Calibration and Validation of an Oral Fluid-Based Sensitive/Less-Sensitive Assay to Distinguish Recent From Established HIV-1 Infections

Anne M. Sill,¹ Kristen Kreisel,¹ Bethany Griffin Deeds,¹ Craig M. Wilson,^{2,3}
Niel T. Constantine,¹ Ligia Peralta,^{1*} Adolescent Trials Network
for HIV/AIDS Interventions

¹*Departments of Pediatrics, Pathology, and Epidemiology, University of Maryland
School of Medicine, Baltimore, Maryland*

²*Department of Pediatrics, University of Alabama at Birmingham, Birmingham, Alabama*

³*Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama*

Sensitive/less-sensitive (S/LS) serum-based serologic methods have been developed to measure human immunodeficiency virus (HIV) incidence by distinguishing recent from established infections. Such methods require venipuncture. The goal of this study was to develop an alternative to serum-based S/LS testing using oral fluid (OF) as the testing medium. Serum/OF pairs were collected from 342 patients attending 15 Adolescent Trials Network (ATN) clinical sites. The sera were tested with the use of the dilutional Vironostika (DV; Biomerieux, Durham, NC) S/LS assay (DV_{SOD} = 1.0) as the reference against which an OF LS assay was calibrated using 40 of the OF pairs. Receiver operating characteristic (ROC) curve analyses pinpointed the OF LS test parameters that maximized concordance with the serum-based DV. Validation of the calibrated OF LS included testing of the remaining 302

serum/OF pairs. During calibration the maximum concordance with the DV was 95.2% and 89.5% for 21 recent and 19 established samples, respectively, at a 1:50 OF sample dilution and an optical density (OD) cutoff of 0.280. When applied to the validation sample set (N = 302), the concordance was 73.6% for the recent samples and 89.6% for the established samples. The OF LS assay showed a good concordance with the serum-based reference S/LS assay. It presents an alternative to invasive specimen collection, and has the potential for increasing test compliance in young subjects. However, because of the uncertainty of the performance characteristics of the serum-based S/LS assay with which it was compared, further validation of the OF LS using seroconversion sample pairs is needed. *J. Clin. Lab. Anal.* 21: 40–45, 2007. © 2007 Wiley-Liss, Inc.

Key words: HIV; saliva; incidence; S/LS; adolescents

INTRODUCTION

The ability to distinguish recent from established human immunodeficiency virus (HIV) infections is important for estimating HIV-1 incidence, referring newly infected patients to early-intervention programs, and evaluating the effectiveness of prevention programs (1–3). Since the first sensitive/less sensitive (S/LS) serologic assay was developed to identify recently infected persons (<6 months postseroconversion), S/LS technology has been used in many studies in the United States and worldwide to estimate HIV incidence. This assay is based on the concept that HIV antibody

Grant sponsor: National Institute of Child Health and Human Development, NIH; Grant number: U01 HD40533; Grant sponsor: National Institute on Drug Abuse; Grant sponsor: National Institute of Mental Health; Grant sponsor: National Institute on Alcohol Abuse and Alcoholism.

*Correspondence to: Ligia Peralta, M.D., Associate Professor of Pediatrics and Epidemiology, Department of Pediatrics, University of Maryland School of Medicine, 2nd floor, 737 Lombard Street, Baltimore, MD. E-mail: lperalta@peds.umaryland.edu

Received 18 July 2006; Accepted 30 October 2006

DOI 10.1002/jcla.20144

Published online in Wiley InterScience (www.interscience.wiley.com).

titers rise during the early months of infection, and the detection of low levels of antibody indicates recent infection (4). The technology uses serum or plasma for testing, and thus requires invasive sample collection methods.

Currently, the most widely used S/LS tests are the Vironostika HIV-1 Microelisa (Biomerieux, Durham, NC), which is also known as the dilutional Vironostika (DV), and the BED CIE test (Calypse Biomedical, Alameda, CA). However, the DV is targeted to be removed from the market, and a recent report by the UNAIDS Program Secretariat indicated that the BED-CIE test should not be used to estimate incidence or measure trends until further evaluations are conducted (5). In addition, the DV S/LS test has been shown to have only 79.2% sensitivity for identifying recent infection samples and 94.4% sensitivity for identifying established infections (6), which makes its utility questionable.

Diagnostic and blood-screening HIV tests that use oral fluid (OF) as a sample medium were developed in the 1990s, and have resulted in FDA approval of an enzyme-linked immunosorbent assay (ELISA), a Western blot, and a rapid test. The advantages of using OF instead of blood are obvious and have been described elsewhere (7). In addition, we previously determined that patients prefer the collection of OF for HIV testing (8). Therefore, we sought to exploit an oral fluid test to develop, calibrate, and validate it as an S/LS enzyme Immunoassay (EIA) method to distinguish recent from established HIV infections for the purposes of identifying to serum-based S/LS assays.

MATERIALS AND METHODS

Population and Sample Collection

A total of 358 HIV-positive adolescents (12–24 years old) who were able to give informed consent and were newly engaged in care at 15 Adolescent Medicine Trials Network (ATN) clinical sites in the United States and Puerto Rico participated in this study. Blood was collected via standard venipuncture procedures, and the serum was separated and stored frozen until it was tested. OF pairs were collected with the use of an Orasure collection device (Orasure Technologies, Bethlehem, PA), stored at room temperature for up to 21 days until they were tested, and then frozen.

Characterization of Samples

Of the 358 patients enrolled, 342 were deemed eligible for the study, and their serum and OF pairs were suitable for testing. Seropositivity for HIV was verified on serum and OF pairs with the use of the Vironostika Microelisa System (serum-sensitive assay) and the

Vironostika OF Microelisa System (OF-sensitive assay), respectively, according to the manufacturer's instructions. All positive sera were then tested using the DV LS assay as the reference, as described elsewhere (9), to classify the samples as being from subjects with recent or established HIV infection. The DV LS test was originally used with an SOD of 0.75; however, since in 2003 the Centers for Disease Control (CDC) recommended the use of a Standardized Optical Density (SOD) of 1.0 (10) (J. Mei, CDC, personal communication), that is what we used in this study. At the time of this study, only the DV S/LS test (and not the HIV-1 Incidence BED CIE EIA) was available for comparison.

Calibration of the OF LS Test

OF pairs from the first 21 recent and 19 established HIV-1 infected individuals (originally selected as 20 and 20, respectively, using the original DV LS SOD of 0.75), as classified by the reference DV LS using sera at the 1.0 SOD, were used to calibrate the Vironostika OF Microelisa System to act as an S/LS test (OF LS). The parameters that were evaluated to optimize the OF LS for maximum concordance with the DV LS included multiple OF optical density (OD) cutoffs. Values below and above these cutoffs indicated a recently or chronically infected patient, respectively. Different OF sample dilutions (1:10, 1:25, 1:50, 1:60, 1:80, and 1:100) were used. Three dilutions were selected (data not shown) that yielded the best concordance with the DV LS for further comparison. A receiver operating characteristic (ROC) curve (11) analysis was used to pinpoint the optimum OF LS OD cutoff for recent samples and established samples at each of the sample dilutions. The OD cutoff and the sample dilution that yielded the highest concordance rate with the DV LS were chosen for use in the validation phase.

Recalibration of the OF LS Test

The optimum parameters defined in the calibration phase were applied to a larger sample set that included the 21 recent and 19 established calibration samples and one-third of the samples in the validation sample set (randomly selected). The resulting sample set included 52 recent samples and 89 established samples. The purpose of this recalibration was to determine whether assessing a larger sample set would result in a change in the parameters of the OF LS for better concordance with the DV LS.

Validation and Reproducibility of the OF LS Test

The validation sample set (N = 302) did not include the samples from the calibration phase. We evaluated the concordance rates of the DV LS and OF LS on

91 recent and 211 established infection samples (as determined by the DV LS) using the OF LS parameters determined during the calibration/recalibration phase. Reproducibility studies for both the DV LS and OF LS tests were conducted using the standard DV LS procedure and the optimum OF LS dilution and cutoff that had been selected. We tested three samples (one recent, one borderline, and one established as classified by the DV LS) using both the DV LS and OF LS 10 times per day for three days, and using the same kit lots of each assay. Interday variation (3 samples $\times$ 10 runs over 3 days, same kit lot) and within-run variation (3 samples $\times$ 10 runs on the same day, same kit lot) were assessed to measure test reproducibility.

Statistical Analysis

All statistical analyses were performed with the use of SPSS (version 12.01). The concordance between the OF LS and the serum-based DV LS was determined by means of cross tabulation and the Kappa statistic, and 95% confidence intervals (CIs) were calculated around the concordance rates for recent and established samples. At each of the three selected sample dilutions, ROC analyses of the OF LS OD cutoffs determined the one OF LS OD cutoff that gave maximum combined “sensitivity” and “specificity” and the smallest difference between the sensitivity and specificity as compared to the reference DV classifications. The sample dilution at which the ROC-determined cutoffs best discriminated

between recent and established infections compared to the reference DV LS was chosen as the optimum dilution. The reproducibility of each assay was determined by calculating the coefficients of variation (CVs) of values obtained during replicate testing.

RESULTS

Calibration

The ROC and cross-tabulation analyses revealed that the highest combined concordance for recent and established samples (as classified by the DV LS at the 1.0 SOD cutoff) was found at the OF LS sample dilution of 1:50 and the OF LS OD cutoff of 0.280 (Fig. 1) using the small sample size of 40. As shown in Tables 1 and 2, this resulted in a concordance of 95.2% and 89.5% for recent and established infection samples, respectively ($\kappa = 0.849$).

Recalibration

The ROC and cross-tabulation analyses of the larger sample set (52 recent and 89 established infection samples) reiterated the findings of the calibration phase that the highest combined concordance for recent and established infection samples was found at the 1:50 dilution and the OF LS OD cutoff of 0.280. Concordance was found to be 86.5% and 88.8% for recent and established samples, respectively (Table 2).

Fig. 1. Calibration of the OF LS test. ROC analysis of the OF LS against the DV LS. Following the selection of the optimum OF dilution of 1:50, iterative ROC analyses reveal optimal concordance of the OF LS against the DV LS at the OD cutoff of 0.280. Points at the farthest top-left of the graph show optimal sensitivity (concordance) for recent samples and sensitivity (concordance) for established samples.

TABLE 1. Concordance Rates at Different Oral Fluid Sample Dilutions*

Oral fluid dilution	% Concordance (recent)	% Concordance (established)
1:50	95.2	89.5
1:60	100.0	47.0
1:80	100.0	53.0

*During assay calibration, different oral fluid dilutions were assessed to determine the best concordance rates with the DV LS test.

Validation

Using the validation sample set of 302 serum/OF pairs, we measured the concordance of the OF LS and the DV LS assays at the optimum parameters selected during the calibration phase (OF LS dilution = 1:50; OF LS OD cutoff = 0.280). Concordance on this sample set was now determined to be 73.6% and 89.6% ($\kappa = 0.636$) for recent and established samples, respectively. Although the 95% CIs overlapped substantially, the absolute concordance for the recent samples fell outside of the 95% CIs found in the calibration phase, while the concordance for established samples was found to be within that determined during the calibration phase.

Reproducibility

The average within-run CV of the DV LS was 0.31%, and the average within-run CV of the OF LS assay was 5.26%. The average interday CVs were 25.71% and 25.40% for the OF LS and DV LS assays, respectively (data not shown).

DISCUSSION

With the parameters selected during the calibration phase, the concordance rates between the OF LS and the reference DV LS assays found in the validation sample set were 74% and 90% for recent and established samples, respectively. Since we realized that the small sample size of the calibration sample set ($N = 21$ recent and 19 established samples) might not provide ample confidence and power to definitively determine the optimum parameters, we performed a recalibration analysis and reiterated the conclusions made during the calibration phase. Thus, the recalibration confirmed the findings of the calibration phase (i.e., optimal performance was found at an OF LS sample dilution of 1:50 and OD cutoff of 0.280). The subsequent validation of the OF LS yielded less impressive indices of concordance; however, the significant overlap in the 95% CIs of the validation and recalibration concordance rates indicated no significant differences in the rates of concordance. That is, the overlap in the CIs

TABLE 2. Concordance Rates During Calibration, Recalibration, and Validation of the OF LS Against the DV LS

Specimen set	N	Procedure	Oral fluid dilutions	Parameters tested	Concordance recent (95% CI)	Concordance established (95% CI)
Calibration ($N = 40$) Recalibration 40 (calibration samples + 101 validation samples)	21 Recent	ROC analysis, Concordance	1:10, 1:25,	OD cutoff 0.270	90.5%	89.5%
	19 Established		1:50, 1:60, 1:80, 1:100	OD cutoff 0.280	95.2% (77.2%–98.9%)	89.5% (68.3%–96.8%)
	52 Recent	ROC analysis, Concordance	1:50	OD cutoff 0.290	95.2%	84.2%
	89 Established			OD cutoff 0.270	82.7%	91.0%
Revalidation samples (all 302 validation samples)	91 Recent	Concordance Analysis	1:50	OD cutoff 0.280	86.5% (74.7%–93.2%)	88.8% (80.5%–93.7%)
	211 Established			OD cutoff 0.290	86.5%	86.5%
				OD cutoff 0.280	73.6% (77.0%–91.0%)	89.6% (80.4%–89.8%)

suggested that the OF LS assay had been optimally calibrated during the initial calibration.

It should be noted that the DV S/LS test was previously shown to possess only 79.2% sensitivity for identifying recent infection samples and 94.4% sensitivity for identifying established infections (6). It is difficult to determine the accuracy of the OF LS test for correctly classifying infections in a concordance study because it is not possible to determine which test is producing the most accurate classifications. Although the concordance between the DV LS and OF LS assays may be acceptable, given the uncertainty of the accuracy of the DV LS test, we cannot conclude that the OF LS assay is better or worse than the DV LS. Our goal in this study was to show a good concordance between the OF LS test and the reference DV LS serum test, which would suggest that OFs can yield results equivalent to those of the reference test. The results did show a good concordance. However, one can only make a definitive conclusion about the “accuracy” of both tests by comparing the tests using samples that have been collected from persons with known times of infection (serum and OF conversion panels), as opposed to relying on samples that have been classified by another, less than perfect S/LS test.

Although we cannot recommend the use of the OF LS test at this time, our study shows that OF as a sample medium has the potential to yield results similar to those obtained with a serum-based S/LS test, as evidenced by a fairly good concordance with the reference serum S/LS assay. If the OF LS technology is shown to be accurate (with recalibration if necessary using appropriate panels), it would be an ideal tool for use in the developing world, since it would minimize the need for invasive collection techniques and would promote greater compliance among subjects. In addition, an HIV oral fluid rapid assay is available and licensed in the US, that offers the potential advantage for use if successfully developed as an S/LS assay. However, it is difficult to obtain seroconversion panels that contain both serum and OF because they require sequential collections from persons who have been identified as having interval or incident seroconversion.

In conclusion, our results are encouraging and suggest that development of an OF LS test should be pursued. Further studies using OF from persons with known times of HIV infection are essential and will facilitate the development of an OF S/LS assay that can offer important advantages over serum-based methods.

ACKNOWLEDGMENTS

The ATN for HIV/AIDS Interventions is funded by grant U01 HD40506 from the National Institutes of

Health through the National Institute of Child Health and Human Development (Audrey Smith Rogers and Leslie Serchuck), with supplemental funding from the National Institute on Drug Abuse (Nicolette Borek), the National Institute of Mental Health (Andrew Forsyth and Pim Brouwers), and the National Institute on Alcohol Abuse and Alcoholism (Kendall Bryant). This research study was scientifically reviewed by the ATN's Therapeutic Research Leadership Groups. The authors acknowledge the ATN Coordinating Center at the University of Alabama at Birmingham, which is funded by grant U01 HD40533 (Craig Wilson and Cindy Partlow); the ATN Data and Operations Center at WESTAT, Inc. (Jim Korelitz and Barbara Driver); the Community Advisory Board; and the subjects who participated in the study.

The following ATN sites participated in this study: University of South Florida (Patricia Emmanuel, M.D., Jorge Lujan-Zilberman, M.D., and Silvia Callejas, RN), Children's Hospital of Los Angeles (Marvin Belzer, M.D., Cathy Salata, RN, and Diane Tucker, RN, MSN), Children's Hospital National Medical Center (Lawrence D'Angelo, M.D., Connie Trexler, RN, PNP, and Stephanie Crane, MPH), Children's Hospital of Philadelphia (Brett Rudy, M.D., Mary Tanney, MPH, and Adrienne DiBenedetto, BSN), John H. Stroger Jr. Hospital and the CORE Center (Lisa Henry-Reid, M.D., Kelly Bojan DsN, NP, and Rachel Jackson, NP), University of Puerto Rico (Irma Febo, M.D., Ileana Blasini, M.D., and Evelyn Rivera, RN, MPH), Montefiore Medical Center (Donna Futterman, M.D., Marina Catallozzi, M.D., and Elizabeth Enriquez-Bruce, M.D.), Mount Sinai Medical Center (Linda Levin-Carmine, M.D., Mary Geiger, RN, MPH, and Angela Lee, PA, MPH), University of California–San Francisco (Barbara Moscicki, M.D., and Coco Auerswald, M.D.), Tulane University Health Sciences Center (Sue Ellen Abdalian, M.D., Leslie Kozina, RN, CCRC, and Trina Jeanjacques), University of Maryland (Ligia Peralta, M.D., Edwin Colocho, M.D., and Leonel Flores, M.D.), University of Miami School of Medicine (Lawrence Friedman, M.D., Donna Maturo, MSN, ARNP, and Hanna Major-Wilson, MSN, ARNP), Children's Diagnostic and Treatment Center (Ana Puga, M.D., Esmine Leonard, BSN, and Amy Inman, BS), Children's Hospital Boston (Cathryn Samples, M.D., Ellen Cooper, M.D., and Helen Mahoney-West, MSN), and University of California–San Diego (Stephen Spector, M.D., Rolando Viani, M.D., MTP, and Lisa Stangl, RN, NP). Four of these sites utilized their General Clinical Research Center/Pediatric Clinical Research Center for this study. These centers were supported by grants from the General Clinical Research Center Program of the National Center for Research

Resources, National Institutes of Health, Department of Health and Human Services, as follows: Children's National Medical Center, M01-RR13297; University of Pennsylvania/Children's Hospital of Philadelphia, M01 RR00240; Tulane Medical Center, M01 RR 05096-10; and University of California–San Diego, M01 RR00827.

REFERENCES

1. Constantine NT, Zink H. HIV testing technologies after two decades of evolution. *Indian J Med Res* 2005;121:519–538.
2. Dobbs T, Parekh BS. Detecting recent human immunodeficiency virus type 1 infection: why and how? *MLO Med Lab Obs* 2003; 35:12–14, 16, 19–20 passim; quiz 24–25.
3. Rutherford GW, Schwarcz SK, McFarland W. Surveillance for incident HIV infection: new technology and new opportunities. *J Acquir Immune Defic Syndr* 2000;25(Suppl 2):S115–S119.
4. Janssen RS, Satten GA, Stramer SL, et al. New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes. *JAMA* 1998;280:42–48.
5. UNAIDS Reference Group on Estimates, Modeling and Projections statement on the use of the BED-assay for the estimation of HIV-1 incidence for surveillance or epidemic monitoring 2005. <www.epidem.org/Publications/BED%20statement.pdf>
6. Constantine NT, Sill AM, Cleghorn FR, et al. Improved classification of recent HIV-1 infection by employing a two-stage sensitive/less-sensitive test strategy. *J Acquir Immune Defic Syndr* 2003;32: 94–103.
7. Constantine NT, Saville R, Dax EM. Retroviral testing and quality assurance: essentials for laboratory diagnosis. Ann Arbor: Malloy. 2005. Malloy Printers, ISBN: 0-9785982-0-2;2005.
8. Peralta L, Constantine N, Griffin-Deeds B, Martin L, Ghalib K. Evaluation of youth preferences for rapid and innovative human immunodeficiency virus antibody tests. *Arch Pediatr Adolesc Med* 2001;155:838–843.
9. Rawal BD, Degula A, Lebedeva L, et al. Development of a new less-sensitive enzyme immunoassay for detection of early HIV-1 infection. *J Acquir Immune Defic Syndr* 2003;33: 349–355.
10. Kothe D, Byers RH, Caudill SP, et al. Performance characteristics of a new less sensitive HIV-1 enzyme immunoassay for use in estimating HIV seroincidence. *J Acquir Immune Defic Syndr* 2003;33:625–634.
11. Begg CB. Statistical methods in medical diagnosis. *CRC Crit Rev Med Inform* 1986;1:1–22.