JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 1981, p. 356-368 Vol. 13, No. 2
0095-1137/81/020356-13$02.00/0

Assessment of Laboratory Improvement by the Center for Disease Control Diagnostic Immunology Proficiency Testing Program

ROGER N. TAYLOR* AND KAREN M. FULFORD

Diagnostic Immunology Section, Proficiency Testing Branch, Center for Disease Control, Atlanta, Georgia 30333

The data accumulated from 1969 to 1979 in the Diagnostic Immunology portion of the Center for Disease Control Proficiency Testing Program were evaluated for evidence of change in performance among the participating laboratories. Evidence of improved performance was found for the rubella, rheumatoid factor, tularemia, quantitative immunoglobulin (immunoglobulin G, A, and M), and hepatitis B tests. No evidence of change was detected for the streptococcal enzyme, C-reactive protein, infectious mononucleosis, antinuclear antibodies, *Salmonella* and *Brucella* agglutinins, and syphilis tests. Data obtained from other tests were inadequate to determine trends. In most tests, deficiencies were identified which could be corrected and thereby could improve performance. It is pointed out that proficiency testing not only improves laboratory performance, but also can be used to evaluate performance levels, identify method, standard, or performance deficiencies, educate, estimate impact of possible changes, serve as external quality control, and document changes.

How does proficiency testing relate to laboratory improvement? Under the Clinical Laboratories Improvement Act of 1967, the Center for Disease Control (CDC) Licensure and Proficiency Testing Program was given the responsibility for measuring and improving clinical laboratory performance in the United States. The goal of the Proficiency Testing portion of the program is to improve performance by analyzing test results, detecting deficiencies, evaluating methods, and disseminating reports and other pertinent information to the participants.

Proficiency testing was designed to measure the competence of the analyst at the bench, but it also measures other factors which can influence performance, such as the following: (i) choice of analytical method; (ii) availability of physical or methodological standards; (iii) proficiency of the person who prepared the standards; (iv) reliability of the reagents; (v) instrument design, manufacture, and maintenance; (vi) adequacy of the internal quality control program; (vii) correctness of test interpretation; (viii) accuracy of accessioning samples and recording results; (ix) effectiveness of the proficiency testing process.

Proficiency testing is most valuable as a laboratory improvement tool when the samples are treated as routine patient samples rather than given special treatment. Under these conditions, the test results will reflect the level of performance achieved with patient samples and, consequently, will be of more value to the laboratory.

Proficiency testing can also function as an external quality control program. It is one component of a total quality control system that ensures quality performance in a laboratory; other components are (4) personnel qualification standards, internal quality control programs, preventive maintenance of instruments and equipment, ongoing education and training programs, periodic re-evaluation and updating of procedures, and safety programs.

The Diagnostic Immunology portion of the CDC Proficiency Testing Program functions as follows. Specimens prepared by CDC personnel are distributed to licensed and some nonlicensed (special study and reference laboratories) participants. A few laboratories located outside the United States (primarily World Health Organization and Pan American Health Organization laboratories) also participate. In instructions accompanying the specimens, laboratories are asked to have regular laboratory staff members test the specimens in a routine manner. Each laboratory is asked to perform only those tests that its staff offers to the public. Participation in the program is mandatory for laboratories that provide interstate testing services.

In this report, data obtained from the Diagnostic Immunology Program are summarized for the purpose of delineating the overall trends and

changes that became apparent after the accumulation of 10 years of test data was evaluated.

MATERIALS AND METHODS

Most of the sera of plasma used for specimen preparation were purchased from commercial suppliers under government contract. Other sources donated pools of human serum of known reactivity in specific tests. Some animal antisera have been used. The hepatitis B surface antigen (HBsAg)-reactive samples were obtained primarily from blood donors who were found to have the antigen in their blood.

The Diagnostic Immunology Section of the Proficiency Testing Branch or an appropriate CDC specialty laboratory, or both, tested the sera for acceptability for use in the program. The Diagnostic Immunology Section or the Virology Section, Phoenix Laboratories Division, determined HBsAg reactivity for all sera after 1973. Only sera negative for HBsAg by radioimmunoassay were used for specimens other than those to be tested for HBsAg. Specimens obtained as plasmas were defibrinated with calcium chloride or thrombin.

Details of specimen preparation and tabulations of results for each survey are included in quarterly summary analyses (10–13, 18–22) annual critiques (14, 23, 24), and professional journal publications (3, 16, 17, 25–27). Briefly, specimens were adjusted to the desired reactivity, filtered through sterile membrane filters, and dispensed into suitable vials or tubes. Many of the specimens were lyophilized. The adequacy of samples was confirmed independently by the Diagnostic Immunology Section, by other CDC specialty laboratories, and by reference laboratories. A continuous quality control program ensures that all specimens satisfy preestablished criteria for sterility, antibody titer and stability, and between-vial variability.

Each specimen shipment was packaged and mailed in accordance with postal regulations and included appropriate instructions and report forms. Completed reports were to be postmarked within 2 weeks of the initial shipping date. Responses were compiled and graded, and individual performance rankings were reported to participants within 3 to 4 weeks after responses were received. Acceptable responses were determined from reference laboratory results. Overall response data, which were evaluated and compiled in summary analyses or published as separate reports, were later sent to all participants.

Periodically, special surveys were prepared to evaluate analytes being considered for inclusion in the routine proficiency testing program. These surveys were performed by similar methods except that grades were not determined.

RESULTS

Table 1 shows the Diagnostic Immunology Program composition and participation levels from the beginning of the program to the present. The rapid growth of the program is readily apparent. In 1969 the program consisted of 5 analytes; the 1980 program contains 39 analytes. The number of laboratories participating in the program has increased tremendously. For some tests the number of participant laboratories in 1980 is almost 10 times the number in 1969. The number of challenges per full-service laboratory increased from 79 to 362 per year.

Table 2 shows the average geometric standard deviations (S_G) as a measure of interlaboratory precision achieved by the participants for each analyte by year. The last column is the average for the analyte. For most analytes there has been no significant change with time. For many analytes there are insufficient data to determine whether a change has occurred. The effect of increased precision can be determined from Table 3, which shows the highest and lowest S_G for each analyte and the effect the S_G has on the 95% limits around a representative value. For a sample with a geometric mean of 1,000 mg of immunoglobulin G per 100 ml, an S_G of 1.23 gives a 2-standard deviation range (95% limits) of 661 to 1,513. When the S_G is decreased to 1.09, the range becomes 841 to 1,188. This indicates that more results were closer to the geometric mean; that is, the interlaboratory precision was better. The smaller S_G values are not always the most recent value, as can be seen from Table 2.

Interlaboratory comparability of rheumatoid factor results improved with the introduction of a rheumatoid factor reference preparation into the Proficiency Testing Program (3, 17). Figure 1 shows the results that were achieved in one survey through the use of such a reference preparation. Nonstandardized results had a range of 12 twofold dilutions, but standardized results had a range of 7 twofold dilutions, with 94% of the results within one dilution of the median. Figure 2 shows the improvement in interlaboratory precision over 10 years. Before use of the standard material, the average S_G was 2.84; after, it was 2.41. This improvement was also documented by evaluating the percentage of results within one twofold dilution of the median (Fig. 3). Before introduction of the standard, 74% of the results were in this range; after introduction of the standard, 84% were in this range.

Significant improvement was shown in the performance of the tube test for tularemia ($P < 0.01$), in which the S_G decreased from 2.62 in 1969 to 1.83 in 1979. Performance with the slide test did not noticeably improve, but because of the improvement with the tube tests, overall performance showed some improvement. The overall S_G decreased from 2.50 to 2.08 ($P < 0.05$). The major source of variation in the bacterial agglutination tests is still the antigen, a problem which has been observed since the beginning of the program. The need to develop standard antigens for bacterial agglutination tests and to use

TABLE 1. *Number of laboratories and samples provided by year and analyte, diagnostic immunology, 1969–1980*

Analyte	No. of laboratories (no. of samples)											
	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980[a]
Syphilis serology	89 (60)	97 (60)	103 (60)	107 (60)	297 (60)	259 (60)	325 (40)	340 (40)	363 (40)	494 (40)	553 (40)	651 (16)
Rubella	54 (5)		103 (5)	59 (5)	183 (5)	243 (4)	259 (8)	283 (8)	318 (8)	334 (8)	380 (8)	488 (8)
Brucella	79 (7)	187 (5)		231 (3)		360 (5)	403 (6)	393 (2)	418 (4)	399 (2)	456 (8)	519 (16)
Tularemia	84 (6)	124 (5)		162 (2)				221 (2)	249 (2)	236 (2)	264 (8)	326 (16)
Rickettsial antibody	75 (1)							359 (1)		353 (2)	416 (4)	173 (16)
Antistreptolysin O		176 (5)	162 (10)	238 (5)	285 (15)	364 (3)	390 (6)	354 (6)	388 (4)	338 (4)	382 (8)	616 (8)
Rheumatoid factor		204 (5)	213 (5)	219 (10)	340 (5)	407 (5)	410 (6)	437 (5)	506 (12)	540 (4)	591 (4)	908 (12)
Infectious mononucleosis			213 (5)	316 (10)		414 (5)	424 (5)	459 (5)	521 (2)	493 (4)	624 (4)	979 (12)
Hepatitis B			94 (10)	153 (10)	238 (10)	239 (10)	289 (10)	264 (10)	276 (10)	273 (10)	285 (10)	383 (10)
Immunoglobulin G					127 (3)		169 (3)	333 (7)	330 (4)	334 (7)	363 (4)	446 (4)
Immunoglobulin A					127 (3)		169 (3)	333 (7)	330 (4)	334 (7)	363 (4)	446 (4)
Immunoglobulin M					127 (3)		169 (3)	333 (7)	330 (4)	334 (7)	363 (4)	447 (4)
Mycotic serology					17 (1)					72 (5)		109 (4)
Toxoplasma						72 (5)	96 (5)	144 (2)	167 (2)	167 (2)	174 (8)	229 (8)
C-reactive protein											411 (8)	653 (12)
Antinuclear antibodies							278 (4)	337 (4)	382 (2)	380 (2)	327 (4)	464 (8)
Salmonella (group D)							363 (1)	365 (1)	394 (2)	418 (2)	317 (4)	485 (16)
Antideoxyribonuclease B								23 (6)	41 (4)	40 (4)	65 (8)	73 (8)
Multiple streptococcal enzymes								98 (6)	96 (4)	181 (4)	176 (8)	308 (8)
Complement C3								304 (5)	287 (2)	269 (2)	359 (2)	393 (4)
Complement C4								164 (5)	154 (2)	197 (2)	282 (4)	376 (4)
Alpha-1-antitrypsin								245 (5)	230 (2)	212 (2)	213 (2)	291 (4)
Alpha-2-macroglobulin								47 (5)				58 (4)
Haptoglobin								213 (5)	212 (2)	212 (2)		270 (4)
Transferrin								82 (5)				127 (4)
Ceruloplasmin								115 (5)	86 (2)	86 (2)		133 (4)
Carcinoembryonic antigen								125 (5)	104 (2)	115 (2)	100 (2)	150 (4)
Immunoglobulin D										42 (5)		120 (4)
Immunoglobulin E										54 (5)		176 (4)
Cytomegalovirus antibody											90 (4)	153 (8)
Herpesvirus antibody											92 (4)	133 (8)
Antimitochrondrial antibody											81 (2)	151 (8)
Antideoxyribonucleic acid antibody											100 (4)	174 (8)
Anti-smooth muscle antibody											100 (2)	131 (8)

Antihyaluronidase												75 (8)
Hemolytic complement												54 (4)
Alpha-fetoprotein												90 (40)
Estrogen receptor assay												74 (20)
Progesterone receptor assay												73 (20)
No. of challenges per laboratory[b]	79	80	95	105	106	97	104	161	122	145	186	362

[a] Estimated.
[b] Challenges = the sum of the number of samples for all of the analytes.

those that are available persists.

Performance with the immunoglobulin quantitation tests has improved. The S_G for immunoglobulin G decreased from 1.23 to 1.09 ($P < 0.05$); the S_G for immunoglobulin A decreased from 1.44 to 1.13 ($P < 0.01$); and the S_G for immunoglobulin M decreased from 1.35 to 1.17 ($P = 0.05$). Improved performance is indicated in Table 4 by decreases in the percentage of results outside the acceptable range.

Table 5 shows the percentage of participants using various serum treatments for rubella hemagglutination inhibition testing. The percentage using the nonstandard kaolin serum treatment for hemagglutination inhibition decreased from 65 to 22%. This indicates improvement in performance because the percentage of results outside the acceptable limits for the kaolin procedure has consistently been almost double that for the standard procedures (8, 9). The reason for some of the fluctuation from year to year is the differences in width of the acceptable ranges (see last column).

The trends in test use for syphilis serology are shown in Table 6. Although syphilis is not listed in Table 2, the performance (based on grades) with these tests seems to have been maintained at a high level.

The variation in substrate sensitivity for antinuclear antibody detection is a problem in standardizing this test (1). The relative sensitivities of substrates used in the indirect immunofluorescence tests for antinuclear antibodies were obtained by ranking the geometric mean titers reported for the 1975 through 1978 surveys. The relative sensitivity of each of the substrates, ranked from highest to lowest, is as follows: rat liver (most sensitive); mouse brain and mouse kidney; human cell line and mouse liver; rat kidney, human kidney, and mouse cell line; human leukocytes; and rat brain (least sensitive). Within each group the substrates showed no significant difference in sensitivity. The geometric mean titers reported with the most sensitive substrate are typically two to eight times higher than those reported with the less sensitive substrates. Obviously, this makes comparisons of results among laboratories difficult. Selection of a single substrate as a standard or use of a standard serum should improve comparability of results with this test.

The importance of adhering to recommended procedures for streptococcal enzyme tests was demonstrated by proficiency testing (5–7). Antistreptolysin O test results indicated that some of the variation in results was because many of the participants were reconstituting the streptolysin reagent with water at room temperature.

TABLE 2. *Average interlaboratory precision by year and analyte*

Analyte	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	$\bar{x}$[a]
Rubella	2.29		1.92	1.88	1.90	2.03	1.90	1.86	1.91	1.88	1.89	1.95
Streptococcal antibodies												
Antistreptolysin O		1.52	1.80	1.47	1.69	1.54	1.55	1.77	1.52	1.88	1.65	1.64
Antideoxyribonuclease B								1.62	1.44	1.40	1.46	1.48
Multiple enzymes					2.26			1.86	1.90	1.74	1.80	1.82
Rheumatoid factor												
Slide						2.16		4.46	1.74	2.83	2.84	2.81
Latex						2.54	2.31	3.10	1.92	2.32	1.91	2.35
Total		3.05	2.66	3.16	2.41	2.87	2.60	3.10	2.50	2.35	2.39	2.71
Infectious mononucleosis												
Slide								4.38	2.36[b]	3.90	3.95	4.08
Heterophile			2.46	2.51		2.06	2.31	2.38	2.11	2.31	2.26	2.30
Ox cell			3.01	2.94		2.48	2.30	2.10	3.97	1.95	2.82	2.68
Antinuclear antibodies												
Indirect immunofluorescence							2.62	2.59	3.08	2.30	2.76	2.67
Salmonella												
Slide								2.07	2.36	2.32	2.56	2.33
Tube								2.26	1.84	2.43	2.41	2.34
Total								2.16	2.14	2.46	2.58	2.34
Brucella												
Slide	2.75	2.06		1.75		2.17		2.17	2.16	1.94	2.06	2.13
Tube	2.29	2.09		2.29		2.50		2.40	2.48	2.47	2.21	2.33
Total	2.69	2.32		2.05		2.32	2.20	2.26	2.33	2.27	2.10	2.23
Tularemia												
Slide	2.13	2.11		2.26				1.84	1.90	2.12	2.33	2.08
Tube	2.62	2.20		2.38				1.88	1.96	1.83	1.83	2.10
Total	2.50	2.27		2.35				1.88	1.94	2.02	2.08	2.15
Immunoglobulin G					1.23		1.23	1.23	1.17	1.17	1.09	1.19
Immunoglobulin A					1.44		1.37	1.23	1.28	1.22	1.13	1.28
Immunoglobulin M					1.35		1.35	1.28	1.30	1.30	1.17	1.29
Toxoplasma												
Indirect immunofluorescence						3.28	2.70	2.86	3.20	2.94	2.37	2.89
Passive hemagglutination						2.41	2.91	1.97	2.97	2.35	2.24	2.48
Complement C3												
β1A										1.40	1.18	1.29
β1A/β1C										1.28	1.22	1.25
Complement C4										1.30	1.26	1.28
Alpha-1-antitrypsin										1.20	1.29	1.24
Weil-Felix												
Slide										—[c]	1.80	
Tube										—[c]	2.38	
Total										2.26	2.10	2.18
Herpesvirus												
Indirect immunofluorescence											3.78	
Complement fixation											1.92	
Cytomegalovirus												
Indirect immunofluorescence											2.41	
Complement fixation											2.44	
Rickettsial antibodies												
Complement fixation										2.31		
Haptoglobin										1.32		
Ceruloplasmin										1.68		
Carcinoembryonic antigen								1.59		2.06		

[a] Average geometric standard deviation.
[b] Heterophile equivalent only.
[c] Not available.

TABLE 3. *Change in limits with change in interlaboratory precision*

Analyte	Selected value	Interlaboratory precision (S_G)	S_G^2 limits (95% limits)
Rubella	16	2.29	3–84
		1.86	5–55
Toxoplasma			
Indirect immunofluorescence	256	3.28	24–2,754
		2.37	45–1,438
Passive hemagglutination	256	2.97	29–2,258
		1.97	66–994
Heterophile	56	2.51	9–353
		2.06	13–238
Ox cell hemolysin	40	3.97	2–630
		1.95	10–152
Antinuclear antibodies	40	3.08	4–379
		2.30	8–212
Rheumatoid factor	80	3.10	8–769
		2.39	14–457
Antistreptolysin O	170	1.88	48–601
		1.47	79–367
Antideoxyribonuclease B	170	1.62	65–446
		1.44	82–352
Immunoglobulin G	1,000	1.23	661–1,513
		1.09	841–1,188
Immunoglobulin A	210	1.44	101–435
		1.13	164–268
Immunoglobulin M	150	1.35	82–273
		1.17	110–205
Complement C3 (β1A)	50	1.40	26–98
		1.18	36–70
Complement C3 (β1A/β1C)	100	1.28	61–164
		1.22	67–149
Complement C4	30	1.30	18–51
		1.26	19–48
Alpha-1-antitrypsin	300	1.29	180–499
		1.20	208–432
Brucella	160	2.69	22–1,158
		2.05	38–672
Tularemia	160	2.50	26–1,000
		1.88	45–566
Salmonella	160	2.58	24–1,065
		2.14	35–733
Weil-Felix	160	2.26	31–817
		2.10	36–706

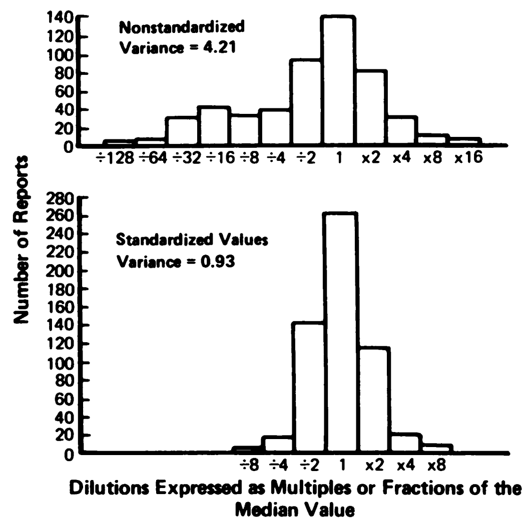

FIG. 1. *Distribution of rheumatoid factor test results.*

This produced significantly elevated titers compared with results obtained when cold water was used (26). Departure from the recommended method of mixing slide tests for multiple streptococcal antibodies was shown to be a factor in some erroneous results. Improvement in performance occurred after consultation with the reagent manufacturer and modification of the instructions for performing the test.

HBsAg results are divided according to the sensitivity category (generation) into which the test belongs. A first-generation test is one capable of detecting those HBsAg-positive sera designated "A" in the Food and Drug Administration, Bureau of Biologics, reference panel (2); a second-generation test detects sera designated "A" and "B"; and a third-generation test detects all HBsAg-positive sera in the panel ("A", "B," and "C"). Agar gel diffusion is a first-generation test; rheophoresis, counter-immunoelectrophoresis, complement fixation, and some reverse passive latex agglutination tests are second-generation tests; and radioimmunoassay, reverse passive hemagglutination, enzyme immunoassay, and some reverse passive latex agglutination tests are third-generation tests.

Samples containing low-reactivity levels of HBsAg may or may not be detected by second-generation tests. These kinds of samples are deliberately included in CDC hepatitis surveys to emphasize the importance of using third-generation tests by demonstrating their superior sensitivity over second-generation tests. The exclusive use of a second-generation test has been the most common reason for erroneous results.

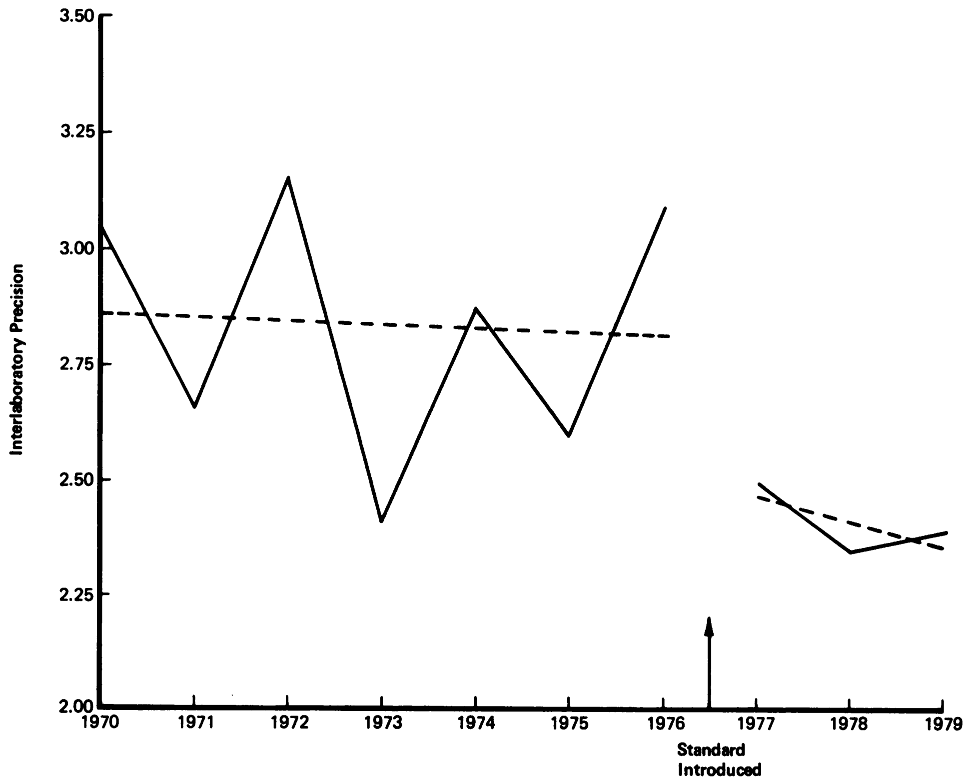

FIG. 2. *Interlaboratory comparability of rheumatoid factor results.*

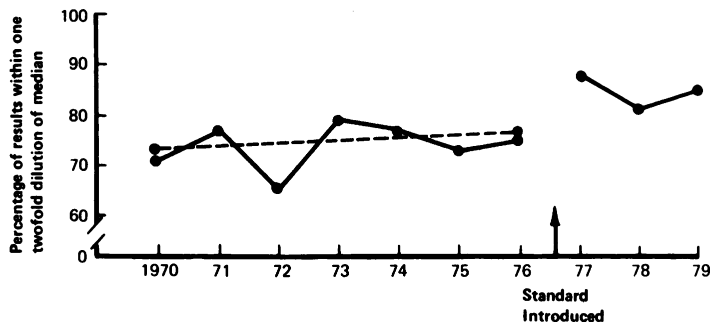

FIG. 3. *Performance in CDC diagnostic immunology testing for rheumatoid factor from 1970 to 1979.*

Table 7 shows the difference in the number and percentage of laboratories correctly detecting HBsAg in proficiency testing samples according to the sensitivity of the test used. Figure 4 and Table 8 show the changes that have occurred in the sensitivities and numbers of tests used by participants to detect HBsAg since the beginning of the CDC Proficiency Testing Program. In addition, Table 8 shows the changes in use of individual tests. At the beginning of the program, 52.1% of the participants were using first-generation tests, but the number diminished quickly as other tests became available, until less than 1% were using first-generation tests by 1979.

Second-generation tests were used by 97.9% of the participants in July 1971, and this level was maintained until April 1973, when the percentage started to decline. By 1979 it had dropped to less than 10%. Of the second-generation tests, the counterimmunoelectrophoresis test has always been the most popular. It was used by 79.8% of the participants in the beginning of our program, and that level was maintained or in-

TABLE 4. *Performance of laboratories in immunoglobulin quantitation*

Immunoglobulin	Results (%) outside acceptable range		
	1976	1977	1978
Immunoglobulin G	83 (219)[a]	37 (306)	11.5 (313)
Immunoglobulin A	59 (278)	42.5 (313)	20.5 (328)
Immunoglobulin M	59 (278)	31 (315)	23 (328)

[a] Number of participants is given in parentheses.

TABLE 5. *Performance of laboratories in quantitation of rubella hemagglutination inhibition antibody*

Yr	% of laboratories using kaolin	% of laboratories with results outside acceptable range		Avg width of acceptable range (2-fold dilutions)
		Kaolin method	Standard methods (8, 9)	
1971	59	11.9	3.4	2.4
1972	35	13.6	8.0	2.0
1973	65	52.0	29.8	1.0
1974	60	64.7	36.2	1.0
1975	54	7.8	6.7	3.0
1976	56	39.0	20.0	2.1
1977	44	59.7	31.9	1.0
1978	28	31.5	19.5	1.8
1979	22	54.7	34.8	1.0
Avg		37.2	21.1	1.7

creased until after the radioimmunoassay tests were licensed (July 1972) and became more prevalent. The complement fixation test has slowly declined in popularity, with 18.1% of the laboratories using it in 1971, but none are using it at present. Rheophoresis was first used by participants in our program in February 1972 but was never used by more that about 10% of the laboratories. The latex tests were first used by participants in October 1974, and the levels of use are still quite low. After their licensure, radioimmunoassay tests were quickly incorporated into the armamentarium of laboratory tests, and now they are used by about 90% of the laboratories. Essentially all of the laboratories in the CDC program now use at least one third-generation test. The Food and Drug Administration requirement for third-generation testing, which was proposed in July 1974, probably stimulated the use of radioimmunoassay by laboratories with the ability to do so and was undoubtedly instrumental in encouraging adoption of the reverse passive hemagglutination tests. At the present time, the reverse passive hemagglutination, enzyme-linked immunosor-

TABLE 6. *Syphilis serology test use by year*

Test[a]	% (no.) of laboratories using test														
	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Hinton	5.8 (4)	4.1 (3)													
Mazzini	8.7 (6)	8.1 (6)													
Kline	10.1 (7)	10.8 (8)													
Kolmer	30.4 (21)	27.0 (20)	17.3 (14)	14.3 (12)	7.9 (7)										
KRP	59.4 (41)	43.2 (32)	27.2 (22)	21.4 (18)	14.6 (13)										
VDRL slide	92.8 (64)	93.2 (69)	98.8 (80)	97.6 (82)	97.7 (87)	94.8 (92)	96.1 (99)	95.3 (102)	79.5 (236)	65.3 (169)	68.3 (222)	60.3 (205)	61.4 (223)	48.0 (237)	42.0 (232)
RPR card	13.0 (9)	16.2 (12)	14.8 (12)	16.7 (14)	21.3 (19)	27.8 (27)	31.1 (32)	36.4 (39)	31.3 (93)	50.5 (131)	50.2 (163)	50.6 (172)	60.3 (219)	67.4 (333)	70.9 (392)
USR	15.9 (11)	14.9 (11)	11.1 (9)	9.5 (8)	9.0 (8)	8.2 (8)	8.7 (9)	7.5 (8)	5.4 (16)	6.2 (16)	4.3 (14)	3.8 (13)	4.1 (15)	3.2 (16)	2.5 (14)
FTA-ABS		37.8 (28)	58.0 (47)	72.6 (61)	74.2 (66)	69.1 (67)	72.8 (75)	72.0 (77)	48.5 (144)	55.2 (143)	44.9 (146)	45.3 (154)	48.8 (177)	36.4 (180)	33.3 (184)
ART					7.9 (7)	13.4 (13)	20.4 (21)	18.7 (20)	14.8 (44)	14.7 (38)	11.4 (37)	8.8 (30)	11.0 (40)	7.5 (37)	6.0 (33)
AFTA							3.9 (4)	6.5 (7)	5.1 (15)	7.3 (19)	4.9 (16)				
MHA-TP										2.7 (7)	7.4 (24)	5.9 (20)	7.2 (26)	5.3 (26)	5.2 (29)
RST													3.6 (13)	2.0 (10)	1.8 (10)
Total no. of participants	69	74	81	84	89	97	103	107	297	259	325	340	363	494	553

[a] KRP, Kolmer Reiter protein; VDRL, Venereal Disease Research Laboratory; RPR, rapid plasma reagin; USR, unheated serum reagin; FTA-ABS, fluorescent treponemal antibody-absorption; ART, automated reagin test; AFTA, automated fluorescent treponemal antibody; MHA-TP, Microhemagglutination assay-*Treponema pallidum*; RST, reagin screen test.

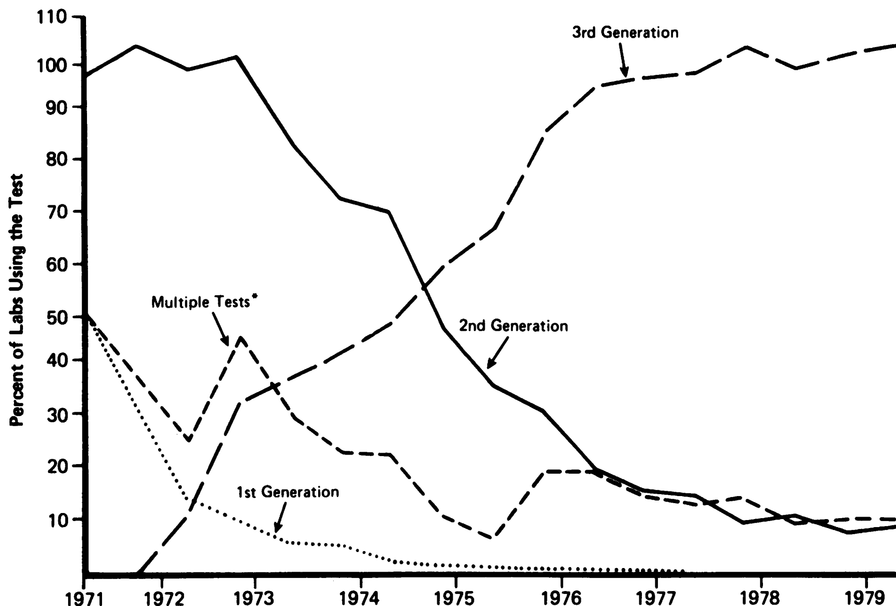

FIG. 4. *Tests used by participants in the CDC Proficiency Testing Program for HBsAg. *Multiple tests= number of tests used expressed as the percentage in excess of 100% of the number of laboratories: [(number of tests − number of laboratories)/number of laboratories] × 100.*

TABLE 7. *Sensitivity of second- and third-generation tests for detecting HBsAg-positive sera*

Yr	Results on Proficiency Testing samples			
	2nd-generation tests		3rd-generation tests	
	No. of laboratories	% Correct	No. of laboratories	% Correct
1975	114	79	206	99
1976	90	60	238	99
	53	62	251	95
1977	48	66	274	98
	48	80	273	99
1978	27	97	281	99
	33	79	276	96
1979	5	60	322	99
Avg % correct		72		98

bent assay, and some latex tests are the only alternatives for HBsAg testing available to laboratories that do not have radioimmunoassay capability.

Probably as a result of improvement in the third-generation tests, as well as increased use of these tests, the number of laboratories performing multiple tests has decreased. Initially, there were 141 tests used in 94 laboratories, or a 50% excess of tests over laboratories. As the number of first-generation tests decreased, so did the number of multiple tests, but when the radioimmunoassay tests were introduced, the number of multiple tests increased because the standard second-generation tests were retained until the laboratorians could develop some confidence in the new procedures. After April 1973, the number of multiple tests resumed its decline, until 1975 when the level stabilized at about 10% excess. Because of the greater sensitivity of the third-generation tests, there can be little doubt that the changes from first- and second-generation tests to third-generation tests represent significant laboratory improvement. The decrease in multiple testing is also an improvement in terms of cost, and although some of the participants used two third-generation tests, no significant differences were observed in their overall results compared with results from laboratories that used only one third-generation test.

A summary of the changes which have occurred in laboratory performance is shown in Table 9. Of the 29 analytes listed, there was evidence of improvement in 7 and no apparent change in 10. Of the 10 with no change, 4 had relatively high performance levels compared with other serological tests. For 12 of the analytes there were insufficient data to determine

TABLE 8. *Changes in tests used by participants in the CDC Proficiency Testing Program for HBsAg*

Date	% of participating laboratories using test[a]									
	1st-generation test (AGD)	2nd-generation tests					3rd-generation tests			Multiple testing[c]
		Rheo	CEP	CF	RPLA	Total[b]	RIA	RPHA	Total[b]	
July 1971	52.1	0.0	79.8	18.1	0.0	97.9	0.0	0.0	0.0	150.0
Feb. 1972	33.6	2.6	87.9	13.8	0.0	104.3	0.0	0.0	0.0	137.9
Oct. 1972	15.3	4.2	87.4	7.4	0.0	98.9	11.0	0.5	11.6	125.8
Apr. 1973	11.1	9.9	88.0	3.4	0.0	101.7	30.9	2.6	33.5	146.3
Oct. 1973	7.0	8.6	73.7	2.9	0.0	86.0	36.6	1.2	37.9	130.9
Apr. 1974	6.3	7.5	64.8	2.0	0.0	74.7	40.7	2.0	42.7	123.7
Oct. 1974	2.7	10.7	53.6	1.8	4.9	71.9	44.6	4.0	48.7	123.1
Apr. 1975	2.5	6.2	33.9	1.5	6.6	48.9	51.5	8.4	59.9	111.3
Oct. 1975	1.6	4.6	24.3	1.0	7.6	37.5	57.2	10.5	67.7	106.9
Apr. 1976	1.1	3.6	17.9	0.0	0.7	32.8	73.4	13.5	86.9	120.8
Sept. 1976	0.8	1.9	12.6	0.0	5.7	20.2	82.1	13.7	95.8	119.1
Apr. 1977	0.4	2.1	8.6	0.0	6.4	17.1	83.2	14.6	97.9	115.0
Aug. 1977	0.0	1.8	6.6	0.0	7.3	15.8	83.9	14.7	98.2	113.9
May 1978	0.7	1.5	5.6	0.0	3.0	10.0	88.1	15.2	104.1	114.8
Sept. 1978	0.0	—[d]	—	0.0	—	11.9	84.8	14.4	99.3	110.8
Apr. 1979	0.0	—	1.7	0.0	5.5	9.0	90.0	12.5	102.4	111.1
Aug. 1979	0.0	—	2.1	0.0	4.6	9.6	90.7	13.2	103.9	111.0

[a] AGD, Agar gel diffusion; Rheo, rheophoresis; CEP, counterimmunoelectrophoresis; CF, complement fixation; RPLA, reverse passive latex agglutination; RIA, radioimmunoassay; RPHA, reverse passive hemagglutination.

[b] Total includes other tests not specifically listed.

[c] [(Number of tests − number of laboratories)/number of laboratories] × 100.

[d] —, Not available.

whether changes had occurred. For most of the analytes, valuable information was derived from the proficiency testing program which permitted improvement in laboratory performance or could permit it in the future.

DISCUSSION

The syphilis serology evaluation survey which was conducted in 1934 was a predecessor of proficiency testing (29). The evaluation survey differs from a proficiency test in that the primary purpose of the former is to obtain statistical data concerning the sensitivity, specificity, and reproducibility of the tests, whereas the primary purpose of the latter is to measure the level of proficiency of the laboratories. In 1936 the Venereal Disease Research Laboratory developed a proficiency testing program in its efforts to assure good laboratory results in syphilis serology. Another evaluation survey of serological tests for syphilis was completed just before World War II. After World War II, interest was again aroused in measuring the proficiency of laboratories and the reliability of tests. In 1947 and 1948, the first College of American Pathologists survey was conducted but the results were not published. The first formal proficiency testing program was provided by the College of American Pathologists in 1949. In that same year, the American Association of Bioanalysis also provided their first proficiency testing program. These early evaluation and proficiency testing surveys emphasized the need for better training and education for laboratorians, better standard reference materials and methods, and continuous surveillance.

The classic (and probably the last) evaluation survey was conducted in 1956–1957 by CDC (29). In this study, 1,298 sera were collected from individual donors in various disease categories, and each serum was then tested by 38 test procedures in 20 laboratories. The results indicated the sensitivity, specificity, and reproducibility of each test. Because of the logistical problems, cost, and other limitations, it is unlikely that this type of evaluation survey will be conducted in the future, but proficiency testing can provide similar data (15).

In 1967, the Clinical Laboratories Improvement Act, which required that all laboratories involved in interstate testing be licensed, was passed. One of the licensure requirements is satisfactory performance in a proficiency testing service. To provide this testing, CDC expanded the Laboratory Improvement program, which had previously been assigned to the Bureau of Laboratories.

Since that time, the CDC Proficiency Testing

TABLE 9. *Changes in laboratory proficiency as measured by the CDC program*

Analyte	No. of yr in program	Status	Comments
Rubella hemagglutination inhibition	11	Improvement but more needed	Shift to standard serum treatments. Method differences still cause most interlaboratory variations.
Antistreptolysin O	10	No change evident	Interlaboratory comparability better than most serological tests. Adherence to published procedures could improve results further.
Antideoxyribonuclease B	4	No change evident	Interlaboratory comparability better than most serological tests. Adherence to published procedures could improve results further.
Multiple streptococcal enzymes	7	No change evident	Interlaboratory comparability better than most serological tests. Adherence to published procedures could improve results further.
Rheumatoid factor	10	Improvement	Resulted from introduction of reference material.
C-reactive protein	5	No change evident	Methods differ substantially in sensitivity. Need to be standardized.
Infectious mononucleosis	11	No change evident	Quantitative results of slide tests and ox cell hemolysin test need to be standardized.
Antinuclear antibodies	5	No change evident	Substrate differences cause most interlaboratory differences. Adoption of standard substrate or use of a standard serum would probably result in substantial improvement.
Salmonella agglutination	11	No change evident	Standard antigen and antiserum needed.
Brucella agglutination	11	No change evident	Use of available standards needs to be encouraged.
Tularemia agglutination	11	Improvement	Improved interlaboratory comparability with the tube. Standard antigen and antiserum are still needed.
Rickettsial antibodies	2	Unknown	Inadequate data
Toxoplasma antibodies	6	No change evident	
Immunoglobulins G, A, and M	7	Improvement	Interlaboratory comparability improved. Some standardization problems. Manufacturer's secondary standards not comparable.
Immunoglobulin D	2	Unknown	Inadequate data
Immunoglobulin E	2	Unknown	Inadequate data
Complement (C3, C4)	2	Unknown	Inadequate data
Alpha-1-antitrypsin	2	Unknown	Inadequate data

TABLE 9—*Continued*

Analyte	No. of yr in program	Status	Comments
Alpha-2-macroglobulin	2	Unknown	Inadequate data
Haptoglobin	2	Unknown	Inadequate data
Ceruloplasmin	2	Unknown	Inadequate data
Transferrin	2	Unknown	Inadequate data
Carcinoembryonic antigen	4	Unknown	Inadequate data
Mycotic serology	1	Unknown	Inadequate data
Syphilis serology	30 (?)	No change	High proficiency level maintained.
Hepatitis B	9	Improvement	Shift to almost exclusively third-generation tests and decrease in the frequency of multiple testing.

Program and state and private programs have grown substantially not only in numbers of participants, but also in the extent and complexity of the programs. For example, in 1969 there were 89 laboratories in CDC's syphilis serology program and 152 laboratories in the nonsyphilis serology program. In 1979, there were over 500 laboratories in the syphilis program, about 375 in the hepatitis program, and well over 1,000 in the immunology program. Also, the number of analytes in the program increased from 5 in 1969 to 39 at the present time (Table 1).

The Diagnostic Immunology portion of the CDC Proficiency Testing Program includes a wide variety of immunological tests. The tests that are included depend on (i) the number of laboratories that provide the tests, (ii) the number of tests performed per year, (iii) the importance of the test in diagnosis or treatment of disease, (iv) the degree of difficulty in obtaining useful results, and (v) the availability of resources.

Proficiency testing programs can be used to stimulate improved performance by means other than strict evaluation. Data obtained through surveys can be used to estimate results that might be expected with patient specimens in situations where reactivity levels are comparable to the samples of the survey. Documentation of the improvement in comparability of results obtained with standardized methods or reagents can be used by laboratories and manufacturers to increase the usefulness of serological test data.

Standardization in serological tests is encouraged by the Proficiency Testing Program. Reference laboratories are selected from laboratories which use standard methods when they are available, and participants are urged to consider adopting them for use in their laboratories. Use of standards or reference materials is also recommended when they are available. The results from laboratories using standardized methods or reference materials are generally more comparable than results from other laboratories.

It appears from analysis of data in this report that proficiency testing per se did not result in a significant general improvement, but substantial improvement resulted in those areas given special consideration. In many cases the reason for poor performance is beyond the control of the laboratorian and sometimes is even beyond the control of the reagent manufacturers. For example, when there is lack of standardization, laboratorians and manufacturers can do very little to improve interlaboratory comparability. In these areas, professional consensus needs to be obtained on reference methods and materials.

In some areas it appears that even though there was no evidence of improvement, the current performance level may satisfy medical requirements. Since no such limits have been clearly specified, we do not know if these needs have been met. Again, professional consensus needs to be obtained on performance levels required for medical application of the data from immunologic tests.

Laboratories other than those directly involved in the Proficiency Testing Program have also benefited from the program. Some proficiency testing efforts may have resulted in improved laboratory performance that is not detectable by conventional evaluation methods. For example, a quality control monograph was distributed to try to stimulate the establishment of better quality control procedures in laboratories and thereby eliminate some aberrant re-

sults on patient samples (28). The publication of a comparison between the kinetic and endpoint methods for serum protein quantitation by radial immunodiffusion may have resulted in change to quicker methods (18). Information about how laboratories are performing with commercial products has been made available to manufacturers, and improvements have been made in some products and package inserts. The total impact of proficiency testing surveys has not yet been measured.

LITERATURE CITED

1. **Cavallaro, J. J., D. F. Palmer, and B. E. Bigazzi.** 1976. Immunofluorescence detection of autoimmune disease. Immunol. Ser. no. 7. Center for Disease Control, Atlanta.
2. **Food and Drug Administration.** 1975. Test for hepatitis B surface antigen. Fed. Regist. **40**:29706–29712.
3. **Fulford, K. M., R. N. Taylor, and V. A. Przbyszewski.** 1978. Reference preparation to standardize results of serological test for rheumatoid factor. J. Clin. Microbiol. **7**:434–441.
4. **Griffin, C. W., and R. N. Taylor.** 1979. Proficiency testing—is it worthwhile? (editorial) Clin. Microbiol. Newsl. **1**:4-5.
5. **Kaplan, E. J., B. F. Anthony, S. S. Chapman, E. M. Ayoub, and L. W. Wannamaker.** 1970. The influence of the site of infection on the immune response to group A streptococci. J. Clin. Invest. **49**:1405–1414.
6. **Klein, G. C.** 1976. Immune response to streptococcal infection, p. 264-273. *In* N. R. Rose and H. Friedman (ed.), Manual of clinical immunology. American Society for Microbiology, Washington, D.C.
7. **Klein, G. C., and W. L. Jones.** 1971. Comparison of streptozyme test with the antistreptolysin O, antideoxyribonuclease B, and antihyaluronidase tests. Appl. Microbiol. **2**:157–259.
8. **Palmer, D. F., K. L. Herrmann, and J. J. Cavallaro.** 1974. Modified rubella hemagglutination-inhibition test, p. 57–83. *In* D. F. Palmer, J. J. Cavallaro, K. Herrmann, J. A. Stewart, and K. W. Walls (ed.), Serodiagnosis of: toxoplasmosis, rubella, cytomegalic inclusion disease, and herpes simplex. Center for Disease Control, Atlanta.
9. **Palmer, D. F., K. L. Herrmann, R. E. Lincoln, M. V. Hearn, and J. M. Fuller.** 1970. Standardized rubella hemagglutination-inhibition tests. Immunol. Ser. no. 2. Procedural guide. Center for Disease Control, Atlanta.
10. **Pope, V., S. A. Larsen, and R. N. Taylor.** 1979. Proficiency testing summary analysis. Syphilis serology III. Center for Disease Control, Atlanta.
11. **Pope, V., S. A. Larsen, and R. N. Taylor.** 1980. Proficiency testing summary analysis. Syphilis serology IV. Center for Disease Control, Atlanta.
12. **Pope, V., S. A. Larsen, R. N. Taylor, and J. C. Feeley.** 1979. Proficiency testing summary analysis. Syphilis serology I. Center for Disease Control, Atlanta.
13. **Pope, V., S. A. Larsen, R. N. Taylor, and J. C. Feeley.** 1979. Proficiency testing summary analysis. Syphilis serology II. Center for Disease Control, Atlanta.
14. **Pope, V., R. N. Taylor, S. A. Larsen, and J. C. Feeley.** 1979. Proficiency testing summary analysis. WHO and PAHO laboratories. Syphilis serology 1978. Center for Disease Control, Atlanta.
15. **Taylor, R. N., H. Ehrhard, and S. Marcus.** 1976. Evaluation of tests for infectious mononucleosis through proficiency testing. Health Lab. Sci. **13**:34–44.
16. **Taylor, R. N., K. M. Fulford, and A. Y. Huong.** 1978. Comparison of kinetic and end-point diffusion methods for quantitating human serum immunoglobulins. J. Clin. Microbiol. **8**:23–27.
17. **Taylor, R. N., K. M. Fulford, and W. L. Jones.** 1977. Reduction of variation in results of rheumatoid factor tests by use of a serum reference preparation. J. Clin. Microbiol. **5**:42–45.
18. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1979. Proficiency testing summary analysis. Diagnostic immunology 1979. I. Hepatitis B. Center for Disease Control, Atlanta.
19. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1979. Proficiency testing summary analysis. Diagnostic immunology 1979. II. Hepatitis B. Center for Disease Control, Atlanta.
20. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1979. Proficiency testing summary analysis. Diagnostic immunology 1977. I. Center for Disease Control, Atlanta.
21. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1979. Proficiency testing summary analysis. Diagnostic immunology 1979. II. Center for Disease Control, Atlanta.
22. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1979. Proficiency testing summary analysis. Diagnostic immunology 1979. III. Center for Disease Control, Atlanta.
23. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1979. Proficiency testing critique. Diagnostic immunology 1978. Center for Disease Control, Atlanta.
24. **Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski.** 1980. Proficiency testing critique. Diagnostic immunology 1979. Center for Disease Control, Atlanta.
25. **Taylor, R. N., K. M. Fulford, V. A. Przybyszewski, and V. Pope.** 1977. Center for Disease Control Diagnostic Immunology Proficiency Testing Program results for 1976. J. Clin. Microbiol. **6**:224–232.
26. **Taylor, R. N., K. M. Fulford, V. A. Przybyszewski, and V. Pope.** 1978. Center for Disease Control Diagnostic Immunology Proficiency Testing Program results for 1977. J. Clin. Microbiol. **8**:388–395.
27. **Taylor, R. N., K. M. Fulford, V. A. Pryzybyszewski, and V. Pope.** 1979. Center for Disease Control Diagnostic Immunology Proficiency Testing Program results for 1978. J. Clin. Microbiol. **10**:805–814.
28. **Taylor, R. N., A. Y. Huong, K. M. Fulford, V. A. Przybyszewski, and T. L. Hearn.** 1979. Quality control for immunologic tests. DHEW Publ. no. (CDC) 79-8376. Center for Disease Control, Atlanta.
29. **U.S. Department of Health, Education, and Welfare.** (undated) Serology evaluation and research assembly 1956–1957. Public Health Service Publ. no. 650. Bureau of State Services, Communicable Disease Center, Atlanta.