Response to amfAR's Report "Blinded by Our Own Data—Recency Testing in PEPFAR"

June 7, 2022

Jessica Justman, MD
Suzue Saito, PhD
Yen Pottinger, PhD
ICAP at Columbia University

On June 1st, 2022, amfAR published a report on PEPFAR's use of HIV recency testing as part of a larger strategy to focus HIV testing resources to areas and populations with the highest transmission rates. The report argues that HIV recency testing within HIV testing services is too inaccurate to produce meaningful data to describe ongoing transmission patterns and influence programmatic decisions. We respectfully and strongly disagree with the report's premise, details and conclusions.

We provide here example use cases and implementation data from HIV recent infection surveillance in a point by point response. While the COVID-19 pandemic has slowed implementation in most countries, we have nonetheless worked closely with national governments, local and international NGOs, and local PLHIV advocacy groups to scale recent infection surveillance. Our experience has shown both the relevance and importance of recent infection surveillance to country programs and the communities they serve.

ICAP at Columbia University has received funding from PEPFAR through the Centers for Disease Control and Prevention since November of 2018 to provide technical assistance and implementation support to 15 country programs to launch HIV recent infection surveillance by rolling out rapid tests for recent infection (RTRI) in routine HIV testing services and collecting and using the results for program improvement and public health response. To date, we have supported training of >6000 certified testers and activation of over 2500 testing sites. These testers have conducted ~160,000 RTRI tests across these supported programs. We have additionally supported ministries of health in rolling out monthly quality control and semi-annual and annual proficiency testing programs to sustain high quality testing. Finally, we have supported launching of surveillance dashboards at national HIV programs that ministerial departments use along with other program data to make program decisions.

There are five main "challenges" the amfAR report describes about recency testing and the surveillance system developed with it. We provide a response to each below.

1. ***Sensitivity:*** The report notes the following: *"Recency testing has poor sensitivity for recent infections: Studies routinely find that recency testing fails to diagnose individuals with true recent infections (infected less than six months) between 35% and 68% of the time. This is inherent to the methodology applied to detect recent infections and not an issue of implementation that can be resolved by improved training.[1]" [page 2, amfAR report, May 2022]*

[1] Kassanjee R, Pilcher CD, Keating SM, et al., 2014, found LAg-avidity correctly identified samples from individuals infected <6 months ~65% of the time, with variation by HIV subtype, and from individuals infected 6-12 months ~15% of the time, available at: https://doi.org/10.1097/QAD.0000000000000429; Konikoff J, Brookmeyer R, Longosz AF, Cousins MM, Celum C, et al., 2013, found LAg-avidity failed to identify 57% of samples from individuals infected <6 months and 69% <12 months based on an ODn<1.5 threshold, available at: https://doi.org/10.1371/journal.pone.0082772; CEPHIA (Grebe E, Facente SN,

This statement reflects a misunderstanding of the intended use of HIV recency tests for recent infection surveillance. **Sensitivity and specificity are not applicable concepts to HIV recency tests because they are not diagnostic tests.** Recency tests cannot be diagnostic tests because they are based on antibody avidity or binding strength and not based on absence or presence of HIV-specific antibodies. There is substantial inter-personal variability in avidity and thus recency can only be useful at a population level, using an average window period in which the majority of individuals crosses the antibody avidity threshold pre-specified by the test. For the RTRI, that period is 6 months, but to account for a small but important group that crosses the threshold after 6 months, we use 12 months as the cut-off point. Those who are classified as recent acquired HIV <12 months from the diagnosis date while those who are classified as long-term acquired HIV ≥12 months from the diagnosis date.

It is accepted that recent infection surveillance will miss some recent infections (i.e., false long-terms, as described in the amfAR report, page 14, figure 3). This is because there are no known characteristics that are associated with someone with a true recent infection crossing the avidity threshold earlier or later. It is a largely random process. We are, however, focused on ensuring that the RTRI recents that are identified are as accurate as possible, that is to minimize false recents. They are intended to represent the true population of recents in programmatically relevant characteristics, i.e., geographical distribution, age, sex, and key or priority population status). The largest impediment to this effort is the substantial number of re-testers noted in the HIV testing programs—those who present as undiagnosed and ART naïve but are in fact already aware of their HIV positive status and are on antiretroviral therapy IART). In order to appropriately exclude such individuals, viral load testing is conducted on those who are RTRI recent. The inclusion of viral load testing with RTRI is known as the recent infection testing algorithm or RITA. The literature has shown that RITA can correctly classify >90%[2] of RTRI recents and greatly minimizes false recents (see more on this under point 2).

The objectives of recency testing and HIV recent infection surveillance is NOT to capture ALL recent infections. The objectives are 1) to detect areas and subpopulations with high transmission rates, and 2) to provide data to assess and guide national HIV testing strategies. As such it is meant for population surveillance rather than individual diagnostic testing. When implemented at scale, the recent infections that are identified and characterized can be informative to addressing a country's epidemic and its programmatic response (see some examples under point 4).

Hampton D, Cheng C, Owen R, Keating SM, Pilcher CD, Welte A, Busch M, Murphy G & Consortium for the Evaluation and Performance of HIV Incidence Assays), 2019, found the Asante HIV-1 Rapid Recency Assay at a band intensity of 3.0 failed to diagnose many samples as recent, available at: https://doi.org/10.5281/zenodo.3509834; Galiwango RM, Ssuuna C, Kaleebu P, et al., December 2021, found that RTRI failed to diagnose 30% and 50% of samples from individuals infected <6 months at different laboratories, available at: https://doi.org/10.1089/AID.2020.0279.

[2] Rice BD et al., 2020, used clinical records searches and ARV detection (separately) to identify recent mis-classifications 9.1% and 4.2% of the time respectively in Kenya, available at: https://doi.org/10.1002/jia2.25513;

2. ***Specificity:*** The Report notes the following: *"Recency testing has poor specificity for recent infections: Recency testing alone in the field returns false positive results between 15% and 90% of the time as individuals currently on treatment, people who have advanced HIV disease, and other factors affect the accuracy of the test. Even when recency tests are combined with viral load testing—as PEPFAR recommends—results are still inaccurate between 5% and 30% of the time.[3]"[page 4, amfAR report, May 2022]*

   The vast majority of PEPFAR-supported HIV recent infection surveillance utilizes RITA. The 15%-90% false positivity rate, therefore, is not applicable. RITA correctly classifies recent infections with >90% accuracy. The range of 5-30% is also misleading as the upper bound of 30% is derived from very small numbers of recent cases identified in the general population in Population HIV Impact Assessment (PHIA) surveys and is therefore an inherently unstable estimate of false positivity. In most larger clinical facility-based studies, such as by Rice, et al 2020, **RITA recent cases are >90% correctly classified** and hence a useful indicator to flag high transmission areas and subpopulations.

3. **Inter-observer reliability***:* The Report notes the following*: Inter-reliability of recency testing is suspect: Recency testing results must be read carefully to deliver consistent results across different sites. Inconsistency in the reading of results undermines the comparability of the results. However, research has shown that even at a laboratory level using the point-of-care tests PEPFAR is deploying, the inter-reliability of results was only around 70%. PEPFAR's planned program intends for recency testing to be deployed at all testing sites in all PEPFAR program countries rather than being based out of labs, creating enormous opportunity for inconsistencies in the reading of results to go undetected.[4] [page 4, amfAR report, May 2022]*

   The Report cites one small study from a single country involving two laboratories and only 85 samples. By contrast, in 143 trainings supported by ICAP, 6130 individuals have been certified as recency testers after achieving 100% accuracy with at least 13 samples each. The samples used had varying band intensities as well to mimic implementation realities. In implementation of recency surveillance, we have additionally supported 9 rounds of proficiency testing in 3 countries. Of nearly 3000 testers that underwent proficiency testing, 94% of the testers passed 5-panel proficiency testing on the first try. All but 6 have passed in the second try. In addition, the test verification line has performed remarkably well vis-a-vis the national HIV diagnostic

[3] Zhu Q, Wang Y, Liu J, Duan X, et al., 2020, found 16.7% of "recent" results were misclassified based on LAg+viral load <1000 alone, available at: https://doi.org/10.1016/j.ijid.2020.09.1421; Voetsch A, Duong Y, Stupp P, et al., 2021, described how ARV detection assays found 15.7% of "recent" results were misclassified based on LAg+viral load<1000 alone, available at: https://doi.org/10.1097/QAI.0000000000002707; Rice BD et al., 2020, using clinical records searches and ARV detection (separately) identified recent mis-classifications 9.1% and 4.2% of the time respectively in Kenya, available at: https://doi.org/10.1002/jia2.25513; HSE Health Protection Surveillance Centre, 2020, found 62.5% of recent results were misclassified after LAg+viral load<1000 based on clinical records, CD4 count <200, AIDS-defining illness, or prior PEP use, available at: https://www.hpsc.ie/a-z/hivandaids/hivdataandreports/2018reports/HIV_2018_recentinfection.pdf.

[4] Galiwango RM, Ssuuna C, Kaleebu P, et al., December 2021.

algorithm (See Table 1). **We believe the volume (over 6000 individuals), consistency and breath of these implementation data, outweigh the limited data cited by the Report.**

Table 1: RTRI Sensitivity to National HIV Diagnostic Algorithm, ICAP 2021

Testing Quality: Sensitivity >99% Across Countries to Date

	Country A	Country B	Country C[1]	Country D	Country E	Country F	Country G	Country H
Tested by RTRI	1,926	13,749	7,635	1121	23601	15,418	603	5,743
RTRI Invalid	1	0	12	0	14	1	0	0
RTRI Negative	26	16	6	5	105	20	2	28
Valid RTRI Test Results	1,899	13,733	7,522	1,121	23,482	15,397	603	5,743
RTRI Sensitivity (%)	99	99.9%	99.9%	99.6%	99.7%	99.9%	99.7%	99.5%

[1]101 clients had no documented result.

4. **Biased program data:** The Report indicates the following*: "Routine program data are likely to bias results in unpredictable ways: When published studies include recency testing to assess incidence, there are attempts to control for biases in the population and geographic regions that are included in the study. By contrast, such controls are not possible in routine program data. Because of this, high-performing HIV testing programs with good testing uptake among specific populations at greater risk of HIV (e.g., youth, key populations) may appear to be higher transmission areas than similar geographic regions that have poor HIV testing uptake among these populations." [page 4, amfAR report, May 2022]*


NEWS Times of Eswatini – Thursday, January 20, 2022 PAGE 13

HIV stats worst in Kwaluseni Inkhundla

TOP 10 TINKHUNDLA AND TESTING SITES IN RECENT INFECTION TREATMENT ALGORITHM (RITA) FROM JULY 2019 TO OCTOBER 2021

TOP 10 TINKHUNDLA

TOP 10 TESTING SITES

CLIENTS RE-CLASSIFIED LONG-TERM BY REGION AGE AND GENDER

TESTIMONIALS

SOME OF THE ISSUES RAISED BY CONSTITUENTS

Need for more participation of students, staff in campaigns

Eswatini Government

Again, the objective of HIV recent infection surveillance is not estimation of incidence. It is understood that recency testing does not yield broadly generalizable data about HIV disease burden. The objective of recent infection surveillance is to find high transmission areas and subpopulations and to inform HIV prevention and testing strategies. **Recent infection surveillance provides one source of program data that, when combined with other program data, yields important insights that guide adjustments to programs as needed**.

For example, in three sub-Saharan countries, recency tests have demonstrated that out-patient

departments, while not necessarily a high-yield testing location in terms of percent positivity, are an important testing location for recent infections based on testing volume (CROI 2022 Abstract #90).

Additionally, recency testing has shown that community-based HIV testing picks up many retesting clients—those who are already known positive and on ART and are retesting—and thus community-based testing may not be the best testing strategy to identify newly diagnosed individuals (data available upon request).

Further, in Eswatini, the National Program has compared recent infection data by residence vs. place of diagnosis and identified underserved communities in prevention and testing services (See above, local newspaper article).

Finally, we have found that identifying the proportion of long-term infections within newly diagnosed individuals is also important. Across the countries where recent infection surveillance was implemented, it was noted that ~90% of new diagnoses are long-term infections. Certainly, some of them are false long-terms but the majority are not and those long-term infected individuals who are presenting late in the course of HIV will continue to hinder the performance of the national programs to reach and sustain epidemic control. It is important to track surveillance trends and characterize of such individuals and fine tune prevention and testing strategies to facilitate earlier diagnosis (IAS 2021 Abstract #54). These are practical programmatic insights that have been helpful for national programs.

5. **Real-time:** The Report notes the following: *"Recency testing doesn't provide real-time information: Recency based incidence estimation is not a real-time incidence estimation. Because of high false-recent results and the demography of individuals who will initially test recent, recency-based surveillance is more indicative of transmission patterns that were active nearly two years ago, rather than real time. While incidence studies can reflect these realities, a real-time surveillance system premised on the idea of real-time or near-time response is actually operating well behind when transmission was taking place.[5] Additionally, the geographic location where individuals are diagnosed does not necessarily align with where they contracted HIV."[page 4, amfAR Report, May 2022]*

   The authors of the Report demonstrate a misunderstanding of the use of the rapid recency testing. It is NOT being used for HIV incidence surveillance and is not a point-of-care substitute for the LAg assay which is used for incidence estimation in nationally-representative surveys such as the PHIA surveys. The testing kit currently used in PEPFAR programs is the Asante™ HIV-1 Rapid Recency® Assay. This test identifies recent infections that were likely acquired in the prior 6-12 months, and not, as described in the Report, in the prior two years. The two-year time frame has been used in studies looking at LAg avidity testing and other recency assays for

[5] Konikoff J, Brookmeyer R, Longosz AF, Cousins MM, Celum C, et al., 2013, described the "shadow" for HIV recency testing as "a measure of how far back in time incidence is being estimated" based on test characteristics, available at: https://doi.org/10.1371/journal.pone.0082772.

incidence estimation, but NOT for RTRI use for recent infection surveillance. RTRI results are then combined with viral load results that are generally from VL tests conducted no more than 1-2 weeks after the RTRI test. These RITA results are then displayed on dashboards that refresh as data are uploaded. This timeline is certainly **near real time in the context of when HIV testing takes place.** Data are examined at least weekly and interpreted and used by programs on a monthly to quarterly basis during work-planning and data review sessions.

In conclusion, recency data are regarded as **an additional form of program data** and like other program data, recency data are viewed in the context of the volume and completeness of data and are combined with other program data to make interpretations. Finally, as described in point 4, most recent infection surveillance does capture both place of diagnosis and residence information, strengthening recency data and highlighting important program gaps.