	12.5 (7.5–19.1)
(3, 4)	77 (110)	0.6 (0.0–5.9)	7.1 (2.5–15.4)	14.9 (7.8–24.9)	14.3 (7.4–24.1)	6.5 (2.1–14.5)
(4, 5)	35 (45)	0.0 (0.0–8.2)	7.1 (1.2–21.1)	5.7 (0.7–19.2)	5.7 (0.7–19.2)	8.6 (1.8–23.1)
>5	112 (193)	0.0 (0.0–2.6)	6.7 (2.8–13.0)	3.1 (0.8–8.3)	1.3 (0.1–5.6)	0.0 (0.0–2.6)
Elite controllers <sup>c</sup>	31 (89)	12.9 (3.6–29.8)	19.4 (7.5–37.5)	48.4 (30.2–66.9)	29.0 (14.2–48.0)	12.9 (3.6–29.8)
Treated patients <sup>d</sup>	113 (185)	58.8 (49.2–68)	65.9 (56.4–74.6)	76.1 (67.2–83.6)	72.6 (63.4–80.5)	50.0 (40.4–59.6)
By time from infection to treatment (years) <sup>d</sup>						
(0, 0.5)	53 (90)	84.9 (72.4–93.3)	86.8 (74.7–94.5)	92.5 (81.8–97.9)	92.5 (81.8–97.9)	64.2 (49.8–76.9)
≥0.5	53 (88)	27.4 (16.0–41.3)	40.6 (27.3–54.9)	56.6 (42.3–70.2)	49.1 (35.1–63.2)	31.1 (19.1–45.3)
Low viral load <sup>e</sup>	154 (273)	47.1 (39.0–55.3)	56.5 (48.3–64.5)	68.5 (60.5–75.7)	62.7 (54.5–70.3)	40.6 (32.8–48.8)
Low CD4 <sup>+</sup> cell count <sup>f</sup>	124 (214)	0.0 (0.0–2.4)	4.0 (1.3–9.2)	2.4 (0.5–6.9)	0.0 (0.0–2.4)	1.6 (0.2–5.7)

AI, avidity index; FRR, false-recent rate; LAG, limiting antigen avidity; LS-Vitros, less-sensitive Vitros; MDRI, mean duration of recent infection; ODn, normalized optical density; S/C, signal-to-cut-off.

<sup>a</sup>Using an HIV viral lysate-based Western blot assay to identify HIV-positive patients, and  $T=2$  years.

<sup>b</sup>Excluding treated patients and SCOPE elite controllers.

<sup>c</sup>Identified as elite controllers in the SCOPE cohort (virally suppressed in the absence of treatment).

<sup>d</sup>No previous treatment interruptions and treated for at least 3 months.

<sup>e</sup>Viral load at draw below 75 copies/ml.

<sup>f</sup>CD4<sup>+</sup> cell count at draw below 200 cells/ $\mu$ l.

variation of ‘recent’/‘nonrecent’ thresholds, inclusion of supplemental tests (in particular, viral load), and the use of multiple incidence assays, all of which is the subject of ongoing work within and beyond CEPHIA [2,4,37,47,48]. Optimization should also consider the time cut-off  $T$ , to distinguish ‘true-recent’ from ‘false-recent’ results – although  $T$  should be not be too large, the value of  $T$  was increased from 1 year, as used in preliminary analyses [49], to 2 years in this study, to better capture the tails of persisting ‘recent’ results and thus reduce FRRs. Ongoing analyses also include the evaluation of tests for recent infection, using the precision

of the incidence estimator as a summary performance metric [50]. In addition, efforts are being made to capture more detailed information on cohorts’ diagnostic testing protocols and more complete testing histories of patients – providing the required data to further refine estimated infection dates for later analyses of assay results.

The repository of specimens and data that has been generated by CEPHIA provides a unique opportunity to further advance the investigation and refinement of markers of ‘recent’ HIV infection. Specimens and datasets are well maintained, sample sizes are large, specimen

**Table 2. Estimated test properties (and 95% confidence intervals) for each assay, by subtype.**

Recent/nonrecent threshold (unit)	Number of patients (data points)	LAG 1.5 (ODn)	BED 0.8 (ODn)	LS-Vitros 20 (S/C)	Vitros avidity 60 (AI as %)	BioRad avidity 40 (AI as %)
MDRI (days) <sup>a,b,c</sup>						
All specimens	400 (1032)	188 (165–211)	302 (274–331)	306 (274–338)	285 (254–316)	333 (302–363)
Subtype A1	80 (166)	211 (156–275)	363 (288–442)	473 (387–560)	451 (362–539)	364 (289–442)
Subtype B	90 (246)	153 (117–196)	255 (208–308)	232 (173–299)	210 (157–270)	299 (233–371)
Subtype C	181 (454)	177 (150–206)	287 (248–328)	265 (221–311)	240 (202–280)	298 (262–338)
Subtype D	38 (131)	273 (170–387)	328 (227–433)	264 (190–339)	276 (194–357)	467 (354–581)
FRR (%) <sup>a,b,d</sup>						
All specimens	316 (665)	1.3 (0.3–3.2)	7.4 (4.8–10.9)	9.7 (6.6–13.5)	6.5 (4.0–9.8)	6.2 (3.8–9.4)
Subtype A1	37 (106)	2.7 (0.1–14.2)	18.9 (8.0–35.2)	35.1 (20.2–52.5)	35.1 (20.2–52.5)	12.2 (3.8–27.1)
Subtype B	190 (388)	0.5 (0.0–2.9)	4.7 (2.2–8.8)	4.7 (2.2–8.8)	1.3 (0.2–4.2)	2.1 (0.6–5.3)
Subtype C	75 (144)	1.3 (0.0–7.2)	7.3 (2.6–15.7)	8.7 (3.4–17.5)	4.0 (0.8–11.2)	6.7 (2.2–14.9)
Subtype D	11 (18)	9.1 (0.2–41.3)	18.2 (2.3–51.8)	18.2 (2.3–51.8)	18.2 (2.3–51.8)	54.5 (23.4–83.3)

AI, avidity index; FRR, false-recent rate; LAG, limiting antigen avidity; LS-Vitros, less-sensitive Vitros; MDRI, mean duration of recent infection; ODn, normalized optical density; S/C, signal-to-cut-off.

<sup>a</sup>Using an HIV viral lysate-based Western blot assay to identify HIV-positive patients, and  $T=2$  years.

<sup>b</sup>Excluding treated patients and SCOPE elite controllers.

<sup>c</sup>In a test for pair-wise differences in MDRI by subtype, using a z-test, the following pairs provided  $P$  values below 0.05: LAG – B and D; BED – A1 and B; LS-Vitros – A1 and B, A1 and C A1 and D; Vitros avidity – A1 and B, A1 and C, A1 and D; BioRad avidity – B and D, C and D. Estimated SDs of the MDRI estimators are used as proxies for true values, and therefore tests are anticonservative (particularly when sample sizes are small).

<sup>d</sup>In a test for pair-wise differences in FRRs by subtype, using the Fisher-Boschloo test [40], the following pairs provided  $P$  values below 0.05: BED – A1 and B; LS-Vitros – A1 and B, A1 and C; Vitros avidity – A1 and B, A1 and C, B and D; BioRad avidity – A1 and B, A1 and D, B and D, C and D.

background information is recorded, and multiple incidence assays and potential supplemental tests have been measured on the same specimens.

Using other assays, CEPHIA has begun testing the 'evaluation panel', with the aim of evaluating 10 incidence assays in its first phase. A second phase of CEPHIA, known as CEPHIA II and also funded by the Bill and Melinda Gates Foundation (BMGF), was launched in the beginning of 2013. Under CEPHIA II, the repository is being expanded to include nonplasma specimens (such as linked whole blood, oral fluid, urine and stool) that are being prospectively collected through collaborations with various study sites. CEPHIA is also supporting biomarker discovery projects funded by the BMGF and US National Institute of Health (NIH), with a focus on earlier steps in the development pathway. Further updates on CEPHIA activities can be found on the project website (<http://www.incidence-estimation.com/page/cephia>).

## Acknowledgements

G.M., A.W., C.D.P. and M.P.B. conceived the study design and sourced funding. R.K. and A.W. led the data analysis. C.D.P., S.N.F., S.L., M.A.P., J.N.M., E.G.K. and F.M.H. led on specimen acquisition and related data collection. G.M., M.P.B., S.M.K. and E.M. led on assay performance and quality, and assay results reporting. R.K. drafted the paper and all authors assisted in the interpretation of findings, provided input and suggestions for analysis, and critically reviewed the draft and the final submitted version. Funding for this project was provided by the Bill and Melinda Gates Foundation (grant OPP1017716).

The authors acknowledge with thanks Mila Lebedeva and Eve Draper for laboratory testing; Bio-Rad Laboratories, Inc. for providing reagents, as well as Paul Contestable and Ortho Clinical Diagnostics, Inc.; David Matten for database support; the assay developers who supplied training in their assays; and the CEPHIA steering group for their advice and suggestions on the data outputs of this work. Members of CEPHIA and individuals contributing to the overall project, not appearing as authors, are listed in Supplementary Digital Content 1 (<http://links.lww.com/QAD/A565>).

IAVI's work is made possible by generous support from many donors including: the Bill & Melinda Gates Foundation; the Ministry of Foreign Affairs of Denmark; Irish Aid; the Ministry of Finance of Japan; the Ministry of Foreign Affairs of the Netherlands; the Norwegian Agency for Development Cooperation (NORAD); the United Kingdom Department for International Development (DFID); and the United States Agency for International Development (USAID). The full list of IAVI donors is

available at [www.iavi.org](http://www.iavi.org). This report is made possible by the generous support of the American people through USAID. The contents are the responsibility of the International AIDS Vaccine Initiative and do not necessarily reflect the views of USAID or the United States Government. Funding was also provided by the OPEC Fund for International Development (OFID), a development finance institution of the OPEC Member States, established to provide financial support for socio-economic development, particularly for low-income countries.

The San Diego Primary Infection Cohort acknowledges funding by the National Institutes of Health (NIH, grants AI43638, AI74621 and AI106039), and the California HIV-1 Research Program (CHRP, grant RN07-SD-702).

The São Paulo Cohort acknowledges funding by the Brazilian Ministry of Health, Brazilian Program for STD and AIDS (grant 914/BRA/3014-UNESCO) and the São Paulo City Health Department (grant 2004-0.168.922-7).

The SCOPE study received funding from the NIH (grants P30 AI027763 and R24 AI067039).

The AMPLIAR and Options Cohorts also received funding from the NIH (grants P01 AI071713 and R01 HD074511).

## Conflicts of interest

All authors, as members or collaborators of the Consortium for the Evaluation and Performance of Incidence Assays (CEPHIA), are supported by a grant from the Bill and Melinda Gates Foundation (OPP1017716). Specimen collection was funded in part by grants from the US National Institutes of Health (P01 AI071713, R01 HD074511, P30 AI027763, R24 AI067039, R34 MH096606 AI43638, AI74621 and AI106039); California HIV-1 Research Program (RN07-SD-702); Brazilian Program for STD and AIDS, Ministry of Health (914/BRA/3014-UNESCO); and São Paulo City Health Department (2004-0.168.922-7).

C.D.P. has an ongoing grant from Bio-Rad Laboratories, Inc., through the University of California, San Francisco, for the conduct of an unrelated clinical trial. M.P.B. receives ongoing funding from Bio-Rad Laboratories, Inc. and Ortho Clinical Diagnostics, Inc., provided to Blood Systems Research Institute, to enable ongoing evaluations of their respective assays. S.L. has received research funding from Gilead Sciences.

## References

1. Brookmeyer R, Quinn TC. **Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests rates from a cross-sectional survey using early diagnostic tests.** *Am J Epidemiol* 1995; **141**:166-172.

2. Mastro TD, Kim AA, Hallett T, Rehle T, Welte A, Laeyendecker O, *et al.* **Estimating HIV incidence in populations using tests for recent infection: Issues, challenges and the way forward.** *J HIV AIDS Surveill Epidemiol* 2010; **2**:1–14.
3. Incidence Assay Critical Path Working Group. **More and better information to tackle HIV epidemics: towards improved HIV incidence assays.** *PLoS Med* 2011; **8**:e1001045.
4. 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.
5. Kassanjee R, McWalter TA, Barnighausen T, Welte A. **A new general biomarker-based incidence estimator.** *Epidemiology* 2012; **23**:721–728.
6. The Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA). <http://www.incidence-estimation.com/page/cephia>. [Accessed 29 March 2014].
7. Duong YT, Qiu M, De AK, Jackson K, Dobbs T, Kim AA, *et al.* **Detection of recent HIV-1 infection using a new Limiting-Antigen avidity assay: potential for HIV-1 incidence estimates and avidity maturation studies.** *PLoS One* 2012; **7**: e33328.
8. Parekh BS, Kennedy MS, Dobbs T, Pau CP, Byers R, Green 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.
9. Keating SM, Hanson D, Lebedeva M, Laeyendecker O, Ali-Napo NL, Owen SM, *et al.* **Lower-sensitivity and avidity modifications of the Vitros Anti-HIV 1+2 Assay for detection of recent HIV infections and incidence estimation.** *J Clin Microbiol* 2012; **50**:3968–3976.
10. Masciotra S, Dobbs T, Candal D, Hanson D, Delaney K, Rudolph D, *et al.* **Antibody avidity-based assay for identifying recent HIV-1 infections based on genetic systems TM 1/2 plus O EIA.** *Abstract 937 at the 17th Conference on Retroviruses and Opportunistic Infections*, San Francisco, 16–19 February 2010.
11. Bill & Melinda Gates Foundation Letters of Inquiry (LOI). New biomarkers for HIV incidence measurement. <https://docs.gatesfoundation.org/Documents/hiv-incidence-rules-and-guidelines.pdf>. [Accessed 29 March 2014].
12. American Red Cross. <http://www.redcrossblood.org/>. [Accessed 29 March 2014].
13. Blood Centers of the Pacific. <http://www.bloodcenters.org/>. [Accessed 29 March 2014].
14. South African National Blood Service. <http://www.sanbs.org.za/>. [Accessed 29 March 2014].
15. Fundacao Pro-Sangue Hemocentro de Sao Paulo. <http://www.prosangue.sp.gov.br/hemocentros/Default.aspx>. [Accessed 29 March 2014].
16. Jain V, Liegler T, Vittinghoff E, Hartogensis W, Bacchetti P, Poole L, *et al.* **Transmitted drug resistance in persons with acute/early HIV-1 in San Francisco, 2002–2009.** *PLoS One* 2010; **5**:e15510.
17. Winkelstein W Jr, Wiley JA, Padian NS, Samuel M, Shiboski S, Ascher MS, *et al.* **The San Francisco Men's Health Study: continued decline in HIV seroconversion rates among homosexual/bisexual men.** *Am J Public Health* 1988; **78**: 1472–1474.
18. Morris SR, Little SJ, Cunningham T, Garfein RS, Richman DD, Smith DM. **Evaluation of an HIV nucleic acid testing program with automated internet and voicemail systems to deliver results.** *Ann Intern Med* 2010; **152**:778–785.
19. Costa PR, Hartogensis W, Diaz RS, Bacchetti P, Hecht R, Santos BR, *et al.* **Subjects infected with clade C-containing HIV exhibit higher CD4+ T cell activation.** *Poster 289 at the 2014 Conference on Retroviruses and Opportunistic Infections*, Boston, MA, 3–6 March 2014.
20. Price MA, Wallis CL, Lakhi S, Karita E, Kamali A, Anzala O, *et al.* **Transmitted HIV type 1 drug resistance among individuals with recent HIV infection in East and Southern Africa.** *AIDS Res Hum Retroviruses* 2011; **27**:5–12.
21. Hunt PW, Brenchley J, Sinclair E, McCune JM, Roland M, Page-Shafer K, *et al.* **Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with undetectable plasma HIV RNA levels in the absence of therapy.** *J Infect Dis* 2008; **197**:126–133.
22. Kassanjee R, Murphy G, Busch MP, Pilcher CD, McKinney E, Keating S, *et al.* **The performance of candidate assays to detect recent HIV infection for cross-sectional incidence estimation: an independent, comparative evaluation.** *Poster 1056/abstract X-168 at the 20th Conference on Retroviruses and Opportunistic Infections*, Atlanta, GA, 3–6 March 2013.
23. Calypte Biomedical Corporation. AwareTM BEDTM EIA HIV-1 incidence test (IgG-capture HIV EIA): enzyme immunoassay for population estimates of HIV-1 incidence, Cat. No. 98003, 2008. [http://www.calypte.com/pdf/LN-20597%20A\\_%20Aware%20BED%20EIA%20HIV-1\\_2%20Incidence%20Test%20Package%20Insert.pdf](http://www.calypte.com/pdf/LN-20597%20A_%20Aware%20BED%20EIA%20HIV-1_2%20Incidence%20Test%20Package%20Insert.pdf). [Accessed 29 March 2014].
24. Sedia Biosciences Corporation. Sedia (HIV-1 LAg-avidity EIA: single well avidity enzyme immunoassay for detection of recent HIV-1 infection using liquid serum or plasma, Cat. No. 1002, 2013. [http://www.hivincidence.com/uploads/LN-6039-06\\_Package\\_Insert\\_LAg\\_Avidity\\_EIA.pdf](http://www.hivincidence.com/uploads/LN-6039-06_Package_Insert_LAg_Avidity_EIA.pdf). [Accessed 29 March 2014].
25. Sedia Biosciences Corporation. Sedia (HIV-1 LAg-avidity EIA: single well avidity enzyme immunoassay for detection of recent infection, Cat. No. 1002, 20, 2012. [http://www.hivincidence.com/uploads/LN-6039-02\\_Product\\_Insert\\_LAg\\_Avidity\\_EIA\\_sm\\_.pdf](http://www.hivincidence.com/uploads/LN-6039-02_Product_Insert_LAg_Avidity_EIA_sm_.pdf). [Accessed 29 March 2014].
26. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, *et al.* **Dynamics of HIV viremia and antibody seroconversion in plasma donors: Implications for diagnosis and staging of primary HIV infection.** *AIDS* 2003; **17**:1871–1879.
27. Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, *et al.* **Modeling sequence evolution in acute HIV-1 infection.** *J Theor Biol* 2009; **261**:341–360.
28. Bollinger RC, Brookmeyer RS, Mehendale SM, Paranjape RS, Shepherd ME, Gadkari DA, *et al.* **Risk factors and clinical presentation of acute primary HIV infection in India.** *J Am Med Assoc* 1997; **278**:2085–2089.
29. Daar ES, Little S, Pitt J, Santangelo J, Ho P, Harawa N, *et al.* **Diagnosis of primary HIV-1 infection.** *Ann Intern Med* 2001; **134**:25–29.
30. Hecht FM, Busch MP, Rawal B, Webb M, Rosenberg E, Swanson M, *et al.* **Use of laboratory tests and clinical symptoms for identification of primary HIV infection.** *AIDS* 2002; **16**:1119–1129.
31. Powers KA, Miller WC, Pilcher CD, Mapanje C, Martinson FE, Fiscus SA, *et al.* **Improved detection of acute HIV-1 infection in sub-Saharan Africa: development of a risk score algorithm.** *AIDS* 2007; **21**:2237–2242.
32. Schacker T, Collier AC, Hughes J, Shea T, Corey L. **Clinical and epidemiologic features of primary HIV infection.** *Ann Intern Med* 1996; **125**:257–264.
33. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, *et al.* **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nat Med* 1997; **3**:205–211.
34. Lindback S, Thorstensson R, Karlsson AC, von Sydow M, Flamholz L, Blaxhult A, *et al.* **Diagnosis of primary HIV-1 infection and duration of follow-up after HIV exposure.** *AIDS* 2000; **14**:2333–2339.
35. Pilcher CD, Eron JJ Jr, Vemazza PL, Battegay M, Harr T, Yerly S, *et al.* **Sexual transmission during the incubation period of primary HIV infection.** *J Am Med Assoc* 2001; **286**:1713–1714.
36. HIV Modelling Consortium. <http://www.hivmodelling.org/>. [Accessed 29 March 2014].
37. Brookmeyer R, Konikoff J, Laeyendecker O, Eshleman SH. **Estimation of HIV incidence using multiple biomarkers.** *Am J Epidemiol* 2013; **177**:264–272.
38. Efron B, Tibshirani RJ. **An introduction to the bootstrap** (monographs on statistics and applied probability 57). New York: Chapman & Hall /CRC; 1993.
39. Agresti A, Coull BA. **Approximate is better than 'exact' for interval estimation of binomial proportions.** *Am Stat* 1998; **52**:119–126.
40. Mehrotra DV, Chan IS, Berger RL. **A cautionary note on exact unconditional inference for a difference between two independent binomial proportions.** *Biometrics* 2003; **59**:441–450.
41. Killian MS, Norris PJ, Rawal BD, Lebedeva M, Hecht FM, Levy JA, *et al.* **The effects of early antiretroviral therapy and its discontinuation on the HIV-specific antibody response.** *AIDS Res Hum Retroviruses* 2006; **22**:640–647.

42. Wendel SK, Mullis CE, Eshleman SH, Blankson JN, Moore RD, Keruly JC, *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**:e55525.
43. Department of Health and Human Services. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. <http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf>. [Accessed 29 March 2014].
44. South African Department of Health. The South African antiretroviral treatment guidelines, Version 14. <http://www.sahivsoc.org/upload/documents/2013%20ART%20Guidelines-Short%20Combined%20FINAL%20draft%20guidelines%2014%20March%202013.pdf>. [Accessed 29 March 2014].
45. Le Vu S, Velter A, Meyer L, Peytavin G, Guinard J, Pillonel J, *et al.* **Biomarker-based HIV incidence in a community sample of men who have sex with men in Paris, France.** *PLoS One* 2012; **7**:e39872.
46. Masciotra S, McDougal JS, Feldman J, Sprinkle P, Wesolowski L, Owen SM. **Evaluation of an alternative HIV diagnostic algorithm using specimens from seroconversion panels and persons with established HIV infections.** *J Clin Virol* 2011; **52** (Suppl 1):S17–22.
47. Curtis KA, Hanson DL, Kennedy MS, Owen SM. **Evaluation of a multiplex assay for estimation of HIV-1 incidence.** *PLoS One* 2013; **8**:e64201.
48. Laeyendecker O, Brookmeyer R, Cousins MM, Mullis CE, Konikoff J, Donnell D, *et al.* **HIV incidence determination in the United States: a multiassay approach.** *J Infect Dis* 2013; **207**:232–239.
49. Kassanjee R, Murphy G, Pilcher C, Busch M, McKinney E, Keating SM, *et al.* **Evaluation panel analysis results.** *Presentation at satellite session of the 20th Conference on Retroviruses and Opportunistic Infections*, Atlanta, GA, 3–6 March 2013.
50. Kassanjee R, McWalter TA, Welte A. **Defining optimality of a test for recent infection for HIV incidence surveillance.** *AIDS Res Hum Retroviruses* 2014; **30**:45–49.