JOURNAL OF CLINICAL MICROBIOLOGY, Jan. 1977, p. 42-45
Copyright © 1977 American Society for Microbiology
Vol. 5, No. 1
Printed in U.S.A.

# Reduction of Variation in Results of Rheumatoid Factor Tests by Use of a Serum Reference Preparation

ROGER N. TAYLOR,* KAREN M. FULFORD, AND WALLIS L. JONES
*Center for Disease Control, Atlanta, Georgia 30333*

Received for publication 9 September 1976

Standardizing test results for rheumatoid factor by comparing results obtained for an unknown with results obtained for a serum reference preparation decreased variance between laboratories, as measured in the Center for Disease Control proficiency testing program, by 77%. The amount of improvement was also estimated by the type of test and by the manufacturer's product. Standardization resulted in an increase in the number of reported results that were within a twofold dilution of the median value. The percentage increased from 50.3 to 93.7% for the slide tests and from 78.1 to 91.2% for the tube tests. Decrease in variance by manufacturer's product ranged from 94 to 27%. The study demonstrated that adopting a reference serum standard could substantially improve the comparability of rheumatoid factor test results and that proficiency testing programs can be used to estimate improvement which could be expected as a result of standardization.

It was previously shown that proficiency testing can be used to derive evaluation survey data (3). This study was designed to demonstrate that proficiency testing can also be used to estimate the amount of improvement in interlaboratory variability that can be expected if laboratories adopt a common reference preparation for a serological test. Because previous proficiency testing surveys have shown poor comparability between laboratories with the tests for rheumatoid factor (4, 5, 6) and because the laboratory division provisional reference preparation for rheumatoid arthritis (1, 2) was available at the Center for Disease Control, the rheumatoid factor tests were selected for this study.

Communication between laboratories and effective, cooperative research require that results from studies in different laboratories be comparable. This comparability is easier to achieve if a serum standard is adopted.

## MATERIALS AND METHODS

Three serum specimens were prepared for this study. They were included in the regular proficiency testing survey specimens mailed to over 500 participants in the diagnostic immunology proficiency testing program of the Center for Disease Control.

Human sera were obtained from commercial suppliers on government contract. Both the Proficiency Testing Branch of the Licensure and Proficiency Testing Division and the Bacterial Immunology Branch of the Bacteriology Division tested the sera for acceptability. The sera containing rheumatoid factor and the negative human sera used to prepare dilutions were tested for the presence of hepatitis B surface antigen by the Proficiency Testing Branch, using a radioimmunoassay procedure. They were found to be negative.

Two specimens, S6-013 and S6-014, were duplicates prepared from the same serum pool. Duplicates were used to permit measurement of within-laboratory variation. Specimen S6-015 was prepared to have a titer that was twofold higher than the other two specimens and was used as a secondary standard reference serum in this study.

This specimen was compared in triplicate determinations with the international reference preparation of rheumatoid arthritis serum (1, 2) and was determined to contain 250 IU/ml. Specimens S6-013 and S6-014 were similarly tested, and the expected result of 125 IU/ml was obtained on each.

Sera were sterilized by filtration and aseptically dispensed in 1-ml portions into 6-ml bottles. Bottles were stoppered, capped, labeled, and packed in styrofoam containers to be sent by first-class mail with the routine proficiency testing survey samples. The shipment was sent to participants on 5 May 1976.

Adequacy of the samples as proficiency testing specimens was confirmed independently by the Proficiency Testing Branch, Bacterial Immunology Branch, and 11 reference laboratories that had been selected on the basis of demonstrated competence with rheumatoid factor tests on previous proficiency testing surveys. Tests were performed to establish and verify titers, measure within-sample variation, and determine sample stability and sterility.

The raw titers reported by the participants were converted by computer to international units. This was done by dividing the result obtained on specimen S6-013 (or S6-014) by the result obtained on specimen S6-015 and multiplying the quotient by 250 (the number of international units per milliliter in

this specimen). The distribution of results before and after standardization were then compared.

Within-laboratory reproducibility was measured by dividing the result reported on specimen S6-013 by the result reported on specimen S6-014 and tabulating the resulting ratios.

Statistical calculations were made by using previously described methods and notations (7). Variance was calculated by transforming the data to logarithms to base 2.

## RESULTS

Figure 1 and Table 1 show the distribution of results and variances for the raw, unstandardized titers, the standardized results, and the within-laboratory precision. The last category reflects the distribution of results obtained by dividing the titer obtained on specimen S6-013 by the titer obtained on specimen S6-014. These graphs depict the effect of the elimination of bias in rheumatoid factor test results obtained by use of a common standard serum in each laboratory. Because systematic error (bias) was responsible for a major portion of the between-laboratory variability, its elimination resulted in better comparability of data between laboratories without otherwise altering the procedures used by the laboratories.

Total variance $(SD_{log_2})^2$ for the unstandardized titers was 4.21. This was reduced by 77% to 0.97 simply by comparing results to a reference standard. The between-laboratory variance (total variance minus within-laboratory variance) was reduced by 88.0% from 3.68 to 0.44 by this standardization procedure. As Table 1 shows, before standardization the between-laboratory component of variance produced 87% of the total variance (3.68 out of 4.21), but after standardization it contributed only 45% of the total variance (0.44 out of 0.97).

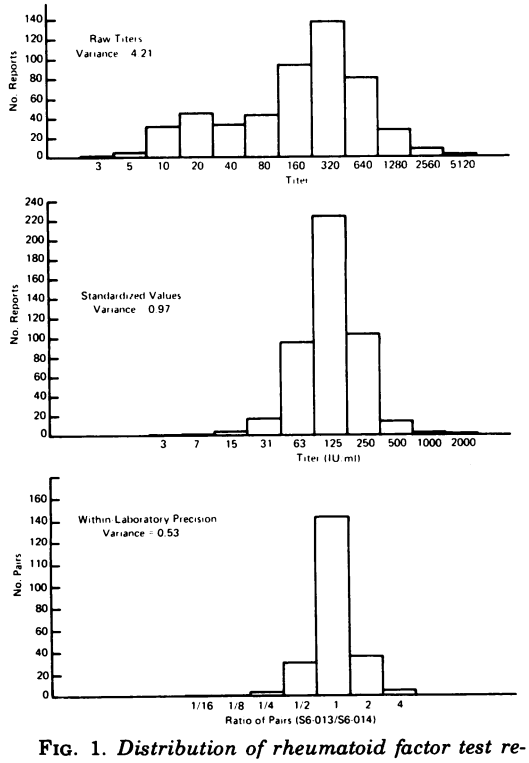


FIG. 1. *Distribution of rheumatoid factor test results.*

TABLE 1. *Comparison of variances for standardized and nonstandardized rheumatoid factor tests*

| Rheumatoid factor tests | Variance $(SD_{log_2})^2$ | | | |
|---|---|---|---|---|
| | Non-standardized | Standardized | Reduction by standardization | % Reduction |
| Slide test | 5.84 | 0.96 | 4.88 | 84 |
| Tube test | 2.14 | 0.75 | 1.39 | 65 |
| Total | 4.21 | 0.97 | 3.24 | 77 |
| Within-laboratory tests[a] | 0.53 | 0.53 | — | — |
| Between-laboratory tests[b] | 3.68 | 0.44 | 3.24 | 88 |

[a] Calculated from results on duplicate samples within the same laboratory.

[b] Total variance minus within-laboratory variance.

The slide test results had larger variance than the tube test results, but they were also reduced the most by the standardization process. Eighty-four and sixty-five percent reductions, respectively, were accomplished.

The geometric mean titer reported for the slide tests was 51, but the geometric mean titer reported for the tube tests was 284, or about 5.5 times higher. In international units, the geometric mean values obtained with the slide and tube tests were 131 and 119, respectively. Thus, the difference in level was effectively eliminated by standardization, and both values were near the 125 IU/ml value that the Bacterial Immunology Branch assigned to those sera.

The histogram of the nonstandardized results shows that the distribution is bimodal in addition to being wider than the standardized distribution. Both the standardized and within-laboratory precision distributions appear to be log normal. Thus, standardization has two beneficial effects: it reduces the between-laboratory variation by eliminating bias and converts the distribution of results to log normal.

Table 2 shows the improvement in comparability obtained by type of test. Before standardization, 15.4% of the laboratories using the slide

tests obtained exactly the median value; after standardization 54.2% obtained it. Only 50.3% of the laboratories were within one twofold dilution of the median result before standardization, but 93.7% were within this range after standardization. Other results were similar but less dramatic. Substantial improvement in overall results was effected both in the tube and slide tests by the simple expedient of using a common reference preparation.

Table 3 indicates the extent that standardization reduced the difference resulting from the use of the various commercial products. Before standardization, the highest geometric mean tube test value reported by the users of a particular commercial test was more than 10 times higher than the geometric mean value reported by the users of the test with the lowest mean. After standardization, the highest geometric mean was less than 1.5 times higher than the lowest. Likewise for the slide tests, before standardization the highest was more than 10 times higher than the lowest, but after standardization it was about 1.75 times higher.

Variance achieved with the tube test was between 7.30 and 0.65 before standardization and between 0.79 and 0.34 after standardization. The percentage decrease in variance resulting from standardization was between 90 and 27. With the slide tests, the variances ranged from 7.99 to 0.79 before standardization and from 0.59 to 0.18 after standardization.

## DISCUSSION

This study confirms the finding of the World Health Organization group (1) that the relative

TABLE 2. *Improvement in comparability of the rheumatoid factor test results by type of test resulting from comparison of titers to a serum standard*

| Results | Slide tests | | Tube tests | | Total | |
|---|---|---|---|---|---|---|
| | Standardized | Nonstandardized | Standardized | Nonstandardized | Standardized | Nonstandardized |
| Median | 54.2%[a] (77/142)[b] | 15.4% (22/143) | 51.0% (173/339) | 33.2% (114/343) | 51.0% (255/500) | 27.1% (137/506) |
| Median ± 1 dilution | 93.7% (133/142) | 50.3% (72/143) | 91.2% (309/339) | 78.1% (268/343) | 91.0% (455/500) | 61.3% (310/506) |
| Median ± 2 dilutions | 97.9% (139/142) | 74.1% (106/143) | 98.5% (334/339) | 91.8% (315/343) | 97.6% (488/500) | 75.1% (380/506) |

[a] Percentage of laboratories within the range.
[b] The numbers in parentheses represent the number of results within the range/total number of results.

TABLE 3. *Effect of standardization of variation by manufacturer*

| Manufacturer | No. of laboratory reports | Before standardization | | | After standardization | | | Percent change in variance |
|---|---|---|---|---|---|---|---|---|
| | | $\bar{x}_G$ | $SD_G$ | Variance[a] | $\bar{x}_G$ | $SD_G$ | Variance[a] | |
| Tube results | | | | | | | | |
| Behring | 21 | 108 | 3.23 | 2.86 | 93 | 1.76 | 0.67 | −77 |
| BCA/Schering | 26 | 312 | 2.74 | 2.12 | 109 | 1.81 | 0.73 | −66 |
| Difco | 29, 27 | 258 | 6.50 | 7.30 | 134 | 1.80 | 0.72 | −90 |
| Hyland | 220, 214 | 309 | 2.05 | 1.08 | 123 | 1.85 | 0.79 | −27 |
| Wampole (R3) | 12 | 63 | 1.57 | 0.65 | 111 | 1.50 | 0.34 | −48 |
| Reagents obtained separately | 10 | 640 | 2.09 | 1.13 | 109 | 1.56 | 0.41 | −64 |
| Slide results | | | | | | | | |
| Behring | 10 | 30 | 1.85 | 0.79 | 144 | 1.34 | 0.18 | −77 |
| BCA/Schering | 6 | 27 | 7.09 | 7.99 | 227 | 1.59 | 0.45 | −94 |
| Hyland | 54 | 119 | 3.20 | 2.82 | 129 | 2.32 | 1.47 | −48 |
| | | | | | 146[b] | 1.69[b] | 0.58[b] | −79[b] |
| Wampole (Rheumaton) | 51 | 16 | 2.31 | 1.46 | 130 | 1.70 | 0.59 | −60 |

[a] Variance = $(SD_{\log_2})^2$.
[b] Minus the results from one laboratory that were three to four dilutions lower than the nearest result.

potencies of a rheumatoid factor specimen estimated against a reference preparation would be more uniform than the titers reported without benefit of this standardization procedure.

Widespread use of a standard reference preparation has the potential of reducing the total variance in laboratory results by about 75%, to a point where the within-laboratory and between-laboratory components of variance are approximately equal. The result would be that it would then become possible to compare with greater confidence the results obtained in one laboratory with those from another laboratory. Further, the disparity between slide and tube tests or between tests in which products from different manufacturers might be used would be reduced. These improvements in comparability can be accomplished without actual change in within-laboratory precision but by adjusting the reporting level in all laboratories to a common reference point.

The fact that standardization results in a distribution of results that is log normal means that comparisons can, therefore, be made by the more sensitive parametric statistical tests than the less sensitive nonparametric tests that should be applied to the bimodally distributed results obtained before standardization.

Standardization reduced variance between laboratories, between types of test, and between manufacturers. The greatest improvement was obtained in the slide tests and in the commercial kits that had the poorest performance before standardization.

The implication of this study is clear: laboratories could substantially improve the comparability of their results by establishing and using a serum reference preparation for rheumatoid factor and by reporting the results in international units. In the absence of widespread availability of a commonly accepted reference preparation, each laboratory may prepare its own secondary reference preparation from a pool of positive serum.

This study has also demonstrated that proficiency testing programs can be used not only for their primary purpose, i.e., to evaluate laboratory performance, but also to evaluate the potential effect of changes in methodology upon performance. Both of these activities are aimed at the ultimate goal of improving the quality of laboratory service provided to the consumer.

## ACKNOWLEDGMENT

The assistance of Alan Huong with the computer programming and data processing is greatly appreciated.

## LITERATURE CITED

1. **Anderson, S. G., M. W. Bentzon, V. Houba, and P. Krag.** 1970. International reference preparation of rheumatoid arthritis serum. Bull. W.H.O. 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. **Taylor, R. N., H. Ehrhard, and S. Marcus.** 1976. Evaluation of tests for infectious mononucleosis through proficiency testing. Health Lab. Sci. 13:34–44.
4. **Taylor, R. N., and K. M. Fulford.** 1974. Rheumatoid factor. IV. Proficiency testing. United States Department of Health, Education, and Welfare, Center for Disease Control, Atlanta.
5. **Taylor, R. N., and K. M. Fulford.** 1975. Diagnostic immunology. II. Proficiency testing. United States Department of Health, Education, and Welfare, Center for Disease Control, Atlanta.
6. **Taylor, R. N., and K. M. Fulford.** 1975. Proficiency testing summary analysis, diagnostic immunology 1975. IV. Antistreptolysin O, rheumatoid factor, C-reactive protein, anti-nuclear antibody. United States Department of Health, Education, and Welfare, Center for Disease Control, Atlanta.
7. **Taylor, R. N., A. Y. Huong, and K. M. Fulford.** 1975. Measurement of variation in serologic tests. United States Department of Health, Education, and Welfare, Center for Disease Control, Atlanta.