JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 1996, p. 2695–2701 Vol. 34, No. 11
0095-1137/96/$04.00+0

Evaluation of a Quality Assurance Program for Quantitation of Human Immunodeficiency Virus Type 1 RNA in Plasma by the AIDS Clinical Trials Group Virology Laboratories

BELINDA YEN-LIEBERMAN,[1] DONALD BRAMBILLA,[2] BROOKS JACKSON,[1†] JAMES BREMER,[3] ROBERT COOMBS,[4] MIKE CRONIN,[5] STEVEN HERMAN,[6] DAVID KATZENSTEIN,[7] SHIELA LEUNG,[2] HSIANG JU LIN,[8] PAUL PALUMBO,[9] SURAIYA RASHEED,[10] JOHN TODD,[11] MARYANNE VAHEY,[12] AND PATRICIA REICHELDERFER[13]*

Cleveland Clinic Foundation, University Hospitals of Cleveland, and Institute of Pathology, Case Western Reserve University, Cleveland, Ohio 44106[1]; New England Research Institute, Watertown, Massachusetts 02172[2]; Rush-Presbyterian-St. Luke's Medical Center, Chicago, Illinois 60612[3]; University of Washington, Seattle, Washington 98144[4]; Organon Teknika Corporation, Durham, North Carolina 27712[5]; Roche Molecular Systems, Branchburg, New Jersey 08876[6]; Stanford University, Stanford, California 94305[7]; Baylor College of Medicine, Houston, Texas 77030[8]; University of Medicine and Dentistry of New Jersey, Newark, New Jersey 07103[9]; University of Southern California School of Medicine, Los Angeles, California 90033[10]; Chiron Corporation, Inc., Emeryville, California 94608[11]; Walter Reed Army Institute of Research, Rockville, Maryland 20850[12]; and National Institutes of Allergy and Infectious Diseases, Bethesda, Maryland 20892[13]

Received 24 April 1996/Returned for modification 10 July 1996/Accepted 9 August 1996

A number of quantitative assays have been developed by using amplification techniques to measure human immunodeficiency virus type 1 RNA in the plasma of infected individuals. The Virology Committee of the AIDS Clinical Trials Group in the Division of AIDS, National Institute of Allergy and Infectious Diseases, has established a quality assurance program (QAP) for quantitative assays of HIV-1 RNA levels in plasma. The primary objective of the QAP was to ascertain that a laboratory could maintain the precision required to have a 90% power to detect a fivefold difference in RNA copy number between two samples in the same batch. To achieve this goal, the QAP required an intra-assay standard deviation of no greater than 0.15 $\log_{10}$ RNA copies per ml. Panels for proficiency testing consisted of coded replicate samples and a common set of standards. To date, 41 laboratories have participated in the program and have used both commercial and in-house assays. We demonstrated that 65% of the laboratories were capable of attaining the necessary level of intra-assay precision. The fitted regressions indicated that the differences among laboratories that used the same kit were generally greater than the differences among population-average regressions for the kits themselves. The use of an external QAP and a common set of standards reduced differences both among laboratories that used the same kit and among laboratories that used different kits. Thus, use of a common set of standards across clinical trial protocols would allow for cross-protocol comparisons.

A number of assays are available for the quantitation of human immunodeficiency virus type 1 (HIV-1) RNA levels in plasma (1, 11, 16, 20, 23, 25, 26, 30, 33, 35). These assays have been used to assess antiretroviral activity (4, 15, 34) and more recently to define aspects of viral pathogenesis, replication kinetics (3), and antiretroviral drug resistance (19). The prevalence of quantifiable levels of HIV-1 RNA in the plasma of most seropositive patients (26), the inverse correlation between plasma HIV-1 RNA levels and $CD4^+$ cell count (28), and the ability of HIV-1 RNA levels to independently predict clinical disease progression (22, 24) suggest that this measure of viral load may be useful as an alternative endpoint in clinical trials (5, 17, 24, 38).

These assays differ in their requirements for sample volume and sample preparation, methods of amplification, and methods of detection. There are strengths and weaknesses to all of the assays; thus, there is probably no one assay which is best suited to all situations. In 1993, a multicenter evaluation of quantitative HIV-1 RNA detection assays demonstrated that several of the procedures were sufficiently reproducible so that an empiric fourfold change could be viewed as being statistically significant (21). Comparison of the individual assay standards with a common set of standards showed some disagreement in the assigned nominal copy number values. Since there were some systematic differences in methodology, it was unclear if a common set of standards could prove useful in aligning the different measurement assay systems.

Multicenter clinical trials are using a variety of assays to measure plasma HIV-1 RNA levels; as such, there is a need to ensure the comparability of the HIV-1 RNA data, within and among trials, obtained from different sites by various methodologies. To address this need, we have developed a quality assurance program to assess the laboratory and kit performance of HIV-1 RNA measurements suitable for clinical studies.

MATERIALS AND METHODS

Participating laboratories. The participating laboratories consisted of 30 laboratories sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), 5 Laboratories sponsored by the National Institute of Child Health and Human Development, the Walter Reed Army Institute of Research (WRAIR) PCR Laboratory, and 4 commercial laboratories. All of the participating laboratories had prior experience with molecular techniques for the detection of

* Corresponding author. Mailing address: NIH/NIAID/DAIDS, 6003 Executive Blvd., Rockville, MD 20776. Phone: (301) 402-0131. Fax: (301) 402-3171.
† Present address: Johns Hopkins University Medical Center, Baltimore, Md.

TABLE 1. Coded panel composition for RNA proficiency panels 2, 3, 4, and 5

Panel round 2		Panel round 3		Panel round 4		Panel round 5	
No. of RNA copies/ml	No.[a]	No. of RNA copies/ml	No.	No. of RNA copies/ml	No.	No. of RNA copies/ml	No.[b]
1,500,000	3	1,375,000	4	1,500,000	2	1,500,000	6/5/5/0
300,000	3	275,000	4	300,000	5	300,000	6/5/5/6
60,000	3	55,000	4	60,000	5	60,000	6/5/5/6
12,000	3	11,000	4	12,000	5	12,000	6/5/5/6
2,400	3	2,200	2	2,400	2	2,400	0/4/4/6
480	3						

[a] Number of replicates.
[b] Replicates are given for Chiron/Digene/NASBA/Roche.

HIV-1 RNA in plasma, and in most instances the technicians in the participating laboratories had undergone training with the specific commercial assay system that they were using. Nine laboratories participated in the first round of testing, 13 laboratories participated in the second round, and 18, 19, and 38 laboratories participated in the third, fourth, and fifth rounds, respectively.

HIV-1 RNA detection. Three of the assay procedures, the Amplicor HIV Monitor test (Roche assay; Roche Diagnostic Systems, Inc., Branchburg, N.J.), the Digene Sharp Signal system (Digene assay; Digene Diagnostics, Inc., Silver Spring, Md.), and the WRAIR in-house procedure (WRAIR assay), are all reverse transcriptase PCR-based assays. All three assays use *gag*-directed primer pairs and have been described previously (20, 23, 33). Two of the assays (WRAIR and Digene) use external standard curves, while the Roche procedure uses an internal standard. Detection of PCR amplification products by the Digene and Roche assays is carried out in microtiter plates with an enzyme-linked immunosorbent assay format. In the Digene assay, the substrate incubation time can vary from 1 to 24 h. In the WRAIR assay, PCR products are hybridized with isotopically labeled oligonucleotide probes and are detected by phosphor image analysis on acrylamide gels.

The Organon Teknika nucleic acid sequenced-based amplification (NASBA) system (NASBA assay; Organon Teknika Corporation, Durham, N.C.) uses isothermal target amplification. The amplification is based on the simultaneous reaction of three enzymatic activities: avian myeloblastosis virus reverse transcriptase, RNase H, and T7 RNA polymerase. The steps of the NASBA process take place in one reaction tube at a single temperature (41°C). Amplification is achieved through the simultaneous activities of the three enzymes. The primers used are from a conserved region of the HIV-1 *gag* gene. Nucleic acid detection occurs by hybridization analysis with ruthenium-labeled probes; this is followed by signal quantitation with an electrochemiluminescence reader (35).

Versions 1.0 and 2.0 of the Chiron branched-chain bDNA assay (Chiron assay; Chiron Corporation, Inc., Emeryville, Calif.) are based on signal amplification rather than target amplification methods. Genomic nucleic acid is captured onto a 96-well microtiter plate by a series of specific oligonucleotide target probes which are complementary to 10 discrete sites on the HIV-1 polymerase gene. Another set consisting of 39 different target probes is used to hybridize the viral genome and the branched DNA amplifier. Multiple copies of an alkaline phosphatase-labeled probe are hybridized to the amplifiers, resulting in enhancement of the signal. Detection occurs by the addition of a chemiluminescent substrate and quantitation of the relative light units (RLUs) emitted by the bound alkaline phosphatase (25). Although the manufacturer recommends that duplicate assays be performed to allow for outlier detection, each specimen was assayed only once. However, outlier detection was still possible because of the replicative design of the panels.

Quality assurance procedures. (i) HIV-1 RNA stock standard. The HIV-1 RNA standards consisted of uninfected human plasma to which HIV-1 was added. A zero-copy-number standard was included with each run to ensure the uninfected nature of the seronegative plasma. These standards were prepared at the Virology Quality Assurance (VQA) laboratory (Rush-Presbyterian-St. Luke's Medical Center). The HIV-1 standards consisted of supernatants obtained by coculturing the peripheral blood mononuclear cells from two HIV-1-infected patients with seronegative donor peripheral blood mononuclear cells as described previously (21). Specifically, the cultures were initiated with a standard volume and were incubated according to the AIDS Clinical Trials Group protocol for qualitative cultures (8). The cultures were harvested, medium was added every third day, and medium containing phytohemagglutin-stimulated peripheral blood mononuclear cells was added once a week. The supernatants were monitored for HIV-1 p24 antigen levels (Coulter, Miami, Fla.) until all culture supernatants had a p24 antigen concentration in excess of 400 pg/ml. The supernatants from the cultures were pooled (approximately 300 ml), distributed into 1-ml aliquots, and frozen in liquid nitrogen.

(ii) Characterization of the standard. The standard stock was sent to Peter Nara (National Cancer Institute) for particle density determination by electron microscopy. The estimated copy number by this procedure was 9.8×10^8 HIV-1 RNA copies per ml. The stock had an absolute HIV-1 p24 antigen level of 40,000 pg/ml, which corresponded to an estimated HIV-1 RNA copy number of 4.0×10^8 copies per ml (5). The HIV-1 RNA stock was also sent to Roche Molecular Systems (Alameda, Calif.) and Chiron Corporation, Inc., for confirmation of the particle count number as determined from the HIV-1 RNA copy number. The stock was also quantitated for plasma HIV-1 RNA levels at VQA by both commercial assays. The results obtained by all methods indicated that the stock concentration contained an average of 5.4×10^8 HIV-1 RNA copies per ml. Thus, the absolute HIV-1 RNA copy number for the undiluted stock was assumed to be 5.4×10^8 copies per ml.

(iii) Composition of panels and standards. The compositions of the panels, sets with unknown HIV-1 RNA copy numbers (Table 1), and standards, sets with known HIV-1 copy numbers (Table 2), varied among the rounds of the proficiency assessment. In round 1, the panel consisted of five coded samples and undiluted plasma from an HIV-1-seropositive patient and a 1:5 and 1:25 dilution of the patient material. In subsequent rounds, only the coded samples described in Table 1 were used. The standards for the Chiron and NASBA assays were 0, 5,000, 10,000, 15,000, 150,000, and 1,500,000 RNA copies per ml for rounds 1 to 4. For all other assays the standards were 0, 750, 1,500, 15,000, 150,000, and 1,500,000 RNA copies per ml for rounds 1 to 4. The 1,500,000-RNA-copy-number standard was eliminated for the Roche assays in round 4 and was replaced with a 750,000-RNA-copy-number standard in round 5, and the 750- and 1,500-RNA-copy number standards were eliminated. The NASBA standards for round 5 consisted of 0, 15,000, 150,000, and 1,500,000 RNA copies per ml.

(iv) Evaluation criteria. The primary goal of this initial proficiency certification was to determine if laboratories could reliably detect fivefold differences in RNA copy numbers. The statistical tests of performance were based on the assumption that the $\log_{10}$ copy number was normally distributed. Two performance criteria were established for "certification" by the quality assurance program. First, the overall intra-assay standard deviation could not be statistically significantly greater than 0.15 $\log_{10}$ copy number. This standard deviation provided a 90% power to detect a fivefold difference in RNA copy number between two samples in the same assay batch at a type 1 error rate of 0.05. Second, the estimates of RNA copy number for samples which differed by fivefold in actual copy number had to be sufficiently different, and the estimates for samples with identical copy numbers had to be sufficiently similar (21). The criteria for comparing estimates are given below. Laboratories performed acceptably by meeting the two criteria given above. If they performed acceptably, they were considered certified to perform HIV-1 RNA quantitation assays. Those laboratories which did not perform acceptably on their initial attempt were allowed to participate in two subsequent rounds of testing to obtain an acceptable result.

Statistical analysis. (i) Estimates of RNA copy numbers per milliliter. Per-

TABLE 2. Known VQA standard compositions for proficiency panels 2, 3, 4, and 5

No. of RNA copies/ml	Assay[a]			
	Panel round 2	Panel round 3	Panel round 4	Panel round 5
1,500,000	C, D, N, R	C, R	N	C, D, N
750,000				R
150,000	C, D, N, R	C, R	N, R	C, D, N, R
15,000	C, D, N, R	C, R	N, R	C, D, N, R
10,000	C, N	C	N	C
5,000	C, N	C	N	C
1,500	D, R	D, R	R	
750	D, R	D, R	R	
0	C, D, N, R	C, R	N, R	C, D, N, R

[a] C, Chiron; D, Digene; N, NASBA; R, Roche.

formance was assessed by using estimates of RNA copy number that were adjusted to the VQA standards. For panels 1 and 2, estimates of the RNA concentration for the Chiron assay were obtained from the VQA standard curve by using a log-log point-to-point regression. For panel 2, the last line segment was extended beyond the standard at 1,500,000 RNA copies per ml to obtain estimates for samples with RLUs that exceeded the RLU for that standard. The curve was not extended below the standard at 10,000 RNA copies per ml. For panels 3 and 4, estimates of the RNA concentration for the Chiron assay were obtained from the VQA standard curve by using a log-log quadratic regression of the mean of the two RLUs per standard on the nominal copy number. Mean RLUs for the negative standards were subtracted from the RLUs for each positive standard prior to curve fitting. VQA-adjusted estimates for all other kits tested in the study were obtained from log-log regressions of estimated copy number on the nominal copy number for the VQA standards. For the Roche assay, estimates were obtained by combining the VQA standards from three plates and by looking at each plate independently. The NASBA assay allowed for the assessment of 10 samples per batch, so a minimum of three batches was necessary to assess the proficiency panel and VQA standards. For panel 4, all laboratories included the full set of standards in the first batch and various subsets in other batches. The standards from the first batch were therefore used to construct the VQA curves.

(ii) Calculation of standard deviations. A standard deviation of the $\log_{10}$ copy number was calculated for each set of replicates in each laboratory. The intra-assay standard deviation was estimated from the mean square error of a one-way analysis of variance in which nominal copy number was the predictor. Samples which were not in the dynamic range of the assay were not used for the calculation. The estimated standard deviation was compared against the target of 0.15 $\log_{10}$ copy number by a chi-square test.

(iii) Fivefold error determinations. The error resulting from the failure to discern a fivefold difference between two values that did differ by fivefold was designated the true difference undetected (TDU). The error that resulted from finding an excessively large difference between the estimates for two samples with the same underlying concentration was designated the false difference detected (FDD). TDUs and FDDs were detected by using Z scores, in which $Z = \log_{10}(y_1/y_2)/(0.15 \cdot 2^{1/2})$, where y_1 was the estimate for the sample with the higher nominal RNA concentration and y_2 was the estimate for the sample with the lower nominal RNA concentration. Under the null hypothesis of no difference in concentrations, Z was normally distributed with a mean of zero and a variance of 1 (32). A TDU was declared if Z was <1.96 for two samples with nominal concentrations that differed by fivefold. For FDDs the two nominal RNA concentrations were the same. A FDD was declared if Z was ≥ 1.96 or Z was ≤ -1.96. TDUs and FDDs were tabulated for each laboratory, and the totals were compared against the maximum allowed values. The number of FDD and TDU errors that can occur by chance alone is a function of the intra-assay standard deviation and the design of the panel. Simulations were used to determine the maximum allowed error rates. Each simulation consisted of a random sample from a multivariate normal distribution with a standard deviation of 0.15 $\log_{10}$ copy number for all observations. The sample size and the expected values for the sample matched those of the design of the panel in question. The numbers of TDUs and FDDs were tabulated for each randomly selected panel, and the distributions of the error rates were estimated from the rates of TDUs and FDDs for 5,000 simulated panels. The maximum acceptable error rates were defined as the rates that were exceeded no more than 5% of the time in the simulations.

(iv) Interlaboratory and interkit estimates of variation. Analysis of covariance was used to assess differences in the estimated RNA concentrations among laboratories that used the same kit. A log-log regression of the estimated concentration on the nominal concentration was fitted to the data from each laboratory. Differences among kits were assessed by using a random effects model, under which the laboratory-specific slopes and intercepts were assumed to be normally distributed. The expected values of the regression parameters define the population-average regression for each kit (9). Differences among kits were detected by testing for differences among the estimated population-average regressions. Both parts of these analyses were limited to the shared dynamic ranges of the assays (10,000 to 1,000,000 RNA copies per ml) and only included data from certified laboratories. A statistical analysis software was used for all data analyses.

RESULTS

Panel composition and assays and laboratories evaluated. The types of HIV-1 RNA assays used by the laboratories and the number of laboratories using each assay are given in Table 3 for each round of proficiency testing. A total of six assays were assessed by 41 laboratories. The composition of the proficiency panel was modified from round to round (Tables 1 and 2) as more information was gained about the performance characteristics of each assay type. For example, the RNA concentrations in the proficiency panels and the VQA standards were modified to accommodate the dynamic range of each assay. Some of this information was used by the manufacturers to make modifications in their assays prior to the subsequent rounds of testing. In round 1 of the proficiency testing, fivefold dilutions of plasma from a seropositive donor were also included; however, the limited supply of this specimen made it impractical for testing at multiple centers, and it was eliminated from subsequent rounds of testing. Round 1 was considered a quality assurance pilot run and was not used for certification purposes; however, the information gained from this round was used to modify the range of copy numbers for the coded samples in round 2.

TABLE 3. Number of laboratories certified by round and combined results for laboratories that were certified

Assay	No. of laboratories certified/total no. in round:				Combined results for all certified laboratories for all rounds[a]			
	2	3	4	5	No. of certified laboratories	SD	% FDD	% TDU
Chiron bDNA, version 1.0	2/4	6/6	NT[b]	3/6	11	0.08	0	0
Chiron bDNA, version 2.0[c]	1/1	1/1	NT	NT	1	0.06	0	0
NASBA	0/1	0/2	5/7	6/9	11	0.15	1.7	6.5
Roche	2/4	3/9	10/12	15/21	30	0.13	5.0	8.0
Digene	1/1	NT	NT	0/2	1	0.09	0	0
WRAIR	1/1	NT	NT	NT	1	0.10	0.04	0.04

[a] Values for the standard deviation, FDD, and TDU are medians across certified laboratories.
[b] NT, not tested.
[c] The Chiron bDNA, version 2.0, was assayed two times by the manufacturer.

One of the first observations in the study was the need to adjust the compositions of the standards and the coded samples to reflect the dynamic range of the assays. The problem encountered for some of the assays when either the standards or the panel was not consistent with the dynamic range was an inability to discern fivefold differences (TDU). For some assays (Chiron and NASBA), this problem was encountered at the lower copy numbers, while for other assays (Roche and WRAIR), this was a problem in the higher-copy-number ranges. Values outside of the dynamic range for a particular assay were not included in the analysis for certification. We determined the values outside the dynamic range using the proficiency panels provided (Tables 1 and 2). Thus, for these specific RNA copy numbers per milliliter assessed, the linear ranges of the assays were determined to be as follows: (i) Chiron version 1.0, 11,000 to 1,500,000 RNA copies per ml; (ii) Chiron version 2.0, 2,400 to 1,500,000 RNA copies per ml; (iii) Digene, 2,400 to 1,500,000 RNA copies per ml; (iv) NASBA, 4,000 to 1,500,000 RNA copies per ml; (v) Roche, 480 to 750,000 RNA copies per ml; and (vi) WRAIR, 2,400 to 750,000 RNA copies per ml.

Another difficulty encountered in establishing the same proficiency panel for multiple assay formats was how and when to use the common external standards. The Chiron assay format could accommodate the six VQA standards and the coded samples in the same 96-well plate when the kit standards were replaced with the VQA standards. Thus, the RNA concentrations were determined from standard curves based on the VQA standards rather than the kit standards. The Roche assay with its internal standard could accommodate the panel and additional standards only when three assay plates were used. The VQA assay standards were included on all three plates.

TABLE 4. Interlaboratory variability: statistical tests of differences in slopes and intercepts among laboratories that used the same kit

Round	Kit	No. of laboratories	*P* value		r^2
			Slope	Intercept	
2	Chiron	2	0.90	0.95	0.96
	Roche	2	0.27	0.21	0.99
3	Chiron	4	<0.01	<0.01	0.98
	Roche	3	0.21	0.40	0.94
4	NASBA	4	0.18	0.08	0.95
	Roche	10	<0.001	<0.001	0.98
5	Chiron	3	0.12	<0.001	0.97
	NASBA	7	<0.001	<0.001	0.97
	Roche	18	0.14	0.27	0.95

The initial assumption was that the standard curves would not vary significantly from plate to plate; however, variation from plate to plate was seen for some, but not all, laboratories (data not shown). The net result was that some laboratories had higher standard deviations when the data from the plates were combined while others had higher standard deviations when the plates were analyzed separately. The TDU and FDD error rates also differed, depending on how the standards were analyzed. Certification was determined by using the standard measurements with the lowest standard deviations. The NASBA assay format is based on a single tube, with a batch composed of 10 to 20 single-tube assays. The VQA standards were assayed by all laboratories by this assay for the first batch of 10 samples and were then assayed intermittently throughout the other batches needed to accommodate the test panel. The data from the first batch were used to calculate the RNA copy number for analysis.

Panel results. The median of the standard deviations, the percent TDUs, and the percent FDDs for all laboratories meeting the certification criteria are presented in Table 3 for each of the respective assays for each round of proficiency testing. The number of laboratories meeting the certification requirements among those participating in a round of proficiency testing is also given in Table 3. Forty-one laboratories participated in the five rounds of testing, for a total of 88 laboratory performance evaluations, of which 56 evaluations (65%) met the specified certification criteria.

The fitted regressions indicated that the differences among laboratories that used the same kit were generally greater than the differences among the population-average regressions for the three commercial kits. Tests for differences in the slopes and intercepts of the log-log regressions of the estimated RNA copy number on the nominal copy number among laboratories that used each kit are provided in Table 4, and fitted regressions are plotted in Fig. 1. Statistically significant interlaboratory variation in either the slope or the intercept of the regression lines was identified in rounds 3, 4, and 5, but not in round 2 and not for all kits in each round. Tests for differences among the population-average regressions for the kits are provided in Table 5. The test for differences in slope and intercept in rounds 3, 4, and 5 was based on VQA standard-adjusted estimates of the RNA concentration. The test for differences in round 2 was based on kit standard-generated estimates because the statistical package failed to converge to a solution with the VQA standard-adjusted data. No significant differ-

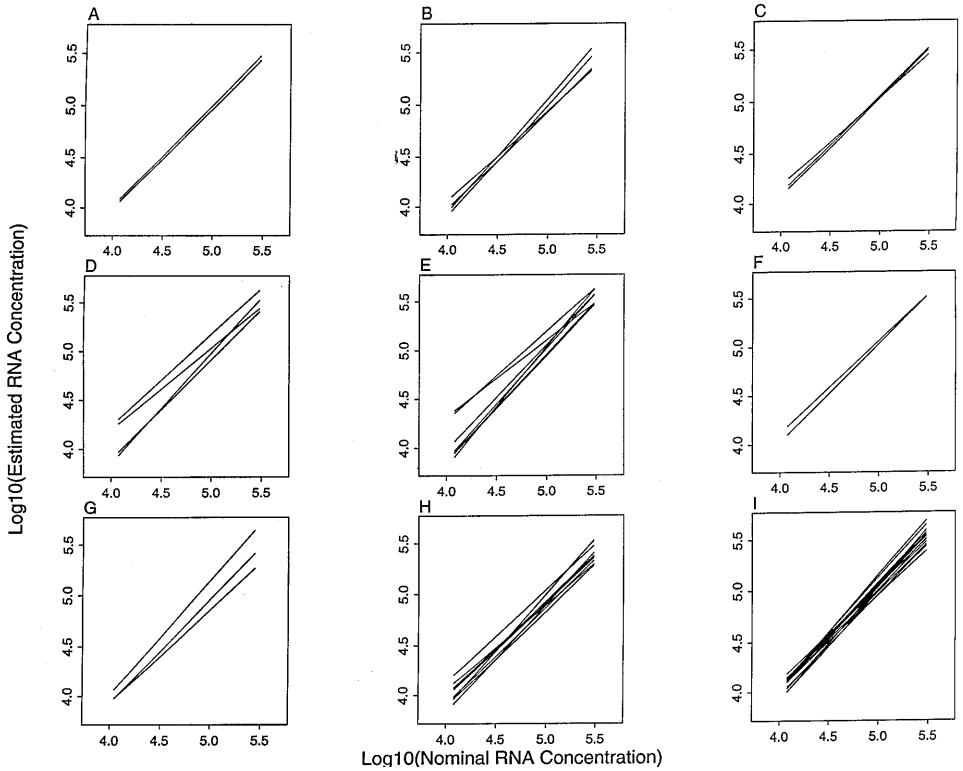

FIG. 1. Fitted regressions of $\log_{10}$ estimated RNA concentration on $\log_{10}$ nominal concentration for all laboratories that used the same kit for all panel rounds. (A to C) Chiron assay for rounds 2, 3, and 5, respectively. (D and E) NASBA assay for rounds 4 and 5, respectively. (F to I) Roche assay for round 2, 3, 4, and 5, respectively.

TABLE 5. Interassay variability: statistical tests of differences in population-averaged slopes and intercepts among kits

Round[a]	Kits	*P* value	
		Slope	Intercept
2	Chiron, Roche	0.10	0.13
2	Chiron, Roche, Digene	0.06	0.10
3	Chiron, Roche	0.75	0.74
4	NASBA, Roche	0.82	0.88
5	Chiron, NASBA, Roche	0.17	0.16

[a] Round 2, test of kit standard values; rounds 3, 4, and 5, tests of VQA standard-adjusted values.

ences in slopes and intercepts among the population-average regressions were observed.

The effect of using a common set of standards was to reduce assay variation both between laboratories (Table 4) and between kits (Fig. 2). Population-average regressions for the kit standard-based estimates are presented in Fig. 2A, and those for the VQA standard-adjusted estimates are presented in Fig. 2B. Differences among the predicted values for the population-average regressions for the VQA standard-adjusted data were <0.25 $\log_{10}$ RNA copies per ml in all rounds and <0.1 $\log_{10}$ RNA copies per ml for most rounds. Differences among the predicted values on the basis of the kit-based estimates were twofold higher (<0.5 $\log_{10}$ RNA copies per ml).

DISCUSSION

This report describes the first quality assurance program for assays for the quantification of HIV-1 RNA levels in plasma for laboratories participating in HIV-1 clinical trials. We have established such programs for HIV-1 coculture (14), p24 and immune complex dissociated p24 antigen detection (10), and, most recently, qualitative HIV-1 DNA PCR (18). Others have indicated that without such programs the sensitivities, specificities, and reproducibilities of these types of assays are poor (2, 6, 31). The primary goal of our study was to ensure laboratory proficiency and provide a set of HIV-1 RNA standards and controls which could be used for all assays. For the purposes of these preliminary proficiency testing rounds, we defined laboratory proficiency as the ability of a laboratory to perform an assay within the linear range of each assay with an acceptable standard deviation of ≤ 0.15 $\log_{10}$ copy number. We also specified that a laboratory must be able to precisely determine fivefold differences in plasma HIV-1 RNA levels in testing well-characterized spiked plasma samples.

Over the course of four rounds of testing involving 41 laboratories, 56 of 88 (65%) laboratory evaluations resulted in acceptable performance by using these criteria. There were differences in the ability of the laboratories to achieve the proficiency goal, in part because of the kits used and the experience of the laboratory technologists. In general, for each successive round of testing, there was an increase in the percentage of laboratories certified.

There were many difficulties associated with the establishment of a single proficiency testing panel for diverse assay formats. The inherent differences in the assays may have biased the results that were observed. For example, for two of the assays (Roche and NASBA), the proficiency panel superimposes an external standard curve on a kit-specific internal standard. For both of these assays, the use of internal standards was specifically selected to control for differential enzymatic activity, yet we essentially circumvented that control by attempting to relate the results to a separately assayed external standard. It is not surprising that the Chiron assay gave the most consistent results with the external standards, because it used an external standard curve as the standard for the kit. What was remarkable was that the performance of the other two assay formats was so reproducible compared with that of kits that used an external standard. The variation seen in the Roche assay between plates and in the NASBA assay between tubes was not consistent for all laboratories and therefore may have reflected operator bias more than assay bias. We could not determine plate variation for the Chiron assay. However, the implication of these findings for the analysis of multiple samples supports batch testing of all samples from the same patient on the same plate (Roche and Chiron) or in the same batch run (NASBA). We recognize that this may not be possible for real-time individual patient management; thus, interpretation of changes in viral load by these assays for individual patients should probably include the outer variability limits of the assays, minimally, >0.5 $\log_{10}$, exclusive of patient variabil-

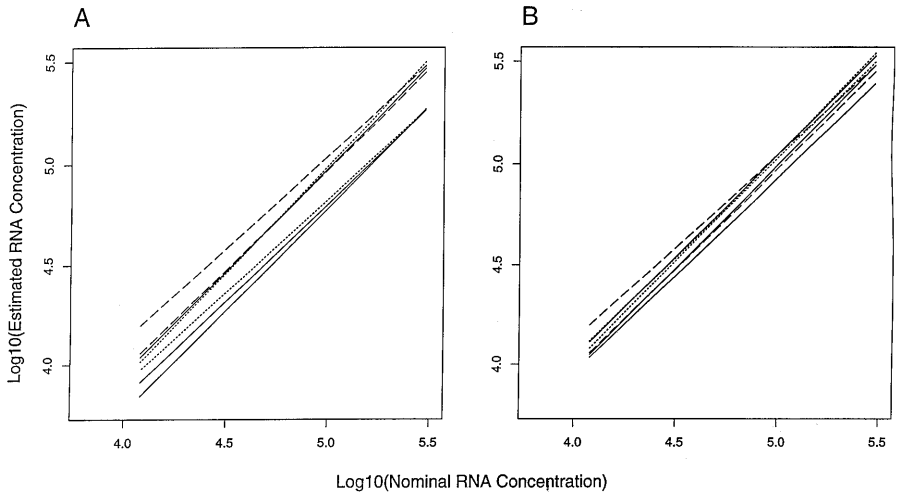

FIG. 2. Population-average regressions (see Materials and Methods) of $\log_{10}$ estimated RNA concentration on $\log_{10}$ nominal concentration for all panel rounds. (A) Kit standard-estimated concentrations. (B) VQA standard-adjusted estimated concentrations. ——, Roche ($n = 4$); – – –, Chiron ($n = 3$); . . . ., NASBA ($n = 3$).

ity, and should be assessed by more than one measurement over time.

The observation that interkit variation when using a common set of standards was not significant allows for flexibility in the selection of a methodology to accommodate the strengths of the assays and the needs of the clinical setting. Thus, use of a common set of standards across clinical trial protocols would allow for cross-protocol comparisons. The significantly different values obtained by different laboratories with the same kit, however, suggests that the certification criteria for the present study were inadequate for assessing individual patient HIV-1 RNA levels obtained from more than one laboratory.

In many of the certification runs there were clusters of errors (either TDUs or FDDs) at either end of the dynamic range for each assay. This resulted in imprecise assessment of copy numbers which were not realized when the errors were averaged for the entire dynamic range. For example, in one certified laboratory, there was a 22% TDU error rate in values from 2,400 to ≤12,000 RNA copies per ml, a 0% TDU error rate in values from >12,000 to ≤60,000 RNA copies per ml, and a 33% TDU error rate in values from >60,000 to ≤300,000 RNA copies per ml. Thus, one in three pairs of samples with RNA copy numbers of >60,000 or <300,000 would have been viewed as having the same value. We are in the process of evaluating other statistical methods to define the accuracy of a given value at a given nominal concentration.

The proficiency testing that we have implemented is based on a semiannual determination of the ability of a laboratory to perform the assay accurately enough to detect a fivefold difference in plasma HIV-1 RNA levels averaged over the dynamic range of the assay. The proficiency testing does not, however, account for the ability of a laboratory to give an absolute value to a standard on either a one-time or ongoing basis. Knowledge of this accuracy for all the assays is a necessary prerequisite for real-time sequential testing of patient plasma specimens. An accurate measurement of HIV-1 RNA copy number is also necessary for defining the levels of RNA associated with an increased risk of disease progression (12, 13, 22) and vertical transmission (7, 36). The lack of a common set of standards for use in either diagnostic or clinical research laboratories to account for interlaboratory or interassay comparisons and the lack of standard specimen collection and anticoagulant usage make the assignment of absolute HIV-1 RNA levels difficult at present. The assignment of an absolute HIV-1 RNA level should be based on quality assurance and extensive understanding of assay performance.

We used only purified stocks for performance evaluation in the present studies. We have not assessed variability owing to individual patient variation over time (37) or factors which may be present in individual patient plasma which would influence assay performance. The abilities of the various kits to determine the same HIV-1 RNA concentration in patient samples have been addressed with a small number of patients (27, 29). The data suggested that there may be variations among kits in assessing the values for an individual patient but that these kit variations are not consistent among patients. Larger studies are needed to validate this finding and its impact on individual patient management and the outcomes of clinical trials. Furthermore, we have evaluated the assays only with regard to the HIV-1 subtype B found primarily in the United States and Europe, and the results may not apply to other viral subtypes. We are in the process of acquiring that information so that our results may be more generalizable to the worldwide AIDS epidemic. Our estimates of variation, precision, and accuracy reflect the very best circumstances, and caution should be used in the interpretation of any one value for any individual patient at a single point in time.

In summary, we have established a quality assurance program necessary for defining performance standards of operation for the quantification of HIV-1 RNA levels in plasma. Multiple laboratories using several different types of assays were capable of analyzing coded samples with minimal variation, supporting the use of each assay in multicenter clinical trials. Use of a common set of standards allowed for direct kit-to-kit and laboratory-to-laboratory comparisons. Variations among laboratories in some cases were significant, whereas variations among kits were not. Additional methods for assessing proficiency performance and assessment of clinical samples from patients are needed before ascertaining the accuracy of any given assay performed by a given laboratory.

ACKNOWLEDGMENTS

We acknowledge the following individuals for their technical assistance: Colleen Starkey (Cleveland Clinic Foundation); Mark Winters (Stanford); Tim Spahlinger, Janice Parsons, and Nancy Knoble (Case Western Reserve); Sandra Barrick (WRAIR); Joan Dragavon, Willa Lee, and Reggie Sampoleo (University of Washington); and Zhiliang Li (University of Southern California).

This work was supported under NIAID contract NO-AI-35172 and NIAID grants AI-25879, AI-27664, and AI-94028.

REFERENCES

1. **Aoki-sei, S., R. Y. Yarchoan, S. Kageyama, D. T. Hoekzema, J. M. Pluda, K. M. Wyvill, S. Broder, and H. Mitsuya.** 1992. Plasma HIV-1 viremia in HIV-1 infected individuals assessed by polymerase chain reaction. AIDS Res. Hum. Retroviruses **8:**1263–1270.
2. **Busch, M. P., D. R. Henrard, I. K. Hewlett, W. F. Mchaffey, J. S. Epstein, J.-P. Allain, T. H. Lee, J. W. Mosley, and the Transfusion Safety Study Group.** 1992. Poor sensitivity, specificity and reproducibility of detection of HIV-1 DNA in serum by polymerase chain reaction. J. Acquired Immune Defic. Syndr. **5:**872–877.
3. **Coffin, J. M.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis and therapy. Science **267:**483–489.
4. **Collier, A. C., R. W. Coombs, M. A. Fischl, P. R. Skolnik, D. Northfeldt, P. Boutin, C. J. Hooper, L. D. Kaplan, P. A. Volberding, L. G. Davis, D. R. Henrard, S. Weller, and L. Corey.** 1993. Combination therapy with zidovudine and didanosine compared to zidovudine alone in human immunodeficiency virus type one infection. Ann. Intern. Med. **119:**786–793.
5. **Coombs, R. W.** 1994. HIV-1 burden as a marker of disease progression and clinical response to therapy in AIDS. Clin. Lab. Med. **14:**301–311.
6. **Defer, C., H. Agut, A. Garbarg-Chenon, M. Moncany, F. Morinet, D. Vignon, M. Mariotti, and J. J. Lefrere.** 1992. Multicentre quality control of polymerase chain reaction for detection of HIV DNA. AIDS **6:**659–663.
7. **Dickover, R. E., E. M. Garrity, S. Herman, M. Sim, S. Plaeger, P. J. Boyer, M. Keller, A. Deveikis, E. R. Stiehm, and Y. Bryson.** 1996. Identification of levels of maternal HIV-1 RNA associated with risk of perinatal transmission: effect of maternal zidovudine treatment on viral load. JAMA **275:**599–605.
8. **Division of AIDS, National Institute of Allergy and Infectious Diseases.** 1994. ACTG virology manual for HIV laboratories. Publication NIH-94-3828. U.S. Department of Health and Human Services, Washington, D.C.
9. **Feldman, H. A.** 1988. Families of lines: random effects in linear regression analysis. J. Appl. Physiol. **64:**1721–1732.
10. **Hammer, S., C. Crumpacker, R. D'Aquila, B. Jackson, J. Lathey, D. Livnat, and P. Reichelderfer.** 1993. Use of virologic assays for detection of human immunodeficiency virus in clinical trials: recommendations of the AIDS Clinical Trials Group Virology Committee. J. Clin. Microbiol. **31:**2557–2564.
11. **Henrard, D. R., W. F. Mehaffey, and J. P. Allain.** 1992. A sensitive viral capture assay for detection of plasma viremia in HIV-infected individuals. AIDS Res. Hum. Retroviruses **8:**47–52.
12. **Henrard, D. R., J. F. Phillips, L. R. Muenz, W. A. Blattner, D. Wesner, M. E. Eyster, and J. J. Goedert.** 1995. Natural history of HIV-1 cell-free viremia. JAMA **274:**554–558.
13. **Hogervost, E., S. Jurriaans, F. de Wolf, A. van Wijk, A. Wiersma, M. Valk, M. Roos, B. van Gemen, R. Coutinho, F. Miedema, and J. Goudsmit.** 1995. Predictors for non- and slow progression in human immunodeficiency virus (HIV) type 1 infection: low viral RNA copy numbers in serum and maintenance of high HIV-1 p24-specific but not V3-specific antibody levels. J. Infect. Dis. **171:**811–821.
14. **Hollinger, F. B., J. W. Bremer, L. E. Myers, J. W. M. Gold, L. McQuay, and the NIH/NIAID/DAIDS/ACTG Virology Laboratories.** 1992. Standardization of sensitive human immunodeficiency virus coculture procedures and estab-

lishment of a multicenter quality assurance program for the AIDS Clinical Trials Group. J. Clin. Microbiol. **30:**1787–1794.
15. **Holodniy, M., D. A. Katzenstein, D. M. Israelski, and T. C. Merigan.** 1991. Reduction in plasma human immunodeficiency virus ribonucleic acid after dideoxynucleoside therapy as determined by the polymerase chain reaction. J. Clin. Invest. **88:**1755–1759.
16. **Holodniy, M., D. A. Katzenstein, and S. Sengupta.** 1991. Detection and quantification of human immunodeficiency virus RNA in patient serum by use of the polymerase chain reaction. J. Infect. Dis. **163:**862–866.
17. **Hooper, C., S. Welles, R. D'Aquila, A. Japour, V. Johnson, D. Kuritzkes, B. Jackson, V. DeGruttola, C. Crumpacker, D. Richman, J. Kahn, S. Kwok, J. Todd, P. Reichelderfer, and R. Coombs.** 1994. HIV-1 RNA level in plasma and association with disease progression, zidovudine sensitivity phenotype and genotype, syncytium-inducing phenotype, $CD4^+$ cell count and clinical diagnosis of AIDS, p. 67. *In* Third International Workshop on Drug Resistance.
18. **Jackson, J. B., J. Drew, H. J. Lin, P. Otto, J. W. Bremer, F. B. Hollinger, S. M. Wolinsky, the ACTG PCR Working Group, and the ACTG PCR Virology Laboratories.** 1993. Establishment of a quality assurance program for human immunodeficiency virus type 1 DNA polymerase chain reaction assays by the AIDS Clinical Trials Group. J. Clin. Microbiol. **31:**3123–3128.
19. **Larder, B. A., S. D. Kemp, and P. R. Harrigan.** 1995. Potential mechanism for sustained efficacy of AZT-3TC combination therapy. Science **269:**696–699.
20. **Lin, H. J., M. Haywood, and F. B. Hollinger.** 1996. Application of a commercial kit for detection of PCR products to quantification of human immunodeficiency virus type 1 RNA and proviral DNA. J. Clin. Microbiol. **34:**329–333.
21. **Lin, H. J., L. E. Myers, B. Yen-Lieberman, F. B. Hollinger, D. Henrard, C. J. Hooper, R. Kokka, S. Kwok, S. Rasheed, M. Vahey, M. A. Winters, L. J. McQuay, P. L. Nara, P. Reichelderfer, R. W. Coombs, and J. B. Jackson.** 1994. Multicenter evaluation of methods for the quantitation of plasma HIV-1 RNA. J. Infect. Dis. **170:**553–562.
22. **Mellors, J. W., L. A. Kingsley, C. R. Rinaldo, J. A. Todd, B. S. Hoo, R. Kokka, and P. Gupta.** 1995. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Ann. Intern. Med. **122:**573–579.
23. **Mulder, J., N. McKinney, C. Christopherson, J. Sninsky, L. Greenfield, and S. Kwok.** 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: application to acute retroviral infection. J. Clin. Microbiol. **32:**292–300.
24. **O'Brien, W. A., P. M. Hartigan, D. Martin, J. Esinhart, A. Hill, S. Benoit, M. Rubin, M. S. Simberkoff, J. D. Hamilton and the Veterans Affairs Cooperative Study Group on AIDS.** 1996. Changes in plasma HIV-1 RNA and $CD4^+$ lymphocyte counts and the risk of progression to AIDS. N. Engl. J. Med. **334:**426–431.
25. **Pachl, C., J. A. Todd, D. G. Kern, P. J. Sheridan, S. Fong, M. Stempien, B. Hoo, D. Besemer, T. Yeghiazarian, B. Irvine, J. Kolberg, R. Kokka, P. Neuwald, and M. S. Urdea.** 1995. Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay. J. Acquired Immune Defic. Syndr. Hum. Retrovirol. **8:**446–454.
26. **Piatak, M., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes, K.-C. Luk, B. H. Hahn, G. M. Shaw, and J. D. Leifson.** 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. Science **259:**1749–1754.
27. **Revets, H., D. Marissens, S. De Wit, P. Lacor, N. Clumeck, S. Lauwers, and G. Zissis.** 1996. Comparative evaluation of NASBA HIV-1 RNA QT, Amplicor-HIV Monitor, and Quantiplex HIV RNA assay, three methods for quantification of human immunodeficiency virus type 1 RNA in plasma. J. Clin. Microbiol. **34:**1058–1064.
28. **Schnittman, S. M., J. J. Greenhouse, C. Miltiades, C. Psallidopoulos, M. Baseler, N. P. Salzman, A. S. Fauci, and H. C. Lane.** 1990. Increasing viral burden in $CD4^+$ T cells from patients with human immunodeficiency virus (HIV) infection reflects rapidly progressive immunosuppression and clinical disease. Ann. Intern. Med. **113:**438–443.
29. **Schuurman, R., and the Delta Virology Group.** 1996. Multicenter comparison of three commercial methods for plasma HIV-1 RNA quantification using spiked human plasma and clinical plasma samples, abstr. 4542. *In* Abstracts of the 3rd Conference on Retroviruses and Opportunistic Infections.
30. **Semple, M., C. Loveday, I. Weller, and R. Tedder.** 1991. Direct measurement of viraemia in patients infected with HIV-1 and its relationship to disease progression and zidovudine therapy. J. Med. Virol. **35:**38–45.
31. **Sheppard, H. W., M. S. Ascher, M. P. Busch, P. R. Sohmer, M. Stanley, M. C. Luce, J. A. Chimera, R. Madej, G. C. Rodgers, C. Lynch, H. Khayam-Bashi, E. L. Murphy, B. Eble, W. Z. Bradford, R. A. Royce, and W. Winlestein, Jr.** 1991. A multicenter proficiency trial of gene amplification (PCR) for the detection of HIV-1. J. Acquired Immune Defic. Syndr. **4:**277–283.
32. **Snedecor, G. W., and W. G. Cochran.** 1967. Statistical methods. Iowa State University Press, Ames.
33. **Vahey, M., and M. T. Wong.** 1995. A quantitative liquid hybridization polymerase chain reaction methodology employing storage phosphor technology, p. 313–338. *In* C. Diffenbach and G. Dveksler (ed.), The Cold Spring Harbor laboratory manual for PCR. Cold Spring Harbor Press, Cold Spring Harbor, New York.
34. **Vahey, M. T., D. L. Mayers, K. F. Wagner, R. C. Y. Chung, O. S. Weislow, S. Zhou, D. S. Burke, and RV43 Study Group.** 1994. Plasma HIV RNA predicts clinical outcome on AZT therapy, abstr. 253B. *In* Abstracts of the Tenth International Conference on AIDS.
35. **van Gemen, B., T. Kievits, R. Schukkink, D. van Strijp, L. T. Malek, R. Sooknanan, H. G. Huisman, and P. Lens.** 1993. Quantification of HIV-1 RNA in plasma using NASBA during HIV-1 primary infection. J. Virol. Methods **43:**177–188.
36. **Weiser, B., S. Nachman, P. Tropper, K. H. Viscosi, R. Grimson, G. Baxter, and G. Fang.** 1994. Quantitation of human immunodeficiency virus type 1 during pregnancy: relationship of viral titer to mother-to-child transmission and stability of viral load. Proc. Natl. Acad. Sci. USA **91:**8037–8041.
37. **Winters, M. A., L. B. Tan, D. A. Katzenstein, and T. C. Merigan.** 1993. Biological variation and quality control of plasma human immunodeficiency virus type 1 RNA quantitation by reverse transcriptase-polymerase chain reaction. J. Clin. Microbiol. **31:**2960–2966.
38. **Yerly, S., L. Kaiser, B. Mermillod, et al.** 1995. Response of HIV RNA to didanosine as a predictive marker of survival. AIDS **9:**159–163.