JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 1990, p. 1799–1807
0095-1137/90/081799-09$02.00/0

Vol. 28, No. 8

Indirect Immunofluorescence Test Performance and Questionnaire Results from the Centers for Disease Control Model Performance Evaluation Program for Human Immunodeficiency Virus Type 1 Testing

ROGER N. TAYLOR,* THOMAS L. HEARN, WILLIAM O. SCHALLA, AND RONALD O. VALDISERRI

Public Health Practice Program Office, Centers for Disease Control, Atlanta, Georgia 30333

Received 15 November 1989/Accepted 10 May 1990

Results from laboratories performing indirect immunofluorescence (IIF) testing for human immunodeficiency virus type 1 antibody and participating in the Centers for Disease Control Model Performance Evaluation Program in 1988 are presented. Approximately 90% of all laboratories receiving specimen panels or questionnaires furnished results to the Centers for Disease Control. In September 1988, 111 reports were received from IIF laboratories from 34 states and nine countries; most of these laboratories did IIF testing in conjunction with other antibody tests. Hospital laboratories were the most common type of laboratory participating in the program. Laboratories that performed IIF employed fewer personnel and performed testing less frequently than did laboratories that performed enzyme immunoassays or Western blot (immunoblot) tests and were less likely to use a commercial test kit. Most of the laboratories that referred specimens for IIF testing sent them to the state laboratory. The analytic specificity for the Model Performance Evaluation Program specimens was 98.5% when indeterminate results on a negative specimen were considered correct (negative) and 89.6% when indeterminate results on a negative specimen were considered incorrect; analytic sensitivity was 94.8% when indeterminate results on a positive specimen were considered correct (positive) and 91.4% when indeterminate results on a positive specimen were considered incorrect. When indeterminate results were considered correct, all types of laboratories (blood bank, state, hospital, independent, and other) had analytic specificities over 96%, and all manufacturers had analytic specificities above 95%. All types of laboratories had analytic sensitivities over 92%, and analytic sensitivities were above 94% for all manufacturers and reagent sources except Cellular Products. Comparison of percentages of correct responses between IIF and Western blot assays on those samples for which there was good agreement on the target interpretation revealed no significant differences. Both individual donor and diluted materials were included in the evaluations; the diluted donor material presented the greatest testing difficulty. Within-survey reproducibility was about 93% overall and by specimen type. Between-survey reproducibility was about 81% for negative and indeterminate specimens and 88.5% for positive specimens, for an overall between-survey reproducibility of 84.3%. Differences in performance were noted when results were compared by type of laboratory and test manufacturer.

Since the initial diagnosis of acquired immunodeficiency syndrome in 1981 and the subsequent identification of human immunodeficiency virus type 1 (HIV-1) as the etiologic agent (2, 3, 6, 10, 11), the efforts of many laboratories have been directed at developing tests for detecting antibodies to this virus. The laboratory and clinical diagnosis of HIV-1 infection has presented a unique challenge to both medical and public health professionals that includes the need to develop and implement systems for ensuring the high quality and reliability of HIV-1 test results. Clinical and public health management and control strategies for acquired immunodeficiency syndrome require that these tests be readily available and reliable. They are essential to the success of surveillance and prevention programs (7).

The advent of serologic testing for HIV-1 has made it possible to screen blood and plasma for antibodies to HIV-1 and to provide counseling and testing to persons with increased risks of infection. Expansion of screening into populations with relatively low HIV-1 prevalence rates makes the monitoring of the specificity, sensitivity, and reproducibility of commercially available HIV-1 antibody kits even more important. It is increasingly important in this rapidly expanding field to monitor the use of screening and confirmatory testing to ensure that the frequency of both false-positive and false-negative results is minimized. Monitoring the quality of laboratory testing over time is also important because of the expansion and changes in the types of facilities that provide HIV-1 antibody testing, in the reasons that tests are performed, in the knowledge about the clinical manifestation of HIV-1 infection, and in the technology being used for testing.

As part of the U.S. Public Health Service response to this need, a proficiency testing program for laboratories performing HIV-1 antibody testing was instituted by the Centers for Disease Control (CDC) in March 1985, shortly after the first commercial enzyme immunoassay (EIA) for detecting HIV-1 antibody was licensed by the Food and Drug Administration (18). In 1986, the program was substantially enhanced to permit the evaluation of preanalytic and postanalytic stages in the HIV-1 antibody-testing process (steps occurring before and after the analysis), and the new program was designated the Model Performance Evaluation Program (MPEP) (16a; W. O. Schalla, T. L. Hearn, C. W. Griffin, and R. N. Taylor, Clin. Microbiol. Newsl. **10**:156–159, 1988).

* Corresponding author.

The goals of the program are to develop appropriate methods for defining and evaluating quality laboratory testing systems (including test selection, sample collection, and reporting and interpreting test results); to determine the analytic quality of HIV-1 antibody testing as currently practiced in private and public health laboratories; to evaluate the effect of testing quality on patient and public health (that is, to determine whether test results meet the needs of physicians and public health officials); and to develop strategies for identifying and correcting laboratory errors and other impediments to achieving high quality.

The most commonly used HIV-1 antibody assays are (i) EIA or enzyme-linked immunosorbent assay, (ii) Western blot (WB) or immunoblot tests, and (iii) the indirect immunofluorescence (IIF) test or immunofluorescence assay. The EIAs are primarily used as screening tests. The occurrence of false-positive results necessitates the confirmation of repeatedly EIA-positive specimens by other, more specific methods. Both WB and IIF are used as supplemental tests to confirm positive screening results, and both have advantages and disadvantages (4, 9, 12, 13, 15–17). Although the MPEP evaluated performance with all of these tests, this paper will report only the results of evaluations of IIF test performance obtained during 1988.

MATERIALS AND METHODS

More than 12,000 laboratories were invited to participate in this evaluation program, and about 1,400 enrolled. Most of the nonparticipating laboratories do not perform HIV-1 antibody testing. Enrollment in the MPEP is voluntary and free of charge. Participants are identified through letters to laboratories, manufacturers, and professional organizations as well as through other means of communication. All participants perform HIV-1 antibody tests, although the number and types they offer vary.

The MPEP is a performance evaluation program and not a proficiency-testing program. Proficiency-testing programs focus on measuring the analytic performance of individual laboratories primarily for regulatory or licensure purposes or as a supplement to internal quality control, with only secondary interest in overall test performance and in performance by testing variables. While the MPEP is interested in individual laboratory performance, its scope is broad and includes defining the current composite performance levels, identifying variables that correlate with levels of performance, implementing intervention, and measuring the effects of these activities. In addition, the goal of the performance evaluation program is to measure the performance of the whole testing process instead of limiting measurements to the analytic component (14). Because of these differences in focus, the activities are specified as performance evaluations instead of proficiency testing.

To evaluate the quality of analytic testing and to assist laboratories in improving their performance, panels of specimens for HIV-1 antibody testing are mailed to MPEP participants two or three times a year. The participants are asked to process these specimens in the same manner that they normally process patient specimens. The first two shipments were made in May and September of 1988; each panel contained 22 specimens (11 pairs of blind replicates), which ranged from nonreactive through weakly reactive to strongly reactive for HIV-1 antibody. The panels were identical in composition, but the specimen numbers were scrambled and recorded. To reduce the possibility of collaboration among participants and to permit the use of undiluted single-donor plasma, eight different subsets of 22 specimens each were prepared for each shipment. Some of the 53 specimens (48 single-donor specimens, numbered 01 through 48, and 5 pools, numbered P1 through P5) were common to all the subsets. Specimens 01 through 08 were negative, and equal portions of each were used to prepare pool P1. Specimens 09 through 16 and 17 through 24 were positives, on the basis of reference laboratory results, and equal portions were used to make pools P2 and P3, respectively. Specimens 25 through 27, 31 and 32 were indeterminate; specimens 28 through 30 were positive. No pool was prepared from specimens 25 through 32. Specimens 33 through 40 and 41 through 48 were positives, and equal portions were used to make pools P4 and P5, respectively. Specimens 29 through 31, 33, 37 through 39, and 41 through 48 consisted of positive donor materials diluted with HIV-1 antibody-negative human serum. All other specimens were undiluted specimens from single donors. This design permitted the measurement of both within- and between-survey reproducibility; i.e., laboratories tested identical specimens in both surveys.

The specimens, collected and prepared for CDC by an outside contractor, were obtained from blood collection centers and consisted of both single-donor and pooled-donor material. Before they were used in a panel, all candidate specimens were tested at CDC and by the contractor using EIAs and WBs from several manufacturers and employing the Association of State and Territorial Public Health Laboratory Directors (ASTPHLD) WB interpretation criteria (1, 5a). In addition, the specimens were validated by 36 candidate reference laboratories who also participate in the program. Candidate reference laboratories are a pool of laboratories with good performance histories that are selected according to laboratory type, testing volume, and testing practices and that are used on a rotating basis as reference laboratories for the tests they offer.

Participants reported their results by mail to CDC on specially designed forms and subsequently received three performance reports. The first report included the reference laboratory results, grouped by test method and kit manufacturer; the second was a summary tabulation of all participant results; and the third was an analysis in which results were grouped for each evaluation specimen by test kit manufacturer, test method and method variables. All participants were identified with a unique, confidential code number. Individual performance reports were not made.

The performance evaluation data largely reflect the analytic aspects of the HIV-1 antibody-testing process. To learn more about other variables associated with HIV-1 testing quality, the MPEP performs separate biannual questionnaire surveys of participants. Participants are requested to complete questionnaires designed to describe their laboratories, their HIV-1 test methods and procedures, the number of tests performed per test method, and their testing process, which includes the purpose of testing (e.g., screening, diagnosis, or research), the source of specimens, how specimens are treated, and how test results are reported to the clinician (19). To maintain confidentiality of the data, laboratory-identifying information such as name and address is not linked at CDC to the data obtained from either the questionnaire or the specimen test results. The data do, however, contain unique code numbers that are necessary for trend analyses and corrective interventions. The results of these questionnaires are linked by this confidential code number to performance results to provide the ability to correlate questionnaire variables with analytic performance.

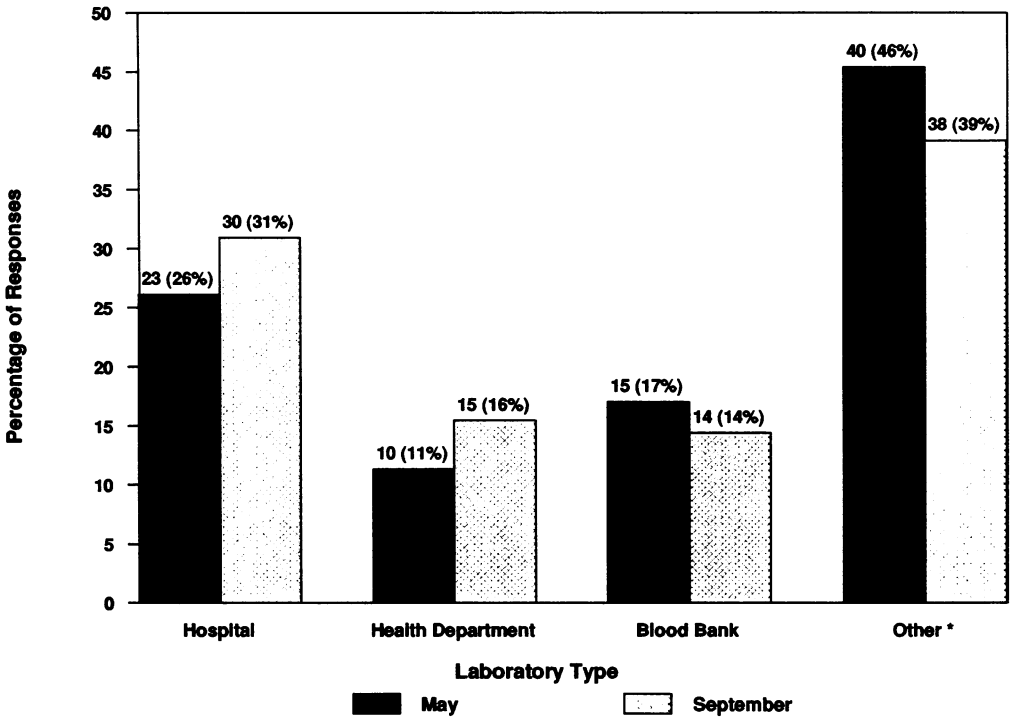

FIG. 1. Types of laboratories doing IIF testing according to 1988 CDC questionnaires. Other includes multiple responses.

Sensitivity and specificity were calculated by a method similar to that of Galen and Gambino (8), but to avoid the possible confusion of the sensitivities calculated in these evaluations with the same estimates derived from testing clinical specimens, we used the terms analytic sensitivity and analytic specificity (5). Because the specimens were selected to maximize the ability to discriminate between good and poor tests, the distribution of reactivities did not approximate that of clinical specimens, nor did it contain the likely confounding variables that would be required to estimate clinical parameters (17). Analytic sensitivity was defined as the percentage of positive reports on samples designated positive on the basis of the reference laboratories' results, and analytic specificity was defined as the percentage of negative reports on samples designated negative on the basis of reference laboratories' results. Because sensitivities and specificities were affected by how indeterminate results were handled, and because laboratories did not resolve indeterminates, sensitivities and specificities were calculated both by considering them correct and by considering them incorrect.

Reproducibility, which is also known as concordance (20), was defined as the percentage of identical results on paired specimens (i.e., the percentage with positive results on both specimens, with indeterminate results on both specimens, or with negative results on both specimens). This is the procedure used in previous proficiency-testing programs (5, 18). For between-survey reproducibility, the results of the first of the duplicate specimens in the May survey were compared with those of the first of the duplicate specimens in the September survey, and the results of the second of the duplicate specimens in the May survey were compared with those of the second of the duplicate specimens in the September survey.

The results of both the performance evaluations and the survey questionnaires are presented in this paper. Cross-correlational analysis (comparing performance with questionnaire variables) will be reported separately.

RESULTS

Currently, 1,416 laboratories are enrolled in the MPEP. Approximately 90% of all laboratories receiving specimen panels or questionnaires voluntarily reported results. Most are located in the United States, although laboratories in three U.S. territories and 82 other countries also participate in this program. Each of the 50 states has laboratories enrolled in the MPEP, with the states of California, Texas, Illinois, Florida, Pennsylvania, and New York accounting for 41% of all U.S. participants. On the May questionnaire, 107 (9.1%) laboratories reported that they perform IIF testing; the number increased to 111 (10.6%) with the September questionnaire. Among these, laboratories from 27 states, Washington, D.C., and 12 countries responded to the May questionnaire, and laboratories from 34 states, Guam, Washington, D.C., and 9 countries responded to the September questionnaire. The states with the largest number of laboratories responding to the September questionnaire were California, with 14; New York, with 8; and Texas, with 6. In all other cases, five or fewer IIF laboratories responded per state or country.

Hospital laboratories were the most common type of laboratory performing IIF tests, constituting 26.1% of the May questionnaire respondents and 30.9% of the September respondents (Fig. 1). Health departments and blood banks ranged from 11.4 to 17.0% of the respondents. Many laboratories reported more than one type of laboratory (e.g., both hospital and blood bank).

Almost two-thirds of the laboratories that performed IIF testing in the performance evaluation program also performed EIA and WB (64.4% in May and 63.2% in September) (Fig. 2). Nearly one-third of the laboratories performed

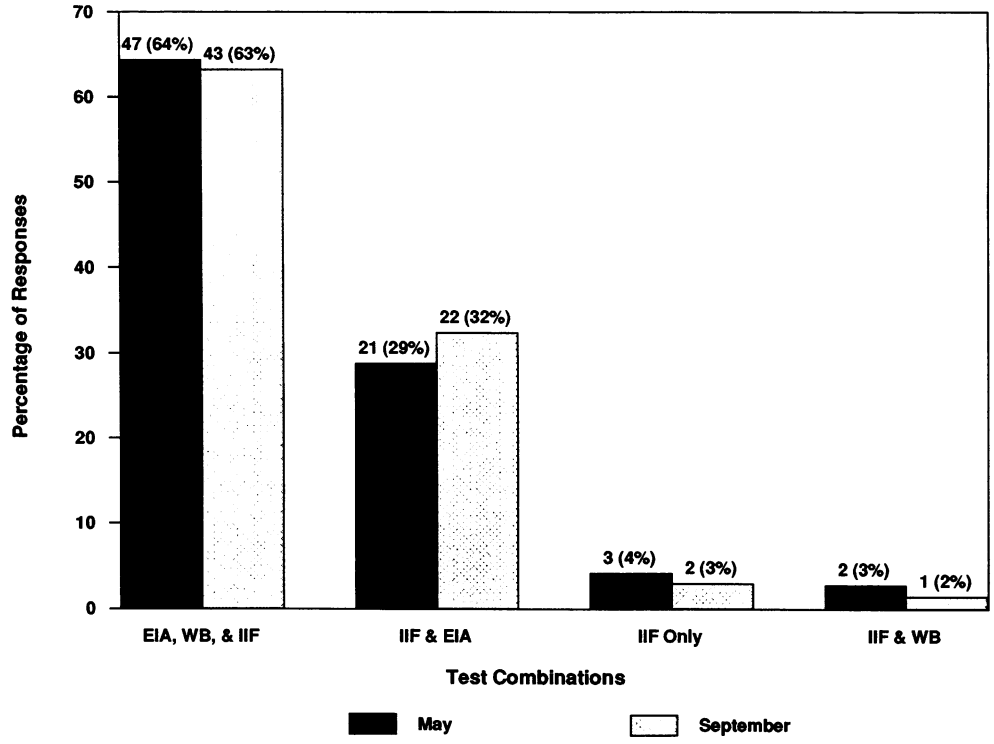

FIG. 2. Tests used in combination with IIF test according to 1988 CDC questionnaires.

a combination of EIA and IIF testing (28.8% in May and 32.4% in September). Two laboratories in September and three in May did only IIF testing; one laboratory in September and two in May did IIF testing in conjunction with WB testing. Of the 63 IIF participants responding to the question about the laboratory testing algorithm on the September 1988 questionnaire, 39 (61.9%) indicated that they use the IIF test in conjunction with the WB after obtaining positive EIA results. One-third (21 laboratories) indicated that they used only the IIF test to confirm positive EIA results. Two laboratories reported that they only performed IIF tests, and one laboratory reported that it performed only IIF and WB testing.

The number of personnel used in HIV-1 testing is shown in Fig. 3. Of the 85 IIF laboratories responding to this question, 52 (61.2%) employed only one or less than one full-time person for IIF testing. All 85 used five or fewer people. Comparison with laboratories doing EIA and WB testing showed that IIF laboratories usually use fewer personnel. In addition, IIF laboratories typically do testing less frequently than do EIA and WB laboratories. Most laboratories (51 of 83, or 61.4%) perform testing only 1 day a week or less (Fig. 4). None of the laboratories did testing on more than 5 days a week. Only seven IIF laboratories (8.4%) did testing 5 days a week, whereas 31 of 235 (13.2%) WB laboratories and 493 of 1,033 (47.7%) EIA laboratories performed testing on 5 or more days per week.

In-house test reagents and kits from Electronucleonics were the most commonly used by IIF performance evaluation participants (Fig. 5). In-house reagents were used by 17 of 46 participants according to the May questionnaire and by 14 of 48 according to the September questionnaire. Electronucleonics kits were used by 11 of 46 participants responding in May and by 18 of 48 participants responding in September. In-house and Electronucleonics reagents were commonly reported on the questionnaires, but noncommercial products (prepared by the main or central state public health laboratory) were also popular among these participants. From 14.8 to 25.0% of the questionnaire responses indicated that one of these three kits or reagents was used. A wide variety of test kits from outside the United States were reported by several laboratories.

The manufacturer's testing protocol for IIF was followed by 23 (28%) of 82 laboratories responding to this question in May and by 47 (42%) of the 111 laboratories responding in September. The state health department protocol was followed by 37 laboratories (45%) responding in May and by 38 laboratories (34%) responding in September. The percentage of laboratories using the manufacturer's testing protocol for IIF testing was lower than for EIA and WB testing because fewer IIF laboratories used complete kits by one manufacturer. Unlike EIA and WB testing, IIF testing is frequently performed with in-house reagents, noncommercial reagents, or tests consisting of components from various manufacturers.

A total of 106 laboratories indicated in the September 1988 questionnaire that they refer specimens to other laboratories for IIF testing and reported the type of laboratory to which such specimens are referred. Most of the laboratories (57, or 53.8%) referred specimens to the state laboratory for IIF testing. The only other frequently used referral site was independent laboratories (23 referrals, or 21.7%). Hospitals, blood banks, and other laboratories were recipients of the remaining (24.5%) specimens referred for IIF testing.

For the combined May and September 1988 performance evaluation results on the negative pool (P1) and the individual negative specimens (01 through 08), the laboratories (reference and participant) reported negative interpretations in 431 (89.6%) of 481 results and only 43 (8.9%) and 7 (1.5%) were reported as indeterminate and positive, respectively

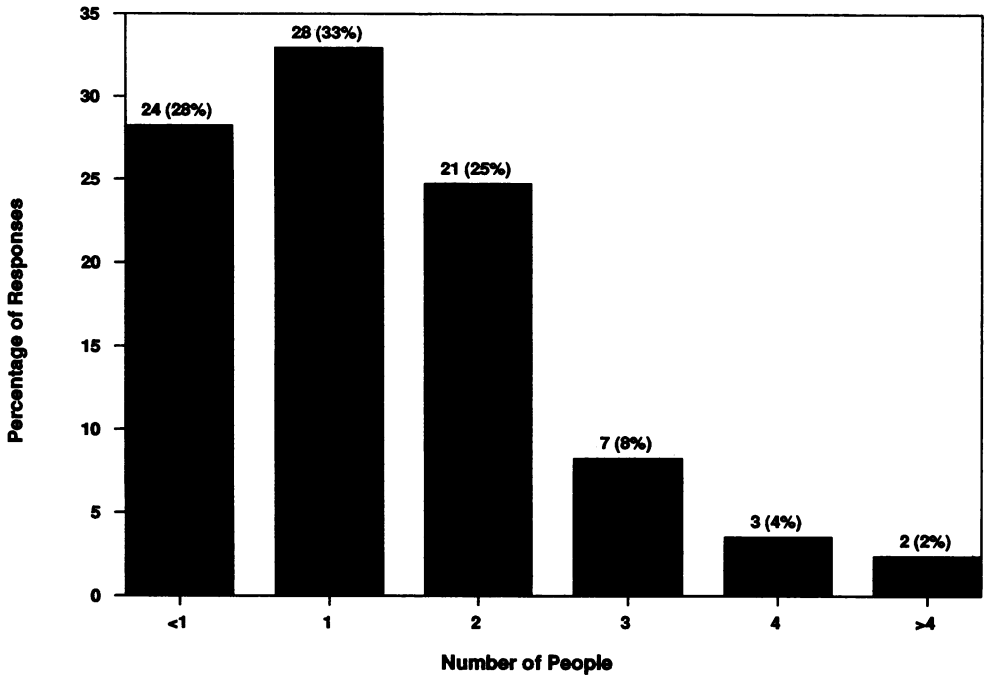

FIG. 3. Numbers of personnel performing tests according to September 1988 CDC questionnaire.

(Table 1). This gives a calculated analytic specificity for these specimens of 98.5% when indeterminate results are considered correct and 89.6% when indeterminate results are considered incorrect.

Comparison of results by laboratory type revealed that all types had analytic specificities over 96% when indeterminate results were considered correct. All but those that did not report type of laboratory had analytic specificities over 85% when indeterminate results were considered incorrect.

Laboratories using test kits manufactured by Electronucleonics (10 results) or in-house (27 results) accounted for 86% of the indeterminate interpretations. One positive interpretation each was reported by laboratories using kits manufactured by Biotech, Cellular Products, and an unspecified manufacturer. Four positive interpretations were reported by laboratories using in-house reagents. The analytic specificities for all manufacturers were above 95% when indeterminate results were considered correct, and all but one were

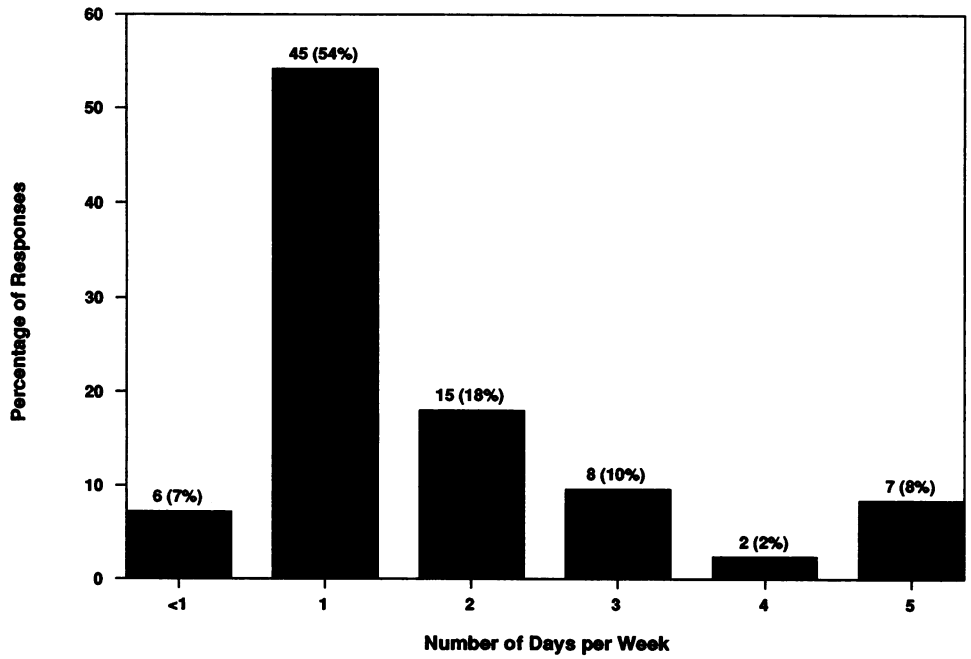

FIG. 4. Frequency of IIF HIV-1 testing per week according to September 1988 CDC questionnaire.

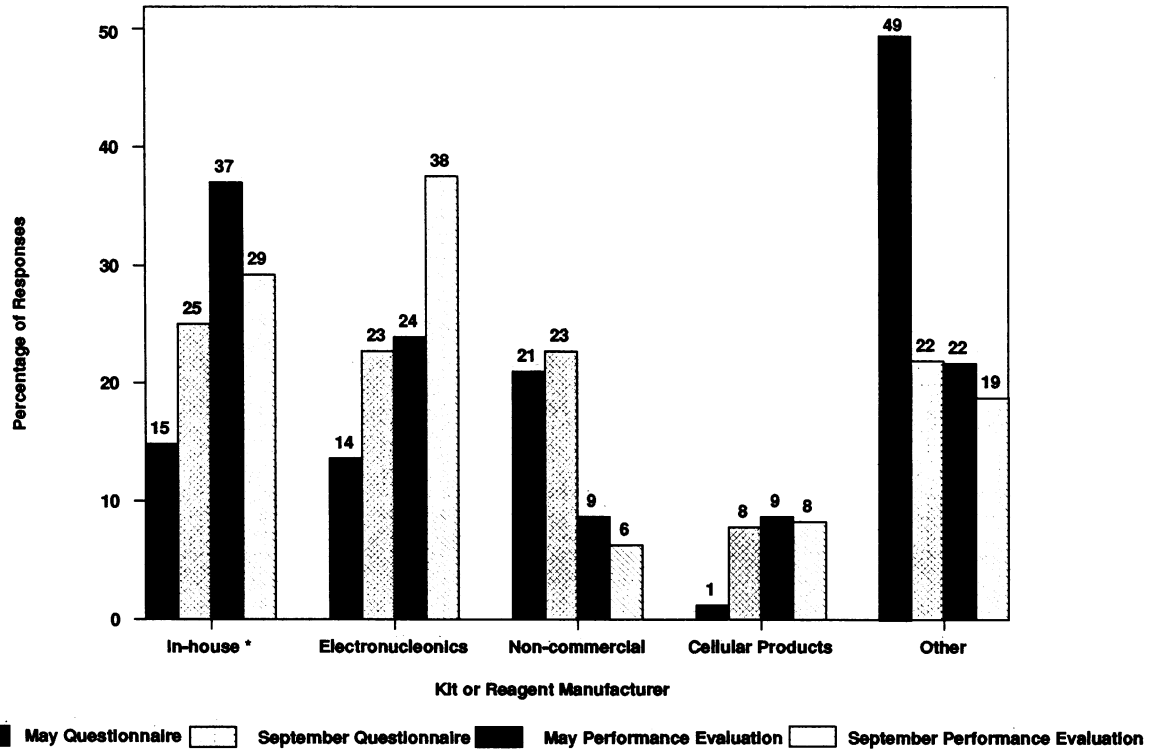

FIG. 5. IIF test reagents or kits used by MPEP participants in 1988. Blank fields in performance evaluations were considered to be in-house; some laboratories gave multiple responses.

TABLE 1. Analytic specificities and sensitivities of IIF tests

Laboratory or test	No. of results	Analytic specificity[a] when indeterminate results are considered:		No. of results	Analytic sensitivity[b] when indeterminate results are considered:	
		Correct	Incorrect		Correct	Incorrect
Laboratory type						
Hospital	68	97.1	86.8	150	98.0	98.0
Health department	179	99.4	89.9	650	92.9	89.4
Blood bank	36	100.0	88.9	102	100.0	96.1
Independent	77	96.1	85.7	159	93.7	88.1
Other[c]	113	99.1	94.7	274	96.4	92.7
Not reported	8	100.0	75.0	26	92.3	92.3
Total	481	98.5	89.6	1,361	94.8	91.4
Test manufacturer or type						
Biotech	22	95.5	81.8	50	100.0	96.0
Cellular Products	29	96.6	93.1	72	79.2	76.4
Electronucleonics	96	100.0	89.6	241	99.2	97.5
In-house	122	96.7	89.3	256	95.3	94.9
Noncommercial	8	100.0	100.0	44	95.5	95.5
Other	49	98.0	93.9	112	96.4	87.5
Not reported	155	100.0	88.4	586	93.9	89.2
Total	481	98.5	89.6	1,361	94.8	91.4

[a] Analytic specificity, Percentage of tests on negative specimens reported as negative.

[b] Analytic sensitivity, Percentage of tests on positive specimens reported as positive.

[c] Includes multiple responses.

above 88% when indeterminate results were considered incorrect.

Among the 1,361 results reported by the laboratories for the undiluted positive donor specimens (09 through 24, 28, 34 through 36, and 40) and for pools P2 and P3 (previously tested by CDC and interpreted as WB positive), 46 (3.4%) were interpreted as indeterminate and 71 (5.2%) were interpreted as negative. This gives a calculated analytic sensitivity on these specimens of 94.8% when indeterminate results are considered correct and 91.4% when indeterminate results are considered incorrect.

Comparison of results by laboratory type revealed that all types had analytic sensitivities over 92% when indeterminate results were considered correct, and all had analytic sensitivities over 88% when indeterminate results were considered incorrect.

Laboratories using test kits manufactured by Cellular Products had the lowest analytic sensitivities (76.4%). The analytic sensitivities for all manufacturers except Cellular Products were above 94% when indeterminate results were considered correct and above 87% when indeterminate results were considered incorrect.

The diluted donor material that CDC tested by WB and interpreted as positive (29, 30, 33, 37 through 48, P4, and P5) presented the greatest IIF testing difficulty. Among the 1,063 interpretations reported by the laboratories, 666 (62.7%) were negative and 144 (13.5%) were indeterminate.

Within-survey reproducibility (Tables 2 and 3) for these two performance evaluation surveys was 92.6% for negative specimens, 91.8% for indeterminate specimens, and 93.5% for positive specimens, for an overall within-survey reproducibility of 92.8%. Between-survey reproducibility (Tables 4 and 5) was 80.5% for negative specimens, 81.6% for

TABLE 2. Within-survey reproducibility for IIF tests by laboratory type according to 1988 CDC surveys

Laboratory type	No. (%) of:			
	Negatives	Indeterminates	Positives	Total
Hospital	17 (88.2)	9 (77.8)	22 (96.2)	48 (89.9)
Health department	34 (94.2)	29 (100.0)	63 (91.8)	126 (94.3)
Blood bank		1 (100.0)	2 (100.0)	3 (100.0)
Independent	15 (86.7)	10 (80.0)	15 (97.4)	40 (89.0)
Not indicated	2 (100.0)	3 (100.0)	4 (100.0)	9 (100.0)
Other[a]	26 (96.2)	21 (90.5)	33 (91.8)	80 (92.9)
Total	94 (92.6)	73 (91.8)	139 (93.5)	306 (92.8)

[a] Includes multiple responses.

TABLE 4. Between-survey reproducibility for IIF tests by laboratory type according to 1988 CDC surveys

Laboratory type	No. (%) of:			
	Negatives	Indeterminates	Positives	Total
Hospital	5 (55.0)	4 (87.5)	8 (96.9)	17 (82.4)
Health department	17 (83.2)	15 (81.8)	31 (84.9)	63 (83.7)
Blood bank	2 (89.0)	2 (80.0)	2 (91.0)	6 (86.7)
Independent	6 (77.3)	5 (80.0)	6 (92.7)	17 (83.5)
Other[a]	12 (87.5)	10 (80.0)	15 (89.6)	37 (86.3)
Total	42 (80.5)	36 (81.6)	62 (88.5)	140 (84.3)

[a] Includes multiple responses.

indeterminate specimens, and 88.5% for positive specimens, for an overall between-survey reproducibility of 84.3%.

DISCUSSION

The fact that about 90% of the laboratories receiving sample panels or questionnaires reported results suggests strong support and willingness on the part of most of the laboratories to participate in programs designed to improve the quality of HIV-1 antibody testing. The questionnaires were designed to create profiles of HIV-1 antibody-testing laboratories and their testing practices. The profiles will be used to help document and understand changes and trends in HIV-1 testing and to identify potential barriers to high-quality testing. By correlating the data obtained from the questionnaires with the results of the specimen analyses, CDC will look for variables that might be used to predict testing quality.

Only about 8% of the 1,416 laboratories in the MPEP perform IIF testing. The distribution of IIF laboratories among the states is similar to the total distribution of laboratories, with IIF laboratories in about three-fourths of the states. Over one-fourth of the IIF laboratories were hospital laboratories, with many laboratories being identified as more than one type of laboratory. Almost two-thirds of the laboratories who did IIF testing also performed EIA and WB. Nearly one-third did a combination of EIA and IIF testing. The IIF laboratories employed fewer personnel and did testing less frequently on a weekly basis than did EIA or WB laboratories.

The fact that most of the laboratories doing IIF testing do not use a commercially available kit prompted the ASTPHLD HIV committee to recommend the production and commercial distribution of standardized IIF reagents and kits for use in supplemental testing (1). The need for such standardization is suggested in this paper by the high percentage of laboratories using in-house kits or tests consisting of components from various manufacturers. ASTPHLD made two other recommendations related to the use of in-house or unlicensed reagents that CDC supports. The first was that in-house procedures and reagents should be used only as an adjunct to licensed products and that laboratories using in-house reagents, tests, or procedures should document that their test is equivalent to currently available licensed tests for the same purpose. The second was that, whenever available, licensed procedures should be the standard against which similar unlicensed procedures are compared, and where no licensed procedure exists, testing should conform to a generally accepted protocol.

Among our participants, when specimens found to be positive by screening tests are referred for supplemental testing by IIF, more than half the time the laboratory to which they are referred is the state laboratory. This differs from EIA and WB testing, for which independent laboratories are the most common recipients of referred specimens.

Analytic specificity (the percentage of reports on negative specimens that were reported as negative) for these specimens was 98.5% when indeterminate results were considered and 89.6% when indeterminate results were considered incorrect. Comparison of results by type of laboratory showed that all types had analytic specificities over 96% when indeterminate results were considered correct. All but those which did not report the laboratory type had analytic specificities over 85% when indeterminate results were considered incorrect. The analytic specificities for all manufacturers were above 95% when indeterminate results were considered correct, and all but one were above 88% when they were considered incorrect.

Analytic sensitivity (the percentage of reports on positive specimens that were reported as positive) was 94.8% when

TABLE 3. Within-survey reproducibility for IIF tests by test manufacturer or type according to 1988 CDC surveys

Test manufacturer or type	No. (%) of:			
	Negatives	Indeterminates	Positives	Total
Biotech	1 (50.0)	1 (0.0)	1 (100.0)	3 (50.0)
Cellular Products	7 (92.9)	4 (100.0)	8 (86.9)	19 (91.9)
Electronucleonics	17 (97.1)	13 (92.3)	25 (97.7)	55 (96.2)
In-house	55 (93.7)	48 (93.7)	83 (93.1)	186 (93.4)
Noncommercial	2 (100.0)	3 (100.0)	7 (100.0)	12 (100.0)
Other	12 (83.3)	4 (75.0)	15 (88.7)	31 (84.8)
Total	94 (92.6)	73 (91.8)	139 (93.5)	306 (92.8)

TABLE 5. Between-survey reproducibility for IIF tests by test manufacturer or type according to 1988 CDC surveys

Test manufacturer or type	No. (%) of:			
	Negatives	Indeterminates	Positives	Total
Biotech	1 (89.0)	1 (100.0)	1 (91.0)	3 (93.3)
Cellular Products	4 (75.0)	2 (100.0)	4 (75.8)	10 (80.3)
Electronucleonics	7 (80.6)	7 (87.1)	11 (95.8)	25 (89.1)
In-house	24 (77.2)	24 (78.2)	38 (87.9)	86 (82.2)
Noncommercial	1 (100.0)		2 (91.5)	3 (94.3)
Other	5 (95.0)	2 (75.0)	6 (85.8)	13 (87.7)
Total	42 (80.5)	36 (81.6)	62 (88.5)	140 (84.3)

indeterminate results were considered correct and 91.4% when indeterminate results were considered incorrect. Comparison of results by type of laboratory showed that all types had analytic sensitivities over 92% when indeterminate results were considered correct. All but independent laboratories and health departments had analytic specificities over 88% when indeterminate results were considered incorrect. The analytic sensitivities for all manufacturers except Cellular Products were above 94% when indeterminate results were considered correct and were above 87% when indeterminate results were considered incorrect.

Comparison of percentages of correct responses between IIF and WB on those samples for which there was good agreement on the target interpretation revealed no significant differences.

Within-survey reproducibilities on these two performance evaluation surveys were about 93% overall and by specimen type. This is approximately the same as the 94.7% which was reported by Vercauteren et al. (20). Within-survey reproducibilities seemed to be about the same regardless of the specimen reactivity level, but the laboratories seemed to be able to reproduce results between surveys better when they tested positive specimens. Between-survey reproducibilities were about 81% for negative and indeterminate specimens and 88.5% for positive specimens, for an overall between-survey reproducibility of 84.3%. Because there are more potential sources for error between surveys than there are within surveys, it was expected that between-survey reproducibility would be lower than within-survey reproducibility; that is what was seen in this study. There did not seem to be a difference in either within- or between-survey reproducibility among the types of laboratories or test manufacturers.

Results from proficiency testing and performance evaluations should be interpreted cautiously, because data from these programs measure the performances of participating laboratories under field conditions using samples with different distributions of reactivity levels and cannot be used to measure the clinical sensitivity or specificity of a given test. Specimens provided in proficiency testing and performance evaluation surveys are often pooled human plasma specimens with known levels of reactivity or dilutions of a single reactive plasma in negative serum. They are rarely fresh serum specimens from a person with documented disease status. Some specimens are selected because they exhibit nonspecific reactivity or are otherwise difficult to test and interpret; they are not typical of the specimens that will be handled by the participating laboratories (5).

Another reason for cautious interpretation of proficiency testing and performance evaluation results comes from the fact that the target interpretations are based on less-than-perfect reference standards. Allowance must be made for the imperfect sensitivities and specificities of the tests used to establish the target interpretation (17).

In the terms used by Schwartz et al. (17), we have attempted an estimate of test effectiveness (performance under average conditions) on these specimens, which is a more practical measure of performance than test efficacy (performance under ideal conditions). As they point out, the effectiveness of a test is always less than its efficacy. The sensitivities and specificities reported in the package inserts of commercial test kits are substantially higher than those from proficiency testing and performance evaluation surveys.

The CDC Model Performance Evaluation Program plans to continue to provide performance evaluation data concerning HIV test procedures and products and will expand its focus to address quality assurance issues related to testing for other human retroviruses. The importance of this activity has been addressed by ASTPHLD (1). We concur with Schwartz et al. (17) that current test performance could be improved by better standardization of test procedures and institution of mandatory proficiency testing. Laboratories are encouraged to participate in proficiency-testing programs as well as in the CDC performance evaluation program.

As a supplementary test to confirm repeatedly positive EIA results, the IIF tests offer several practical advantages. The IIF tests are already being performed in many laboratories, especially public health laboratories, and consequently these laboratories can capitalize on existing expertise and equipment to adapt to the HIV-1 procedure. The IIF tests are usually less expensive than WB procedures. Finally, the IIF tests are valuable in helping to determine the status of sera that are still indeterminate after EIA and WB. The performance data for the IIF tests suggest that, even though testing is often performed well, improvements could be made.

ACKNOWLEDGMENTS

We thank Sharon Blumer, David Cross, Ron Fehd, Russell Gerber, Angela Glaude, Charles Griffin, Carol Johnson, Patrick Kelly, Gerri Mast, Harold Muir, Pam Robinson, Robert Schwartz, June Sumpter, and Ross Wood for providing clerical, data processing, and technical support; for reviewing the manuscript; or for providing helpful suggestions.

LITERATURE CITED

1. **Association of State and Territorial Public Health Laboratory Directors.** 1989. Fourth consensus conference on testing for human retroviruses: report and recommendations. Association of State and Territorial Public Health Laboratory Directors, Kansas City, Mo.
2. **Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, R. Vezinet-Braun, C. Rouzioux, W. Rosenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**:868–870.
3. **Broder, S., and R. C. Gallo.** 1984. A pathogenic retrovirus (HTLV-III) linked to AIDS. N. Engl. J. Med. **51**:1292–1297.
4. **Carlson, J. R., J. Yee, S. H. Hinrichs, M. L. Bryant, M. B. Gardner, and N. C. Pedersen.** 1987. Comparison of indirect immunofluorescence and Western blot for detection of anti-human immunodeficiency virus antibodies. J. Clin. Microbiol. **25**:494–497.
5. **Centers for Disease Control.** 1988. Update: serologic testing for antibody to human immunodeficiency virus. Morbid. Mortal. Weekly Rep. **36**:833–840.

5a. **Centers for Disease Control.** 1989. Interpretation and use of the Western blot assay for serodiagnosis of human immunodeficiency virus type 1 infections. Morbid. Mortal. Weekly Rep. **38**:1–7.

6. **Curran, J. W., H. W. Jaffe, A. Hardy, W. M. Morgan, R. M. Selik, and T. J. Dondero.** 1988. Epidemiology of HIV infection and AIDS in the United States. Science **239**:610–616.
7. **Dondero, T. J., M. Pappaianou, and J. W. Curran.** 1988. Monitoring the levels and trends of HIV infection: the Public Health Service's HIV surveillance program. Am. J. Public Health **103**:213–220.
8. **Galen, R. S., and S. R. Gambino.** 1975. Beyond normality: the predictive value and efficiency of medical diagnosis. John Wiley & Sons, Inc., New York.
9. **Gallo, D., J. L. Diggs, G. R. Shell, P. L. Dailey, M. N. Hoffman, and J. L. Riggs.** 1986. Comparison of detection of antibody to the acquired immune deficiency syndrome virus by enzyme immunoassay, immunofluorescence, and Western blot methods.

J. Clin. Microbiol. **23**:1049–1051.

10. **Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, G. White, P. Foster, and P. D. Markham.** 1984. Frequent detection and isolation of cytopathic retrovirus (HTLV-III) from patients with AIDS and at risk for AIDS. Science **224:** 500–503.
11. **Gallo, R. C., P. S. Sarin, E. P. Gelmann, M. Robert-Guroff, E. Richardson, V. S. Kalyanaraman, D. Mann, G. D. Sidhu, R. E. Stahl, S. Zolla-Pazner, J. Leibowitch, and M. Popovic.** 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science **220**:865–867.
12. **Goedert, J. J.** 1986. Testing for human immunodeficiency virus. Ann. Intern. Med. **105**:609–610.
13. **Hedenskog, M., S. Dewhurst, C. Ludvigsen, F. Sinangil, L. Rodriguez, Y. T. Wu, and D. J. Volsky.** 1986. Testing for antibodies to AIDS-associated retrovirus (HTLV-III/LAV) by indirect fixed cell immunofluorescence: specificity, sensitivity, and applications. J. Med. Virol. **19**:325–334. (Erratum, **20**:390.)
14. **Inhorn, S. L., and B. V. Addison (ed.).** 1987. Proceedings of the 1986 institute on critical issues in health laboratory practice: managing the quality of laboratory test results in a changing health care environment. Du Pont Company, Wilmington, Del.
15. **Lennette, E. T., S. Karpatkin, and J. A. Levy.** 1987. Indirect immunofluorescence assay for antibodies to human immunodeficiency virus. J. Clin. Microbiol. **25**:199–202.
16. **McHugh, T. M., D. P. Stites, C. H. Casavant, J. R. Carlson, M. P. Busch, and J. A. Levy.** 1987. Evaluation of the indirect immunofluorescence assay as a confirmatory test for detecting antibodies to the HIV. Diagn. Immunol. **4**:233–240.

16a. **Schalla, W. O., T. L. Hearn, R. N. Taylor, E. Eavenson, R. O. Valdiserri, and J. D. K. Essien.** 1990. Centers for Disease Control Model Performance Evaluation Program: assessment of the quality of laboratory performance for human immunodeficiency virus type 1 (HIV-1) antibody testing. Public Health Rep. **105**:167–171.

17. **Schwartz, J. S., P. E. Dans, and B. P. Kinosian.** 1988. Human immunodeficiency virus test evaluation, performance, and use. J. Am. Med. Assoc. **259**:2574–2579.
18. **Taylor, R. N., and V. A. Przybyszewski.** 1988. Summary of the Centers for Disease Control human immunodeficiency virus (HIV) performance evaluation surveys for 1985 and 1986. Am. J. Clin. Pathol. **89**:1–13.
19. **Valdiserri, R. O., R. N. Taylor, T. L. Hearn, W. O. Schalla, H. W. Muir.** 1990. Centers for Disease Control perspective on quality assurance for HIV-1 antibody testing: model performance evaluation program. Arch. Pathol. Lab. Med. **114**:263–267.
20. **Vercauteren, G., G. van der Groen, and P. Piot.** 1987. Comparison of enzyme immunoassays and an immunofluorescence test for detection of antibody to human immunodeficiency virus in African sera. Eur. J. Clin. Microbiol. **6**:132–135.