Designing HIV Testing Algorithms Based on 2015 WHO Guidelines Using Data from Six Sites in sub-Saharan Africa

Cara S. Kosack[1], Leslie Shanks[1], Greet Beelaert[2], Tumwesigye Benson[3], Aboubacar Savane[4], Anne Ng'ang'a[5], Bita Andre[6], Jean-Paul B. N. Zahinda[7], Katrien Fransen[2], and Anne-Laure Page[8#]

[1]Médecins sans Frontières, Amsterdam, Netherlands; [2]Institute of Tropical Medicine, Antwerp, Belgium; [3]Ministry of Health Uganda, Kampala, Uganda; [4]Laboratoire National de Reference, Conakry, Guinea; [5]National AIDS and Sexually Transmitted Infections Control Programme, Nairobi, Kenya; [6]Regional Delegation of Public Health for the Littoral Region, Cameroon; [7]Programme National de Lutte contre le Sida et les IST (PNLS), Democratic Republic of Congo; [8]Epicentre, Paris, France.

Running head: Designing HIV Testing Algorithms

Address correspondence to Anne-Laure Page, anne-laure.page@epicentre.msf.org

Abstract

Our objective was to evaluate the performance of HIV testing algorithms based on WHO recommendations, using data from specimens collected at six HIV testing and counselling sites in sub-Saharan Africa (Guinea, Conakry; Kitgum and Arua, Uganda; Homa Bay, Kenya; Douala, Cameroun; Baraka, Democratic Republic of Congo). A total of 2780 samples, including 1306 HIV-positive, were included in the analysis. HIV testing algorithms were designed using Determine as a first test. Second and third rapid diagnostic tests (RDT) were selected based on site-specific performance, adhering where possible to the WHO-recommended minimum requirements of sensitivity and specificity of ≥99%. The threshold for specificity was reduced to 98% or 96% if necessary. We also simulated algorithms consisting of one RDT followed by a simple confirmatory assay. The positive predictive values (PPV) of the simulated algorithms varied from 75.8%-100% using strategies recommended for high-prevalence settings; 98.7%-100% using strategies recommended for low-prevalence settings; and 98.1%-100% using a rapid test followed by a simple confirmatory assay. Although we were able to design algorithms that met the recommended PPV of ≥99% in five of six sites using the applicable high prevalence strategy, options were often very limited due to sub-optimal performance of individual RDTs and to shared false-reactive results. These results underscore the impact of the sequence of HIV tests and of shared false-reactivity on algorithm performance. Where it is not possible to identify tests that meet WHO-recommended specifications, the low-prevalence strategy may be more suitable.

Introduction

HIV rapid diagnostic tests (RDTs) are the main diagnostic tools for HIV screening and diagnosis in resource-constrained settings (1). Given the potential for severe medical, psychological and social impacts of HIV misdiagnosis and evidence of elevated false positive results from some settings, it is imperative that HIV diagnosis is confirmed to be both sensitive and specific (2).

In 2012 and 2015, the World Health Organization (WHO) published revisions of the HIV testing guidelines with different recommendations for low (<5%) and high (≥5%) HIV prevalence settings (1, 3, 4). These recommendations call for the sequential use of up to three different serological assays, including RDTs, for final HIV diagnosis. Whereas a first non-reactive test result is sufficient to provide a final negative results in both settings, two or three reactive assays are needed to provide a final HIV-positive results in high and low-prevalence settings, respectively (Figure 1). The guidelines stipulate that each of the three RDTs should have a sensitivity of at least 99%, while for specificity the first RDT should have at least 98% and the second and third RDTs at least 99%; overall the combination should be designed to minimize the potential for shared false-reactivity. Different strategies for high- and low-prevalence settings were developed based on mathematical models using three theoretical assays assumed to meet the criteria described above to achieve an overall positive predictive value (PPV) of at least 99% (1). To date, however, these recommendations and the performance of the resulting algorithms have not been validated using real data from different field contexts.

Several factors could influence the design and performance of these algorithms. Although WHO-prequalified HIV RDTs met the minimum recommended sensitivity and specificity in the prequalification evaluations, several reports from different countries indicate much poorer performance in real-world settings (5–12). Moreover, little is known about shared false-reactivity

among different RDTs (13). The use of the same antigen preparations to produce different tests, which is occurring with increasing frequency due to re-branding or re-labelling arrangements among test manufacturers (1), can lead to shared cross-reactivity, though this may not be the only cause. Even low levels of shared cross-reactivity, or marginally substandard performance by one RDT, could have a meaningful impact on the performance of an algorithm.

Given concerns about false positivity raised by previous findings, over the period of 2011-2015 we conducted an evaluation of eight HIV RDTs and two simple confirmatory assays differentiating antibodies against several viral proteins (14). We used specimens collected at six HIV testing and counselling (HTC) centers in sub-Saharan Africa, the most affected region by HIV/AIDS with approximately 70% of the total number of people living with HIV worldwide (15). Consistent with the aforementioned reports (5–12), this study revealed lower-than-expected specificity for most of the tests and important variations by specimen origin (14). Here, we have used these data to validate the performance of simulated algorithms developed according to the latest WHO recommendations. Additionally, we explored the possibility of using algorithms incorporating simple confirmatory assays that could be suitable for use in low- and middle-income countries.

Methods

Study setting

Samples collected at voluntary or provider-initiated HTC service programs in six public health care clinics and hospitals in Sub-Saharan Africa between August 2011 and January 2015 were used for this study: the Centre Communautaire Matam in Conakry, Guinea; Madi Opei Clinic

and Kitgum Matidi Clinic in Kitgum, Uganda; Homa Bay District Hospital in Homa Bay, Kenya; Arua District Hospital in Arua, Uganda; Nylon Hospital in Doula, Cameroun; and Baraka Hospital in Baraka, South-Kivu, DRC. The details of the HIV testing algorithm used at each site is provided elsewhere (16). A minimum of 220 positive and 220 negative specimens, as classified by the algorithm used on site, were prospectively collected as described previously (16). All frozen plasma samples were then sent to the AIDS reference laboratory at the Institute for Tropical Medicine (ITM), Antwerp, Belgium, for characterization with a standard reference algorithm (Figure 1) and for testing with eight RDTs and two simple confirmatory assays.

Reference method for HIV diagnosis

All plasma samples were tested at ITM using a fourth-generation ELISA (Vironostika® HIV Uni-Form II Ag/Ab, bioMérieux, France) followed by a Line-Immunoassay (LIA, i.e. INNO-LIA™ HIV I/II Score, Innogenetics NV, Ghent, Belgium and an antigen-enzyme-immunoassay (Ag-EIA, i.e. INNOTEST HIV Antigen mAb, Innogenetics NV, Ghent, Belgium) and in-house DNA PCR when applicable, as described in Figure 1.

HIV rapid diagnostic tests

All eight HIV RDTs and two simple confirmatory assays were performed at ITM on all collected plasma samples from the six study sites, as reported elsewhere (14). All tests were performed by six trained laboratory technicians. Each test was read by two technicians, who were blinded to the results of the other reader and to the reference standard result. When the two readers gave discordant results, a third reader was consulted to solve the discrepancy. The details of the tests, as well as their performance per origin of specimens in our evaluation, are presented elsewhere (14).

Simulated algorithms

Results of the RDTs performed at ITM were used to construct simulated algorithms using the WHO-recommended testing strategies for high- (≥5%) and low- (<5%) prevalence settings, as described in Figure 1 (A and B). We could not simulate the repetition of the tests for discordant RDT1+ RDT2- results, nor retesting 14 days later, as recommended by WHO. All simulations used the RDT Determine as the first test. For RDT2 and RDT3, we selected all assays that met WHO recommendations, i.e. sensitivity ≥99% and specificity ≥99%, based on their individual performance estimates, compared to the reference algorithm, per origin of specimens (14). For sites where fewer than two tests met these criteria, we expanded the criteria to tests that had specificity estimates ≥98%, or ≥96%. We also ensured that assays RDT2 and RDT3 had higher specificity than RDT1 in all the algorithms simulated here.

In addition, we simulated a testing strategy using an RDT as screening test, followed by a simple confirmatory assay (Figure 1C). For the screening test we used all RDTs that met the WHO recommendations for the first assay, i.e. sensitivity ≥99% and specificity ≥98%.

Statistical analysis

STATA version 13.1 (StataCorp, College Station, Texas, USA) was used to carry out data analysis.

As for any performance evaluation, results of the simulated algorithms were compared to those of the reference algorithm, considered as the gold standard. We performed an inverse-probability weighted analysis to adjust for the initial sampling strategy, which under-represented negative samples by the onsite algorithm. For each participant, the weight was calculated as the inverse of

the probability of inclusion in the study, i.e. the total number of clients with a similar onsite result during the study period divided by the number of included participants with similar results. Since all tests included in this evaluation were antibody tests that are not expected to detect acute infections, we excluded samples classified as acute infections with the reference algorithm, i.e. positive with a fourth-generation EIA, negative or indeterminate with LIA and positive with the antigen test (Figure 1). We also excluded from all analyses samples with indeterminate results by the reference algorithm. Samples with an inconclusive result with a specific simulated algorithm were excluded from the estimates of sensitivity, specificity and predictive values of this specific algorithm, and their number and proportion are reported separately.

Ethics

The study was approved by the MSF Ethical Review Board and by ethics committees in the five countries where the samples were collected. All participants provided written informed consent.

Results

From August 2011 to January 2015, a total of 2785 samples collected at the six HTC sites (between 437 and 500 samples at each site) were sent to the reference laboratory. The HIV positivity rate by site ranged from 8.0% to 37.1% (Table 1). More information on the characteristics of clients included in the study are provided elsewhere (16). Using the reference algorithm, 1306 were classified as HIV-positive (including one positive for HIV-2) and 1474 as HIV-negative. Three samples with inconclusive reference results and two classified as acute infections were excluded from the analysis.

153 The performance of the HIV RDTs and simple confirmatory assays individually and by origin of
154 specimens is described elsewhere (14). Of a total of 438 specimens that gave at least one false
155 reactive result, the majority gave a false-reactive result with only one of the eight RDTs (n=295),
156 81 with two RDTs, 41 with three RDTs, 15 with four RDTs, four with five RDTs, and two with
157 six RDTs. All RDTs exhibited some shared false reactive results with each of the seven other
158 RDTs, with the exception of SD Bioline and Stat-Pak (Table 2).

159 For only one site, Conakry (Guinea), could we identify at least two RDTs to be used as second or
160 third test with sensitivity and specificity estimates ≥99%, as recommended by WHO. Using the
161 testing strategy for high-prevalence settings with Determine as the first test and these assays as
162 second and third tests, the PPV of the algorithms ranged from 98.3%-100% (Table 3). For three
163 other sites (Douala, Cameroun; Kitgum, Uganda; and Homa Bay, Kenya), only one test met the
164 WHO criteria, necessitating the use of tests with specificity ≥98% as RDT2 and RDT3, and
165 resulting in PPVs ranging from 92.7%-100%. For the remaining two sites (Arua, Uganda and
166 Baraka, DRC), one test met the WHO criteria, but all others had specificities <98%,
167 necessitating the use of tests with specificity between 96% and 98%. The PPV of the resulting
168 algorithms ranged from 75.8%-99.6%. Detailed results are presented in Table 3.

169 Using the WHO strategy for low-prevalence settings, most simulated algorithms showed PPVs
170 ≥99%, even for the two sites (Arua, Uganda and Baraka, DRC) where tests with specificity
171 between 96 and 98% were included in the algorithms (Table 4). The proportion of inconclusive
172 results remained low at <1% for most algorithms, but rose to 2.5% at sites where tests with
173 specificity between 96% and 98% were included in the algorithms.

174 We also evaluated a simplified version of a reference algorithm, using a rapid test meeting
175 criteria for RDT1 as screening assay followed by a simple confirmatory assay. The PPV of these

Journal of Clinical Microbiology

algorithms ranged from 98.1%-100%, with the proportion of inconclusive results ranging from 0%-0.5% (Table 5).

Discussion

WHO-recommended HIV testing strategies were developed based on models using theoretical RDTs with high sensitivity and specificity and no shared cross-reactivity. Here, we have used the results of a large multi-center evaluation of individual RDTs to estimate the performance of HIV testing algorithms using real data from six sub-Saharan African HTC sites. To our knowledge, this is the first study that evaluates the performance of algorithms based on the new WHO recommendations; all other such studies published to date focus on strategies using either two tests or a third test as tiebreaker (7, 9, 11, 17–20). Though WHO has never recommended the use of a tiebreaker due to the associated risk of generating false-positive results, this strategy is still widely used and not adapted since in the different countries (21).

Several algorithms simulated here based on the strategy for high-prevalence settings resulted in a PPV <99%, even when RDTs with high specificity were used as second and third tests, due to shared false-reactive results among the tests used. In particular, a general trend of shared false-reactive results between Determine and Vikia could explain the finding that combinations using these two tests with samples from Conakry resulted in a sub-optimal PPV of 98.3%, despite the fact that each test used at this site had an estimated specificity of ≥99%. Although we could not identify a similar trend of shared false-reactive results between Determine and SD Bioline, the level of false-reactivity was high with samples from Kitgum, leading to a PPV of only 92.7% for algorithms using these tests for Kitgum despite the acceptable specificity of SD Bioline (98.6%) on specimens from this site. A larger sample size is needed to investigate whether this represents

Journal of Clinical Microbiology

a local phenomenon or a random occurrence. In the absence of reliable knowledge on the source of antigen preparations and of a good understanding of the mechanisms underlying false-reactive results, only raw data from RDT evaluation studies using samples from local sites can provide the necessary information to avoid shared false-reactive results.

For sites where only one test had a specificity of >99% and tests with specificity between 96% and 98% had to be included in the algorithms, the PPV of algorithms using the strategy for high prevalence settings varied widely depending on the order of the second and third tests. In both sites (Arua, Uganda and Baraka, DRC), only algorithms using the highly specific test STAT-PAK as the second test reached or approached the threshold, while all other combinations gave PPVs below 95%. These results underscore the importance of the order of the RDTs in the algorithm, and of using the test with the highest specificity as the second (and not third) test when employing a three-test strategy in the absence of two highly specific tests.

The strategy recommended for low-prevalence settings, which requires three reactive RDTs to establish a diagnosis of HIV infection, generally led to algorithms with very high PPV. For Baraka, DRC, where none of the high-prevalence algorithms achieved a PPV $\geq$ 99%, this was the only strategy that reached the threshold. In addition, since this strategy considers a discordant result (RDT1+; RDT2-) as negative, it is important to ensure that the NPV, together with the PPV, is >99%, as it was for the algorithms simulated here. This suggests that the low-prevalence HIV testing strategy may be suitable for use not only in settings with low HIV prevalence, but wherever HIV RDTs are known to have specificity issues.

We also propose a testing strategy that, similar to a reference algorithm, relies on a sensitive screening assay followed by a simple confirmatory assay. One of these confirmatory assays, the ImmunoComb, has shown good correlation with Western Blot in evaluations in the DRC and

Journal of Clinical Microbiology

Ethiopia when used to confirm a two-RDT algorithm positive result, but is no longer produced (11, 22). Another option, the Geenius assay, has generally shown performance results sufficient for recommending it as an alternative to existing confirmatory assays such as Western Blot or immunoblots (23–29). However, here we found that the use of these confirmatory assays did not consistently ensure PPVs ≥99% in the different combinations tested, particularly for the two sites where RDTs showed high false-reactivity. Given the added complexity and cost of the Geenius confirmatory assay, we conclude that it does not compare favourably with the three-RDT combination recommended by WHO for use in these settings.

One of the limitations of this study is that Determine was used as the first assay in all algorithms we simulated. We used Determine for the same reasons it is currently used as the first test in most algorithms: its relative low cost and very high sensitivity. Another limitation is that our sampling strategy under-represented negative clients according to the onsite algorithm, resulting in a collection of specimens that is not representative of the population screened. To account for this verification bias, we conducted a weighted analysis aimed at mitigating its effect. The inclusion of all specimens with inconclusive results from onsite testing might also explain the high proportion of false-reactive specimens in this study compared to other evaluations, including those for WHO pre-qualification. We believe, however, that these data reflect the reality of HIV testing at HTC sites. Nevertheless, although centralized testing in a reference laboratory had advantages for standardization and comparison of results, it had the disadvantage of not reproducing all aspects of field conditions. In particular, we could not reproduce repeat testing for clients with inconclusive results, which might have an impact on final performance of these algorithms. Finally, we did not illustrate the use of these algorithms in low-prevalence settings, since all specimens came from sites that would classify as high prevalence. A simple

245 calculation using the sensitivity and specificity reported here, together with the prevalence in the
246 setting of interest, could provide useful information on the expected PPV for such settings. In
247 addition, since most of the low-prevalence algorithms achieved a PPV of 100%, which would
248 not be affected by the prevalence, our data supports the recommended strategy for these settings.

249 This attempt to illustrate the process and results of designing an HIV testing strategy using real
250 data offers important lessons for navigating the various obstacles in the process. First, our data
251 underscore the impact of shared false-reactivity on algorithms performance and show that this
252 phenomenon affects most RDT combinations to different degrees. More transparent information
253 from test manufacturers is needed on possible shared false-reactivity due to test re-branding or
254 common sources of antigens. Moreover, results on shared false-reactive results from other
255 studies using a standard panel for the evaluation of different assays would provide useful
256 complementary information. Second, our results demonstrate that data from local evaluations is
257 important for assessing diagnostic accuracy in the specific setting, although often not feasible
258 (30). We also highlight the importance of the order of tests, particularly when using the strategy
259 for high HIV prevalence settings, where the test with highest specificity should be used as the
260 second rather than third assay. Finally, if sufficient information is available and these steps are
261 followed, good RDT-based HIV testing algorithms can be designed, though sometimes only with
262 the strategy recommended for low-prevalence settings.

263

264 **Acknowledgements**

265 The authors thank all study teams for their assistance and support at each site and the participants
266 for providing samples for this study. The authors also thank the laboratory staff at the Institute of
267 Tropical Medicine for sample management and testing. We would also like to thank Birgit

Journal of Clinical Microbiology

268 Schramm and Elisabeth Poulet for critical reading of the manuscript. We are grateful to Sarah
269 Venis and Patricia Kahn for their editing comments.
270 This work was supported by Médecins Sans Frontières Innovation Fund.
271 The authors declare that they have no conflict of interest

References

1. World Health Organisation. 2015. Consolidated guidelines on HIV testing services 2015. World Heal Organ.
2. Shanks L, Klarkowski D, O'Brien DP. 2013. False positive HIV diagnoses in resource limited settings: operational lessons learned for HIV programmes. PLoS One 8:e59906.
3. World Health Organization. 2004. Rapid HIV tests: guidelines for use in HIV testing and counselling services in resource-constrained settings 50.
4. World Health Organisation. 2012. Service Delivery Approaches to HIV Testing and Counselling (HTC): A Strategic HTC Programme Framework.
5. Kagulire SC, Opendi P, Stamper PD, Nakavuma JL, Mills L a, Makumbi F, Gray RH, Shott JP, Serwadda D, Reynolds SJ. 2011. Field evaluation of five rapid diagnostic tests for screening of HIV-1 infections in rural Rakai, Uganda. Int J STD AIDS 22:308–9.
6. Klarkowski D, Glass K, O'Brien D, Lokuge K, Piriou E, Shanks L. 2013. Variation in specificity of HIV rapid diagnostic tests over place and time: an analysis of discordancy data using a Bayesian approach. PLoS One 8:e81656.
7. Gray RH, Makumbi F, Serwadda D, Lutalo T, Nalugoda F, Opendi P, Kigozi G, Reynolds SJ, Sewankambo NK, Wawer MJ. 2007. Limitations of rapid HIV-1 tests during screening for trials in Uganda: diagnostic test accuracy study. BMJ 335:188.
8. Lejon V, Ngoyi DM, Ilunga M, Beelaert G, Maes I, Büscher P, Fransen K. 2010. Low specificities of HIV diagnostic tests caused by Trypanosoma brucei gambiense sleeping sickness. J Clin Microbiol 48:2836–9.
9. Singer DE, Kiwanuka N, Serwadda D, Nalugoda F, Hird L, Bulken-Hoover J, Kigozi G, Malia JA, Calero EK, Sateren W, Robb ML, Wabwire-Mangen F, Wawer M, Gray RH,

Journal of Clinical Microbiology

296 Sewankambo N, Birx DL, Michael NL. 2005. Use of stored serum from Uganda for
297 development and evaluation of a human immunodeficiency virus type 1 testing algorithm
298 involving multiple rapid immunoassays. J Clin Microbiol 43:5312–5.
299 10. Anzala O, Sanders EJ, Kamali A, Katende M, Mutua GN, Ruzagira E, Stevens G, Simek
300 M, Price M. 2008. Sensitivity and specificity of HIV rapid tests used for research and
301 voluntary counselling and testing. East Afr Med J 85:500–4.
302 11. Shanks L, Siddiqui MR, Kliescikova J, Pearce N, Ariti C, Muluneh L, Pirou E, Ritmeijer
303 K, Masiga J, Abebe A. 2015. Evaluation of HIV testing algorithms in Ethiopia: the role of
304 the tie-breaker algorithm and weakly reacting test lines in contributing to a high rate of
305 false positive HIV diagnoses. BMC Infect Dis 15:1–10.
306 12. Kroidl I, Clowes P, Mwalongo W, Maganga L, Maboko L, Kroidl AL, Geldmacher C,
307 Machibya H, Hoelscher M, Saathoff E. 2012. Low specificity of determine HIV1/2 RDT
308 using whole blood in south west Tanzania. PLoS One 7:e39529.
309 13. Klarkowski D, O'Brien DP, Shanks L, Singh KP. 2014. Causes of false-positive HIV
310 rapid diagnostic test results. Expert Rev Anti Infect Ther 12:49–62.
311 14. Kosack CS, Page A-L, Beelaert G, Benson T, Savane A, Ng'ang'a A, Andre B, Zahinda J-
312 PB, Shanks L, Fransen K. 2017. Towards more accurate HIV testing in sub-Saharan
313 Africa: a multi-site evaluation of HIV RDTs and risk factors for false positives. J Int
314 AIDS Soc 19:1–12.
315 15. THE GAP REPORT.
316 16. Kosack CS, Shanks L, Beelaert G, Benson T, Savane A, Ng'ang'a A, Andre B, Zahinda J-
317 PBN, Fransen K, Page A-L. 2017. HIV misdiagnosis in sub-Saharan Africa: Performance
318 of diagnostic algorithms at six testing sites. J Int AIDS Soc 20:21419.
319 17. Baveewo S, Kamya MR, Mayanja-Kizza H, Fatch R, Bangsberg DR, Coates T, Hahn J a,

Wanyenze RK. 2012. Potential for false positive HIV test results with the serial rapid HIV testing algorithm. BMC Res Notes 5:154.

18. Crucitti T, Taylor D, Beelaert G, Fransen K, Van Damme L. 2011. Performance of a rapid and simple HIV testing algorithm in a multicenter phase III microbicide clinical trial. Clin Vaccine Immunol 18:1480–5.

19. Galiwango RM, Musoke R, Lubyayi L, Ssekubugu R, Kalibbala S, Ssekweyama V, Mirembe V, Nakigozi G, Reynolds SJ, Serwadda D, Gray RH, Kigozi G. 2013. Evaluation of current rapid HIV test algorithms in Rakai, Uganda. J Virol Methods 192:25–27.

20. Lyamuya EF, Aboud S, Urassa WK, Sufi J, Mbwana J, Ndugulile F, Massambu C. 2009. Evaluation of simple rapid HIV assays and development of national rapid HIV test algorithms in Dar es Salaam, Tanzania. BMC Infect Dis 9:19.

21. Flynn D, Johnson C, Sands A, Wong V, Figueroa C, Baggaley R. 2012. Uptake of WHO recommended HIV testing strategies: An analysis of national policies on HIV testing services. Poster.

22. Klarkowski DB, Wazome JM, Lokuge KM, Shanks L, Mills CF, O'Brien DP. 2009. The evaluation of a rapid in situ HIV confirmation test in a programme with a high failure rate of the WHO HIV two-test diagnostic algorithm. PLoS One 4:e4351.

23. Moon H-W, Huh HJ, Oh GY, Lee SG, Lee A, Yun Y-M, Hur M. 2015. Evaluation of the Bio-Rad Geenius HIV 1/2 Confirmation Assay as an Alternative to Western Blot in the Korean Population: A Multi-Center Study. PLoS One 10:e0139169.

24. Mor O, Mileguir F, Michaeli M, Levy I, Mendelson E. 2014. Evaluation of the Bio-Rad Geenius HIV 1/2 assay as an alternative to the INNO-LIA HIV 1/2 assay for confirmation of HIV infection. J Clin Microbiol 52:2677–9.

25. Montesinos I, Eykmans J, Delforge M-L. 2014. Evaluation of the Bio-Rad Geenius HIV-1/2 test as a confirmatory assay. J Clin Virol 60:399–401.

26. Malloch L, Kadivar K, Putz J, Levett PN, Tang J, Hatchette TF, Kadkhoda K, Ng D, Ho J, Kim J. 2013. Comparative evaluation of the Bio-Rad Geenius HIV-1/2 Confirmatory Assay and the Bio-Rad Multispot HIV-1/2 Rapid Test as an alternative differentiation assay for CLSI M53 algorithm-I. J Clin Virol 58 Suppl 1:e85-91.

27. Hallen AH, Samuelson A, Nordin M, Albert J, Bogdanovic G. 2014. Evaluation of bio-rad geenius HIV-1 and -2 assay as a confirmatory assay for detection of HIV-1 and -2 antibodies. Clin Vaccine Immunol 21:1192–1194.

28. Abbate I, Pergola C, Pisciotta M, Sciamanna R, Sias C, Orchi N, Libertone R, Ippolito G, Capobianchi MR. 2014. Evaluation in a clinical setting of the performances of a new rapid confirmatory assay for HIV1/2 serodiagnosis. J Clin Virol 61:166–169.

29. Herssens N, Beelaert G, Fransen K. 2014. Discriminatory capacity between HIV-1 and HIV-2 of the new rapid confirmation assay Geenius. J Virol Methods 208:11–5.

30. Plate DK, Rapid HIV Test Evaluation Working Group. 2007. Evaluation and implementation of rapid HIV tests: the experience in 11 African countries. AIDS Res Hum Retroviruses 23:1491–8.

362 **TABLES**

363 **Table 1.** Demographic and clinical characteristics by study site

	Guinea Conakry	Cameroun Douala	Uganda Kitgum	Kenya Homa-Bay	Uganda Arua	DRC Baraka	Total
Tested at site during study period							
Total N	2033	1239	3159	1003	2971	3610	14015
Positive on site, n (%)	574 (28.2)	396 (32.0)	332 (10.5)	372 (37.1)	386 (13.0)	288 (8.0)	2348 (16.8)
Included in the study							
Total N	446	462	437	500	443	497	2785
Positive, n (%)	222* (49.8)	214 (46.3)	213 (48.7)	224 (44.8)	212 (47.9)	221 (44.5)	1306 (46.9)
Negative, n (%)	224 (50.2)	247 (53.5)	222 (50.8)	276 (55.2)	230 (51.9)	275 (55.3)	1474 (52.9)
Acute infection, n(%)	0 (0)	0 (0)	2 (0.5)	0 (0)	0 (0)	0 (0)	2 (0.1)
Indeterminates, n(%)	0(0)	1 (0.2)	0 (0)	0 (0)	1 (0.2)	1 (0.2)	3 (0.1)
Age and sex							
Median age (IQR)	29 (22-39)	31 (25-41)	30 (24-39)	30 (23-40)	29 (23-37)	30 (23-39)	30 (24-39)
Males, n (%)	132 (29.6)	163 (35.3)	176 (40.3)	201 (40.2)	213 (48.2)	177 (35.6)	1062 (38.2)

364

365

366 **Table 2.** Number and proportion of shared false-reactive results using test A1 (in column)
367 followed by A2 (in line)

A1 \ A2	No shared false-reactive results	Determine	Uni-Gold	Genie Fast	Vikia	Stat-Pak	Insti	SD Bioline	First Response
Determine (N=124)	42 (33.9)		11 (8.9)	26 (21.0)	46 (37.1)	6 (4.8)	29 (23.4)	9 (7.3)	23 (18.6)
Uni-Gold (N=39)	11 (28.2)	11 (28.2)		10 (25.6)	4 (10.3)	1 (2.6)	18 (46.2)	5 (12.8)	5 (12.8)
Genie Fast (N=102)	46 (45.1)	26 (25.5)	10 (9.8)		17 (16.7)	6 (5.9)	25 (24.5)	8 (7.8)	19 (18.6)
Vikia (N=61)	11 (18.0)	46 (75.4)	4 (6.5)	17 (27.9)		6 (9.8)	15 (25.6)	3 (4.9)	10 (16.4)
STAT-PAK (N=10)	3 (30.0)	6 (60.0)	1 (10.0)	6 (60.0)	6 (60.0)		4 (40.0)	0 (0.0)	2 (20.0)
INSTI (N=151)	86 (57.0)	29 (19.2)	18 (11.9)	25 (16.6)	15 (9.9)	4 (2.7)		18 (11.9)	18 (11.9)
SD Bioline (N=43)	9 (20.9)	9 (20.9)	5 (11.6)	8 (18.6)	3 (7.0)	0 (0.0)	18 (41.9)		20 (46.5)
First Response (N=142)	87 (61.3)	23 (16.2)	5 (3.5)	19 (13.4)	10 (7.0)	2 (1.4)	18 (12.7)	20 (14.1)	

368 N represents the total number of false reactive by RDT A1
369 The percentage in parenthesis indicates the proportion of false reactive by A2 among N.
370

371

372 **Table 3**. Simulated algorithms with Determine HIV-1/2 combined with other HIV RDTs when
373 used in a serial 3-test algorithm for high prevalence (≥5%) settings.

Specimen origin	2nd test	3rd test	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Inconclusive n (adjusted %)
Conakry, Guinea (N=446)	Uni-Gold	STAT-PAK	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia		100 (98.4-100)	99.3 (98.2-99.8)	98.3 (95.5-99.4)	100 (98.4-100)	0 (0)
	SD Bioline		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia	Uni-Gold	100 (98.4-100)	99.3 (98.2-99.8)	98.3 (95.5-99.4)	100 (98.4-100)	0 (0)
	STAT-PAK		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	SD Bioline		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Uni-Gold	SD Bioline	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia		100 (98.4-100)	99.3 (98.2-99.8)	98.3 (95.5-99.4)	100 (98.4-100)	0 (0)
	STAT-PAK		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
Douala, Cameroun (N=461)	SD Bioline*	STAT-PAK	100 (98.3-100)	100 (98.5-100)	100 (98.3-100)	100 (98.5-100)	2 (0.3)
	STAT-PAK	SD Bioline*	100 (98.3-100)	99.6 (98.3-99.9)	99.1 (96.3-99.8)	100 (98.5-100)	1 (0.1)
Kitgum, Uganda (N=435)	Uni-Gold*	STAT-PAK	96.2 (77.8-99.5)	100 (98.4-100)	100 (98.3-100)	99.5 (96.8-99.9)	0 (0)
	SD Bioline*		100 (98.3-100)	99.1 (96.4-99.8)	92.7 (76.6-98.0)	100 (98.3-100)	0 (0)
	STAT-PAK	Uni-Gold*	96.2 (77.8-99.5)	100 (98.4-100)	100 (98.4-100)	99.5 (96.8-99.9)	0 (0)
	SD Bioline*		100 (98.3-100)	99.1 (96.4-99.8)	92.7 (76.6-98.0)	100 (98.3-100)	0 (0)
	Uni-Gold*	SD Bioline*	100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	4 (1.3)
	STAT-PAK		100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	4 (1.3)
Homa bay, Kenya (N=500)	Uni-Gold	STAT-PAK*	99.6 (96.9-99.9)	99.7 (98.1-100)	99.6 (96.9-99.9)	99.7 (98.1-100)	0 (0)
	STAT-PAK*	Uni-Gold	99.6 (96.9-99.9)	100 (98.7-100)	100 (98.7-100)	99.7 (98.1-100)	2 (0.4)
Arua, Uganda (N=442)	Uni-Gold**	STAT-PAK	100 (98.3-100)	99.1 (96.5-99.8)	95.0 (82.3-98.7)	100 (98.4-100)	1 (0.1)
	Vikia**		100 (98.3-100)	98.1 (95.3-99.2)	89.7 (77.5-95.7)	100 (98.4-100)	0 (0)
	Vikia**	Uni-Gold**	100 (98.3-100)	98.1 (95.3-99.2)	89.7 (77.5-95.7)	100 (98.4-100)	2 (0.8)
	STAT-PAK		100 (98.3-100)	99.9 (99.5-100)	99.6 (97.0-99.9)	100 (98.4-100)	2 (0.8)
	Uni-Gold**	Vikia**	100 (98.3-100)	99.1 (96.5-99.8)	95.0 (92.3-98.7)	100 (98.4-100)	6 (1.6)
	STAT-PAK		100 (98.3-100)	99.9 (99.5-100)	99.6 (97.0-99.9)	100 (98.4-100)	5 (1.6)
Baraka, DRC (N=496)	Uni-Gold**	STAT-PAK	100 (98.3-100)	99.0 (96.6-99.7)	89.2 (71.3-96.5)	100 (98.6-100)	3 (0.1)
	Vikia**		100 (98.3-100)	97.3 (94.5-98.7)	75.8 (60.2-86.7)	100 (98.6-100)	1 (0.0)
	SD Bioline**		100 (98.3-100)	99.5 (97.2-99.9)	94.3 (75.3-98.9)	100 (98.6-100)	3 (0.1)
	Vikia**	Uni-Gold**	100 (98.3-100)	97.2 (94.4-98.6)	75.8 (60.2-86.7)	100 (98.6-100)	6 (1.0)
	STAT-PAK		100 (98.3-100)	99.9 (99.6-100)	98.7 (96.1-99.6)	100 (98.6-100)	6 (1.0)
	SD Bioline**		100 (98.3-100)	99.5 (97.2-99.9)	94.3 (75.3-98.9)	100 (98.6-100)	6 (1.0)
	Uni-Gold**	SD Bioline**	100 (98.3-100)	99.0 (96.6-99.7)	89.2 (71.3-96.5)	100 (98.6-100)	3 (0.5)
	Vikia**		100 (98.3-100)	97.2 (94.4-98.7)	75.8 (60.2-86.7)	100 (98.6-100)	4 (0.5)
	STAT-PAK		100 (98.3-100)	99.9 (99.6-100)	98.7 (96.1-99.6)	100 (98.6-100)	3 (0.5)
	Uni-Gold**	Vikia**	100 (98.3-100)	98.9 (96.5-99.7)	89.2 (71.3-96.5)	100 (98.5-100)	18 (2.5)
	STAT-PAK		100 (98.3-100)	99.9 (99.6-100)	98.7 (96.1-99.6)	100 (98.5-100)	16 (2.4)
	SD Bioline**		100 (98.3-100)	99.5 (97.1-99.9)	94.3 (75.3-98.9)	100 (98.5-100)	19 (2.5)

374 * RDT with specificity estimate comprised between 98.0% and 98.9% for this site
375 ** RDT with specificity estimate comprised between 96.0% and 97.9% for this site
376

377 **Table 4**. Simulated algorithms with Determine HIV-1/2 combined with other HIV RDT when
378 used in a serial 3-test algorithm for low prevalence (<5%) settings.

Specimen origin	2nd test	3rd test	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Inconclusive n (adjusted %)
Conakry, Guinea (N=446)	Uni-Gold	STAT-PAK	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia		100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	4 (0.5)
	SD Bioline		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia	Uni-Gold	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	4 (0.5)
	STAT-PAK		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	SD Bioline		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Uni-Gold	SD Bioline	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia		100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	4 (0.5)
	STAT-PAK		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
Douala, Cameroun (N=461)	SD Bioline*	STAT-PAK	100 (98.3-100)	100 (98.5-100)	100 (98.3-100)	100 (98.5-100)	1 (0.1)
	STAT-PAK	SD Bioline*	99.5 (96.7-99.9)	100 (98.5-100)	100 (98.3-100)	99.8 (98.5-100)	2 (0.3)
Kitgum, Uganda (N=435)	Uni-Gold*	STAT-PAK	96.2 (77.8-99.5)	100 (98.4-100)	100 (98.3-100)	99.5 (96.8-99.9)	0 (0)
	SD Bioline*		100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	4 (1.3)
	STAT-PAK	Uni-Gold*	96.2 (77.8-99.5)	100 (98.3-100)	100 (98.3-100)	99.5 (96.8-99.9)	0 (0)
	SD Bioline*		100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	100 (98.3-100)	4 (1.3)
	Uni-Gold*	SD Bioline*	96.2 (77.8-99.5)	100 (98.3-100)	100 (98.3-100)	99.5 (96.8-99.9)	0 (0)
	STAT-PAK		96.2 (77.8-99.5)	100 (98.3-100)	100 (98.3-100)	99.5 (96.8-99.9)	0 (0)
Homa bay, Kenya (N=500)	Uni-Gold	STAT-PAK*	99.6 (96.9-99.9)	100 (98.7-100)	100 (98.7-100)	99.7 (98.1-100)	1 (0.2)
	STAT-PAK*	Uni-Gold	99.6 (96.9-99.9)	100 (98.7-100)	100 (98.7-100)	99.7 (98.1-100)	0 (0)
Arua, Uganda (N=442)	Uni-Gold**	STAT-PAK	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	2 (0.8)
	Vikia**		100 (98.3-100)	99.9 (99.5-100)	99.6 (97.0-99.9)	100 (98.4-100)	5 (1.6)
	Vikia**	Uni-Gold**	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	6 (1.6)
	STAT-PAK		100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	1 (0.1)
	Uni-Gold**	Vikia**	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	2 (0.8)
	STAT-PAK		100 (98.3-100)	99.9 (99.5-100)	99.6 (97.0-99.9)	100 (98.4-100)	0 (0)
Baraka, DRC (N=496)	Uni-Gold**	STAT-PAK	99.6 (96.8-99.9)	99.9 (99.7-100)	99.6 (97.0-99.9)	99.9 (99.7-100)	5 (0.9)
	Vikia**		99.6 (96.8-99.9)	99.9 (99.6-99.9)	98.7 (96.1-99.6)	99.9 (99.7-100)	16 (2.4)
	SD Bioline**		100 (98.3-100)	100 (98.6-100)	100 (98.3-100)	100 (98.6-100)	3 (0.5)
	Vikia**	Uni-Gold**	99.6 (96.8-99.9)	99.9 (99.7-100)	99.1 (96.6-99.8)	99.9 (99.7-100)	18 (2.5)
	STAT-PAK		100 (98.3-100)	99.9 (99.7-100)	99.6 (97.0-99.9)	100 (98.6-100)	3 (0.1)
	SD Bioline**		100 (98.3-100)	99.9 (99.7-100)	99.6 (97.0-99.9)	100 (98.6-100)	3 (0.5)
	Uni-Gold**	SD Bioline**	99.6 (96.8-99.9)	100 (98.6-100)	100 (98.3-100)	99.9 (99.7-100)	6 (1.0)
	Vikia**		99.6 (96.8-99.9)	100 (98.6-100)	100 (98.3-100)	99.9 (99.7-100)	19 (2.5)
	STAT-PAK		100 (98.3-100)	100 (98.6-100)	100 (98.3-100)	100 (98.6-100)	3 (0.1)
	Uni-Gold**	Vikia**	99.6 (96.8-99.9)	99.9 (99.7-100)	99.1 (96.6-99.8)	99.9 (99.7-100)	5 (0.9)
	STAT-PAK		100 (98.3-100)	99.9 (99.7-100)	98.7 (96.1-99.6)	100 (98.6-100)	1 (0.0)
	SD Bioline**		100 (98.3-100)	100 (98.6-100)	100 (98.3-100)	100 (98.6-100)	4 (0.5)

379 * RDT with specificity estimate comprised between 98.0% and 98.9% for this site
380 ** RDT with specificity estimate comprised between 96.0% and 97.9% for this site
381

382

383 **Table 5.** Simulated algorithms with a rapid test used as screening test followed by a simple
384 confirmatory test for reactive samples.

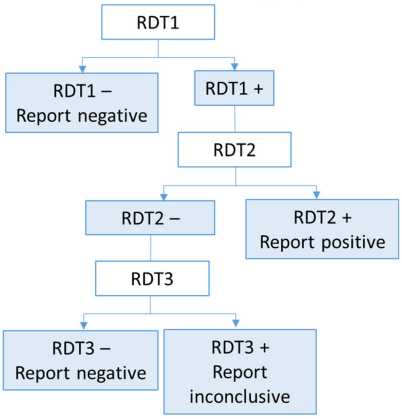
Specimen origin	Screening test	Confirmatory test	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Inconclusive n (adjusted %)
Conakry, Guinea (n=446)	Determine	Immunocomb Combfirm	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	4 (0.5)
	Uni-Gold		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Vikia		100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	3 (0.4)
	Stat-Pak		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	INSTI		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	1 (0.1)
	SD Bioline		100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	100 (98.4-100)	0 (0)
	Determine	Geenius	100 (98.4-100)	99.8 (98.8-100)	99.6 (96.9-99.9)	100 (98.3-100)	4 (0.5)
	Uni-Gold		100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	0 (0)
	Vikia		100 (98.3-100)	99.7 (98.7-99.9)	99.2 (96.6-99.8)	100 (98.3-100)	3 (0.4)
	Stat-Pak		100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	0 (0)
	INSTI		100 (98.3-100)	99.2 (94.6-99.9)	98.1 (87.5-99.7)	100 (98.4-100)	1 (0.1)
	SD Bioline		100 (98.3-100)	99.8 (98.8-100)	99.6 (97.1-99.9)	100 (98.4-100)	0 (0)
Douala, Cameroun (n=461)	Stat-Pak	Immunocomb Combfirm	99.5 (96.7-99.9)	99.8 (98.5-100)	99.5 (96.7-99.9)	99.8 (98.5-100)	1 (0.1)
	SD Bioline		100 (98.3-100)	100 (98.5-100)	100 (98.3-100)	100 (98.5-100)	1 (0.3)
	Stat-Pak	Geenius	99.5 (96.7-99.9)	99.4 (98.0-99.8)	98.6 (95.8-99.6)	99.8 (98.5-100)	0 (0.0)
	SD Bioline		100 (98.3-100)	99.8 (98.5-100)	99.5 (96.8-99.9)	100 (98.5-100)	1 (0.3)
Kitgum, Uganda (n=435)	SD Bioline	Immunocomb Combfirm	100 (98.3-100)	100 (98.4-100)	100 (98.3-100)	100 (98.4-100)	1 (0.4)
	SD Bioline	Geenius	96.2 (77.8-99.5)	100 (98.3-100)	100 (98.3-100)	99.5 (96.8-99.9)	1 (0.4)
Homa Bay, Kenya (n=500)	Uni-Gold	Immunocomb Combfirm	99.6 (96.9-99.9)	100 (98.7-100)	100 (98.4-100)	99.7 (98.1-99.9)	2 (0.3)
	Stat-Pak		99.6 (96.9-99.9)	100 (98.7-100)	100 (98.4-100)	99.7 (98.1-99.9)	2 (0.4)
	Uni-Gold	Geenius	99.6 (96.9-99.9)	100 (98.7-100)	100 (98.4-100)	99.7 (98.1-99.9)	2 (0.3)
	Stat-Pak		99.6 (96.9-99.9)	100 (98.7-100)	100 (98.4-100)	99.7 (98.1-100)	2 (0.4)
Arua, Uganda (n=442)	Stat-Pak	Immunocomb Combfirm	100 (98.3-100)	99.9 (99.5-100)	99.6 (97.0-99.9)	100 (98.4-100)	1 (0.1)
	Stat-Pak	Geenius	100 (98.2-100)	99.9 (99.5-100)	99.6 (97.0-99.9)	100 (98.4-100)	0 (0)
Baraka, DRC (n=496)	Stat-Pak	Immunocomb Combfirm	100 (98.3-100)	99.9 (99.7-100)	99.6 (97.0-99.9)	100 (98.7-100)	2 (0.1)
	Stat-Pak	Geenius	100 (98.3-100)	99.9 (99.7-100)	98.7 (96.1-99.6)	100 (98.7-100)	0 (0)

385
386
387
388
389

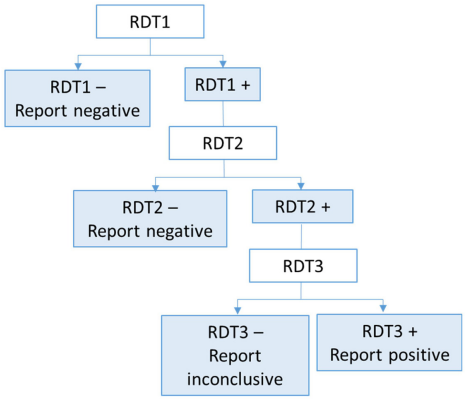
390 **Figure 1.** HIV testing strategies used to simulate algorithms (A to C) and as reference testing
391 algorithm (D).

Journal of Clinical Microbiology

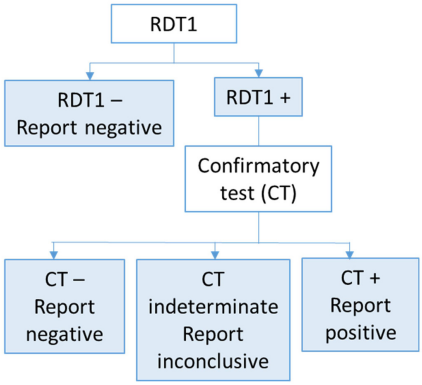
A/ Simulated algorithms – high prevalence settings



B/ Simulated algorithms – low prevalence settings



C/ Simulated algorithms – confirmatory test



D/ Reference testing