Reference Preparation to Standardize Results of Serological Tests for Rheumatoid Factor

KAREN M. FULFORD, ROGER N. TAYLOR,* AND VINCENT A. PRZYBYSZEWSKI

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

Received for publication 22 February 1978

The variance of between-laboratory results for rheumatoid factor tests was reduced through the use of a reference serum preparation. Two separate proficiency testing surveys have shown reduction of 58 to 81% in variance with samples of differing levels and composition. Reduction of variance was seen in data from laboratories in which the same methods and reagents were used, but greater reduction in variance was seen in results from laboratories using different methods and reagents. This study demonstrates that comparability of results of rheumatoid factor tests can be significantly improved through the use of a standard reference serum preparation and that confidence in such results can therefore be increased. A serum rheumatoid factor level of approximately 50 IU/ml is equivalent to a titer of 160 in the Singer-Plotz latex test, which has traditionally been accepted as evidence of rheumatoid arthritis.

Proficiency testing can be used to derive evaluation survey data (4) and to estimate the amount of improvement in interlaboratory comparability that can be expected if laboratories adopt a common reference preparation for a serological test (5). This study was designed to demonstrate the actual change in comparability of laboratory results effected by including a serum reference preparation for rheumatoid factor (RF) in proficiency testing surveys.

MATERIALS AND METHODS

Participants in the diagnostic immunology proficiency testing program of the Center for Disease Control (CDC) were provided with a serum that had been compared in triplicate determinations with the international reference preparation of rheumatoid arthritis serum (1, 2, 5) and had been found to contain 250 IU/ml. The standardized serum was shipped with the January 1977 survey (6), which included two samples (B17-A07 and B17-A08) to be tested for RF. For this survey (1977I), laboratories were asked to report results in international units per milliliter rather than as titers. International units per milliliter were to be calculated by dividing the titer obtained for a proficiency testing sample by the titer obtained for the standardized serum and multiplying by 250 (international units per milliliter in the standardized serum).

A second survey (1977S), containing 10 samples for RF testing and the standardized sample, was provided to selected special study laboratories in September 1977. Participants were asked to report results as both titers and international units per milliliter.

Human sera for the samples were obtained from commercial suppliers on government contract or from noncommercial sources. The sera containing RF and

the negative human sera used to prepare dilutions were tested for the presence of hepatitis B surface antigen by the CDC's Proficiency Testing Branch, using a radioimmunoassay procedure, and were found to be negative.

The 1977S survey included two sets of duplicate samples (B17-S04 and B17-S07; B17-S05 and B17-S09) and several sets of paired samples prepared from the same positive serum. B17-S05 and B17-S09 were dilutions of B17-S03; B17-S08 was a dilution of the serum used for B17-S04 and B17-S07; B17-S10 was a dilution of B17-S06. B17-S01 was a serum that has been used extensively in RF studies. Marion Waller of the Medical College of Virginia, Richmond (8, 9), provided this serum. It gives consistent positive test results for RF but comes from a donor with no other signs of rheumatoid arthritis.

Sera were filtered to remove any bacterial contaminants and aseptically dispensed in 1-ml portions into 6-ml bottles. Bottles were stoppered, capped, labeled and packed in styrofoam containers, and sent by first-class mail.

Adequacy of the samples as proficiency testing specimens was confirmed independently by the Proficiency Testing Branch and the Bacterial Immunology Branch at CDC and by 15 reference laboratories selected on the basis of demonstrated competence with RF tests on previous proficiency testing surveys. Quality-control tests were performed to establish and verify titers, measure within-sample variation, and determine sample stability and sterility.

Results were analyzed by using previously described statistical calculations and notations (7) and were compared with the previously published (5) estimates of the type of improvement that could be expected. Variance was calculated by transforming the data to logarithms to base 2.

A few (less than 8%) of the reported results from

the 1977 survey were excluded from some of the analyses because they were obviously reported as titers. Analyses that included these incorrectly reported titers are indicated. Several laboratories in the 1977S survey reported only titers or international units per milliliter. No conversions were made for these results.

RESULTS

Figure 1 and Table 1 show the distribution of results and variances for the raw, unstandardized titers and the externally standardized results that were suggested as estimates of the kind of improvement in comparability that could

be expected if the results were internally standardized. These results are compared with results actually obtained when the laboratories reported test results in terms of international units per milliliter. It was predicted that the total variance could be reduced by 77%, from 4.21 to 0.97. The actual variance was reduced by 78% to 0.93 in the 1977I survey and by 58 to 81% in the 1977S survey. Figure 2 shows the distribution of results for each type of test on each of the samples that were combined to produce the distribution in the bottom histogram in Fig. 1. Figure 3 shows the distribution of results from

FIG. 1. Distribution of RF test results: nonstandardized, predicted standardized, and actual standardized.

TABLE 1. Comparison of variances ($SD_{\log_2}$)² for RF tests with raw and standardized results

Study	Variance				
	Nonstandardized re- sults	Standardized results		Percent reduction	
		Predicted ^a	Actual	Predicted ^a	Actual
1977I	4.21 ^b	0.97	0.93	77	78
1977S					
Mean ^c	6.23	0.97	2.01	77	68
Range ^c	4.96–8.11		1.07–2.45		58–81

^a Values predicted from initial study (5).
^b Variance from 1976 study (5).
^c Mean and range derived from nine positive samples.

FIG. 2. Distribution of standardized RF test results.

the 1977S survey by comparing results in titers, international units per milliliter, and reproducibility for the two sets of duplicate samples. Table 2 shows the effect of standardization in terms of the percentage of results equal to the median result or within one or two tubes of the median result. Substantial improvements were observed

in all three surveys, but the results were not as dramatic when the survey included RFs from a wider variety of sources.

Differences in results due to reagents were reduced by standardization, but not as much improvement was seen in the 1977S survey as had been predicted by the 1976 survey. The

FIG. 3. Comparison of results in titer, international units per milliliter, and reproducibility.

TABLE 2. Improvement in comparability of RF test results due to standardization

Results	Study ^a				
	1976		1977I	1977S	
	Unstd	Std	Std	Unstd	Std
Median	27.1 ^b	51.0	46.0	17.7 (11.7-31.6) ^c	35.0 (22.0-44.4)
Median ± 1 dilution	61.3	91.0	90.3	47.7 (36.8-50.0)	76.5 (68.2-92.5)
Median ± 2 dilutions	75.1	97.6	98.0	68.8 (62.3-80.3)	92.9 (86.4-96.2)

^a Unstd, Unstandardized results; Std, standardized results.
^b Percentage of results within the range for all samples in survey.
^c Range of percentages of laboratories within the range for nine samples.

ratios of the highest geometric mean ($\bar{x}_G$) by manufacturer for the 1977S survey are given in Table 3. Previously, the corresponding ratios were 10 for the results in titers, and 1.48 to 2 IU/ml for results thus presented. Still, the reduction was 47 to 91%.

The wider distribution seen for the sample results in the 1977S survey may be related to differences in distribution of tests the participants used. In routine shipments, most participants use reagents from one major supplier,

whereas for the special shipment, participants were selected to give a more even distribution of reagents. As a consequence, the distributions of results for the 1977S survey were more obviously bimodal than for previous surveys. The distribution of results by manufacturer (Fig. 4) shows that, in fact, the reagents can be divided into two distinct groups.

Another difference in the 1977S survey was the sera used for samples. In the earlier studies, the same positive serum was diluted for all the

TABLE 3. *Reduction of between-manufacturer bias by standardization for the 1977S Survey*

Sample no.	Titer ($\bar{x}_G$)			IU ($\bar{x}_G$)			% Reduction
	High	Low	Ratio	High	Low	Ratio	
BI7-S01	3,547	238	15	3,084	410	8	47
BI7-S03	5,120	195	26	4,362	353	12	54
BI7-S04	2,032	40	51	1,834	125	15	71
BI7-S06	10,240	905	11	8,724	1,714	5	55
BI7-S07	2,032	50	41	1,834	136	14	66
BI7-S10	557	17	33	385	115	3	91

FIG. 4. *Distribution of results by manufacturer.* ▨, group 1: R3, Rheumaton; □, group 2: BCA, Behring, Difco, Hyland.

specimens, including the standardized sample. For the 1977S survey, several different positive sera were used. The ratios of the $\bar{x}_G$'s obtained with different reagents were not constant for all sera, even in international units per milliliter (Table 4). For example, in comparing Hyland and BCA results, Hyland results were higher on samples B17-S01, B17-S03, B17-S05, and B17-S09, which were prepared from different positive sera. There seem to be differences in reactivity between the various antigens and RFs in different patients. Since the standardized sample was prepared from a single positive serum, some bias is still present in the standardized results. A reference preparation composed of a large number of RF-positive sera might eliminate much of the remaining bias.

Samples B17-S05, B17-S08, and B17-S09 were positive in most tube tests, but were frequently negative in slide tests. Qualitative slide test results for the 1977S survey are given in Table 5. A number of laboratories screen sera with a slide test and titrate only the positive reactors; in so doing, they are missing positive reactions that are too weak to give a positive result in the slide test.

Standardized results can be used to determine levels of sensitivity, such as minimum detectable units (MDU) for quantitative tests or mean effective doses (ED_{50}) for qualitative tests. The approximate ED_{50} 's (determined from log probit graphs) for the tests used in the 1977S survey are given in Table 6, along with the K value determined for the quantitative tests as the ratio of the $\bar{x}_G$ international units per milliliter to $\bar{x}_G$ titer. The K value is defined as the ratio of international units per milliliter to the titer or, in other words, the number of international units detectable per milliliter in undiluted serum. For tests that call for a dilution of the serum to be tested, the K value times the dilution factor is the MDU.

A serum RF level of 50 IU/ml might be considered a positive serological indication of rheumatoid arthritis, since it is approximately equivalent to the traditionally accepted significant titer of 160 in the Singer-Plotz latex test.

DISCUSSION

Several investigators have shown the need to standardize tests for RF (1-3, 8). Singer (3) has thoroughly discussed the problems associated with standardizing these tests. Because of the heterogeneity of the test reagents and commercial products available for RF testing, the use of a standard reference preparation seems to be the logical approach to standardizing RF tests.

From our previous work, we determined that

TABLE 4. Quantitative RF results (international units per milliliter) by reagent manufacturer

Sample No.	Reagent manufacturer												Total			
	BCA		Behring		Hyland		Difco		R3		Rheumaton			Other		
	$\bar{x}_G$	$SD_{G'}^a$	$\bar{x}_G$	SD_G	$\bar{x}_G$	SD_G	$\bar{x}_G$	SD_G	$\bar{x}_G$	SD_G	$\bar{x}_G$	SD_G		$\bar{x}_G$	SD_G	
B17-S01	1,188	4.77	3,084	3.59	1,277	2.08	707	1.80	410	1.63	2,295	1.31	1,616	2.59	1,314	2.79
B17-S03	917	3.31	4,362	2.74	1,042	1.77	536	1.66	353	1.68	852	1.35	1,896	3.93	1,075	2.96
B17-S04	583	2.30	1,834	3.12	500	1.58	329	1.62	125	1.90	267	1.40	556	2.75	438	2.61
B17-S05	52	1.94	186	2.85	88	1.74	69	2.30	N ^b		N		111	3.41	87	2.92
B17-S06	4,000	2.85	8,724	1.98	3,687	2.57	1,714	2.46	4,489	2.24	6,128	1.42	5,222	2.83	4,325	2.48
B17-S07	630	1.99	1,834	3.50	500	1.67	250	1.58	136	1.78	344	1.43	726	1.83	474	2.54
B17-S08	99	4.25	226	3.64	77	1.60	51	2.15	N		N		92	2.96	89	2.90
B17-S09	44	3.30	164	1.85	66	1.93	51	2.58	N		N		125	2.13	72	2.74
B17-S10	157	2.34	385	3.02	224	1.69	207	1.56	205	1.68	115	2.26	213	1.65	197	2.05
No. of labs	9		8		19		10		8		13		15		77	

^a SD_G , Geometric standard deviation.
^b N, Negative.

TABLE 5. Results of qualitative slide tests for RF

Reagent source	No. of labs	% Positive results reported for samples									
		BI7-S01	BI7-S02	BI7-S03	BI7-S04	BI7-S05	BI7-S06	BI7-S07	BI7-S08	BI7-S09	BI7-S10
BCA	9	100	0	100	100	44	100	100	33	44	100
Behring	12	100	0	100	100	67	100	100	58	58	100
Hyland	19	100	5	100	100	53	100	100	37	53	100
Organon	3	100	0	100	100	0	100	100	0	0	0
Rheumaton	20	100	0	100	100	15	100	100	15	20	95
HWD	2	100	0	100	100	100	100	100	100	100	100
R3	2	100	0	100	100	0	100	100	0	0	100
Other	4	100	0	100	100	50	100	100	75	100	100
Total	71	100	1	100	100	41	100	100	35	44	94
$\bar{x}_G$ Titer		1,063	Neg ^a	988	380	106	3,262	401	94	83	180
$\bar{x}_G$ IU/ml		1,314	Neg	1,075	438	87	4,325	474	89	72	197

^a Neg, Negative.

TABLE 6. Performance characteristics of RF tests by manufacturer

Reagent manufacturer	Qualitative tests	Quantitative tests	
	Approx ED ₅₀ ^a (IU/ml)	K value ^b	MDU ^c
BCA	95	0.69 ± 0.20	13.8
Behring	65	0.88 ± 0.13	17.6
Hyland	85	0.40 ± 0.03	8.0
Rheumaton	120	7.01 ± 0.64	7.01
R3	ND ^d	1.74 ± 0.03	34.8
Difco	ND	0.34 ± 0.09	6.8

^a Approximate level for 50% of tests to be positive.^b Ratio of international units per milliliter to titer ± 1 standard deviation.^c Minimal international units per milliliter detectable at recommended first dilution.^d ND, Not determined.

standardizing results of serological tests for RF by comparison with a standard reference preparation potentially could reduce the total variance in laboratory results by about 75%, that is, to a point where the within-laboratory and between-laboratory components of variance were approximately equal. The present study verifies this reduction for proficiency test results. The total variation was reduced 50 to 81% by standardization.

If laboratories used this standardization procedure for patient specimens, the variation in results might likewise be reduced, and results obtained in different laboratories could then be compared with more confidence. In addition, the disparity of results obtained with the various commercial slide and tube tests would be greatly reduced.

It has been argued that the improved perform-

ance observed in the proficiency testing results may not have been caused by standardization of the results but could have resulted from change in composition of program participants; i.e., performance improved as poor laboratories dropped out of the program and better laboratories entered it. In the special survey (1977S), however, the same laboratories reported both the raw titers and their calculated equivalent in international units per milliliter. The reduction in variation in this case was similar to that in the other comparisons; therefore, it seems reasonable to conclude that standardization was the reason for the improvement and that turnover of laboratories did not contribute substantially, if at all, to the observed improvement.

The improvement in laboratory performance documented in this paper was possible because systematic error (bias) was responsible for a major portion of the between-laboratory variability and because reduction of the bias resulted in better comparability of data without otherwise altering the procedures the laboratories used. Not all bias was eliminated, and further studies are needed to determine the reasons for the bias and to develop steps to minimize it. Antigenic or procedural differences and differences in the reactivity of RFs may be involved. The preparation of reagents or of the standard may have to be modified to remove more of the bias. Pooling a large number of positive sera for the standard may increase its usefulness.

A standard reference preparation may also be of value in comparing, evaluating, or controlling commercial kits, reagents, or procedures. Determination of the ED₅₀ for qualitative tests and the MDU for quantitative tests would give a measure of their relative sensitivities.

This study again demonstrates that profi-

ciency testing programs can achieve their ultimate goal of laboratory improvement by means other than the commonly accepted evaluative methods. CDC is actively engaged in ongoing studies to increase the effectiveness and efficiency of its proficiency testing programs. The potential of proficiency testing is just beginning to be appreciated.

LITERATURE CITED

1. Anderson, S. G., M. W. Bentzon, V. Houba, and P. Krag. 1970. International reference preparation of rheumatoid arthritis serum. *Bull. WHO* 42:311-318.
2. Jones, W. L., and G. L. Wiggins. 1973. A study of rheumatoid arthritis latex kits. *Am. J. Clin. Pathol.* 60:703-706.
3. Singer, J. M. 1974. Standardization of the latex test for rheumatoid arthritis serology. *Bull. Rheum. Dis.* 24:762-769.
4. Taylor, R. N., H. Ehrhard, and S. Marcus. 1976. Evaluation of tests for infectious mononucleosis through proficiency testing. *Health Lab. Sci.* 13:34-44.
5. 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.
6. Taylor, R. N., K. M. Fulford, and V. A. Przybyszewski. 1977. Proficiency testing summary analysis—diagnostic immunology 1977. I. Rubella, infectious mononucleosis, tularemia, rheumatoid factor, immunoglobulin quantitation. U. S. Department of Health, Education, and Welfare, Center for Disease Control, Atlanta, Ga.
7. Taylor, R. N., A. Y. Huong, and K. M. Fulford. 1975. Measurement of variation in serologic tests. U. S. Department of Health, Education, and Welfare, Center for Disease Control, Atlanta, Ga.
8. Waller, M. 1971. Present status of the rheumatoid factor. *CRC Crit. Rev. Clin. Lab. Sci.* 2:173-209.
9. Waller, M., and E. C. Toone. 1968. Normal individuals with positive tests for rheumatoid factor. *Arthritis Rheum.* 11:50-55.