A Study of the Accuracy and Precision of Clinical Chemistry Determinations in 170 Canadian Laboratories

David B. Tonks

Lyophilized samples of two serum pools, prepared commercially, were sent to 200 Canadian clinical chemistry laboratories for estimations of total cholesterol, sodium, chloride, glucose, nonprotein or urea nitrogen, total protein, and inorganic phosphorus. The purpose of this evaluation study was to determine if these laboratories, as a group, were performing satisfactorily, and to help individual laboratories to evaluate their performances and detect gross errors in their results.

Values were reported from 170 laboratories. A summary of the complete data is presented in the form of tables, scatter diagrams, and frequency charts. It was found that there was a lack of accuracy and precision in many laboratories. Over 40% of the 3762 values reported fell outside of the allowable limits of errors and therefore were classified as unacceptable. These results indicated a need for improved performances in many of the participating laboratories.

SINCE 1947 a considerable number of evaluation studies of the accuracy of clinical chemistry determinations have been carried out in various countries (*1–5*). They have shown that there is a need for improved performance in many laboratories. Canadian experience has been gained from a number of surveys conducted by Tonks and Allen from 1952 to 1954 (*6, 7*). Although these Canadian studies proved to be useful, they were limited in scope and showed the need for a more

From the Department of Biochemistry, The Hospital for Sick Children, Toronto 2, Ontario. The present address of the author is the Division of Clinical Chemistry, Department of Medicine, The Montreal General Hospital, Montreal 25, Canada.

This paper was presented in part at the Joint Meeting of the Canadian Society for Clinical Chemistry and the American Association of Clinical Chemists, Montreal, Aug. 30, 1960.

The study reported here was a personal research project and was not conducted on behalf of any society, association, company, or institution. It was supported financially by the Warner-Chilcott Laboratories of Canada and the U.S.A. The help and cooperation of the participants is also gratefully acknowledged.

Received for publication Jan. 20, 1961.

representative national study which would include a larger number of laboratories and which would therefore emphasize much more strongly the need for increased accuracy and precision in clinical chemistry determinations. Such a study has been completed and is described here. One hundred and seventy laboratories from across Canada carried out 3762 analyses for eight components in two freeze-dried serum specimens. Allowable limits of error were set for each determination. Using these limits, over 40% of the results were classified as inaccurate and unacceptable. An estimate of precision was also possible since two specimens were analyzed. The results indicated a serious lack of precision as well as accuracy in many laboratories.

Method

Invitations to participate in the survey were sent to 226 laboratories. Included were the laboratories in all Canadian hospitals of 100 beds or over, and a few commercial and government laboratories where clinical chemistry analyses were performed routinely. About 200 affirmative replies were received, and subsequently two specimens, in the form of freeze-dried serums,* were sent to each participant, with complete instructions and report forms. After reconstitution, each laboratory had 10-ml. volumes of the specimens for analysis. Aliquots of the same specimens were sent to each laboratory, but the sample bottles were numbered differently in each case. The bottles had no other markings and were packaged in plain boxes.

Participants were asked to analyze both specimens (A and B) for sodium, chloride, total protein, glucose or total reducing substances, urea nitrogen or nonprotein nitrogen, and phosphorus, and one specimen (B) for cholesterol. However, they were instructed to do only determinations which were performed routinely. Instructions were also given to carry out as many analyses as possible in duplicate. The values obtained, with information concerning the methods used, were sent directly to the author for coding and tabulation. When all data had been received, the stated values for each component, the means of the values reported, and the proposed allowable limits of error were furnished to each participant.

*Special lots of Versatol and Versatol-A, provided by the Warner-Chilcott Laboratories of Canada, were used for the study. These products were prepared in the usual manner by adding weighed amounts of the various components to large pools of serum after these components have been either completely removed or reduced to a known constant level by selective dialysis. Samples of each batch were tested by independent laboratories. They proved to be ideal specimens for the survey since they were stable, had no added preservatives, and contained accurately known concentrations of the various components.

The allowable limits of error were calculated by means of an empirical formula which the author has found to be useful. This formula is based upon the premise that errors should not exceed one-quarter of the normal range. This formula is as follows:

$$\text{Allowable limits of error (in \%)} = \pm \frac{(\frac{1}{4} \text{ of the normal range})}{(\text{mean of the normal range})} \times 100\%.$$

If the normal range for sodium is taken to be 135-145 mEq./L., then the allowable limits of error calculated by this formula are ± 1.8%. The maximum limits for any determination, however, were set at ± 10%, even though in some cases those calculated by the above formula exceeded this figure. For this study, then, the allowable limits of error have been established as follows: for sodium, ± 1.8%; for chloride, ± 2%; for total protein, ± 7%; for glucose, total reducing substances, phosphorus, urea nitrogen, NPN, and cholesterol, ± 10%. These limits are comparable with those used by other authors in similar studies (*1, 4, 8–11*).

An evaluation of the accuracy with which each component was determined was obtained by calculating the percentage of results which fell outside of these allowable limits of error. To evaluate precision, use was made of standard deviations, scatter diagrams, and frequency charts. Also coefficients of variation for the values reported for the two samples were compared with the coefficients calculated from the ratios of the values.

The tabulation of the cholesterol results, which incidentally were from one sample only, presented a problem because of the many different methods used. It was decided to divide the data into four groups as shown in Table 1. Since only the absolute concentration of cholesterol was furnished by the manufacturer, it was necessary to determine "correct" values for the four groups in the author's laboratory. These determined values were used in place of the stated value for all calculations.

Results and Discussion

Of the 276 laboratories which were invited to participate, 170, or 61.7%, completed the study. The results obtained are summarized in Table 1. The last two columns give the percentages of values classified as unacceptable, that is, the percentages of the values which fell outside of the ranges calculated from the suggested allowable limits of

Table 1. Summary of Results

Component	Sample	Stated value*	No. of values reptd.	Mean value	Median value	Range of values	S.D.	Coef. of vari.	Allowable limits of error (%)	Percentage of unaccept. values: Based on stated value	Percentage of unaccept. values: Based on mean
Total red. substances	A	100	239	95.6	96.0	62 –125	9.20	9.2	±10	27.2	27.2
	B	220	230	214	217	148 –285	19.1	8.9	±10	19.1	22.2
"True" glucose	A	86	90	84.6	84.0	63 –115	10.2	12.1	±10	34.4	32.2
	B	197	90	193	190	154 –240	16.8	8.7	±10	20.0	17.8
Inorganic phosphorus	A	3.70	275	3.90	3.90	1.85– 9.00	0.63	16.1	±10	22.0	22.6
	B	8.0	263	7.88	8.10	2.8 – 11.0	1.12	14.2	±10	22.4	25.4
Total protein	A	6.90	320	7.19	7.20	4.15– 10.1	0.63	8.6	± 7	40.0	26.3
	B	4.4	310	4.92	4.90	3.7 – 6.7	0.45	9.2	± 7	71.5	30.3
Sodium	A	139	260	141	140	122 –205	5.98	4.2	± 1.8	53.5	45.4
	B	126	272	126	125	110 –155	5.49	4.4	± 1.8	53.7	53.7
Chloride	A	102	299	101.8	101.5	81.2 –130	4.98	4.9	± 2	41.8	41.8
	B	89	290	92.2	91.8	70.1 –115	4.78	5.2	± 2	75.0	53.5
Urea nitrogen	A	12.1	196	13.5	13.0	3.4 – 40	4.13	31.6	±10	55.0	52.0
	B	29.5	195	31.3	30.0	3.6 – 85	7.25	23.1	±10	47.6	46.1
NPN	A	27	122	33.6	33.5	10.8 – 62.0	7.58	22.5	±10	74.5	64.6
	B	47	121	49.7	49.0	20 – 76	8.30	17.6	±10	47.1	43.0
Cholesterol (S & S)	B	90†	23	99.9	95.0	89.7 –140	13.9	13.9	±10	39.0	30.4
Cholesterol ($FeCl_3$)	B	103†	52	114.8	111	86.5 –180	20.2	17.6	±10	53.9	53.9
Cholesterol (Bloor)	B	106†	139	123.2	118	42.4 –263	34.4	28.0	±10	70.5	61.1
Cholesterol (Direct LB)	B	106†	97	126.5	128	83 –201	26.4	20.9	±10	83.5	67.0
		Average								47.6%	40.9%

*Total red. substances, glucose, P, urea N, NPN, and cholesterol values are expressed as mg./100 ml.; protein as gm./100 ml.; Na and Cl as mEq./L.

†For cholesterol, the values determined by the author by four representative methods, have been used in place of the stated value of 86, for all calculations.

Table 2. CLASSIFICATION OF ALL VALUES ACCORDING TO THEIR MAGNITUDE OF ERROR

	For errors based on content		*For errors based on means*	
Classification of errors	*No. of values*	*% of values*	*No. of values*	*% of values*
Error < allowable error	1971	52.4	2223	59.1
Error > allowable error	1791	47.6	1539	40.9
Error > 1½ × allowable error	1290	34.3	970	25.8
Error > 5 × allowable error	652	17.3	476	12.6
Error > 10 × allowable error	184	4.9	100	2.7

error. These ranges were estimated both from the mean values and from the absolute or stated values of the specimens. For total reducing substances in Specimen A, for example, the stated and mean values were 100 mg.% and 95.6 mg.%, respectively, and since the allowable limits of error were ± 10%, the acceptable range of values were calculated to be 90–110 mg.% and 86.4–104.2 mg.%, respectively. In this particular example, 27.2% of the reported values fell outside of the ranges in both cases. However, in most instances the figures in the two columns are not the same. As one would expect, more values fell outside of the ranges based on the stated values than those based on the mean. But the differences are not large, and the averages of the figures in the two columns are 47.6% and 40.9%, respectively.

Table 1 shows that total reducing substances were estimated most accurately, with inorganic phosphorus a close second. Nonprotein nitrogen and cholesterol proved to be the most difficult to determine correctly. The percentage of unacceptable values based on the mean ranged from 17.8 to 67.0%, indicating a need for a great deal of improvement.

In Table 2 is given the general distribution of the results according to their magnitude of error. The data show that many values were grossly inaccurate.

Figure 1 is a scatter diagram for total reducing substances prepared by plotting Sample A values against their corresponding (obtained by the same laboratory) Sample B values. The values used for plotting in the majority of cases were the averages from duplicate determinations. Figures 2–8 are similar diagrams for all of the other components except cholesterol. These charts show not only the actual values obtained for both specimens but also the correlation between the two sets of values. The vertical and horizontal lines, drawn to outline the acceptable ranges (based on stated values), divide the charts into nine

sections of areas. The number of dots or points in each area is printed on the charts. Each point represents a different laboratory.

The center section contains points plotted from values which were both acceptable, and the upper right-hand section, points plotted from

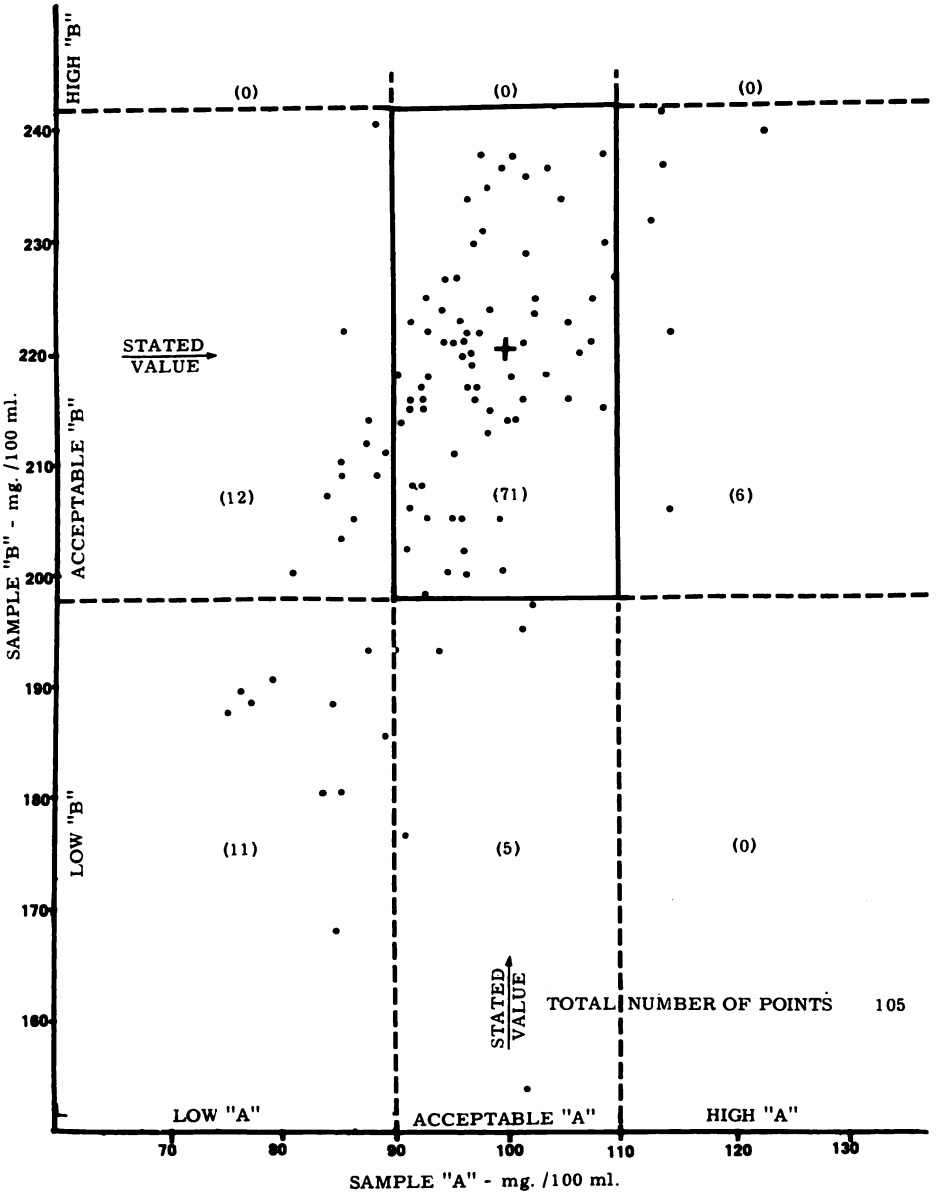

Fig. 1. Scatter diagram for total reducing substances. Plot of Sample A values versus Sample B values.

values which were both too high (that is, too high for Sample A and for Sample B). If errors had been due only to a lack of accuracy, say to improper standardization resulting in Values A and B both being either too high or too low, then all the points not in the central areas would have fallen either into the upper right (high A, high B) or lower left (low A, low B) sections. But 42.7% of all of the points on the

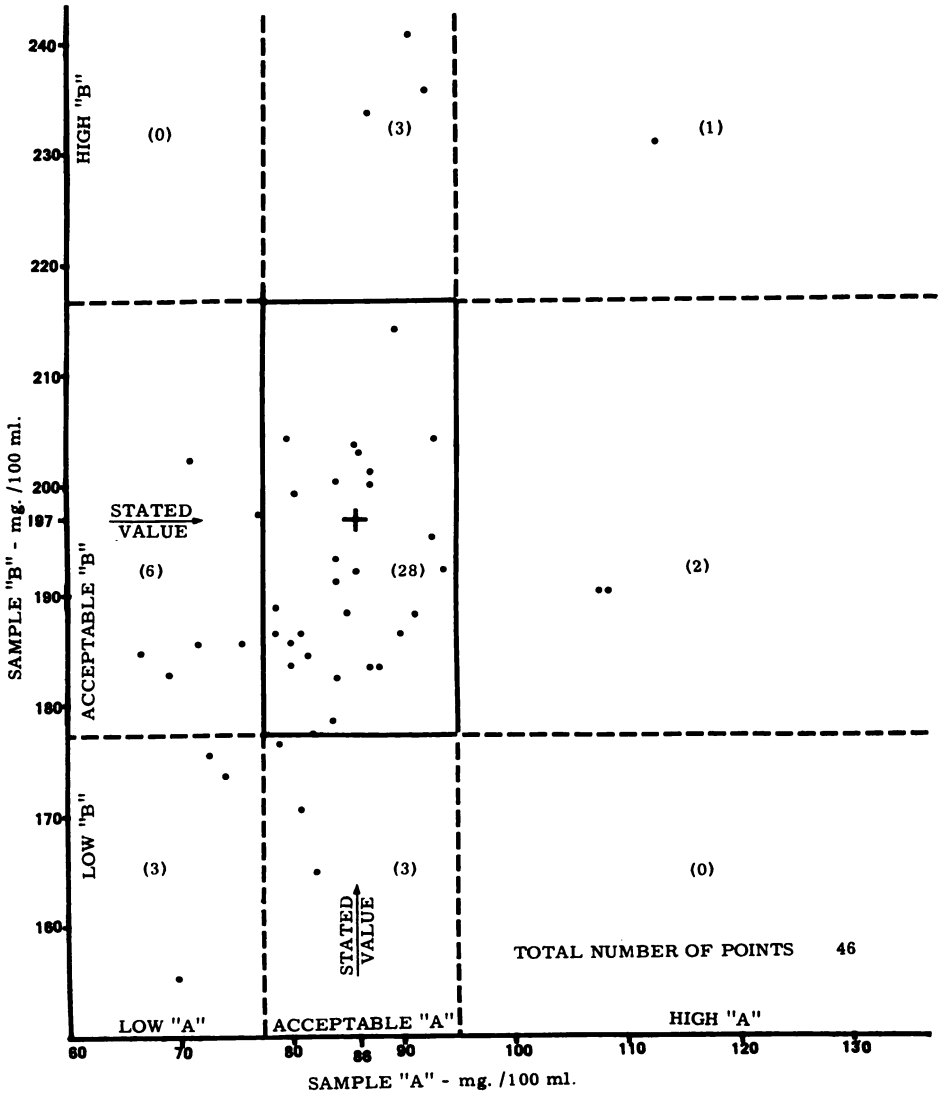

Fig. 2. Scatter diagram for "true" glucose. Plot of Sample A values versus Sample B values.

charts have fallen into the other areas. Therefore it is evident that there was a lack of precision as well as accuracy; and that poor technic, such as inaccurate pipetting, or the use of poor equipment, such as unmatched cuvettes, was responsible for many of the errors.

This conclusion is borne out by the data in Table 3, which compares

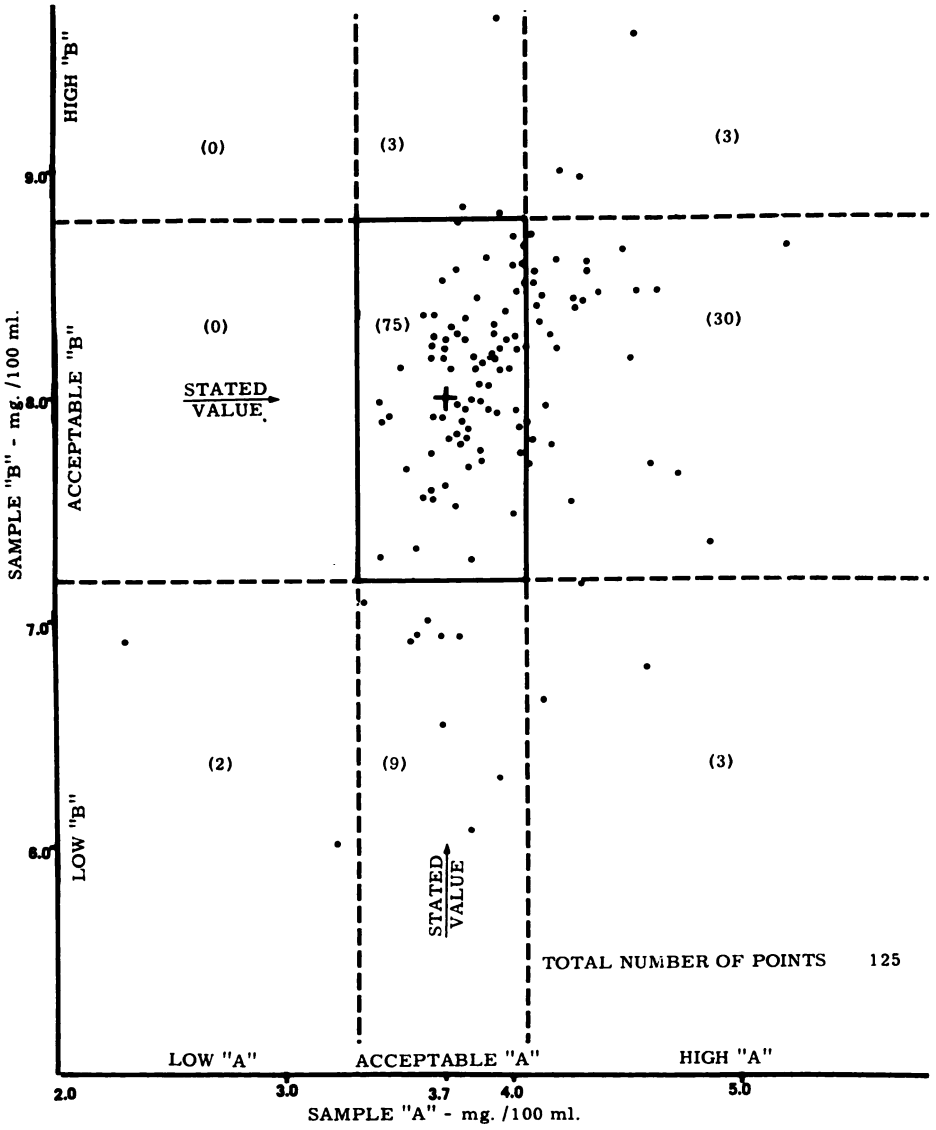

Fig. 3. Scatter diagram for inorganic phosphorus. Plot of Sample A values versus Sample B values.

the coefficients of variation of the reported values for Samples A and B with those of the ratios of the A and B values. The scatter of the ratios is obviously not very different from that of the reported values, again indicating that there was not only a serious lack of accuracy, but also of reproducibility or precision, in many of the laboratories.

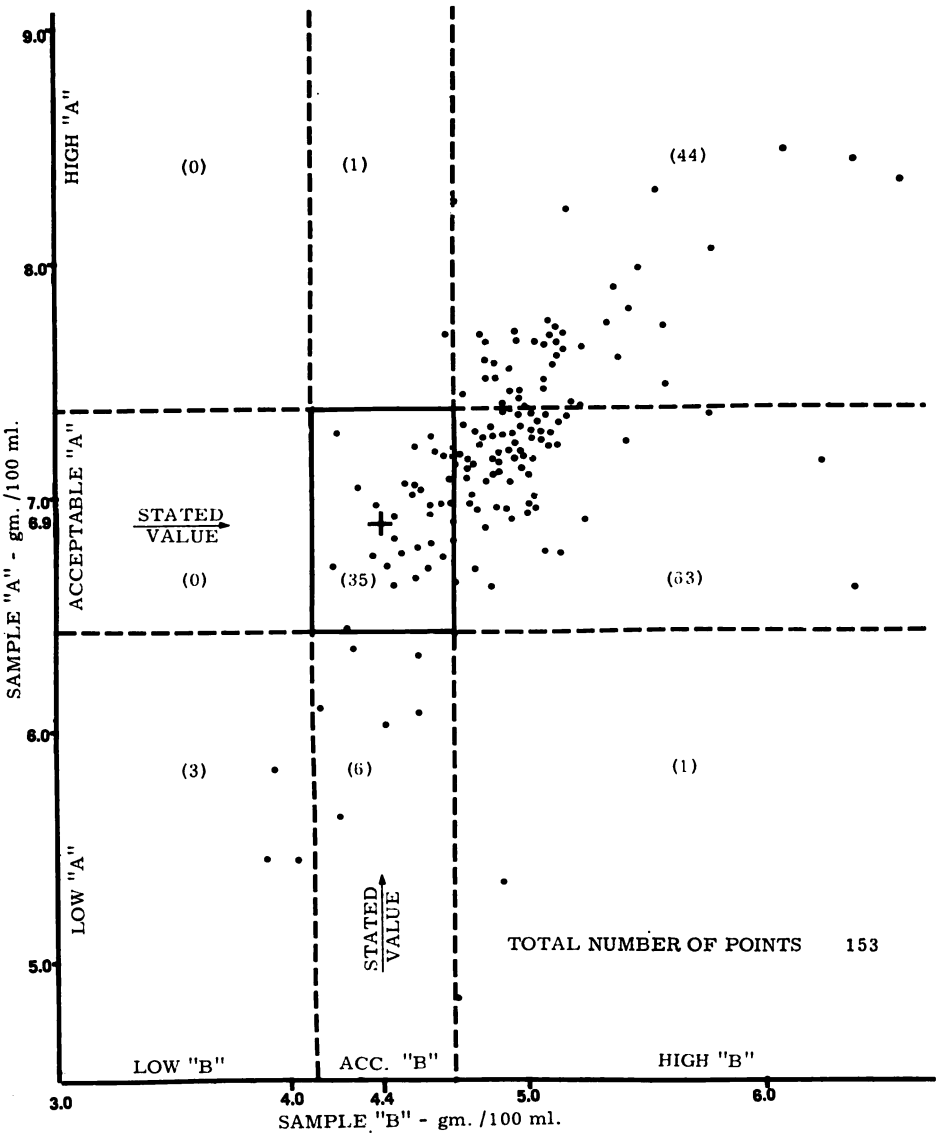

Fig. 4. Scatter diagram for total protein. Plot of Sample B values versus Sample A values.

Figures 9–12 are frequency charts prepared from the cholesterol values. They show the wide scatter of values reported for each of the four groups. The calculated mean values are in each case considerably higher than the absolute value of cholesterol in the specimen and are also considerably higher than the values estimated by the author for

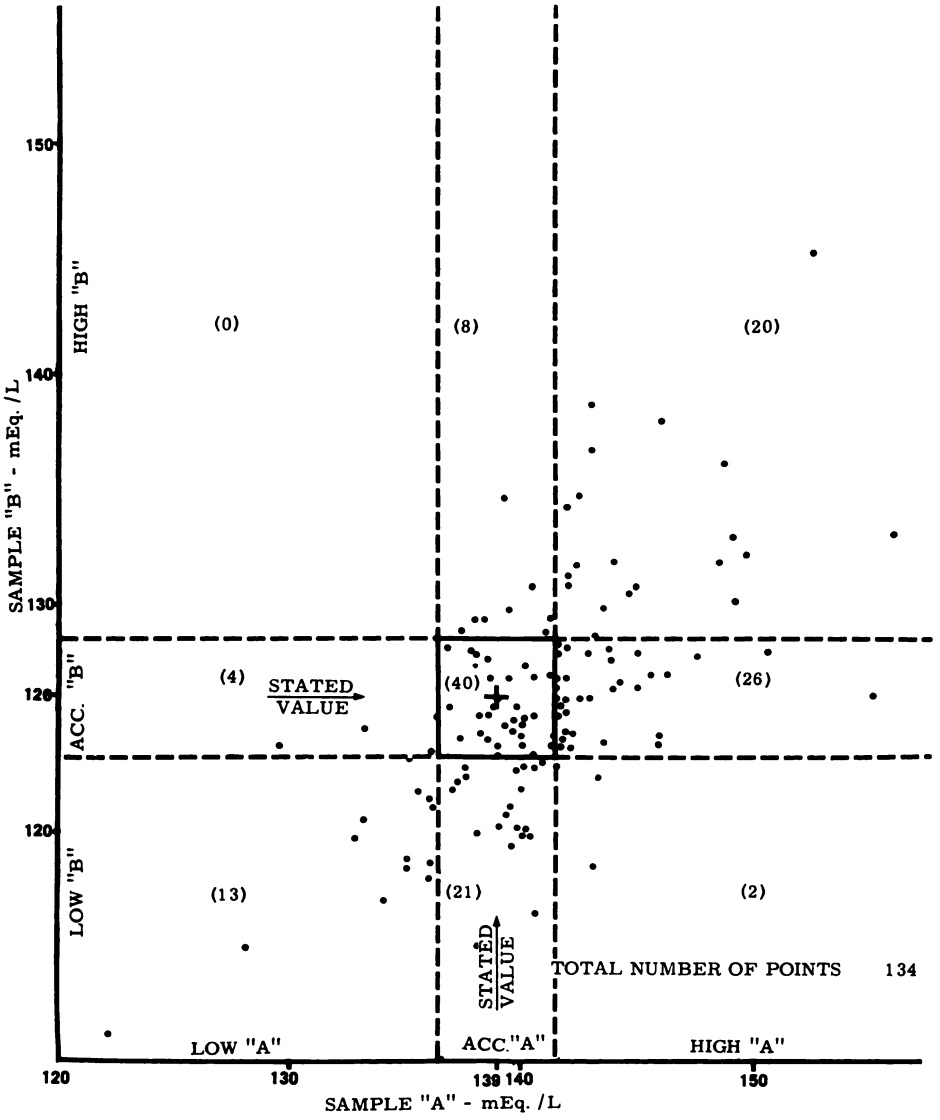

Fig. 5. Scatter diagram for sodium. Plot of Sample A values versus Sample B values.

the four groups. It is obvious that most of the cholesterol methods used lacked specificity.

Although it was not the purpose of this study to obtain information about the various methods used, it is worthwhile reporting that only 180 of the glucose values were obtained by true glucose methods. A large proportion of the other 469 values were estimated by a Folin-Wu

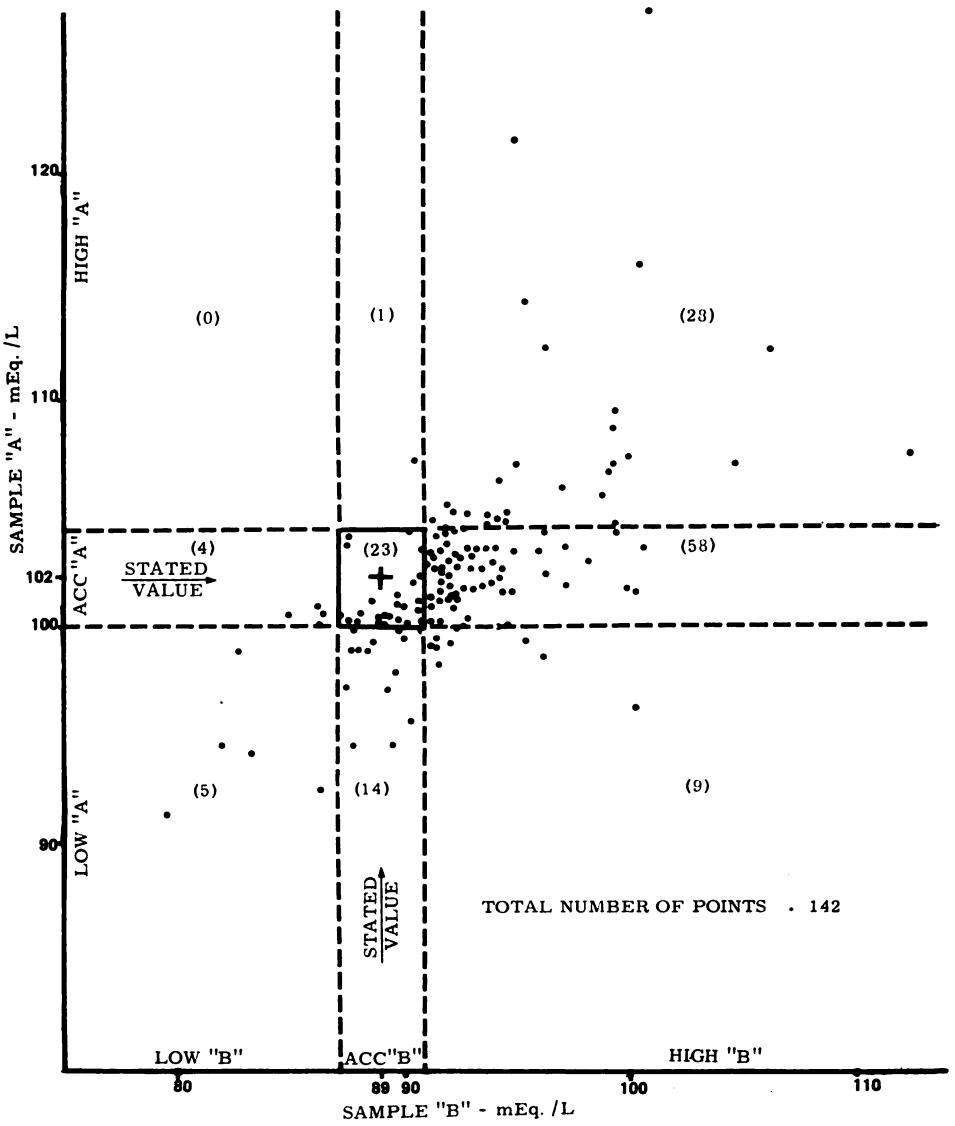

Fig. 6. Scatter diagram for chloride. Plot of Sample B values versus Sample A values.

technic. Almost all of the inorganic phosphorus values were obtained by the Fiske-SubbaRow method, and a majority of the laboratories used a biuret method for the determination of protein. Many different methods were used for cholesterol, with 137 of the 311 values being obtained by the Bloor method or some modification of this method, and 97 by a direct Liebermann-Burchard technic such as the Pearson,

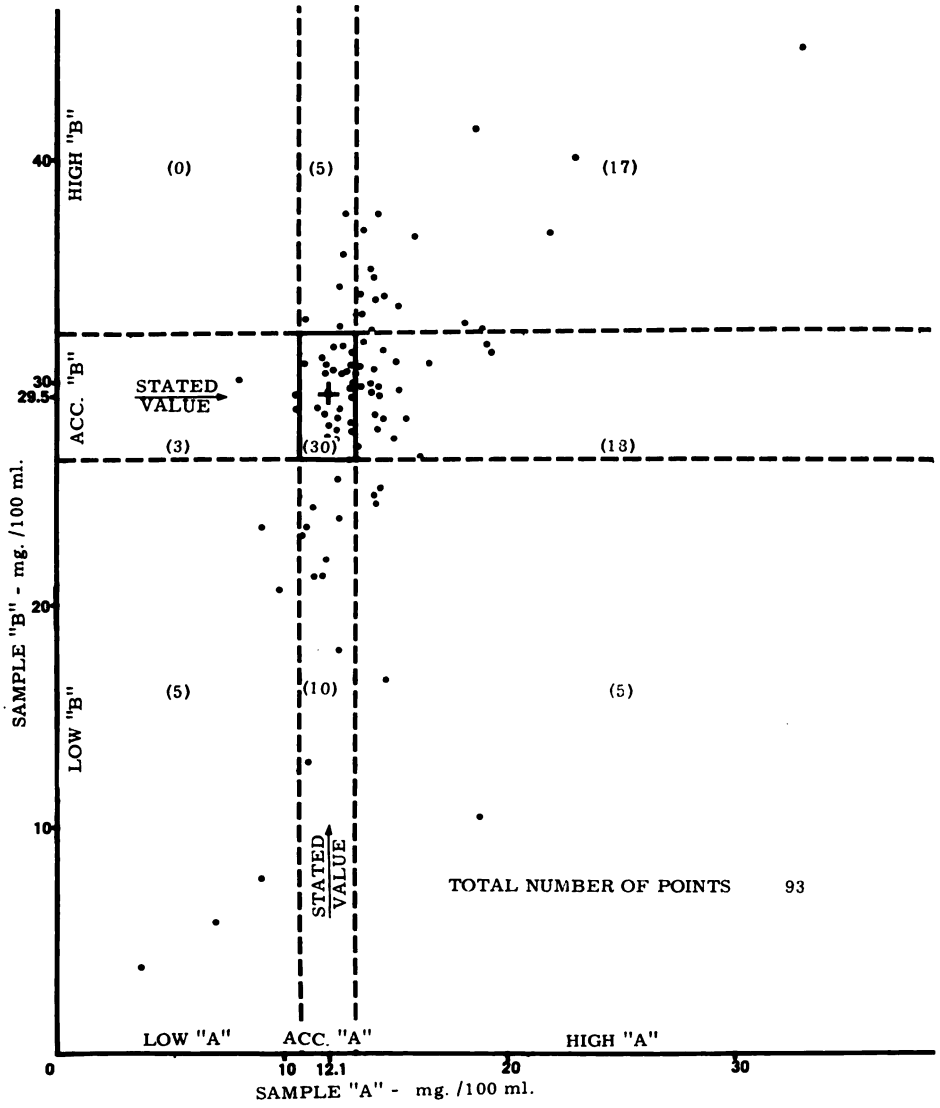

Fig. 7. Scatter diagram for urea nitrogen. Plot of Sample A values versus Sample B values.

Stern, and McGavack method. In the latter group were included the new "kit" methods which have become commercially available and which were used by eight laboratories. In most cases the results obtained by these methods were poor. Relatively few laboratories used

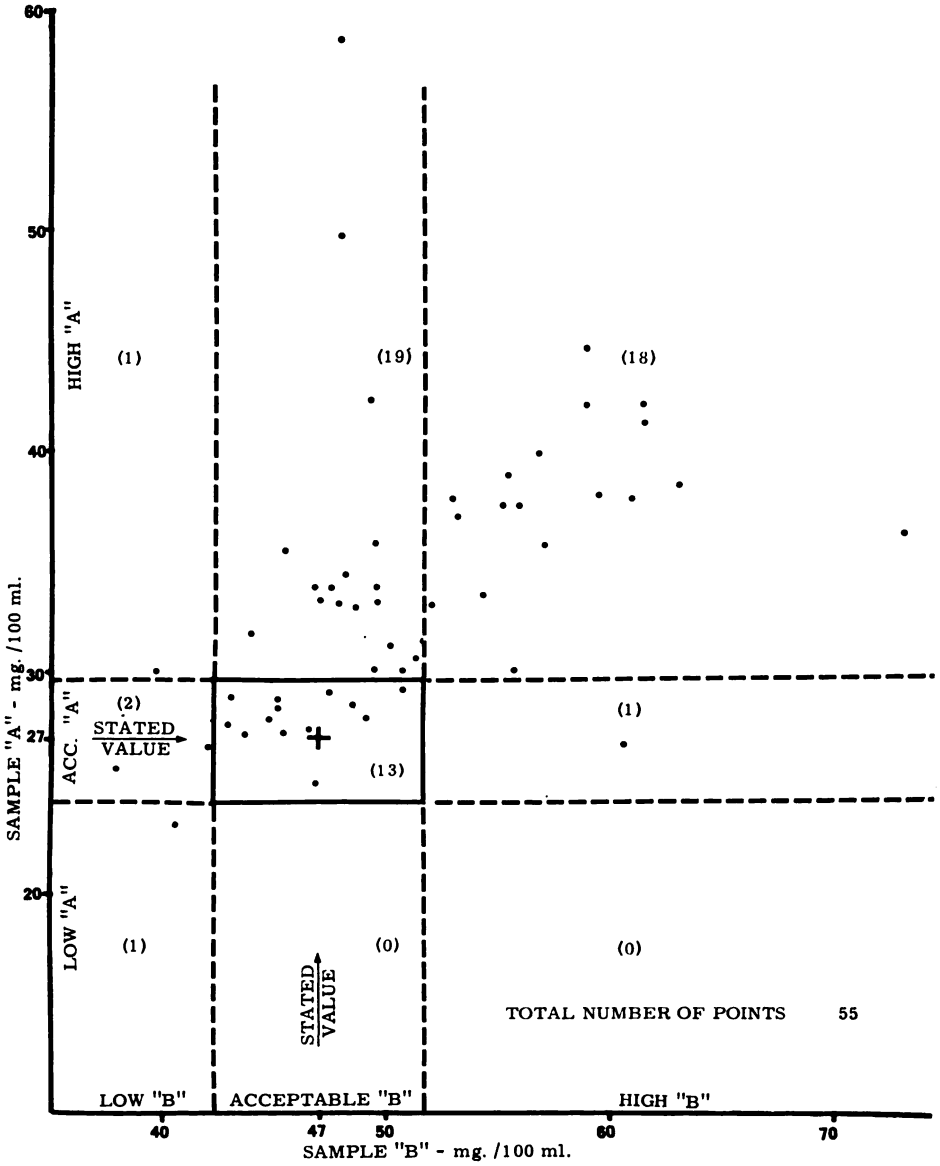

Fig. 8. Scatter diagram for nonprotein nitrogen. Plot of Sample B values versus Sample A values.

Table 3. COMPARISON OF THE SCATTER OF THE RESULTS FOR SAMPLES A AND B WITH THE SCATTER OF THE RATIO A/B,* AS EXPRESSED BY COEFFICIENTS OF VARIATION

	Coefficients of variation†		
Component	*Sample A*	*Sample B*	*Ratio*
Total reducing substances	9.2	8.9	10.1
"True" glucose	12.1	8.7	7.9
Inorganic phosphorus	16.1	14.2	12.2
Total protein	8.6	9.2	7.5
Sodium	4.2	4.4	5.1
Chloride	4.9	5.2	4.1
Urea N	31.6	23.1	22.7
NPN	22.5	17.6	14.3

*Or B/A, if this gave a value greater than 1.

† $\frac{\text{S.D.}}{\text{mean value}} \times 100\% = \text{Coeff. of variation.}$

the Schoenheimer-Sperry method or some modification of it, but this method gave the most accurate values. Ferric chloride methods ran a reasonably close second.

Conclusion

The chief purpose of this study was to determine whether or not there was a definite need for improvement in the performance of Canadian clinical chemistry laboratories. Since 61.7% of the laboratories which were originally contacted completed the survey, the results obtained should be quite representative of this performance. Actually the results should be better than those normally reported by the participating laboratories for several reasons. Very likely most of the participants gave special attention to the survey specimens. Better performances are to be expected from laboratories which take part in surveys than from those that do not take part because of fear of criticism or lack of interest of the personnel. Only relatively large laboratories took part, and generally speaking, we would expect to obtain better performances from large laboratories than from small ones. Therefore, since over 40% of the values obtained in this study were unacceptable by standards which are not overly severe, *it is obvious that there is a definite need for improvement.* Perry in the Maryland surveys (*4*) classified a laboratory as performing satisfactorily only when 90% of the reported values fell within the allowable limits; this is not an unreasonably high standard for an analytical laboratory. Very few laboratories approached this standard in this survey. It

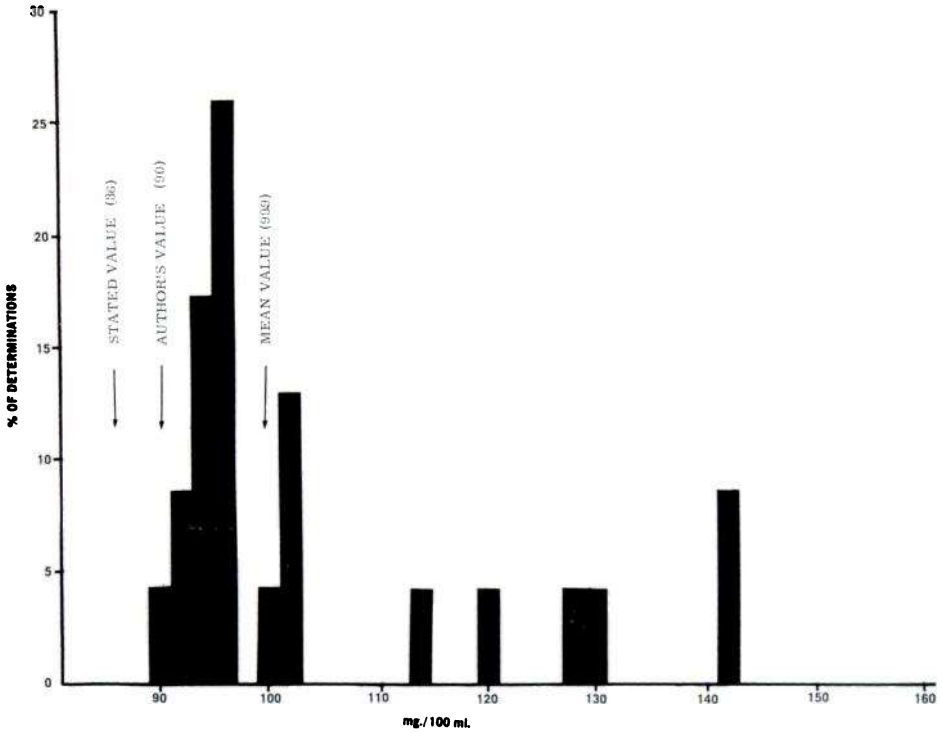

Fig. 9. Frequency chart of cholesterol values. Group 1, Schoenheimer-Sperry methods.

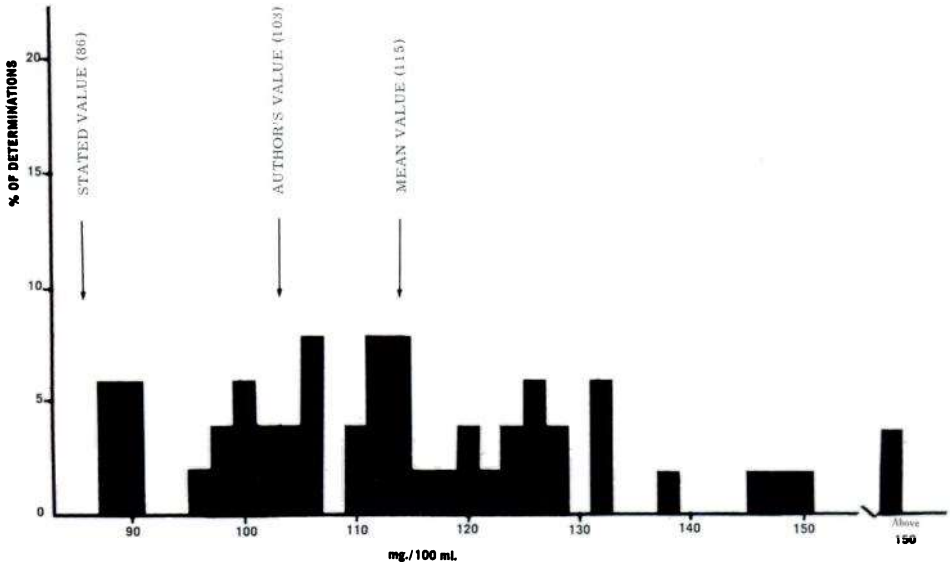

Fig. 10. Frequency chart of cholesterol values. Group 2, ferric chloride methods.

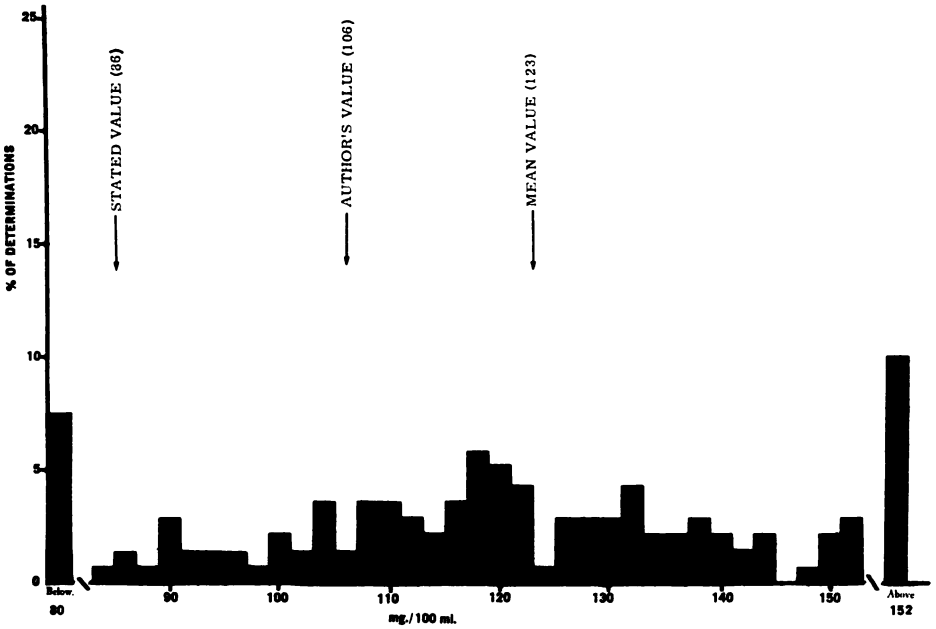

Fig. 11. Frequency chart of cholesterol values. Group 3, Bloor methods.

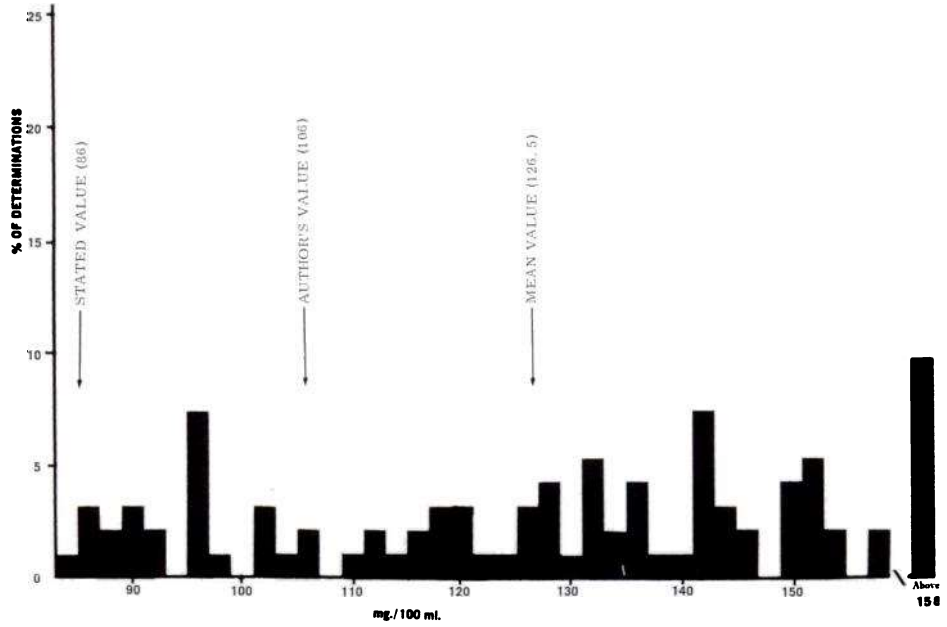

Fig. 12. Frequency chart of cholesterol values. Group 4, direct Liebermann-Burchard methods.

should be pointed out, however, that the results are no worse than those reported in similar surveys in other countries (*5*).

The second purpose of the study was to assist the individual participants, particularly those in the smaller hospitals and in isolated areas, to evaluate their own performances and to enable them to compare their results with those obtained by other laboratories. This purpose has been fulfilled, and it is known that the survey has exerted a considerable influence on the thinking of the participants regarding the accuracy of their analyses and the need for quality control. Also, many gross inaccuracies were detected, and it is certain that serious efforts were made to determine the cause of these errors and to eliminate them.

The data obtained have shown that there was a lack of both accuracy and precision in many laboratories. There is obviously a need for an improvement in the methods of detecting errors and of evaluating precision in the laboratory. An adequate quality-control system and the proper usage of control serums are indispensable for these purposes. It is also useful for every laboratory to take part in periodic studies such as this one in order to obtain an unbiased evaluation of its performance.

References

1. Belk, W. P., and Sunderman, F. W., *Am. J. Clin. Pathol.* **17**, 853 (1947).
2. Shuey, H. E., and Cebel, J., *Bull. U. S. Army Med. Dept.* **9**, 799 (1949).
3. Snavely, J. G., Golden, W. R. C., and Cooper, A. B., *Conn. State Med. J.* **16**, 894 (1952).
4. Maryland State Department of Health. *Laboratory Bulletin for Medical and Public Health Laboratories*, edited by C. A. Perry. Volume 9, March, 1957.
5. Wooton, I. D. P., *Clin. Chem.* **2**, 296 (1956).
6. Tonks, D. B., and Allen, R. H., *Can. Med. Ass. J.* **72**, 605 (1955).
7. Hollands, M., and Tonks, D. B., *Can. J. Med. Tech.* **21**, 88 (1959).
8. King, E. G., and Wooton, I. D. P., *Micro-analysis in Medical Biochemistry*, ed. 3. Churchill, London, p. 5.
9. Beneson, A. S., Thompson, H. L., and Klugerman, M. R., *Am. J. Clin. Pathol.* **25**, 575 (1955).
10. Kingsley, G. R., *The Filter* **18**, 1 (1956).
11. Freier, E. F., and Rausch, V. L., *Am. J. Med. Tech.* **24**, 195 (1958).