Analysis of Viral Diversity in Relation to the Recency of HIV-1C Infection in Botswana

Sikhulile Moyo[1,2☯]*, Alain Vandormael[3☯], Eduan Wilkinson[3], Susan Engelbrecht[1,4], Simani Gaseitsiwe[2,5], Kenanao P. Kotokwe[2], Rosemary Musonda[2,5], Frank Tanser[3], Max Essex[2,5], Vladimir Novitsky[2,5☯], Tulio de Oliveira[3,6,7☯]

1 Division of Medical Virology, Stellenbosch University, Tygerberg, South Africa, 2 Botswana-Harvard AIDS Institute Partnership, Gaborone, Botswana, 3 Wellcome Trust Africa Centre for Health and Population Studies, Dorris Duke Medical Research Centre, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, 4 National Health Laboratory Services (NHLS), Tygerberg Coastal, South Africa, 5 Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America, 6 Research Department of Infection, University College London, London, United Kingdom, 7 College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

☯ These authors contributed equally to this work.
* sikhulilemoyo@gmail.com

Abstract

Background

Cross-sectional, biomarker methods to determine HIV infection recency present a promising and cost-effective alternative to the repeated testing of uninfected individuals. We evaluate a viral-based assay that uses a measure of pairwise distances (PwD) to identify HIV infection recency, and compare its performance with two serologic incidence assays, BED and LAg. In addition, we assess whether combination BED plus PwD or LAg plus PwD screening can improve predictive accuracy by reducing the likelihood of a false-recent result.

Methods

The data comes from 854 time-points and 42 participants enrolled in a primary HIV-1C infection study in Botswana. Time points after treatment initiation or with evidence of multiplicity of infection were excluded from the final analysis. PwD was calculated from quasispecies generated using single genome amplification and sequencing. We evaluated the ability of PwD to correctly classify HIV infection recency within <130, <180 and <360 days post-seroconversion using Receiver Operator Characteristics (ROC) methods. Following a secondary PwD screening, we quantified the reduction in the relative false-recency rate (rFRR) of the BED and LAg assays while maintaining a sensitivity of either 75, 80, 85 or 90%.

Results

The final analytic sample consisted of 758 time-points from 40 participants. The PwD assay was more accurate in classifying infection recency for the 130 and 180-day cut-offs when compared with the recommended LAg and BED thresholds. A higher AUC statistic

confirmed the superior predictive performance of the PwD assay for the three cut-offs. When used for combination screening, the PwD assay reduced the rFRR of the LAg assay by 52% and the BED assay by 57.8% while maintaining a 90% sensitivity for the 130 and 180-day cut-offs respectively.

Conclusion

PwD can accurately determine HIV infection recency. A secondary PwD screening reduces misclassification and increases the accuracy of serologic-based assays.

1.0 Background

Identification of HIV infection recency is crucial for the accurate estimation of HIV incidence, the evaluation of the effectiveness of antiretroviral treatment (ART) programs, and the timely linking of HIV-infected individuals (and their partners) to treatment and care services [1–9]. The timing of infection can also be used to identify the immunological and virological characteristics of individuals who have recently acquired HIV and to characterize individuals who are putative transmitters in linked infections [10–14].

The longitudinal cohort design is currently recognized as the standard approach to identify new HIV infections [15–17]. However, frequent HIV testing at the population level is a logistically challenging, time-consuming, and expensive enterprise. For these reasons, large-scale surveillance programs are typically undertaken on a periodic basis of 12 or more months, making it difficult to ascertain the precise date of an HIV infection. Factors associated with illness, work commitments, temporary or cyclical migration, assumed knowledge of current HIV status, and the stigma associated with a positive status, among others, may decrease the frequency at which an eligible individual is captured for HIV testing [18–20]. On the other hand, the identification of new HIV infections is possible for experimental trials where relatively small cohorts (typically <500 individuals) are routinely tested on a weekly or monthly basis [21–23].

There is growing scientific interest in the use of cross-sectional sampling methods to identify individuals recently infected with HIV. Cross-sectional methods can mitigate the impact of infrequent testing and the high lost-to-follow-up rates that are associated with the longitudinal approach [24–28]. Biomarker data collected from cross-sectional sampling has also shown great promise in the ability to differentiate between recent and established HIV infections. Serological assays, for example, the Calypte Incidence Assay (BED) and Limiting Antigen assay (LAg), depend on the markers of evolution of the host immune response to HIV, such as antibody levels, avidity, isotype and proportion [29–35]. Attention is now turning to the improvement of assay-based methods and the use of multi-assay algorithms (MAA) to better predict HIV infection recency [36].

One area that is receiving increasing attention is the use of a viral diversity measure [11, 13, 37–41]. The majority of HIV infections are caused by the transmission of a single founder virus, resulting in a relatively homogeneous population of viral quasispecies during the early stage of HIV infection [42–45]. Due to the error prone nature of the Reverse Transcriptase (RT) enzyme and the host immune response to pressure, the virus is able to diversify rapidly over time. The approximately linear diversification of HIV in early infection [46] provides a rationale for using viral diversity as a marker for HIV infection recency [11, 39, 47, 48]. One example of a time-dependent, viral-based diversity measure is the pairwise nucleotide diversity (PwD). PwD measures the average number of pairwise nucleotide differences per site in DNA

sequences [11, 37, 38, 43, 49, 50]. Assays based on a measure of PwD should be less sensitive to the variability in immune responses modulated by HIV clade, host genetics and routes of transmission. However, viral-based assays are more challenging and costly to implement.

About 20–25% of HIV infections are caused by the transmission of multiple viral variants [43, 51–53]. The rate of HIV-1 super-infection could be comparable with the rate of primary HIV-1 infection [54], although super-infection is less frequent in the HIV-1C epidemic in South Africa [55]. Ignoring multiplicity of HIV infection could mislead analysis and lead to erroneous conclusions due to increased intra-host diversity in cases with multiple transmitted HIV variants, or in super-infection. Using intra-host viral sequences that represent HIV quasispecies provides an opportunity to identify phylogenetically distinct viral lineages and take into account multiplicity of HIV infection.

In this paper, we use data from a frequently tested longitudinal HIV-1C infection cohort (the "*Tshedimoso*" study from Botswana) for which the exact date of infection is known. We assess the accuracy of the PwD assay to correctly classify HIV infection recency, and compare its performance with the BED and LAg assays. Because of the high cost currently associated with single genome sequencing, we further investigate the use of a MAA (BED plus PwD or LAg plus PwD) to increase accuracy and maintain affordability. We also evaluate the addition of viral load (VL) as a covariate to the MAA algorithm. We discuss the potential of cross-sectional, biomarker information and the use of MAAs as an affordable and accurate alternative to the longitudinal cohort approach.

2.0 Methods

2.1 Participants and specimens

The data comes from 854 time points and 42 study participants enrolled into a primary HIV-1 subtype C infection longitudinal cohort in Botswana (the "Tshedimoso" study) from April 2004 to April 2008 [56, 57]. Recent HIV-1 infections were identified by a positive HIV-1 RNA test combined with a negative HIV-1 serology in double enzyme immunoassay [58] or by applying a 2-step testing algorithm using the Vironostika HIV-1 Plus O Microelisa System (bioMérieux, Durham, NC) [59]. Acutely infected participants had weekly visits for the first 2 months, biweekly visits for the next 2 months and monthly visits for the first year following the date of seroconversion. Participants were then followed-up on a quarterly basis after the first post-seroconversion year. The study design and participant characteristics are described in greater detail elsewhere [56, 59, 60]. This study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Institutional Review Boards of Botswana and the Harvard School of Public Health. All patients provided written informed consent for the collection of samples and subsequent analysis.

2.2 Serological assays and HIV pairwise diversity for recency determination

Blood specimens from 42 participants were used to generate 594 BED (Calypte Aware BED HIV-1 Incidence Test, Calypte Biomedical Corporation; Portland, USA) and 597 LAg (Limiting Antigen Assay, Sedia BioSciences; Portland, USA) test results according to manufacturers' instructions [34, 35]. All available specimens were included for testing with both serological assays. UNAIDS/WHO guidelines for determining infection recency recommend the removal of specimens with evidence of ART use [33, 61, 62], resulting in the exclusion of 49 time points and one participant from our analysis (see S1 Fig of the Supplement).

The intra-host viral sequences representing HIV-1C quasispecies were generated by single genome amplification and sequencing, as described elsewhere [47, 63]. The primary goal of sequencing was analysis of viral diversity and evolution during primary HIV-1C infection [47, 63]. The quarterly time points spanning the period from the earliest sample at enrollment to about 500 days post seroconversion were selected from the available sampling points (S4 Fig). Individuals with acute HIV-1C infection were sampled more frequently than individuals enrolled during Fiebig stages IV-V [64].

The targeted region spanned HIV-1C *env* gp120 V1C5 corresponding to nucleotide positions 6,615 to 7,757 of HXB2. A total of 2,540 single genome amplification sequences were generated from an average of 6 time points per patient and an average of 10 multiple-sequences (quasispecies) per time point. Both viral RNA and proviral DNA were used as templates for amplification and sequencing. Viral sequences were codon-aligned by muscle [65] in MEGA 6.06 [66]. Mean pairwise distances (PwD) were estimated per participant per time point using the Maximum Composite Likelihood model and pairwise deletion of gaps in MEGA 6.06 [66]. The accession numbers of the viral sequences used in this study are KC628761—KC630726.

2.3 Multiplicity of HIV infection

Previous research has shown that multiplicity of infection can result in highly variable PwD values [33, 61, 62]. For this reason, we undertook a phylogenetic analysis to identify and exclude time points with multiple founder variants or potential super-infection. Multiplicity was determined by the branching topology of viral quasispecies (~1,200 bp V1C5 region of HIV-1 env gp120) derived from a single time point of sampling. A total of 2,540 viral sequences from 42 subjects were analyzed with 1322 HIV-1 subtype C V1C5 sequences retrieved from the Los Alamos National Laboratory (LANL) HIV Database (S2 Table). Phylogenetic trees were inferred by the Maximum-Likelihood (ML) using Fasttree v.2.1.8 with a GTR model of nucleotide substitution [67]. Phylogenetic trees were visualized and inspected in FigTree [68]. Monophyletic clustering was interpreted as HIV transmission from a single source including transmission of multiple viral variants from the same source. We excluded 47 time points and one participant with viral quasispecies separated by reference sequence(s), as these were interpreted as HIV transmissions from multiple sources (including potential super-infection). The final sample size was 758 time-points from 40 participants (see S1 Fig of the Supplement).

2.4 Statistical Analysis

We used a receiver operating characteristics (ROC) analysis to compare the accuracy of the BED, LAg and PwD assays to identify HIV infection recency. Frequent and repeated testing of study participants enabled us to identify the *known* instances of a recent HIV infection. Specifically, known HIV infection recency was defined as any specimen obtained within a <130, <180 or <360-day post-seroconversion period. For each BED, LAg and PwD assay, we then classified a specimen as a "recent" infection if it was below a threshold value, or classified the specimen as an "established" infection if it was above this threshold. We refer to these as the *classified* instances of a recent HIV infection [69]. For example, we classified specimens with a BED value ≤0.8 as a recent infection or an established infection otherwise. The recommended threshold values for the BED and LAg assays are 0.8 and 1.5 respectively [34, 70].

The best performing thresholds for the PwD assay have yet to be definitively established. Previous research has suggested that the rate of increase in the pairwise sequence diversity of the HIV-1 *env* gene region is a constant rate of approximately 0.01 per year during early infection [46]. We therefore used these biological guidelines to select PwD thresholds of 0.004, 0.005 and 0.01 for the 130, 180 and 360-day cut-offs respectively. For each threshold, we

obtained the sensitivity (recent infections correctly identified) and the specificity (established infections correctly identified) using maximum likelihood estimates from a logistic regression analysis. Because repeated measurements were taken for each participant over time, we calculated the standard errors and 95% confidence intervals (CI) for these estimates using the Huber-White sandwich estimator [71, 72]. Given the evaluation of multiple test thresholds, we used the highest percentage of specimens correctly classified (CC) as a guide to evaluate the performance of each PwD threshold. The CC is computed as the sum of the recent and established specimens correctly classified divided by the total number of specimens classified.

We next evaluated the predictive performance of combination BED plus PwD screening to determine infection recency, and repeated this procedure for combination LAg plus PwD screening. Specifically, our aim was to determine whether the more affordable BED or LAg assay can be combined with the more sensitive PwD assay to reduce the likelihood of a false-recent classification. We first screened for recent infections using a recommended BED threshold of 0.8 for the 180-day cut-off and a recommended LAg threshold of 1.5 for the 130-day cut-off. The threshold and cut-off combination selected for the analysis are based on the work of Kassanjee et al. and Duong et al. [33–35]. We then used the PwD assay with a threshold of 0.005 to reduce the false-recency rate associated with the primary BED and LAg screening assays. Shaw et al. [73] propose to obtain the relative true-recent rate (rTRR) and the relative false-recent rate (rFRR) of the combined BED (or LAg) and PwD assays with:

$$rTRR = \frac{P(BED = +, \; PWD = +|\; R)}{P(BED = +|R)} \quad \text{and}$$

$$rFRR = \frac{P(BED = +, \; PWD = +|\; \bar{R})}{P(BED = +|\bar{R})}.$$

In the above equations, $BED+$ and $PwD+$ are the specimens classified as recent infections by the respective assay, R denotes the specimens known to be recent infections and $\bar{R}$ denotes the specimens known to be established infections. When considering the use of a second marker to improve predictive performance, it is expected that a high rTRR (sensitivity) is maintained while the rFRR is reduced, such that the rTRR will be close to 1.0 and the rFRR will be substantially less than 1.0 [73]. We evaluate the percentage reduction in the rFRR by the PwD assay at rTRR (sensitivity) levels of 75%, 80%, 85% and 90%. Further, we show how the addition of viral load (VL) information can improve accuracy. Research has shown that VL measurements <1,000 copies/mL are associated with false-recent infections and can identify individuals with viral suppression [74, 75]. We used the methods of Shaw et al. [73], Janes et al. [76] and Pepe et al. [77] to obtain estimates for the rFRR and its 95% confidence intervals. Statistical analyses were undertaken in Stata 13.1.

The mean duration of recent infection (MDRI), the average time being recent while infected for less than time cut-off time (T) was estimated using the Incidence Estimation Tools version 1.0.5.9001 (The *inctools* package in R software version 3.2.4). The T value of 2 years and time points with viral load above 1,000 copies/mL were used for the MDRI calculation.

3.0 Results

All of the 2,540 sequences from the 42 participants in the cohort were classified as subtype C. To account for multiplicity of HIV infection and avoid inflated estimate of HIV pairwise distances, time points with phylogenetically distinct viral lineages (n = 47) were excluded from analysis (see section 2.3 Multiplicity of HIV infection in Methods). The final analytic sample consisted of 758 (BED = 554, LAg = 579, and PwD = 238) time-points from 40 participants

(see S1 Fig of the Supplement for the data flow diagram). Among the study participants, 28 (70%) were female. The median (IQR) age at enrollment was 27 (20–56) years. Participants were followed for a median (IQR) of 45.9 (32.4–53.9) months, with a median (IQR) of 21 (18–27) time points per participant. The mean (SD) and median (IQR) time between tests were 2.0 (±2.9) months and 1.1 (0.92–3.0) months respectively. Table 1 shows the summary statistics for the participant characteristics and covariate measures.

We present the maximum likelihood estimates for the PwD assay in S1 Table of the supplement. Given that there is currently no recommended PwD threshold, we show the sensitivity and specificity estimates for values ranging from 0.0005 to 0.015. For the 130-day cut-off, a PwD threshold of 0.004 gives a sensitivity of 76.2% and a specificity of 79.7%, with 77.8% of the total specimens correctly classified. For the 180-day cut-off, a PwD threshold of 0.005 gives a sensitivity of 74.5% and a specificity of 75.5%, with 74.9% of the total specimens correctly classified. We found that PwD values of 0.0055 and 0.006 performed slightly better than the 0.004 and 0.005 values for both the 130 and 180-day cut-offs, and are biologically plausible given that HIV is known to evolve at a rate of approximately 0.01 per year.

We found the PwD threshold values (reported above) to be more accurate than the recommended LAg = 1.5 and BED = 0.8 threshold values in identifying infection recency. For a 130-day cut-off and a threshold value of 1.5, the LAg assay gives a sensitivity of 71.3% and a specificity of 72.9%, with 72.4% of the total specimens correctly classified. For a 180-day cut-off and a threshold value of 0.8, the BED assay gives a sensitivity of 87.4% and a specificity of 50.2%, with 65.5% of the total specimens correctly classified. For these cut-offs and thresholds, we see that the PwD assay has a higher proportion of specimens correctly classified when compared with the LAg and BED assays.

We also compare the accuracy of the three assays to identify infection recency using the AUC estimate of a ROC graph. An AUC closer to 1.0 indicates a better accuracy, and we show these estimates along with their standard errors and 95% CIs in Table 2. The AUC value for the 130-day cut-off is 0.83 compared with 0.78 for the BED assay and 0.81 for the LAg assay. For the 180-day cut-off, these values are PwD = 0.82, BED = 0.75, and LAg = 0.79 and for the 360-day cut-off these are PwD = 0.78, BED = 0.74, and LAg = 0.72 (see also Fig 1).

We investigated whether MAA could further distinguish recent from established infections. Table 3 shows the ability of the PwD assay to improve predictive accuracy by reducing the relative false-recent rate (rFRR) of the LAg and BED assays. Here, we are specifically interested in the percentage reduction in the rFRR and so we subtract the rFRR estimate from 100%. As an example, we interpret the result for the LAg plus PwD combination screening for the 130-day

Table 1. Participant and covariate characteristics.

Participant Characteristics	n = 40	
Female, N (%)	28	(70)
Age (years), Median (IQR)	27	(20–56)
Time under observation (months), Median (IQR)	45.9	(32.4–53.9)
Difference between time points (months), Median (IQR)	1.1	(0.92–3.0)
Total time points per participant, Median (IQR)	21	(18–27)
Assay time points per participant, Median (IQR)		
BED	14	(10–19)
Lag	14.5	(7–22)
PwD	5	(4–6)
CD4 cells/µl, Median (IQR)	417	(302–569)
Viral load ($\log_{10}$) copies/mL, Median (IQR)	3.9	(2.65–4.73)

doi:10.1371/journal.pone.0160649.t001

Table 2. Area under the curve (AUC) of a receiver-operator characteristics (ROC) graph comparing the accuracy of the PwD, BED, and LAg assays in identifying HIV infection recency.

Assay	AUC	SE	95% CI
130-day cut-off			
PWD	0.83	0.03	0.78–0.88
BED	0.78	0.02	0.74–0.82
LAg	0.81	0.02	0.78–0.85
180-day cut-off			
PWD	0.82	0.03	0.76–0.88
BED	0.75	0.02	0.71–0.79
LAg	0.79	0.02	0.75–0.82
360-day cut-off			
PWD	0.78	0.05	0.68–0.89
BED	0.74	0.03	0.69–0.79
LAg	0.72	0.02	0.68–0.77

The table shows the results for the area under the curve (AUC) of a receiver operating characteristics (ROC) graph. The AUC is an objective measure of the accuracy of a classification schema. The best possible value is 1.0, which represents a 100% sensitivity and 100% specificity of the assay to correctly distinguish recent from established HIV infections. The results show that the PwD assay has the best predictive performance for the three window periods. CI: confidence interval.

doi:10.1371/journal.pone.0160649.t002

cut-off as follows: The PwD assay reduces the rFRR by (100–48) 52% while maintaining a 90% rTRR (sensitivity) of the LAg assay. We can also interpret this result using the upper bound of the 95% CI: the PwD assay reduces the rFRR by *at least* (100–87.5) 12.5% while maintaining a LAG sensitivity of 90%.

Results show that the PwD assay reduces the rFRR by (100–42.2) 57.8%, or that it reduces the rFRR by *at least* (100–62.8) 37.2%, while maintaining a 90% sensitivity of the BED assay. Panel D of S2 Fig provides a graphical illustration of the reduction in the rFRR due to the BED plus PwD combination screening. The panel shows the rFRR estimate (red dot) on the ROC graph that corresponds with a 90% sensitivity (y-axis) and a 42.2% false-recent (x-axis) value. The red bar represents the 95% CI of the rFRR. Panels A-C of S2 Fig show that rFRR estimates at a sensitivity levels of 75%, 80% or 85% respectively, the values of which can be obtained from Table 3.

We further provide a data flow diagram in S3 Fig to demonstrate the procedure used to produce the results for Table 3. There were 217 time points that had values for *both* the PwD and BED assays, of which 134 were known to be recent. We first used a recommended BED threshold of ≤0.8 to classify 168 time points as recent infections. We then used a PwD threshold of ≤0.005 to re-screen these 168 time points in order to improve predictive accuracy. S3 Fig shows a reduction in the number of false-recent infections from 45 to 16 (64%) due to the PwD screening, while maintaining a BED sensitivity of 91.6%. This result differs slightly from that of Table 3, which is interpreted at an exact sensitivity of 90%.

We then show how additional biomarker information can be used to improve the combination screening procedure. Here we hypothesize that treatment naïve participants with viral loads <1000 copies/mL are less likely to be recently infected with HIV. Fig 2 shows the BED plus PwD screening for the 180-day cut-off. The rFRR estimate is 31.6% (95% CI: 11.0–63.1), which shows that the PwD assay and VL information reduces rFRR by 68.4% (or by at least 36.9%) while maintaining a BED sensitivity of 90%. We also show this result for the LAg plus PwD combination screening for the 130-day cut-off in Fig 3.
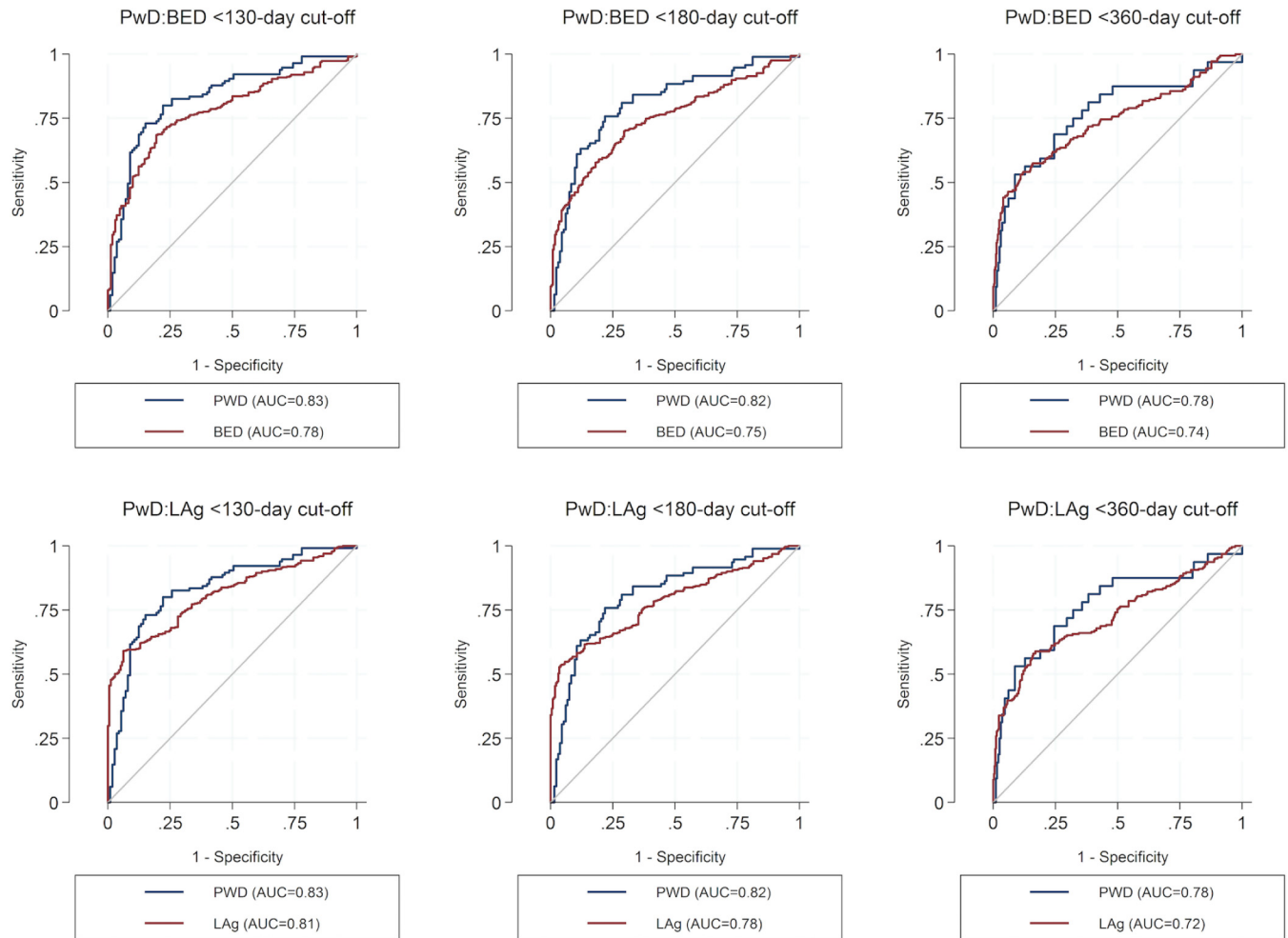
Fig 1. ROC graphs comparing the predictive performance of the PwD, LAg and BED assays for determing HIV infection recency for the 130, 180, and 360-day cut-offs. We used the area under the curve (AUC) of a receiver operator characteristics (ROC) graph to assess the accuracy of the PwD, BED and Lag assays to identify HIV infection recency. The best possible AUC value is 1.0. The ROC graphs are produced by calculating the sensitivity and specificity at different thresholds, which are typically incremented by a fixed value over the minimum and maximum range of the assay. The AUC results show that the PwD assay is the most accurate identifier of infection recency for the three cut-off periods.

doi:10.1371/journal.pone.0160649.g001

Finally, we estimated MDRI's for PwD using a threshold of 0.005, BED and LAg using standard thresholds of 0.8 and 1.5 respectively. PwD had an estimated MDRI of 128 days (95% CI 92–185). BED and LAg had estimated MDRIs of 267 days (95% 212–335) and 129 days (81–190), respectively (S4 Table)

4.0 Discussion and Conclusion

There is an urgent need in HIV research to classify infection recency using accurate, practical and cost effective methods [29, 78–81]. In this study, we evaluate the accuracy of a viral-based assay, HIV pairwise diversity (PwD), to identify participants recently infected with HIV. Our study provides information on the best-performing thresholds for the PwD assay, and compares this assay with two serologic-based assays, BED and LAg. We found that PwD threshold values in the range of 0.005 and 0.006 gave a high sensitivity and specificity for the 130 and

Table 3. Combination assay screening to identify HIV infection recency for the 130 and 180-day cut-offs periods.

	Sensitivity level	Relative False- Recency Rate	95% Lower bound	95% Upper bound
BED+PwD	75	28.3	13.8	47.9
130-day	80	35.0	17.0	48.3
cut-off	85	36.7	19.4	52.1
	90	40.0	23.2	68.5
BED+PwD	75	28.9	11.8	44.6
180-day	80	31.1	13.8	46.2
cut-off	85	31.1	18.1	53.2
	90	42.2	21.1	62.8
LAg+PwD	75	44.0	25.0	68.2
130-day	80	48.0	28.6	68.4
cut-off	85	48.0	27.5	73.3
	90	48.0	30.9	87.5
LAg+PwD	75	42.1	15.8	71.8
180-day	80	42.1	17.4	73.3
cut-off	85	42.1	18.5	71.8
	90	47.4	22.2	83.3

The table shows the reduction in the relative false-recency rate (rFRR) of the BED and LAg assays due to the PwD assay. A BED = 0.8 or LAg = 1.5 threshold was first used to screen the specimens for HIV infection recency. Specimens classified as recent were then re-screened using the PwD assay in order to reduce the rFRR while maintaining a 75%, 80%, 85% or 90% true-recency rate (sensitivity) of the BED or LAg assay. Since we are interested in the reduction of the rFRR by the PwD assay, we subtract this estimate from 100%. The results can be interpreted as follows: for the 180-day cut-off, the PwD assay reduces the rFRR by (100–42.2 =) 57.8% while maintaining a BED sensitivity of 90%. The table also gives the 95% confidence bounds for the reduction in the rFRR. The same result can be interpreted as follows: for the 180-day cut-off, the PwD assay reduces the rFRR by *at least* (100–62.8 =) 37.2% while maintaining a BED sensitivity of 90%.

180-day cut-offs. These values are biologically feasible and consistent with previous work. For example, studies have determined that the mean pairwise sequence diversity of the HIV-1 *env* gene region increases at an approximately constant rate of 0.01 per year during early HIV infection [46]. Other studies using a different measure of HIV diversity, namely proportion of ambiguous sites, found that a threshold ranging from 0.0045 to 0.005 gave a high sensitivity for the 180-day cut-off [11]. Xia et. al. [82] show that a 0.006 diversity cut-off distinguished recent infections with both single and multiple infections.

The results of our study show that the PwD assay can accurately identify recent HIV infections. The PwD assay gave the best performance for the 130, 180, 360-day cut-offs according to the AUC estimates. PwD thresholds of 0.004 and 0.005 correctly classified a higher proportion of specimens when compared with BED and LAg thresholds of 0.8 and 1.5 respectively. We also evaluated a multi-assay algorithm (BED plus PwD or LAg plus PwD) to identify HIV infection recency. Our algorithm first uses an affordable, serologic based assay (BED or LAg) to identify a high proportion of true-recent HIV infections, and then the more sensitive PwD assay to reduce the percentage of specimens misclassified as recent infections. Combination screening significantly improved the classification of HIV infection recency. We found that the PwD assay was able to reduce the relative false-recency rate (rFRR) by approximately 52% while maintaining a LAg sensitivity of 90% for the 130-day cut-off. PwD reduced the rFRR by approximately 58% while maintaining a BED sensitivity of 90% for the 180-day cut-off. Results also show an improvement in accuracy when including biomarker information such as participant viral load (VL).
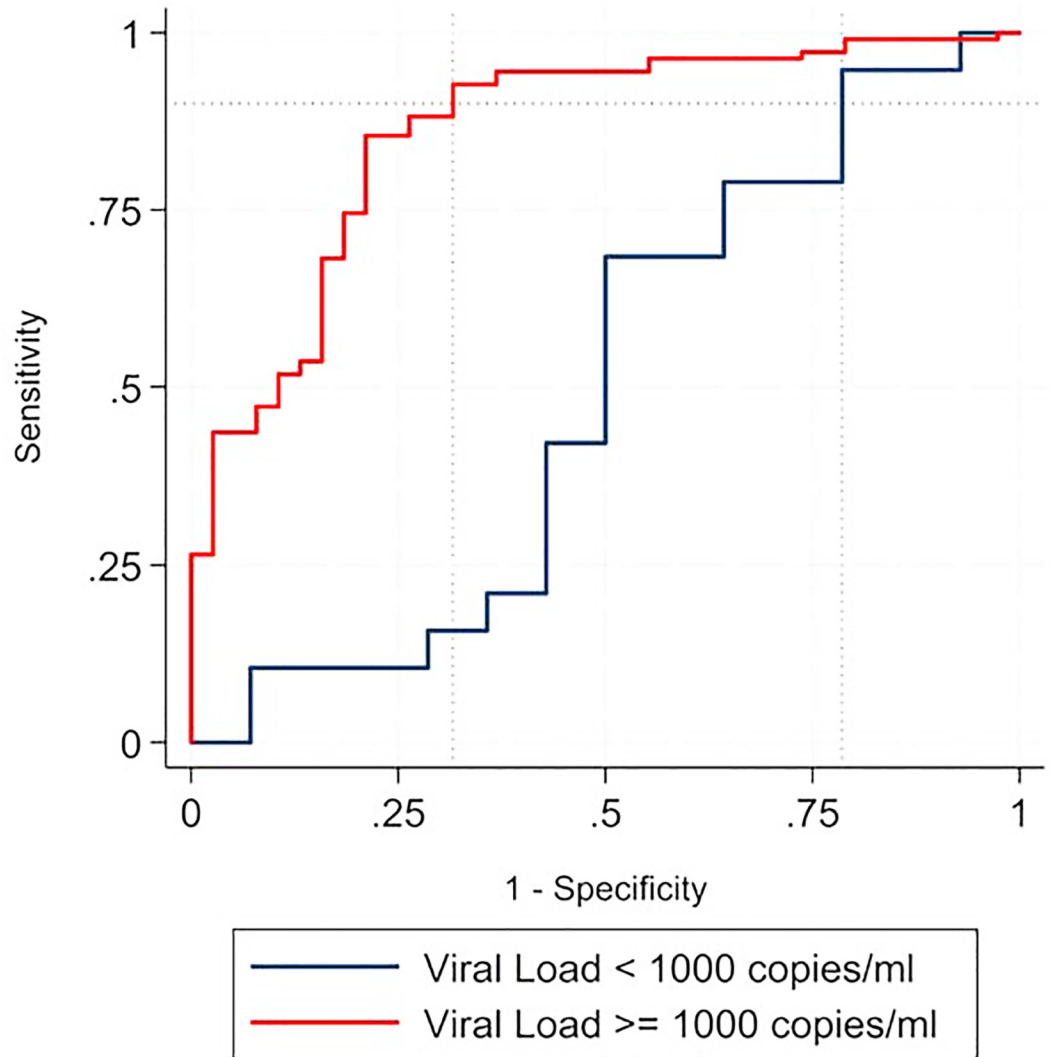
Fig 2. Shows a reduction in the relative false-recency rate (rFRR) when viral load information is added to the combination BED plus PwD screening procedure. The figure shows how additional biomarker information can be used to improve the combination screening procedure for the 180-day cut-off. We hypothesize that treatment naïve participants with viral loads ≤1000 copies/mL are more likely to be recently infected with HIV. Results show an rFRR estimate of 31.6% (95% CI: 11–63.1) at a 90% sensitivity level. Since we are interested in the reduction of the rFRR by the PwD assay, we subtract this estimate from 100%. Thus, the PwD assay reduces the rFRR by 68.4% (or by at least 36.9% given the upper bound of the 95% CI) while maintaining a BED sensitivity of 90% for the subsample of VL >1000 copies/mL specimens. The figure displays both ROC curves for the viral load covariate and the corresponding rFRR estimates (displayed by the dotted vertical lines).

doi:10.1371/journal.pone.0160649.g002

Prior research has shown that the presence or active use of ART can reduce HIV diversity and result in the misclassification of infection recency [62, 83]. The sensitivity of incidence assays can be maintained if auxiliary patient information on ART usage is collected at the same time as the blood specimen. The collection of additional information, such as VL or CD4 counts, has also been shown to improve the performance of bio-marker based assays to detect infection recency [12, 13, 32]. Our study confirms that the inclusion of VL as a covariate in the analysis significantly reduced the false-recent rate of the BED or LAg assays while maintaining a high sensitivity [12, 84, 85]. Collecting VL or CD4 count information may however increase
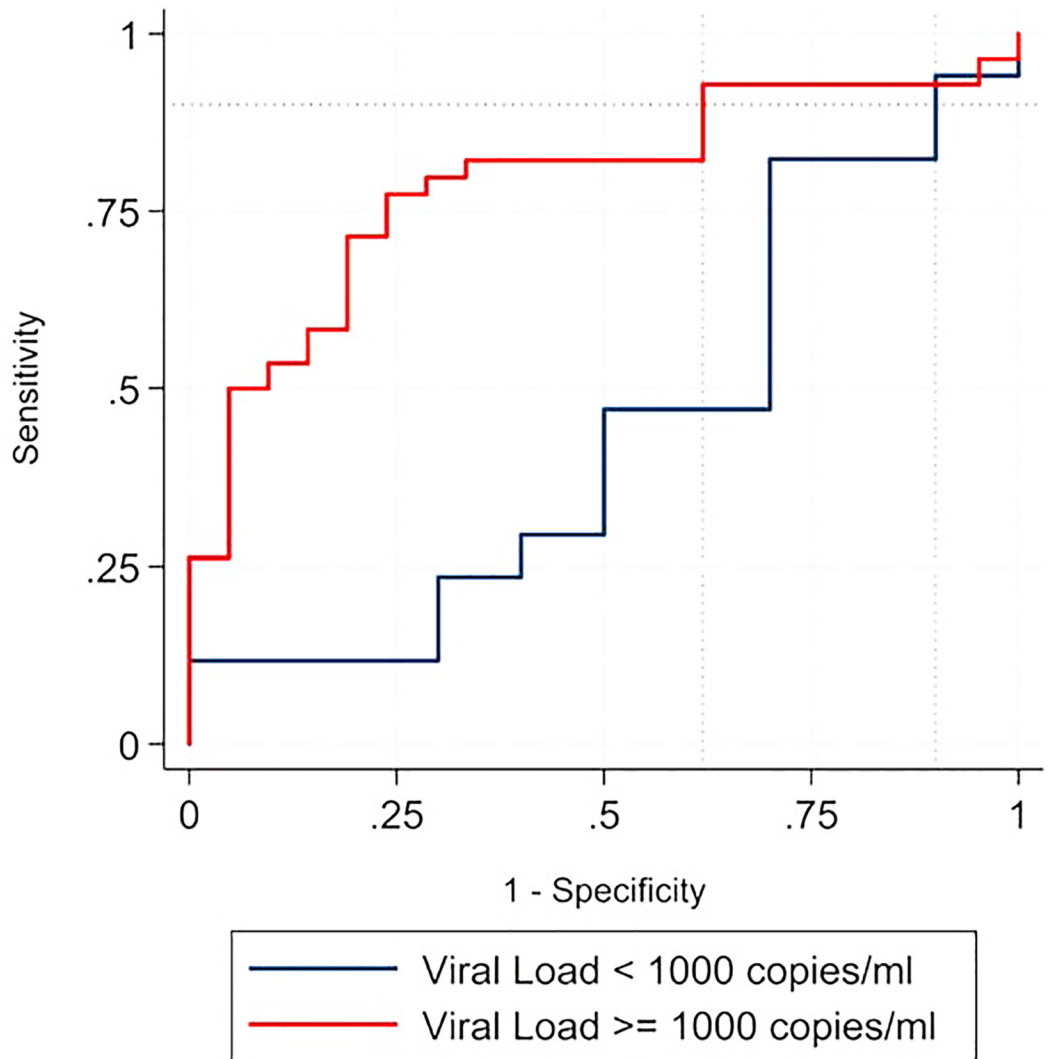
Fig 3. Shows a reduction in the relative false-recency rate when viral load information is added to the combination LAg plus PwD screening. The figure shows how additional biomarker information can be used to improve the combination screening procedure for the 130-day cut-off. We hypothesize that treatment naïve participants with viral loads ≤1000 copies/mL are more likely to be recently infected with HIV. Results show an rFRR estimate of 38.1% (95% CI: 15.8–88.6) at a 90% sensitivity level. Since we are interested in the reduction of the rFRR by the PwD assay, we subtract this estimate from 100%. Thus, the PwD assay reduces the rFRR by 61.9% (or by at least 11.4% given an upper bound of the 95% CI) while maintaining a LAg sensitivity of 90% for the subsample of VL <1000 copies/mL specimens. The figure displays both ROC curves for the viral load covariate and the corresponding rFRR estimates (displayed by the dotted vertical lines).

doi:10.1371/journal.pone.0160649.g003

operational costs. Some of these markers may not be readily available during routine cross-sectional surveys or for previously collected specimens. The PwD MDRI estimates are similar to LAg MDRI recently published [34, 86], although larger sample sets could help to evaluate different thresholds of PwD.

In this paper, we excluded time points with evidence of multiple founder variants or superinfection. Previous research has shown that multiplicity of infection can result in highly variable pairwise distances. PwD values calculated from multi-infection time points are likely to fall outside of the expected range, and do not give an accurate estimate of HIV diversity [61, 87]. Methods to better identify multi-infections in cross sectional sampling are currently being

developed. The PwD assay may be of limited use in men who have sex with men (MSM) [88, 89] due to a high multiplicity of infection. However, more than 80% of all heterosexual HIV infections are seeded by a single founder strain [42–45], which is the main route of transmission in Botswana.

One current limitation associated with the wide-scale use of the PwD assay is the cost of genomic sequencing, which requires expensive laboratory equipment, the training of staff and the technically demanding task of generating single genomes or clonal sequences. The current cost of generating quasispecies from a single time point ranges from $150–200$ compared to $5.29 and $2.35 per test for LAg and BED, respectively. Nevertheless, we argue that the data generated from genome sequencing can address a range of research questions related to the timing of infections in transmission clusters, the number of strains infecting individuals, tropism of the virus and the selection of optimal drug regimens. In this regard, the costs of genome sequencing would be absorbed into a body of research initiatives and questions, rather than used exclusively for the generation of a viral diversity measure. It is also likely that expensive viral-based assays will become a moot point in the near future as the cost of genomics technology continues to decline.

In conclusion, serologic assays and their algorithms have become increasingly popular in recent years because they are based on antibody laboratory tests that are cheaper, quicker and relatively straightforward to implement at the population level [30, 31, 80, 90]. In this study, we show that a measure of HIV diversity can accurately classify infection recency. Our results show that BED plus PwD or LAg plus PwD combination screening has the potential to correctly identify a high proportion of recent HIV infections in a cost-effective manner. The use of bio-marker based assays and cross-sectional data to identify HIV infection recency presents a promising alternative to the resource-intensive approach of a longitudinal cohort design. With continued development, these assays hold the potential to accurately estimate HIV incidence, monitor the spread of the epidemic, evaluate the impact of treatment interventions and inform the design of vaccine and prevention trials.

Supporting Information

S1 Fig. Data flow diagram showing total time points and participants included in the final analysis.
(PNG)

S2 Fig. ROC graphs showing a reduction in the relative false-recency rate (rFRR) of the BED assay by the PwD assay for the <180-day cut-off. The figure gives an example of the reduction in the relative false-recency rate (rFRR) of the BED assay by the PwD assay for the 180-day cut-off. The panels A-D show the ROC curves for the four sensitivity levels. The y-axis is the sensitivity and the x-axis the false-recency rate (1 –specificity); the red point on each graph is the rFRR estimate along with its 95% CI, as shown by the red error bar. The PwD assay reduces the rFRR by 57.8% while maintaining a 90% sensitivity of the BED assay. The ROC graphs show that after performing combination screening, an rFRR estimate can be obtained for any sensitivity value between 0 and 1.0.
(PNG)

S3 Fig. Flow chart of the combination BED plus PwD screening to identify HIV infection recency for the 180-day cut-off. Flow chart showing how the PwD assay can be combined with the BED assay to reduce the likelihood of a false-recent result (i.e., established infections misclassified as recent infections). A recommended BED assay threshold value of 0.8 was used to classify infection recency for the N = 217 specimens. This first screening correctly identified

123 of the 134 recent infections for the 180-day cut-off (true positives), giving a sensitivity of 91.8%. However, 45 of the 83 (54.2%) established specimens were falsely classified as recent. A PwD threshold of 0.005 was then used to screen the subset of specimens classified as recent (n = 168) by the BED assay. Results show that the secondary PwD screening reduces the false-recent infections by 64% (45 to 16 specimens) at a BED sensitivity of 87.8%. (This result differs slightly from that of Table 3, which is interpreted at an exact sensitivity of 90%.)
(PNG)

S4 Fig. Distribution of time-points for the BED, LAg and PwD assays. The figure gives the analysed time points of sampling and sequencing in the study since the known time of seroconversion. Time in days post-seroconversion is shown on the x-axis.
(PNG)

S5 Fig. Spaghetti plots for the BED, LAg and PwD time-points.
(PNG)

S1 Table. Performance of PwD threshold values to determine HIV Infection Recency for 130, 180, and 360-day cut-offs. The table shows the performance of the PwD threshold values to identify HIV infection recency. The range of values were selected according to rate of increase in the pairwise sequence diversity of the HIV-1 *env* gene region, which is approximately a constant rate of 0.01 per year during early infection. For example, a 180-day cut-off corresponds with a PwD value of 0.005. We selected thresholds values in the range of these biological values for each of the cut-off periods. For each threshold we obtained the sensitivity, specificity, their 95% CI, likelihood ratio, and percentage correctly classified. For the 130-day cut-off, a PwD threshold of 0.005 correctly identified 79.37% (95% CI: 62.83–95.9) of the recent infections (sensitivity) and correctly identified 72.57% (95% CI: 61.87–83.26) of the established infections (specificity), giving a percentage correctly classified of 76.15%.
(DOCX)

S2 Table. Accession numbers for the reference sequences used.
(DOCX)

S3 Table. Area under the curve (AUC) for the PwD, BED, and LAg assays for shared time-points (n = 238). Table shows the results for the area under the curve (AUC) of a receiver operating characteristics (ROC) graph for the <130, <180- and <360-day cut-offs. Using only shared time-points (n = 238) significantly reduces the sample size and therefore the performance of the three assays. The performance of the three assays are therefore indistinguishable given the overlap in the confidence intervals of the AUC estimates.
(DOCX)

S4 Table. Mean Duration of Recent Infection for BED, LAg and PwD Assays. Table shows the estimated Mean Duration of Recent Infection (MDRI), average time 'recent' while infected for less than some time cut-off T for the BED, LAg and PwD assays.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: SM EW SE VN TdO SG RMM.

Performed the experiments: SMM KPK.

Analyzed the data: SMM AV.

Contributed reagents/materials/analysis tools: VN ME SMM AV SE EW TdO.

Wrote the paper: SM AV EW SE VN FT ME TdO.

Designed and supervised the primary infection cohort "Tshedimoso": VN ME, Provided laboratory support for the primary infection cohort: SM.

References

1. Brenner BG, Roger M, Routy J-P, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis. 2007; 195(7):951–9. PMID: 17330784

2. Pilcher CD, Eron JJ Jr, Galvin S, Gay C, Cohen MS. Acute HIV revisited: new opportunities for treatment and prevention. J Clin Invest. 2004; 113(7):937–45. PMID: 15057296.

3. Pope M, Haase AT. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. Nat Med. 2003; 9(7):847–52. PMID: 12835704

4. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis. 2005; 191 (9):1403–9. PMID: 15809897.

5. Gray GE, Andersen-Nissen E, Grunenberg N, Huang Y, Roux S, Laher F, et al. HVTN 097: Evaluation of the RV144 Vaccine Regimen in HIV Uninfected South African Adults. AIDS Res Hum Retroviruses. 2014; 30 Suppl 1:A33–4. Epub 2014/10/31. doi: 10.1089/aid.2014.5052a.abstract PMID: 25357829.

6. Lima KO, Salustiano DM, Cavalcanti AM, Leal Ede S, Lacerda HR. HIV-1 incidence among people seeking voluntary counseling and testing centers, including pregnant women, in Pernambuco State, Northeast Brazil. Cad Saude Publica. 2015; 31(6):1327–31. Epub 2015/07/23. doi: 10.1590/0102-311x00186813 PMID: 26200379.

7. Lu X, Kang X, Chen S, Zhao H, Liu Y, Zhao C, et al. HIV-1 genetic diversity and transmitted drug resistance among recently infected individuals at MSM sentinel surveillance points in Hebei province, China. AIDS Res Hum Retroviruses. 2015. Epub 2015/07/23. doi: 10.1089/aid.2015.0160 PMID: 26200883.

8. Lunar MM, Matkovic I, Tomazic J, Vovko TD, Pecavar B, Poljak M. Longitudinal trends of recent HIV-1 infections in Slovenia (1986–2012) determined using an incidence algorithm. J Med Virol. 2015; 87 (9):1510–6. Epub 2015/05/15. doi: 10.1002/jmv.24209 PMID: 25970253.

9. Noguchi LM, Richardson BA, Baeten JM, Hillier SL, Balkus JE, Chirenje ZM, et al. Risk of HIV-1 acquisition among women who use different types of injectable progestin contraception in South Africa: a prospective cohort study. Lancet HIV. 2015; 2(7):e279–e87. Epub 2015/07/15. doi: 10.1016/s2352-3018(15)00058-2 PMID: 26155597; PubMed Central PMCID: PMCPMC4491329.

10. Kassanjee R, McWalter TA, Barnighausen T, Welte A. A New General Biomarker-based Incidence Estimator. Epidemiology. 2012. Epub 2012/05/26. doi: 10.1097/EDE.0b013e3182576c07 PMID: 22627902.

11. Ragonnet-Cronin M, Aris-Brosou S, Joanisse I, Merks H, Vallée D, Caminiti K, et al. Genetic diversity as a marker for timing infection in HIV-infected patients: evaluation of a 6-month window and comparison with BED. J Infect Dis. 2012:jis411.

12. Brookmeyer R, Konikoff J, Laeyendecker O, Eshleman SH. Estimation of HIV incidence using multiple biomarkers. Am J Epidemiol. 2013; 177(3):264–72. Epub 2013/01/11. doi: 10.1093/aje/kws436 PMID: 23302151; PubMed Central PMCID: PMC3626051.

13. Cousins MM, Konikoff J, Laeyendecker O, Celum C, Buchbinder SP, Seage GR 3rd, et al. HIV Diversity as a Biomarker for HIV Incidence Estimation: Including a High Resolution Melting Diversity Assay in a Multi-Assay Algorithm. J Clin Microbiol. 2013. Epub 2013/10/25. doi: 10.1128/JCM.02040-13 PMID: 24153134.

14. Emerson B, Plough K. Detection of acute HIV-1 infections utilizing NAAT technology in Dallas, Texas. J Clin Virol. 2013; 58 Suppl 1:e48–53. Epub 2013/09/04. doi: 10.1016/j.jcv.2013.08.005 PMID: 23999031.

15. Barnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell M. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. AIDS. 2008; 22(1):139–44. doi: 10.1097/QAD.0b013e3282f2ef43 PMID: 18090402

16. Kim JH, Excler JL, Michael NL. Lessons from the RV144 Thai Phase III HIV-1 Vaccine Trial and the Search for Correlates of Protection. Annu Rev Med. 2014. Epub 2014/10/24. doi: 10.1146/annurev-med-052912-123749 PMID: 25341006.

17. Kumwenda N, Hoffman I, Chirenje M, Kelly C, Coletti A, Ristow A, et al. HIV incidence among women of reproductive age in Malawi and Zimbabwe. Sex Transm Dis. 2006; 33(11):646–51. doi: 10.1097/01.olq.0000223283.27142.9f PMID: 16773032

18. Tanser F, Bärnighausen T, Vandormael A, Dobra A. HIV treatment cascade in migrants and mobile populations. Current Opinion in HIV and AIDS. 2015; 10(6):430–8. doi: 10.1097/COH.0000000000000192 PMID: 26352396

19. Visser MJ, Makin JD, Vandormael A, Sikkema KJ, Forsyth BW. HIV/AIDS stigma in a South African community. AIDS care. 2009; 21(2):197–206. doi: 10.1080/09540120801932157 PMID: 19229689

20. Vandormael A, Newell M-L, Bärnighausen T, Tanser F. Use of antiretroviral therapy in households and risk of HIV acquisition in rural KwaZulu-Natal, South Africa, 2004–12: a prospective cohort study. The Lancet Global Health. 2014; 2(4):e209–e15. PMID: 24782953

21. Barnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell ML. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. AIDS. 2008; 22(1):139–44. Epub 2007/12/20. doi: 10.1097/QAD.0b013e3282f2ef43 PMID: 18090402.

22. Novitsky V, Woldegabriel E, Wester C, McDonald E, Rossenkhan R, Ketunuti M, et al. Identification of primary HIV-1C infection in Botswana. AIDS Care. 2008; 20(7):806–11. Epub 2008/07/09. doi: 10.1080/09540120701694055 PMID: 18608056; PubMed Central PMCID: PMCPmc2605733.

23. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, et al. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS One. 2008; 3(4):e1954. Epub 2008/04/17. doi: 10.1371/journal.pone.0001954 PMID: 18414658; PubMed Central PMCID: PMCPmc2278382.

24. Dandona L, Kumar G, Lakshmi V, Ahmed GM, Akbar M, Ramgopal S, et al. HIV incidence from the first population-based cohort study in India. BMC Infect Dis. 2013; 13(1):327. doi: 10.1186/1471-2334-13-327

25. Dandona L, Lakshmi V, Sudha T, Kumar G, Dandona R. A population-based study of human immunodeficiency virus in south India reveals major differences from sentinel surveillance-based estimates. BMC Med. 2006; 4:31. doi: 10.1186/1741-7015-4-31 PMID: 17166257

26. Feldblum P, Latka M, Lombaard J, Chetty C, LChen P, Sexton C, et al. HIV incidence and prevalence among cohorts of women with higher risk behaviour in Bloemfontein and Rustenburg, South Africa: a prospective study. BMJ Open. 2012; 2(1):e000626. doi: 10.1136/bmjopen-2011-000626 PMID: 22331388

27. Geis S, Maboko L, Saathoff E, Hoffmann O, Geldmacher C, Mmbando D, et al. Risk factors for HIV-1 infection in a longitudinal, prospective cohort of adults from the Mbeya Region, Tanzania. J Acquir Immune Defic Syndr. 2011; 56(5):453–9. doi: 10.1097/QAI.0b013e3182118fa3 PMID: 21297483

28. Lopman B, Nyamukapa C, Mushati P, Mupambireyi Z, Mason P, Garnett G, et al. HIV incidence in 3 years of follow-up of a Zimbabwe cohort-1998-2000 to 2001–03: contributions of proximate and underlying determinants to transmission. Int J Epidemiol. 2008; 37(1):88–105. doi: 10.1093/ije/dym255 PMID: 18203774

29. Sharma UK, Schito M, Welte A, Rousseau C, Fitzgibbon J, Keele B, et al. Workshop summary: Novel biomarkers for HIV incidence assay development. AIDS Res Hum Retroviruses. 2012; 28(6):532–9. Epub 2011/12/31. doi: 10.1089/aid.2011.0332 PMID: 22206265; PubMed Central PMCID: PMC3358102.

30. Le Vu S, Pillonel J, Semaille C, Bernillon P, Le Strat Y, Meyer L, et al. Principles and uses of HIV incidence estimation from recent infection testing—a review. Euro surveillance: bulletin europeen sur les maladies transmissibles = European communicable disease bulletin. 2008; 13(36):537–45.

31. Hallett TB. Estimating the HIV incidence rate: recent and future developments. Curr Opin HIV AIDS. 2011; 6(2):102–7. Epub 2011/04/21. doi: 10.1097/COH.0b013e328343bfdb PMID: 21505383; PubMed Central PMCID: PMC3083833.

32. Rosenberg NE, Pilcher CD, Busch MP, Cohen MS. How can we better identify early HIV infections? Curr Opin HIV AIDS. 2014. Epub 2014/11/13. doi: 10.1097/coh.0000000000000121 PMID: 25389806.

33. Kassanjee R, Pilcher CD, Keating SM, Facente SN, McKinney E, Price MA, et al. Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. AIDS. 2014; 28

(16):2439–49. doi: 10.1097/QAD.0000000000000429 PMID: 25144218; PubMed Central PMCID: PMC4210690.

34. Duong YT, Kassanjee R, Welte A, Morgan M, De A, Dobbs T, et al. Recalibration of the Limiting Antigen Avidity EIA to Determine Mean Duration of Recent Infection in Divergent HIV-1 Subtypes. PLoS One. 2015; 10(2):e0114947. Epub 2015/02/25. doi: 10.1371/journal.pone.0114947 PMID: 25710171.

35. 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(3):e33328. doi: 10.1371/journal.pone.0033328 PMID: 22479384

36. Moyo S, Wilkinson E, Novitsky V, Vandormael A, Gaseitsiwe S, Essex M, et al. Identifying Recent HIV Infections: From Serological Assays to Genomics. Viruses. 2015; 7(10):5508–24. doi: 10.3390/v7102887 PMID: 26512688

37. Andersson E, Shao W, Bontell I, Cham F, Cuong DD, Wondwossen A, et al. Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys. Infect, Genet Evol. 2013; 18:125–31.

38. Kouyos RD, von Wyl V, Yerly S, Böni J, Rieder P, Joos B, et al. Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. Clinical infectious diseases. 2011:ciq164.

39. Cousins MM, Laeyendecker O, Beauchamp G, Brookmeyer R, Towler WI, Hudelson SE, et al. Use of a high resolution melting (HRM) assay to compare gag, pol, and env diversity in adults with different stages of HIV infection. PLoS ONE. 2011; 6(11):e27211. Epub 2011/11/11. doi: 10.1371/journal.pone.0027211 PMID: 22073290; PubMed Central PMCID: PMC3206918.

40. Allam O, Samarani S, Ahmad A. Hammering out HIV-1 incidence with Hamming distance. AIDS. 2011; 25(16):2047–8. doi: 10.1097/QAD.0b013e32834bac66 PMID: 21997490

41. Hall N. Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol. 2007; 210(Pt 9):1518–25. Epub 2007/04/24. doi: 10.1242/jeb.001370 PMID: 17449817.

42. Joseph SB, Swanstrom R, Kashuba AD, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. Nat Rev Microbiol. 2015. Epub 2015/06/09. doi: 10.1038/nrmicro3471 PMID: 26052661.

43. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proceedings of the National Academy of Sciences. 2008; 105(21):7552–7. doi: 10.1073/pnas.0802203105

44. Parker ZF, Iyer SS, Wilen CB, Parrish NF, Chikere KC, Lee F-H, et al. Transmitted/Founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5. J Virol. 2013; 87(5):2401–11. doi: 10.1128/jvi.02964-12 PMID: 23269796

45. Parrish NF, Wilen CB, Banks LB, Iyer SS, Pfaff JM, Salazar-Gonzalez JF, et al. Transmitted/Founder and Chronic Subtype C HIV-1 Use CD4 and CCR5 Receptors with Equal Efficiency and Are Not Inhibited by Blocking the Integrin α4β7. PLoS Pathog. 2012; 8(5):e1002686. doi: 10.1371/journal.ppat.1002686 PMID: 22693444

46. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol. 1999; 73(12):10489–502. Epub 1999/11/13. PMID: 10559367; PubMed Central PMCID: PMCPMC113104.

47. Novitsky V, Lagakos S, Herzig M, Bonney C, Kebaabetswe L, Rossenkhan R, et al. Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. Virology. 2009; 383(1):47–59. doi: 10.1016/j.virol.2008.09.017 PMID: 18973914

48. Kearney M, Maldarelli F, Shao W, Margolick JB, Daar ES, Mellors JW, et al. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. J Virol. 2009; 83(6):2715–27. Epub 2009/01/01. doi: 10.1128/jvi.01960-08 PMID: 19116249; PubMed Central PMCID: PMCPmc2648286.

49. Giorgi E, Funkhouser B, Athreya G, Perelson A, Korber B, Bhattacharya T. Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. BMC Bioinformatics. 2010; 11 (1):532. doi: 10.1186/1471-2105-11-532

50. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. J Virol. 2008; 82(8):3952–70. doi: 10.1128/jvi.02660-07 PMID: 18256145

51. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. J Virol. 2009; 83(8):3556–67. Epub 2009/02/06. JVI.02132-08 [pii] doi: 10.1128/JVI.02132-08 PMID: 19193811.

52. Novitsky V, Wang R, Margolin L, Baca J, Rossenkhan R, Moyo S, et al. Transmission of Single and Multiple Viral Variants in Primary HIV-1 Subtype C Infection. PLoS One. 2011; 6(2):e16714. PMCID: PMC3048432. Epub 2011/03/19. doi: 10.1371/journal.pone.0016714 PMID: 21415914.

53. Kiwelu IE, Novitsky V, Margolin L, Baca J, Manongi R, Sam N, et al. HIV-1 subtypes and recombinants in Northern Tanzania: distribution of viral quasispecies. PLoS One. 2012; 7(10):e47605. PMCID: PMC3485255. doi: 10.1371/journal.pone.0047605 PMID: 23118882; PubMed Central PMCID: PMC3485255.

54. Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, Ricklefs SM, et al. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. J Infect Dis. 2012; 206 (2):267–74. doi: 10.1093/infdis/jis325 PMID: 22675216; PubMed Central PMCID: PMC3415936.

55. Redd AD, Mullis CE, Wendel SK, Sheward D, Martens C, Bruno D, et al. Limited HIV-1 superinfection in seroconverters from the CAPRISA 004 Microbicide Trial. J Clin Microbiol. 2014; 52(3):844–8. doi: 10.1128/JCM.03143-13 PMID: 24371237; PubMed Central PMCID: PMC3957790.

56. Novitsky V, Wang R, Margolin L, Baca J, Moyo S, Musonda R, et al. Dynamics and timing of in vivo mutations at Gag residue 242 during primary HIV-1 subtype C infection. Virology. 2010; 403(1):37–46. Epub 2010/05/07. S0042-6822(10)00236-9 [pii] doi: 10.1016/j.virol.2010.04.001 PMID: 20444482.

57. Novitsky V, Wang R, Kebaabetswe L, Greenwald J, Rossenkhan R, Moyo S, et al. Better control of early viral replication is associated with slower rate of elicited antiviral antibodies in the detuned enzyme immunoassay during primary HIV-1C infection. J Acquir Immune Defic Syndr. 2009; 52(2):265–72. Epub 2009/06/16. doi: 10.1097/QAI.0b013e3181ab6ef0 PMID: 19525854.

58. Rawal BD, Degula A, Lebedeva L, Janssen RS, Hecht FM, Sheppard HW, et al. Development of a new less-sensitive enzyme immunoassay for detection of early HIV-1 infection. J Acquir Immune Defic Syndr. 2003; 33(3):349–55. PMID: 12843746.

59. Novitsky V, Woldegabriel E, Kebaabetswe L, Rossenkhan R, Mlotshwa B, Bonney C, et al. Viral load and CD4+ T-cell dynamics in primary HIV-1 subtype C infection. Journal of acquired immune deficiency syndromes (1999). 2009; 50(1):65.

60. Novitsky V, Woldegabriel E, Wester C, McDonald E, Rossenkhan R, Ketunuti M, et al. Identification of primary HIV-1C infection in Botswana. NIHMSID # 79283. AIDS Care. 2008; 20(7):806–11. PMID: 18608056. doi: 10.1080/09540120701694055

61. Park SY, Goeken N, Lee HJ, Bolan R, Dube MP, Lee HY. Developing high-throughput HIV incidence assay with pyrosequencing platform. J Virol. 2014; 88(5):2977–90. doi: 10.1128/JVI.03128-13 PMID: 24371062; PubMed Central PMCID: PMC3958066.

62. Marinda ET, Hargrove J, Preiser W, Slabbert H, van Zyl G, Levin J, et al. Significantly diminished long-term specificity of the BED capture enzyme immunoassay among patients with HIV-1 with very low CD4 counts and those on antiretroviral therapy. J Acquir Immune Defic Syndr. 2010; 53(4):496–9. Epub 2010/03/23. PMID: 20306555.

63. Novitsky V, Wang R, Rossenkhan R, Moyo S, Essex M. Intra-Host Evolutionary Rates in HIV-1C env and gag during Primary Infection. Infect Genet Evol. 2013;s1567–1348(13):77–4. Epub 2013/03/26. doi: 10.1016/j.meegid.2013.02.023 PMID: 23523818.

64. 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(13):1871–9. PMID: 12960819.

65. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004; 5:113. PMID: 15318951.

66. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30(12):2725–9. doi: 10.1093/molbev/mst197 PMID: 24132122; PubMed Central PMCID: PMC3840312.

67. Price MN, Dehal PS, Arkin AP. FastTree 2 –Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010; 5(3):e9490. doi: 10.1371/journal.pone.0009490 PMID: 20224823

68. Rambaut A. FigTree; http://tree.bio.ed.ac.uk/software/figtree/. 1.1.2 ed2008.

69. Fawcett T. An introduction to ROC analysis. Pattern Recog Lett. 2006; 27(8):861–74.

70. Dobbs T, Kennedy S, Pau CP, McDougal JS, Parekh BS. Performance characteristics of the immunoglobulin G-capture BED-enzyme immunoassay, an assay to detect recent human immunodeficiency virus type 1 seroconversion. J Clin Microbiol. 2004; 42(6):2623–8. PMID: 15184443.

71. Huber PJ, editor The behavior of maximum likelihood estimation under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability; 1967: University of California Press.

72. White H. Maximum likelihood estimation of misspecied models. Econometrica. 1982; 50:1–25.

73. Shaw PA, Pepe MS, Alonzo TA, Etzioni R. Methods for assessing improvement in specificity when a biomarker is combined with a standard screening test. Statistics in biopharmaceutical research. 2009; 1(1):18–25. PMID: 20054437

74. Konikoff J, Brookmeyer R, Longosz AF, Cousins MM, Celum C, Buchbinder SP, et al. Performance of a limiting-antigen avidity enzyme immunoassay for cross-sectional estimation of HIV incidence in the United States. PLoS One. 2013; 8(12):e82772. Epub 2014/01/05. doi: 10.1371/journal.pone.0082772 PMID: 24386116; PubMed Central PMCID: PMCPMC3873916.

75. Yu L, Laeyendecker O, Wendel SK, Liang F, Liu W, Wang X, et al. Short Communication: Low False Recent Rate of Limiting-Antigen Avidity Assay Among Long-Term Infected Subjects from Guangxi, China. AIDS Res Hum Retroviruses. 2015; 31(12):1247–9. Epub 2015/09/04. doi: 10.1089/aid.2015.0097 PMID: 26331573; PubMed Central PMCID: PMCPmc4663635.

76. Janes H, Longton G, Pepe M. Accommodating covariates in ROC analysis. The Stata Journal. 2009; 9(1):17. PMID: 20046933

77. Pepe M, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. The Stata journal. 2009; 9(1):1. PMID: 20161343

78. UNAIDS. WHO/UNAIDS Technical Update on HIV incidence assays for surveillance and epidemic monitoring. Geneva, Swtizerland: UNAIDS; 2013. Available: http://www.unaids.org/en/resources/documents/2015/HIVincidenceassayssurveillancemonitoring.

79. Guy R, Gold J, Calleja JMG, Kim AA, Parekh B, Busch M, et al. Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: a systematic review. The Lancet infectious diseases. 2009; 9(12):747–59. doi: 10.1016/S1473-3099(09)70300-7 PMID: 19926035

80. Incidence Assay Critical Path Working Group. More and better information to tackle HIV epidemics: towards improved HIV incidence assays. PLoS Med. 2011; 8(6):e1001045. Epub 2011/07/07. doi: 10.1371/journal.pmed.1001045 PMID: 21731474; PubMed Central PMCID: PMC3114871.

81. Kim AA, Hallett T, Stover J, Gouws E, Musinguzi J, Mureithi PK, et al. Estimating HIV Incidence among Adults in Kenya and Uganda: A Systematic Comparison of Multiple Methods. PLoS ONE. 2011; 6(3):e17535. Epub 2011/03/17. doi: 10.1371/journal.pone.0017535 PMID: 21408182; PubMed Central PMCID: PMC3049787.

82. Xia X-Y, Ge M, Hsi JH, He X, Ruan Y-H, Wang Z-X, et al. High-Accuracy Identification of Incident HIV-1 Infections Using a Sequence Clustering Based Diversity Measure. PLoS ONE. 2014; 9(6):e100081. doi: 10.1371/journal.pone.0100081 PMID: 24925130

83. Chaillon A, Le Vu S, Brunet S, Gras G, Bastides F, Bernard L, et al. Decreased Specificity of an Assay for Recent Infection in HIV-1-Infected Patients on Highly Active Antiretroviral Treatment: Implications for Incidence Estimates. Clinical and Vaccine Immunology: CVI. 2012; 19(8):1248–53. doi: 10.1128/CVI.00120-12 PMC3416072. PMID: 22718132

84. Cousins MM, Konikoff J, Sabin D, Khaki L, Longosz AF, Laeyendecker O, et al. A comparison of two measures of HIV diversity in multi-assay algorithms for HIV incidence estimation. PLoS One. 2014; 9(6):e101043. Epub 2014/06/27. doi: 10.1371/journal.pone.0101043 PMID: 24968135; PubMed Central PMCID: PMCPmc4072769.

85. Moyo S, LeCuyer T, Wang R, Gaseitsiwe S, Weng J, Musonda R, et al. Evaluation of the false recent classification rates of multiassay algorithms in estimating HIV type 1 subtype C incidence. AIDS Res Hum Retroviruses. 2014; 30(1):29–36. Epub 2013/08/14. doi: 10.1089/AID.2013.0055 PMID: 23937344; PubMed Central PMCID: PMC3887420.

86. Kassanjee R, Pilcher CD, Keating SM, Facente SN, McKinney E, Price MA, et al. Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. Aids. 2014; 28(16):2439–49. Epub 2014/08/22. doi: 10.1097/qad.0000000000000429 PMID: 25144218.

87. Janes H, Herbeck JT, Tovanabutra S, Thomas R, Frahm N, Duerr A, et al. HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders. Nat Med. 2015; 21(10):1139–41. Epub 2015/09/01. doi: 10.1038/nm.3932 PMID: 26322580; PubMed Central PMCID: PMCPMC4598284.

88. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, et al. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. PLoS Pathog. 2010; 6(5):e1000890. Epub 2010/05/21. doi: 10.1371/journal.ppat.1000890 PMID: 20485520.

89. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, et al. Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol. 2010; 84(12):6241–7. Epub 2010/04/09. JVI.00077-10 [pii] doi: 10.1128/JVI.00077-10 PMID: 20375173.

90. Murphy G, Parry JV. Assays for the detection of recent infections with human immunodeficiency virus type 1. Euro Surveill. 2008; 13(36). Epub 2008/09/09. PMID: 18775293.