Infection Staging and Incidence Surveillance Applications of High Dynamic Range Diagnostic Immuno-Assay Platforms

Eduard Grebe, PhD,* Alex Welte, PhD,* Jake Hall, BSc,† Sheila M. Keating, PhD, MSPH,‡§ Shelley N. Facente, MPH,§ Kara Marson, MPH,§ Jeffrey N. Martin, MD, MPH,§ Susan J. Little, MD,‖ Matthew A. Price, PhD,§¶ Esper G. Kallas, MD, PhD,# Michael P. Busch, MD, PhD,‡§ Christopher D. Pilcher, MD, PhD,§ and Gary Murphy, PhD,† on behalf of the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA)

Background: Custom HIV staging assays, including the Sedia HIV-1 Limiting Antigen (LAg) Avidity EIA and avidity modifications of the Ortho VITROS anti-HIV-1+2 and Abbott ARCHITECT HIV Ag/Ab Combo assays, are used to identify "recent" infections in clinical settings and for cross-sectional HIV incidence estimation. However, the high dynamic range of chemiluminescent platforms allows differentiating recent and long-standing infection on signal intensity, and this raises the prospect of using unmodified diagnostic assays for infection timing and surveillance applications.

Methods: We tested a panel of 2500 well-characterized specimens with estimable duration of HIV infection with the 3 assays and the unmodified ARCHITECT. Regression models were used to estimate mean durations of recent infection (MDRIs), context-specific false-recent rates (FRRs) and correlation between diagnostic signal intensity and LAg measurements. Hypothetical epidemiological scenarios were constructed to evaluate utility in surveillance applications.

Results: Over a range of MDRIs (reflecting recency discrimination thresholds), a diluted ARCHITECT-based RITA produced lower FRRs than the VITROS platform (FRR ≈ 0.5% and 1.5%, respectively at MDRI ≈ 200 days), and the unmodified diagnostic ARCHITECT produces incidence estimates with comparable precision to LAg (relative SE ≈ 17.5% and 15%, respectively at MDRI ≈ 200 days). ARCHITECT S/CO measurements were highly correlated with LAg optical density measurements (r = 0.80), and values below 200 are strongly predictive of LAg recency and duration of infection less than 1 year.

Conclusions: Low quantitative measurements from the unmodified ARCHITECT obviate the need for additional recency testing, and its use is feasible in clinical staging and incidence surveillance applications.

Key Words: infection timing, infection staging, incidence, recent infection, diagnostic assays, staging assays

(*J Acquir Immune Defic Syndr* 2017;76:547–555)

INTRODUCTION

Laboratory assays for the detection, staging, and management of HIV have been a significant priority over the course of the HIV/AIDS epidemic. In particular, the idea of using a test for "recent" HIV infection to generate incidence estimates from cross-sectional surveys has attracted substantial attention and investment.[1–11] A variety of candidate immunological and virological markers have been investigated, with most early applications focused on the expansion of the dynamic range of serological assays used to identify recent infection (ie, increasing the length of time an incident case could still be classified as recently infected). In conventional plate reader–based enzyme-linked immunosorbent assay platforms, high-diagnostic sensitivity requires a rapid rise of signal strength with increasing antibody titre and avidity, so that patients' readings rapidly traverse the range of quantifiable detection as the infection progresses. Customizations to expand the dynamic range and facilitate the introduction of a reproducible recent/nonrecent threshold have variously taken the form of dilution, incubation time reduction, or antibody–antigen binding degradation (sometimes in combination) to produce markers indicative of antibody titre, HIV-specific proportion, or avidity.

Individuals who are virally suppressed (either through effective endogenous control or by antiretroviral treatment) tend to undergo at least partial seroreversion, leading to "false" recent classifications under protocols designed for cross-sectional incidence estimation. Inspired by the hypothesis that antibody quality reverts less than antibody titre or proportion, the notion of a "2-well avidity modification" has gained some traction. In this case, a specimen is subjected to 2 runs on the diagnostic platform under different conditions: an "untreated" reaction well (ie, close to standard conditions for diagnostic use) and a "treated" reaction well, in which conditions are altered to restrict antibody–antigen binding. The ratio of the signals generated by "treated" vs. "untreated" reaction wells —usually named an "avidity index"—is interpreted as a measure of antibody binding capacity. Avidity typically increases over time after infectious exposure as the antibody response matures.

Two-well avidity modifications of 2 modern chemiluminescent platforms with high intrinsic dynamic range, the Ortho VITROS anti-HIV-1+2 Assay (VITROS)[12] and the Abbott ARCHITECT HIV Ag/Ab Combo Assay (ARCHITECT),[13] have been proposed. These use "chaotropic" agents —ie, agents that interfere with antibody–antigen binding—in the "treated" reaction well to inhibit the formation of antibody–antigen complexes or disrupt complexes after formation. Although not relying on specimen dilution, (which is an alternative, nonavidity approach to dynamic range expansion), the use of an additional reagent in the modified run requires some dilution of samples in the "untreated" run to match input volumes and ensure that the resulting signals are comparable.

Performance of the VITROS and ARCHITECT avidity assays for use in cross-sectional incidence estimation has been investigated by the CEPHIA collaboration, as part of an independent evaluation of leading candidate HIV incidence assays.[9,10] During application of the proposed procedures for recency ascertainment to the CEPHIA "evaluation panel," these 2-well avidity protocols additionally produced results from diluted (1:10) but otherwise unmodified runs. These previously unreported data and a new evaluation of the unmodified Abbott ARCHITECT HIV Ag/Ab Combo Assay [ie, applied according to the manufacturer's Instructions for Use (IFU) in diagnostic applications] form the subject of this work. We additionally report, for comparison, results from the Sedia HIV-1 Limiting Antigen (LAg) Avidity assay on the same specimen panel.

Until now, no practical recent/nonrecent infection classification scheme has been constructed using only markers available through unmodified diagnostic platforms and algorithms. This is a noteworthy opportunity, given that the Western blot, at times widely used as a "confirmatory" diagnostic test, provides a compelling picture of immune response evolution.[14] In many settings, it has been noted that "indeterminate" Western blot patterns tend to evolve into unambiguously HIV-positive patterns within a few weeks. However, lack of consistent production and quantitation of Western blot band intensity has been one key limitation preventing use of the Western blot as a staging assay. Similarly, the cost and complexity of using customized recency assays in disease surveillance have limited adoption.[15] A recent analysis of the Bio-Rad Geenius HIV1/2 supplemental assay,[16] although using a platform available in diagnostic settings requires a research-use-only modification of the cartridge reader software to facilitate the extraction of quantitative band intensities, which form the basis of the recency classification. The potential to interpret data routinely available from diagnostic assays to classify infections as recent or nonrecent may improve the feasibility of incidence surveillance in many settings.

In clinical settings the only well-known early infection staging scheme is "Fiebig staging,"[14] based on unmodified commercial assays, but using only the qualitative results. Fiebig stages are based on discordance between tests of varying diagnostic sensitivity—specifically, the average time from infectious exposure to test conversion. Given progress in the reduction of diagnostic delays (window periods), such a staging system has utility only in the very early stages of infection, and inferences about the duration of infection can in practice only be made when specific diagnostic platforms are used. Many of these assays are no longer commercially available.

We report the first evidence that certain unmodified fourth-generation chemiluminescent diagnostic assays can provide meaningful infection staging information applicable to incidence surveillance and clinical interpretation. The latter application is already being practiced in several countries where custom recency staging assays (such as LAg) are applied to specimens from newly diagnosed persons. This involves additional laboratory resources and longer turnaround times, as specimens are usually reflexed to a small number of centralized specialist laboratories for recency ascertainment.

METHODS

The CEPHIA Specimen Repository and Evaluation Panel

The CEPHIA specimen repository houses over 25,000 HIV-1–positive specimens. Specimen details and background clinical data (obtained from contributing clinical cohort studies) are stored in a harmonized form in a research database. As previously described,[9,10,16] the CEPHIA Evaluation Panel consists of 2500 plasma specimens, selected to allow a full independent assessment of promising tests for recent infection, including estimation of test properties relevant to HIV incidence estimation. The specimens were obtained from 932 unique subjects (1–13 specimens per subject). Most specimens were obtained from subjects infected with subtype B (53% of specimens), C (27%), A1 (12%), and D (6%). The panel further contained multiple blinded aliquots of 3 control specimens (25 replicates of each), with antibody reactivity characteristic of recent, intermediate, and longstanding infection, to allow evaluation of the reproducibility of assay results.

The majority (67%) of subjects contributing specimens to the panel had sufficient clinical data to produce Estimated Dates of Detectable Infection (EDDIs), which are obtained by systematically interpreting diverse diagnostic testing histories by means of "diagnostic delays," ie, the average time elapsed from exposure to a first positive result on the assay in question.[11] The distribution of times since EDDI and subtype composition of specimens in the evaluation panel are shown in Figure 1.

The UCSF Human Research Protection Program and IRB (formerly CHR, #10-02365) approved the study procedures.

Laboratory Procedures

For the present analysis, data for the VITROS Avidity, ARCHITECT Avidity, and LAg assays were generated as previously described,[9,10] with the exception that the data

FIGURE 1. Distribution of times since estimated date of detectable infection (EDDI) in the CEPHIA evaluation panel, in 3-month intervals.

from the CEPHIA evaluations of the VITROS Avidity and the ARCHITECT Avidity assays were re-analyzed using only the "untreated" (diluted) run (ie, the sample diluted 1:10 with phosphate-buffered saline). Additional testing of the CEPHIA evaluation panel was undertaken according to the manufacturer's IFU for the Abbott ARCHITECT HIV Ag/Ab Combo assay. Data for the LAg assay are included for comparison.

Specimens were independently tested in CEPHIA laboratories—Blood Systems Research Institute, San Francisco, CA (VITROS Avidity) and National Infection Service, Public Health England, London, United Kingdom (ARCHITECT Avidity, unmodified ARCHITECT, LAg)—by technicians blinded to specimen background data.

The procedures used in the testing of the 2 avidity modifications and LAg have been previously described.[9] The unmodified ARCHITECT HIV Ag/Ab Combo assay is a 2-step immunoassay, which detects both the presence of HIV p24 antigen and antibodies to HIV-1 and HIV-2 in human serum and plasma. For this evaluation, we investigated anti–HIV-1 antibody detection and recommend that recency interpretations only be applied to confirmed HIV-1 positive specimens. All specimens in the CEPHIA evaluation panel are from HIV-1–infected individuals and confirmed antibody positive. In the first step, the sample, assay diluent, and paramagnetic microparticles are combined. HIV p24 antigen and anti-HIV-1/anti-HIV-2 antobodies present in the sample bind to the HIV p24 monoclonal (mouse) antobody and HIV-1/HIV-2 antigen-coated microparticles. After washing, the HIV p24 antigen and anti–HIV-1/HIV-2 antibodies bind to the acridinium-labelled conjugates (HIV-1/HIV-2 antigens, synthetic peptides, and HIV p24 antibody). After another wash cycle, pretrigger and trigger solutions are added to the reaction mixture. The resulting chemiluminescent reaction is measured as relative light units. The assay incorporates a number of controls and calibrator specimens, which are used to define a percycle "cutoff" value, and results are then reported as a signal-to-cutoff ratio (S/CO). Regular maintenance of the ARCHITECT platform is required to ensure accuracy of results. For the purpose of this evaluation, specimens were tested and analyzed in singleton, but for diagnostic purposes the manufacturer suggests specimens be tested in duplicate.

Statistical Analysis

For use in incidence surveillance, the performance of recent infection tests is summarized by 2 parameters, the mean duration of recent infection (MDRI) and the false-recent rate (FRR).

MDRI denotes the average amount of time that individuals spend exhibiting the "recent" biomarker, while infected for less than some cutoff time (denoted T, 2 years in this work). This captures the defining biological aspects of the recency test and should be more than approximately half a year to yield informative incidence estimates from feasibly sized surveys, even in high-incidence settings. As described previously,[3,4] MDRI was estimated by fitting a linear binomial regression model for the probability of testing recent as a function of estimated time since infection, $P_R(t)$, using
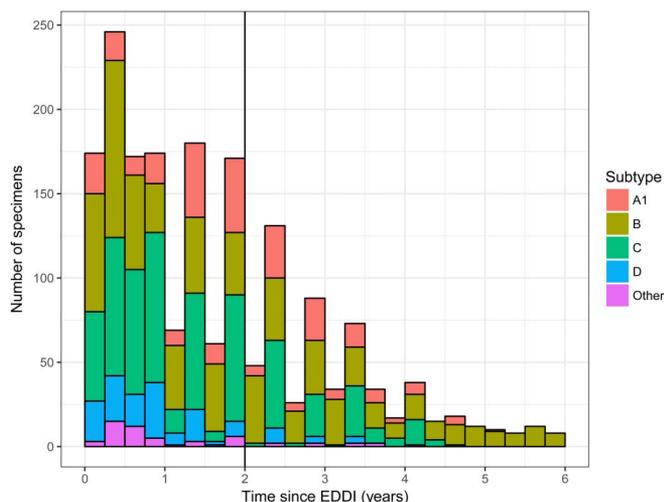
a logit link function, and a cubic polynomial in time (since estimated date of detectable infection). The function was fit to data points up to 800 days postinfection. The MDRI was then obtained by integrating the function from 0 to T. Confidence intervals were obtained by resampling subjects in 10,000 bootstrap iterations. Because the estimation procedure does not rely on longitudinal biomarker progression within individuals, the bootstrapping procedure resamples subjects rather than individual data points (with replacement) to account for the nonindependence of measurements on specimens drawn from the same individual.

The FRR is the proportion of those individuals who are infected for longer than the cutoff time T, but who nevertheless produce a recent result on the test. For surveillance, values of FRR above approximately 1%–2% are highly vulnerable to bias and artifacts during the incidence estimation procedure.

FRR is inevitably context dependent. To estimate FRR, 3 hypothetical epidemiological scenarios were constructed and the FRR in untreated and treated individuals estimated and weighted according to the treatment coverage specified in each scenario. The primary epidemiological scenario used in this study (scenario A) can be summarized as follows: (1) HIV prevalence of 30%; (2) incidence of 1.5 cases per 100 person-years (PY); (3) treatment coverage (defined as proportion of HIV-infected individuals who are on ART, all of whom are assumed to be virally suppressed) of 80%, and (4) predominant subtypes distributed as in the evaluation panel. This scenario is one in which precision would be expected to be fairly good on all assays. To evaluate sensitivity of performance metrics to epidemiological context, 2 further scenarios were specified, designed to resemble a medium-prevalence generalized epidemic and a concentrated epidemic, respectively. Scenario B is summarized as follows: (1) HIV prevalence of 10%, (2) incidence of 0.5 cases/100 PY, (3) treatment coverage of 60%, and (4) predominantly subtype C infections. Scenario C is summarized as follows: (1) HIV prevalence of 15%, (2) incidence of 1.5 cases/100 PY, (3) treatment coverage of 90%, and (4) predominantly subtype B infections.

To estimate the FRR in untreated individuals, the function $P_R(t)$ was fit to data from all times postinfection, and weighted according to the probability density function for times since infection in the untreated population. The distribution of times since infection was parameterized as a Weibull survival function, with the shape and scale parameters chosen to produce the desired treatment coverage in a population with the specified incidence and prevalence, and normalized to the specified recent incidence. The FRR in treated subjects, $P(R|tx)$ is simply the binomially estimated probability that treated subjects infected for longer than T would produce a recent result. In the CEPHIA Evaluation Panel, all treated subjects were virally suppressed, resulting in an estimate of $P(R|tx) = 0$ in all cases where a supplemental viral load threshold is applied. In real-world populations, it is likely that a certain (unknown) proportion of treated subjects would be virally unsuppressed and that the FRR in treated subjects would therefore be nonzero.

$$FRR = c \cdot P(R|tx) + (1 - c) \cdot \int_{T}^{\infty} \rho(t)P_R(t) \ dt,$$

where c is the treatment coverage, $\rho(t) = \frac{f(t)}{\int_{T}^{\infty} f(t)dt}$ and $f(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$.

For surveillance applications, utility is defined in terms of the SE on the incidence estimator.[17] To assess this, MDRI and context-specific FRR were calculated for a range of "recent/nonrecent" discrimination thresholds, and the variance on the incidence estimator was calculated for the specified epidemiological contexts, assuming demonstrative simple random samples of 10,000 individuals in scenarios A and B, and 3000 in scenario C. A key difference between a previously constructed scenario[10] and this analysis is in the HIV-positive case definition, which previously required a fully developed Western blot, and in this case is expanded to all individuals who test positive on a qualitative nucleic acid test (NAT) assay (threshold of detection 30 HIV-RNA copies/mL). This has ramifications for MDRI, which was estimated using EDDIs in which detectable infection is defined as positivity on a hypothetical viral load assay with a detection threshold of 1 copy/mL. Reported MDRI estimates must be interpreted against this reference standard, and for the purposes of incidence estimation, MDRI was adjusted to account for the sensitivity of the HIV screening algorithm (ie, 4.8 days shorter than reported MDRI to account for the diagnostic delay of the NAT screening assay).

MDRI, FRR, and the variance of the incidence estimate in the hypothetical scenarios were computed using the R package *inctools*.[18]

Linear regression was used to assess the correlation between ARCHITECT S/CO readings and LAg optical density (ODn) readings on the untreated subset of the evaluation panel. Binomial logistic regression models were used to assess the predictive value of ARCHITECT S/CO values for (1) duration of infection ≤ 1 year, and (2) LAg ODn values ≤ 1.5 (the conventional recency discrimination threshold). The logistic regression models had the following form:

$$g(P_R) = \beta_0 + \beta_1 X + \beta_2 X^2$$

with P_R the probability of recency (duration of infection ≤ 1 year or LAg ≤ 1.5), g() the logit link function, and X the ARCHITECT S/CO measurements.

RESULTS

Figure 2 shows, in 4 panels customized to the apparent dynamic range of each assay, the MDRI (y-axis) as a function of recent/nonrecent discrimination threshold (x-axis, range chosen to yield MDRI values from approximately 50 to approximately 400 days). A viral load threshold of 100 HIV RNA copies/mL is applied in all cases. MDRI is tuneable by adjusting the recent/nonrecent discrimination threshold, but FRR increases with increasing MDRI.

ARCHITECT S/CO measurements are highly correlated with LAg normalized ODn measurements (Pearson correlation coefficient of 0.796). Low ARCHITECT
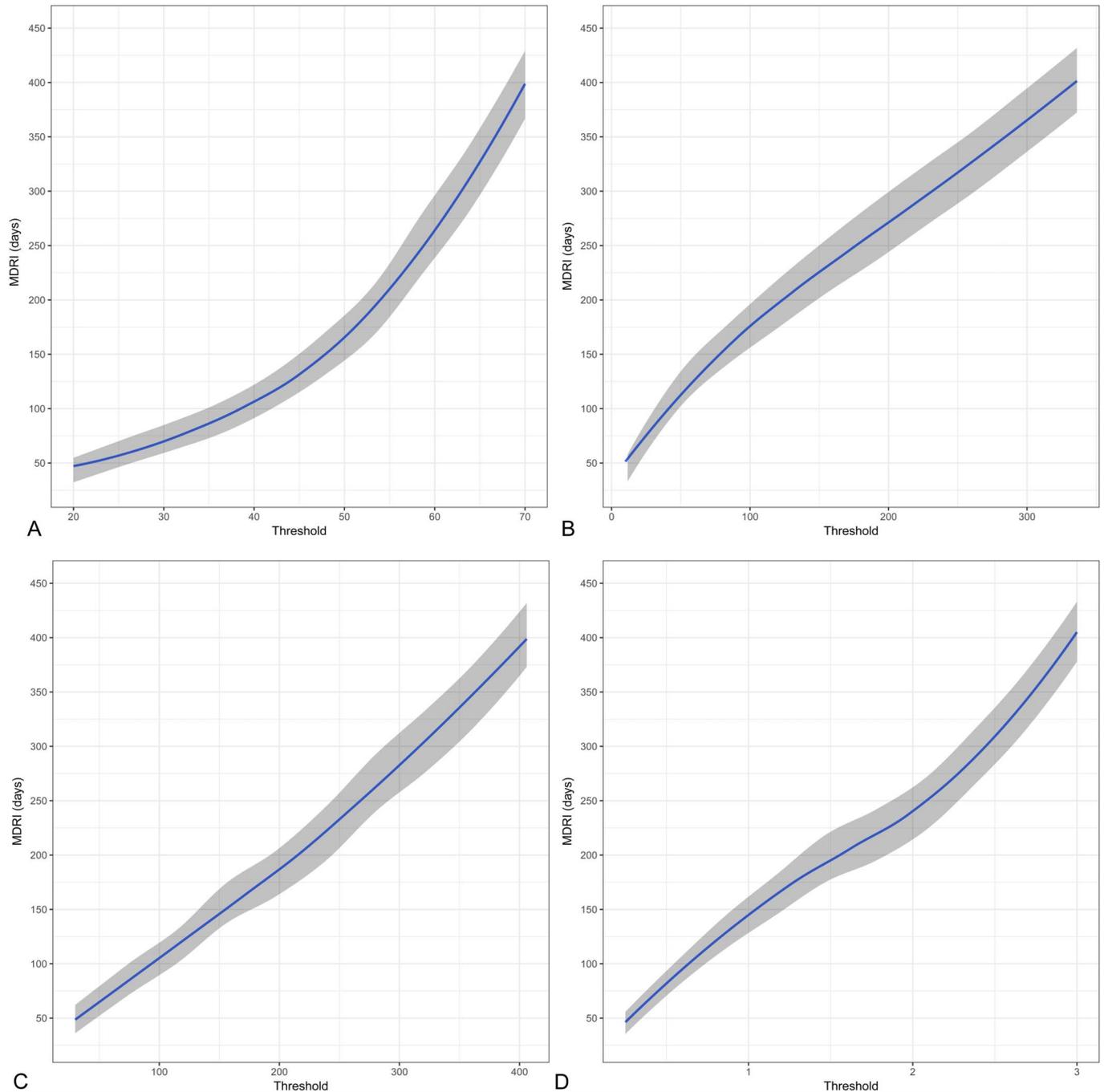
FIGURE 2. Mean duration of recent infection (*y*-axis), as a function of recency discrimination threshold (*x*-axis, range chosen to yield MDRI values from approximately 50–400 days). A, Minimally diluted ("untreated well") of the Ortho VITROS anti–HIV-1+2 assay avidity modification. B, Minimally diluted ("untreated well") of the Abbott ARCHITECT HIV Ag/Ab Combo assay avidity modification. C, The completely unmodified (manufacturer IFU) Abbott ARCHITECT HIV Ag/Ab Combo assay. D, (As a reference) the Sedia HIV-1 limiting antigen avidity EIA (LAg).

measurements are highly "predictive" of low LAg measurements, as shown in Figure 3, a scatterplot and linear regression model for LAg ODn vs. ARCHITECT S/CO (panel A) and binomial logistic regression for the probability of obtaining an ODn ≤1.5 (the conventional recency discrimination threshold) and duration of infection less than 1 year (panel B). Table 1 further reports that 78% of untreated specimens in the evaluation panel with S/CO readings <200
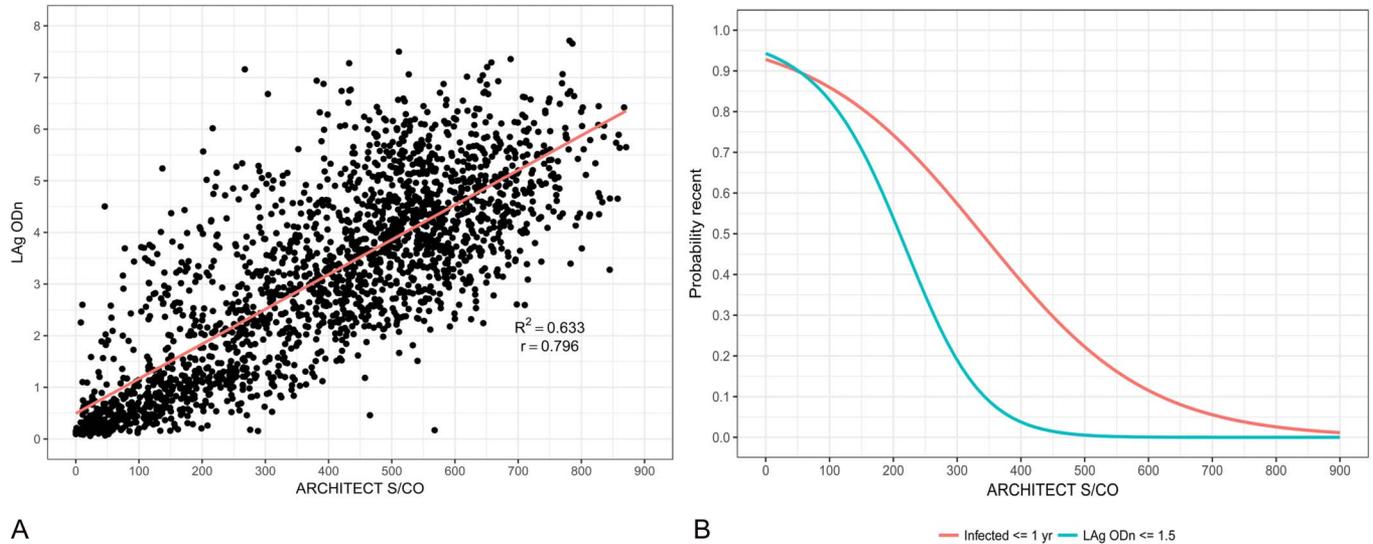
FIGURE 3. Low ARCHITECT S/CO measurements "predict" low LAg ODn measurements and short duration of infection. A, Scatterplot of ARCHITECT (*x*-axis) and LAg (*y*-axis) readings, and linear regression model. B, Binomial logistic regression models for the probability of obtaining an LAg "recent" result (ODn ≤ 1.5) and probability of duration of infection ≤1 year.

and viral load measurements >100 copies/mL produce LAg ODn values ≤1.5, and 87% of these specimens were drawn within 1 year of EDDI. Low quantitative readings from the diagnostic assay, together with the above-threshold viral load, therefore appear to obviate the need for additional staging assays.

Figure 4 summarizes performance metrics for the minimally diluted "untreated wells" of the VITROS and ARCHITECT Avidity assays, the unmodified ARCHITECT diagnostic assay, and LAg as part of an RITA in surveillance applications. Panel A shows context-specific FRR in epidemiological scenario A, against MDRI (encoding recency discrimination threshold) and panel B shows the relative SE (RSE) on the incidence estimate in scenario A against MDRI. Comparison of the diluted VITROS and ARCHITECT platforms shows a more rapid rise in FRR for the VITROS platform, reaching more than 5% when MDRI is approxi-

mately 1 year. For this reason, the ARCHITECT platform was selected for further evaluation in entirely unmodified form (manufacturer's IFU). The results indicate that an RITA based on the unmodified ARCHITECT produces a marginally higher FRR than the diluted version, and that its performance for surveillance purposes (as captured in the precision of incidence estimates) is very similar to that of a RITA based on the widely used LAg assay. In the range of thresholds explored (MDRI ranging from approximately 50 days to approximately 400 days), the lowest RSE on the incidence estimate (in scenario A) achieved with a LAg-based RITA was 13.6%. With the ARCHITECT Avidity (untreated well) it was 16.0%, with the VITROS Avidity (untreated well) 22.7%, and with the unmodified ARCHITECT 17.0%.

Table 1 shows performance metrics for a range of S/CO thresholds. MDRI, context-specific FRR, and RSE on the incidence estimate under epidemiological scenarios A, B, and

TABLE 1. Performance Characteristics of the Unmodified ARCHITECT Assay

S/CO Threshold <	VL Threshold >	MDRI* (95% CI), Days	Scenario A		Scenario B		Scenario C		Proportion LAg ODn ≤ 1.5, %	Proportion Time since EDDI < 1 yr, %
			FRR†, %	RSE‡, %	FRR†, %	RSE‡, %	FRR†, %	RSE‡, %		
50	100	71 (57 to 86)	0.0	24.7	0.0	35.8	0.0	38.7	94	90
100	100	105 (90 to 121)	0.1	20.4	0.1	29.9	0.2	32.3	91	91
150	100	150 (132 to 169)	0.1	17.4	0.2	25.4	0.3	27.2	85	91
200	100	186 (165 to 208)	0.4	17.0	0.5	24.4	0.6	25.2	78	87
250	100	232 (207 to 257)	0.7	17.0	0.9	23.9	0.6	22.5	72	85
300	100	281 (254 to 310)	1.3	19.0	3.6	37.1	0.7	20.4	63	81
350	100	337 (307 to 366)	1.7	18.3	5.0	39.4	1.1	19.0	57	77

*MDRI relative to hypothetical viral load assay with detection threshold of 1 copy/mL.
†Context-specific FRR, estimated for the specified epidemiological scenario.
‡Relative SE on the incidence estimate for the specified epidemiological scenario. MDRI of RITA adjusted for screening with a qualitative NAT assay (detection threshold 30 copies/mL).
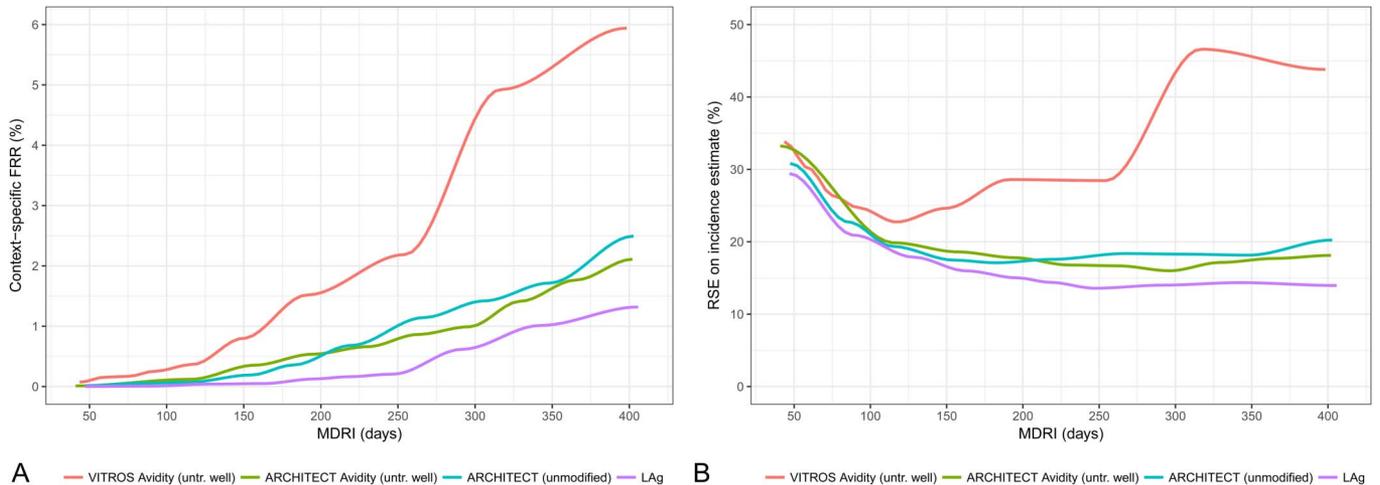
FIGURE 4. Context-specific FRR and RSE on incidence estimate in a demonstrative epidemiological context (30% HIV prevalence, 1.5% per annum incidence, 80% treatment coverage) and using an unstructured sample of 10,000 respondents. A, Context-specific FRR by MDRI. B, Precision of incidence estimate by MDRI.

C. For estimating the RSE on the incidence estimate, MDRI was adjusted for the sensitivity of a NAT-based screening algorithm. In scenario A, an S/CO threshold of 200–250 appears close to optimal, yielding MDRIs of 186 days (95% CI: 165 to 208) and 232 days (207 to 258), respectively, context-specific FRRs of 0.4% and 0.7%, and RSEs on the incidence estimate of 17.0%. In scenario B, an S/CO threshold of 250 produced the highest precision in the incidence estimate, and in scenario C, a threshold of 350 (RSEs of 23.9% and 19.0%, respectively). Precision declines at low S/CO thresholds as a result of shorter MDRI and at high thresholds as a result of higher FRRs. Lower thresholds may be useful for clinical staging applications.

DISCUSSION

In Figure 2 we can see that, although the actual ranges of plausible thresholds vary considerably between platforms, the chemiluminescent diagnostic assays have sufficient dynamic range to support simple threshold-based definitions of recent infection. For surveillance applications, such as case-based surveillance, where central laboratories apply recency staging assays to a large number of cases identified in a health system, blood banking applications, and large population-based surveys, the use of custom recency staging assays appear to offer little benefit over interpretation of the information contained in routinely available diagnostic results. Optimal threshold choice depends on the specific epidemiological and methodological context of application, but would involve a similar trade-off between MDRI and FRR as that demonstrated in Table 1 and Figure 4. This was evident in all 3 epidemiological scenarios, with both MDRI and context-specific FRR increasing with higher thresholds, and different thresholds producing the "optimal" trade-off (ie, highest precision incidence estimate). Thresholds should therefore be chosen to optimize performance in the specific context of any surveillance application.

It is worth noting that even state-of-the-art staging algorithms exhibit disappointing performance in surveillance applications, with precise incidence estimates emerging only at extraordinarily high sample sizes or high values of incidence.[10,19] Recency staging data is unlikely ever to be the only source of incidence-related information—meaning that it should be used in conjunction with estimation procedures that also use the demographic (age and time) structure of prevalence data and mortality.[20,21]

In interpreting data for individual use, it is important to consider an inherent limitation of any diagnostic test: the potential for false positive results. Although an HIV diagnosis would not be based on the outcome of a single test, the lack of confirmatory tests for "recent" infection means that a single test is often used for recency determination. It is likely that a false positive or nonspecific reaction would give a low S/CO reading, and thus would be more likely to be misclassified as a recent infection. Consideration must be given to confirming the antibody-positive status of the individual. In addition, the very nature of fourth generation combined antigen/antibody tests implies that antigen-only positive samples may also be detected. These results, which are almost certainly truly recent infections, may also generate a low S/CO depending on a number of factors, including the amount of antigen present and whether this antigen has complexed with developing antibody, and should not be discarded over the absence of antibody reactivity, as long as HIV infection is confirmed by, eg, NAT. Before interpreting a test for recency, it is critical to ensure that the HIV diagnosis is confirmed by local testing algorithms and criteria that identify nonspecific reactivity and differentiate antibody and antigen reactivity.

One must not dismiss the performance of the unmodified ARCHITECT demonstrated in the present analysis, simply because it is not quite as good as what can currently be achieved by other staging algorithms based on more complex protocols. Substantial logistical and cost advantages, in both routine contexts and population-based surveys, could

be achieved using a single serological assay for diagnostic and staging purposes, especially in the light of the high-throughput, automated nature of these platforms. However, it is recognized that platforms like ARCHITECT and VITROS are expensive, and the benefits of HIV recency testing in conjunction with diagnostic testing would be most apparent in settings where these are already available in laboratories and used for multiple diagnostic or monitoring assays. In settings where routine diagnosis is performed using point-of-care rapid tests, and laboratory confirmation is the exception, these advantages may be less relevant.

Throughout Western Europe, the United States, Canada, Australia, and in many other countries, the standard of laboratory diagnosis for HIV infection includes testing using a fourth-generation diagnostic assay (ie, able to detect HIV p24 antigen and anti–HIV antibody simultaneously). This testing is now available using large automated platforms, such as VITROS and ARCHITECT, as part of a wide panel of microbiological, virological, and chemical markers. Consequently, such platforms have been widely adopted at local laboratories to provide diagnostic support to facilities in their catchment areas. By contrast, LAg testing is restricted to only a few sites (1 in the United Kingdom). Using the United Kingdom as an example, anti-HIV-1 positive specimens identified locally, often by testing on ARCHITECT or VITROS, are aliquoted and shipped to a central laboratory for LAg testing, and reporting from the central laboratory back to the local laboratory, which in turn passes the results on to the requesting clinic. This process adds a minimum of 2 weeks to the determination of HIV recency status. Interpreting recency from the ARCHITECT or VITROS result would reduce delays and costs, by avoiding additional shipping and testing. In many international studies, specimens are collected and shipped to central laboratories for HIV confirmatory testing. Our study suggests that if such confirmatory testing is performed using platforms such as ARCHITECT or VITROS, the recency status could be determined contemporaneously—negating the need for further recency testing.

Estimating time since infection based on HIV biomarker progression is made difficult by the complex nonlinear growth exhibited by immune markers of disease progression. The absence of a robust model for capturing intersubject variability leads to an inability to formally measure the precision of the timing estimate. Nevertheless, simply "looking up" the assay result on the threshold–MDRI curve, and noting that "*X* reactivity level of this diagnostic assay is one that subjects, on average, spend *Y* time beneath," provides clinically meaningful information. In principle, the most natural and coherent way to provide realistic estimates for individual-level times since infection would involve the use of testing histories and other contextually derived "prior information" in a Bayesian framework. This would require a high level of confidence that the chosen model of biomarker growth correctly captures complex intersubject variability.

Critical clinical decisions—such as whether to offer expedited antiretroviral treatment—could be based on the transformation, through a calibration curve, of a test result into a timescale (only loosely interpreted as an estimate of time since infection). Currently, a simple "likely recent/non-recent"

result is usually fed back to clinician and patient, qualified by a mean time meant by recent, with caveats outlining a number of known confounders of assay performance. Our proposal uses the specific assay result, rather than simply dichotomizing results into a recent/nonrecent classification based on a single threshold that may be far from the particular patient's result. It therefore offers an interpretation that is much more informative.

This analysis provides the first evidence that unmodified diagnostic assays can provide meaningful staging information applicable in clinical and surveillance settings. It must be noted that these assays are primarily intended for HIV diagnosis and that these additional interpretations are "off-label" uses of the assay. Previous data have shown that HIV incidence assays, that often work at the limits of detection, are relatively robust to specimen handling conditions,[22] and CEPHIA evaluations have not shown anomalous results attributable to specimen handling. However, the possibility does exist that specimen collection methods, handling, and storage—even within the manufacturer's IFU—affect the quantitative results of diagnostic assays, and therefore infection staging interpretations of these results.

Tests that intrinsically serve as part of both a diagnostic algorithm and a staging algorithm offer obvious practical advantages and, perhaps more significantly, create a genuine market advantage. This may stimulate investment in the otherwise unattractive sector of staging assays.[15] Although customized staging assays will remain useful in many contexts, the ability to extract staging information from diagnostic assays constitutes a significant advance for the fields of clinical staging and incidence surveillance.

REFERENCES

1. Brookmeyer R, Quinn TC. Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests. *Am J Epidemiol.* 1995;141:166–172.
2. 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;22:945–952.
3. McWalter TA, Welte A. A comparison of biomarker based incidence estimators. *PLoS One.* 2009;4:e7369.
4. Guy R, Gold J, Calleja JMG, et al. Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: a systematic review. *Lancet Infect Dis.* 2009;9:747–759.
5. Mastro TD, Kim AA, Hallett T, 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.
6. Busch MP, Pilcher CD, Mastro TD, et al. Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation. *AIDS.* 2010;24:2763–2771.
7. Kassanjee R, McWalter TA, Bärnighausen T, et al. A new general biomarker-based incidence estimator. *Epidemiology.* 2012;23:721–728.
8. Laeyendecker O, Brookmeyer R, Cousins MM, et al. HIV incidence determination in the United States: a multiassay approach. *J Infect Dis.* 2013;207:232–239.
9. Kassanjee R, Pilcher CD, Keating SM, et al. Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. *AIDS.* 2014;28:2439–2449.
10. Kassanjee R, Pilcher CD, Busch MP, et al. Viral load criteria and threshold optimization to improve HIV incidence assay characteristics. *AIDS.* 2016;30:2361–2371.
11. Murphy G, Pilcher CD, Keating SM, et al. Moving towards a reliable HIV incidence test— current status, resources available, future directions and challenges ahead. *Epidemiol Infect.* 2016;145:925–941.
12. Chawla A, Murphy G, Donnelly C, et al. Human immunodeficiency virus (HIV) antibody avidity testing to identify recent infection in newly diagnosed HIV type 1 (HIV-1)-seropositive persons infected with diverse HIV-1 subtypes. *J Clin Microbiol.* 2007;45:415–420.
13. Suligoi B, Rodella A, Raimondo M, et al. Avidity Index for anti-HIV antibodies: comparison between third- and fourth-generation automated immunoassays. *J Clin Microbiol.* 2011;49:2610–2613.
14. Fiebig EW, Wright DJ, Rawal BD, 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.
15. Morrison C, Homan R, Averill M, et al. *Assays to Measure HIV Incidence: Updated Global Market Assessment and Estimated Public Health Value.* Geneva: FHI 360 and FIND; 2017.
16. Keating SM, Kassanjee R, Lebedeva M, et al. Performance of the Bio-Rad Geenius HIV1/2 supplemental assay in detecting "recent" HIV infection and calculating population incidence. *J Acquir Immune Defic Syndr.* 2016;73:581–588.
17. Kassanjee R, McWalter TA, Welte A. Short communication: defining optimality of a test for recent infection for HIV incidence surveillance. *AIDS Res Hum Retroviruses.* 2014;30:45–49.
18. Welte A, Grebe E, McIntosh A, et al. *Inctools: Incidence Estimation Tools.* R package version 1.0.10. 2016. Available at: https://cran.r-project.org/package=inctools.
19. Kassanjee R. *Characterisation and Application of Tests for Recent Infection for HIV Incidence Surveillance [PhD thesis].* Johannesburg, South Africa: School of Computational and Applied Mathematics, University of the Witwatersrand; 2014.
20. Mahiane GS, Ouifki R, Brand H, et al. A general HIV incidence inference scheme based on likelihood of individual level data and a population renewal equation. *PLoS One.* 2012;7:e44377.
21. Grebe E, Huerga H, Van Cutsem G, et al. *Cross-Sectionally Estimated Age-specific HIV Incidence Among Young Women in a Rural District of Kwazulu-Natal, South Africa [Internet].* Durban, South Africa: 21st International AIDS Conference; 2016. Available at: http://programme.aids2016.org/PAGMaterial/eposters/0_6506.pdf.
22. Laeyendecker O, Latimore A, Eshleman SH, et al. The effect of sample handling on cross sectional HIV incidence testing results. *PLoS One.* 2011;6:e25899.