Precision in a clinical chemistry laboratory

J. D. ACLAND¹ AND S. LIPTON²

From the Department of Chemical Pathology, University of Sheffield, England, and the Department of Statistics, University of New South Wales, Kensington, New South Wales, Australia

SYNOPSIS The assessment of accuracy and precision in a routine clinical chemistry laboratory is discussed. A figure and tables are presented which relate the chance of making incorrect clinical decisions to the precision of an analytical method.

The question of accuracy and precision in clinical chemistry is frequently discussed (see, for instance, Thompson and Jones, 1965; Mitchell, 1966). It is a waste of time in routine practice to strive for a degree of accuracy or precision which is greater than that required by clinicians. The object of this communication is to enable the performance of a laboratory method to be judged quantitatively in relation to clinical needs.

'Accuracy' is the extent to which a method measures what it sets out to measure. Determinations by an accurate method will not show systematic error, *i.e.*, they will not differ consistently and in the same direction from the result which should be obtained. 'Precision' is the reproducibility of a method. A precise method has a small random error, *i.e.*, there is little variation in results when the sample is run repeatedly. A method may be fairly precise but inaccurate, *e.g.*, the determination of blood 'sugar' by the method of Folin and Wu.

In practice, systematic error may result either from the use of inaccurate methods or from technical faults, such as the deterioration of standards or mistakes made in preparing solutions of reagents. Inaccurate methods should be replaced if possible by accurate methods, but the systematic method errors can be allowed for when the normal clinical range for a particular determination is defined. Systematic errors caused by technical faults are detectable by means of control charts.

The present communication is concerned with the random error of laboratory methods. The performance of a method is assessed by calculating the chance that diagnostic mistakes may occur as a result of method error. These diagnostic mistakes are of

two types (Fig. 1): first, method error may cause the analytical result from a normal sample to fall outside the normal range; secondly, it may cause the analytical result from an abnormal sample to fall within the normal range. The chance of making either of these mistakes depends (1) on the size of the method error compared with that of the normal clinical range and (2) on the difference between the values of the normal limits and the true analytical value of the sample. Mistakes are most likely to occur when the true sample value is close to (either within or outside) the limits. As the true sample value moves away from the limits, the chance of mistakes decreases at a rate which depends on the precision of the method. The greater the precision, the more rapidly the chance of mistakes decreases on either side of the limits.

The standards of precision achieved, or aimed at, in a routine clinical laboratory are thus directly related to the proportion of clinical mistakes which would be expected to occur within areas of uncertainty, defined in terms of the normal ranges for particular constituents.

STATISTICAL METHOD

In the account which follows, σ_1 is the standard deviation of repeated determinations on a single sample and measures the precision of the method, whereas σ_2 is the standard deviation of accurately determined values in a population of healthy subjects, and is related to the normal clinical range. The use of these standard deviations in this context implies that both the method errors and the accurate values for healthy subjects are distributed according to the Gaussian (= 'normal' in the statistical sense) distribution. Small deviations from the Gaussian distribution are unlikely to cause serious quantitative errors. Large deviations from the Gaussian can usually be dealt with by the logarithmic transformation discussed by Gaddum (1945 a, b). It has been found that in non-Gaussian populations the logarithms of the actual analytical values usually follow a Gaussian distribution (Wootton, King, and Smith, 1951; Wootton, 1964). If insufficient determinations on healthy

¹Present address: Chemical Pathology Department, Central Middlesex Hospital, Park Royal, London.

²Present address: Department of Statistics, Queensland University Brisbane, Queensland, Australia.

Received for publication 8 March 1967.

FIG. 1. Clinical mistakes caused by method error. Specimens 1 and 4 are abnormal, specimens 2 and 3 are normal. Some determinations on specimen 2 give a result outside the normal range and some determinations on specimen 4 give a result within the normal range. $(\mu_1)_{1-4}$ = true analytical result for specimens 1-4 respectively. μ_2 = midpoint of normal clinical range.

FIG. 2. Graphs for determining the percentage of clinical mistakes caused by method error at different true analytical values for a particular constituent. The normal clinical range is defined by 90% limits. Analytical results are standardized with reference to the normal clinical range. μ_1 = true analytical value. μ_2 = mid-point of the normal clinical range. σ_1 = standard deviation of method. σ_2 = standard deviation of accurate values for healthy individuals.

$$K = \frac{(\mu_2 - \mu_1)}{\sigma_2} \text{ is the standardized analytical result. } r = \frac{\sigma_1}{\sigma_2}.$$

individuals have been made to determine whether or not the population is Gaussian, the assumption that it is Gaussian is not likely to introduce serious additional error.

The representation of the normal clinical range in terms of a standard deviation (after converting the population of normal values to a Gaussian distribution if necessary) has an advantage in relation to the different criteria which may be used in constructing such ranges. Wootton *et al.* (1951) recommend the use of two ranges, one enclosing 98% of the values obtained from normal individuals and the second enclosing 80% of such values. Results falling outside the 80% limits but inside the 98% limits are considered suspicious. Bodansky (1963) recommends the use of approximate 95% limits (mean ± 2 standard deviations). Wootton (1964) recommends 90% limits. From tables of the Gaussian distribution (Fisher and Yates, 1963) the standard deviation, σ_2 , may be obtained by dividing the 98% range by 4.653, the 95% range by 3.920, the 90% range by 3.290, and the 80% range by 2.563. In this way, the same value of σ_2 is obtained in whatever way the normal range has been defined.

In Fig. 2, the percentage of (a) 'normal' specimens classified as 'abnormal', and (b) 'abnormal' specimens classified as 'normal' are plotted for different ratios of σ_1 to σ_2 and for different true values of the analytical result on a particular sample. The normal range is defined by 90% limits in Figure 2. True sample values (μ_1) are expressed as positive deviations from the mid-point of the normal range, μ_2 , and are measured in terms of σ_2 . The

quantity $\frac{(\mu_2 - \mu_1)}{\sigma_2}$ which is used for this purpose is a 'normal equivalent deviate' (see Gaddum, 1945a). Figure 2 gives the percentage of mistakes for positive values of $\frac{(\mu_2 - \mu_1)}{\sigma_2}$. For negative values of $\frac{(\mu_2 - \mu_1)}{\sigma_2}$, the minus sign is disregarded and $\frac{(\mu_2 - \mu_1)}{\sigma_2}$ treated as if it were positive.

The chances that method error will cause the analytical result from an abnormal specimen whose true value is above the upper limit of normal to fall below the lower limit of normal, or *vice versa*, are negligible even in the extreme conditions of 80% limits and $\sigma_1 / \sigma_2 = 1$. In a normal specimen, analytical mistakes may cause the observed result to fall either above the upper limit or below the lower limit of normal. Both these types of mistakes are included in the total percentage of mistakes. The chances that both types of mistakes may occur at different times in repeated determinations on one particular specimen increase as $\frac{(\mu_2 - \mu_1)}{\sigma_2}$ approaches zero, but are negligible except in extreme conditions (80% limits and $\sigma_1 / \sigma_2 = 0.75$ or 1).

A formal statistical derivation and data for normal ranges based on 80%, 95%, and 98% limits are given in the Appendix.

The use of Fig. 2 is best demonstrated by means of examples:

EXAMPLE 1 In a routine laboratory, the normal clinical range (90% limits) for the plasma calcium concentration

is taken to be 8.5 - 10.5 mg./100 ml. (Gaussian distribution). (a) Suppose that the use of grade A instead of grade B volumetric glassware reduces the standard deviation of the method from 0.11 to 0.07. How great an influence will the use of grade A glassware have on the numbers of clinical mistakes? (b) Suppose that the use of an automated method for plasma calcium reduces the standard deviation of the method from 0.11 (manual) to 0.03 (automated) mg./100 ml. How great an influence will the introduction of the automated method have on the numbers of clinical mistakes? It is assumed that both methods are equally accurate.

σ_1 (manual) = 0.11 mg./100 ml.

σ_1 (manual—grade A glassware) = 0.07 mg./100 ml.

σ_1 (automated) = 0.03 mg./100 ml.

$\sigma_2 = \frac{10.5 - 8.5}{3.290} = 0.61$ mg./100 ml.

$\mu_2 = 9.5$ mg./100 ml.

Procedure Take suitable values of μ_1 for plasma calcium and calculate $\frac{\mu_2 - \mu_1}{\sigma_2}$. Read the percentage of mis-

takes for each value of μ_1 from Fig. 2 using the graph for the value of σ_1 / σ_2 which is the closest to the actual value, i.e., $\sigma_1 / \sigma_2 = 1/5$ for the manual method and $\sigma_1 / \sigma_2 = 1/10$ for the manual method with grade A glassware. The value of σ_1 / σ_2 for the automated method is approximately 1/20. When $\sigma_1 / \sigma_2 = 1/20$, the percentage of mistakes is negligible unless μ_1 is very close to the normal limit and no curve has been included for this value of σ_1 / σ_2 .

Answer

Plasma Ca (mg./100 ml.) (μ_1)	8.2	8.3	8.4	8.5	8.6	8.7	8.8
$\frac{\mu_2 - \mu_1}{\sigma_2}$	2.13	1.97	1.80	1.65	1.48	1.31	1.15

Percentage of mistakes ¹							
Manual	0	5	21	50	21	5	0
Manual—grade A glassware	0	0	5	50	5	0	0

Plasma Ca (mg./100 ml.) (μ_1)	10.2	10.3	10.4	10.5	10.6	10.7	10.8
$\frac{\mu_2 - \mu_1}{\sigma_2}$	-1.15	-1.31	-1.48	-1.65	-1.80	-1.97	-2.13

Percentage of mistakes ¹							
Manual	0	5	21	50	21	5	0
Manual—grade A glassware	0	0	5	50	5	0	0

¹ = uncorrected for grouping error.

Plasma calcium concentration is reported to the nearest 0.1 mg./100 ml. This means for instance that plasmas with true calcium concentrations between 8.45 and 8.54 mg./100 ml. will all be reported as containing 8.5 mg./100 ml. This procedure is equivalent to classifying the results into groups with an interval of 0.1 mg./100 ml. between groups. When the standard deviation of the method is approximately one quarter of the interval between groups or less, the number of mistakes caused by method error will not be so significant.

The adoption of grade A glassware would therefore reduce the numbers of clinical mistakes caused by random method error within the range ± 0.2 mg./100 ml. on either side of the upper and lower limits of the normal range. The adoption of the automated method would eliminate such mistakes altogether.

EXAMPLE 2 The 90% limits of the clinical range for serum cholesterol in a routine laboratory are 150 and 350 mg./100 ml. The logarithms of the concentrations are distributed in a Gaussian manner. The standard deviation of the method is 15 mg./100 ml. Results are reported to the nearest whole number. Calculate (a) the proportion of normal specimens classified as abnormal when the true serum cholesterol concentration is 320, 330, 340, and 350 mg./100 ml., and (b) the proportion of abnormal specimens classified as normal which would occur with true serum cholesterol concentrations of 360, 370, and 380 mg./100 ml.

Procedure Take the mid-point of the normal range and call it μ_2 .¹ Divide the normal range by the appropriate factor and call the result σ_2 . Then proceed as in the previous example.

$$\begin{aligned}\sigma_1 &= 15 \text{ mg./100 ml.} \\ \sigma_2 &= \frac{350-150}{3.290} = 61 \text{ mg./100 ml.} \\ \mu_2 &= 250 \text{ mg./100 ml.}\end{aligned}$$

Answer

Serum cholesterol (mg./100 ml.)	320	330	340	350
Percentage normal classified as abnormal	> 0 and < 7	> 5 and < 16	> 21 and < 31	50
Serum cholesterol (mg./100 ml.)	360	370	380	
Percentage abnormal classified as normal	> 21 and < 31	> 5 and < 16	> 0 and < 7	

DISCUSSION

Bodansky (1963) has pointed out that laboratory precision is important in follow-up studies on chemical changes in individual patients, as well as in assessing the chances of diagnostic error. Greater precision may be demanded for follow-up studies than for a simple diagnostic test, because the range of variation within a single individual may be smaller than the normal clinical range for all subjects. The precision necessary for follow-up studies can be assessed by comparing the size of the method error with the size of the variations normally encountered in serial determinations on the same individual, taking into account diurnal rhythms and day-to-day changes. The measurement of normal variability has been fully discussed elsewhere in relation to changes in the count of circulating eosinophils (Acland and Gould, 1956). The statistical method which was used is applicable to any quantitative clinical determination.

Previous attempts have been made to relate laboratory precision to clinical needs. Tonks (1963) proposed a criterion of the combined accuracy and

precision of a determination, the 'allowable limits of error', defined as $\pm \frac{1}{4}$ of the normal clinical range expressed as a percentage of the mean of the normal range. He laid down the empirical rule that all clinical determinations on a sample of known composition should fall within the allowable limits on either side of the true value, but did not discuss the methods by which normal clinical ranges may be established. Sparapani and Berry (1965) applied Tonks' method to quality control in routine laboratory practice. They considered only the precision of a method and equated Tonks' allowable limits to ± 1 standard deviation (determined from method errors).

Barnett (1966) argued that the allowable limits should ideally be made to correspond to ± 2 or 3 standard deviations, which would be expected to include approximately 95% or 99.7% of method errors respectively. In reply, Berry (1966) stated that estimates of standard deviations in routine laboratory practice tend to be too large because they are based on small samples and may include observations made while methods are subject to technical errors (subsequently corrected). In his experience, using a pooled serum sample, 75 to 80% of the errors on control charts fell within ± 1 standard deviation of the mean. However, if method errors are consistently distributed in the manner stated by Berry, it would be preferable to calculate a more accurate estimate of the standard deviation from the results on the control chart.

It is assumed in the present communication that random method errors form a Gaussian distribution and that σ_1 is an accurate measure of the dispersion of random errors in routine practice. Figure 2 and the tables in the Appendix then provide a quantitative basis for fixing the standards of laboratory precision which are required for diagnostic purposes.

We wish to thank Dr. Arthur Jordan for suggesting this problem to us and Mr. A. S. Foster for drawing the figures. The work was done while one of us (S.L.) was on study leave at the University of Sheffield.

REFERENCES

- Acland, J. D., and Gould, A. H. (1956). *J. Physiol. (Lond.)*, **133**, 456.
- Barnett, R. N. (1966). *Amer. J. clin. Path.*, **45**, 329.
- Berry, R. E. (1966). *Ibid.*, **45**, 330.
- Bodansky, O. (1963). *Clin. Chem.*, **9**, 1.
- Fisher, R. A., and Yates, F. (1963). *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. Oliver and Boyd, Edinburgh and London.
- Gaddum, J. H. (1945a). *Nature, (Lond.)*, **156**, 463.
- (1945b). *Ibid.*, **156**, 747.
- Mitchell, F. L. (1966). *Proc. Ass. clin. Biochem.*, **4**, 38.
- Sparapani, A., and Berry, R. E. (1965). *Amer. J. clin. Path.*, **43**, 591.
- Thompson, G. S., and Jones, E. S. (1965). *J. clin. Path.*, **18**, 443.
- Tonks, D. B. (1963). *Clin. Chem.*, **9**, 217.
- Wootton, I. D. P. (1964). *Micro-analysis in Medical Biochemistry*, 4th ed. Churchill, London.
- , King, E. J., and Smith, J. M. (1951). *Brit. med. Bull.*, **7**, 307.

¹ μ_2 is not the mean of the lognormal distribution and σ_2 is not its standard deviation. σ^2/σ_2 is $1/4$, and the percentage of mistakes falls between that for $\sigma^2/\sigma_2 = 1/2$ and that for $\sigma^2/\sigma_2 = 1/8$.

APPENDIX

STATISTICAL DERIVATION

The statistical problem is to find the probabilities that a random variate, X , falls within or outside a fixed range. X represents the actual determination on a particular specimen and is assumed to have a Gaussian distribution with mean μ_1 (being the 'true' but unknown analytical result) and standard deviation σ_1 . If we represent the normal range by the segment of line AB

then A is the point $\mu_2 - \sigma_2 N$ and B the point $\mu_2 + \sigma_2 N$; N is the appropriate percentage point of the standardized Gaussian distribution according to the way the range is defined (e.g., $N = 1.64$ for a 90% range). If μ_1 lies to the left of A or the right of B then the specimen is clinically abnormal. If, however, the actual determination X lies in AB the specimen is classified as normal. Hence the probabilities, γ , of misclassification of abnormal specimens as normal are given by

$$\gamma = \Pr[\mu_2 - \sigma_2 N < X < \mu_2 + \sigma_2 N] \text{ for } \begin{cases} \mu_1 > \mu_2 + \sigma_2 N \\ \mu_1 < \mu_2 - \sigma_2 N \end{cases}$$

This can be written in terms of standard normal deviates, so that the usual tables can be consulted, as

$$\gamma = \Pr\left[\frac{\mu_2 - \mu_1 - \sigma_2 N}{\sigma_1} < Z < \frac{\mu_2 - \mu_1 + \sigma_2 N}{\sigma_1}\right]$$

where Z is Gaussian with mean 0 and standard deviation 1. Putting $K = (\mu_2 - \mu_1)/\sigma_2$ and $r = \sigma_1/\sigma_2$ we can write the probabilities as

$$\gamma = \Pr\left[\frac{K - N}{r} < Z < \frac{K + N}{r}\right] \text{ for } |K| > N$$

($|K| > N$ is equivalent to $K > N$ or $K < -N$). Similarly the probabilities of classifying normal specimens as abnormal (μ_1 lies in AB but X falls to the left of A or to the right of B) are given by

$$\gamma = 1 - \Pr\left[\frac{K - N}{r} < Z < \frac{K + N}{r}\right] \text{ for } |K| < N.$$

TABLES FOR ASSESSING THE EXPECTED PERCENTAGE OF CLINICAL MISTAKES CAUSED BY METHOD ERROR IN RELATION TO THE TRUE ANALYTICAL VALUE FOR A PARTICULAR CONSTITUENT

The tables give the percentage of clinical mistakes for positive values of $\frac{(\mu_2 - \mu_1)}{\sigma_2}$. For negative values of $\frac{(\mu_2 - \mu_1)}{\sigma_2}$, the minus sign is disregarded and $\frac{(\mu_2 - \mu_1)}{\sigma_2}$ treated as if it were positive. μ_1 = true analytical value. μ_2 = mid-point of normal range. σ_1 = standard deviation of method. σ_2 = standard deviation of accurate values for healthy individuals. $K = \frac{(\mu_2 - \mu_1)}{\sigma_2}$. $r = \frac{\sigma_1}{\sigma_2}$. N is the appropriate percentage point of the standardized Gaussian distribution according to the way the normal range is defined.

TABLE I

80% LIMITS : $N = 1.28$

r	'Normal' Result from Abnormal Specimen														'Abnormal' Result from Normal Specimen					
	4.0	3.8	3.6	3.4	3.2	3.0	2.8	2.6	2.4	2.2	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2
1	/	1	1	2	3	4	6	9	13	18	24	30	37	45	47	40	33	28	24	21
$\frac{1}{2}$	/	/	/	/	1	2	4	7	11	17	25	33	44	46	36	26	19	13	10	
$\frac{2}{3}$	/	/	/	/	/	/	/	/	1	3	7	15	26	41	44	29	17	9	4	2
$\frac{3}{4}$	/	/	/	/	/	/	/	/	/	2	6	17	36	41	41	20	7	2	/	/
$\frac{4}{5}$	/	/	/	/	/	/	/	/	/	/	/	5	27	34	34	8	1	/	/	/
$\frac{5}{6}$	/	/	/	/	/	/	/	/	/	/	/	/	12	21	21	/	/	/	/	/

TABLE II

95% LIMITS : $N = 1.96$

r	'Normal' Result from Abnormal Specimen											'Abnormal' Result from Normal Specimen								
	4.0	3.8	3.6	3.4	3.2	3.0	2.8	2.6	2.4	2.2	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2
1	2	3	5	7	11	15	20	26	33	41	48	44	36	29	22	17	13	9	7	5
$\frac{1}{2}$	/	1	1	3	5	8	13	20	28	37	48	42	32	23	16	10	6	4	2	1
$\frac{2}{3}$	/	/	/	/	1	2	5	10	19	32	47	37	24	13	6	3	1	/	/	/
$\frac{3}{4}$	/	/	/	/	/	/	1	3	9	24	45	32	14	5	1	/	/	/	/	/
$\frac{4}{5}$	/	/	/	/	/	/	/	/	1	12	42	21	4	/	/	/	/	/	/	/
$\frac{5}{6}$	/	/	/	/	/	/	/	/	/	1	34	5	/	/	/	/	/	/	/	/

TABLE III
98% LIMITS : $N = 2.33$

r	K	'Normal' Result from Abnormal Specimen									'Abnormal' Result from Normal Specimen											
		4.0	3.8	3.6	3.4	3.2	3.0	2.8	2.6	2.4	2.2	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2	0.0
1		5	7	10	14	19	25	32	39	47	45	37	30	23	18	13	9	6	4	3	2	2
$\frac{1}{2}$		1	2	5	8	12	19	26	36	46	43	33	24	17	11	7	4	2	1	1	/	/
$\frac{1}{3}$		/	/	1	2	4	9	17	29	44	40	25	14	7	3	1	/	/	/	/	/	/
$\frac{1}{4}$		/	/	/	/	/	2	8	11	42	35	16	6	1	/	/	/	/	/	/	/	/
$\frac{1}{5}$		/	/	/	/	/	/	1	9	36	26	5	/	/	/	/	/	/	/	/	/	/
$\frac{1}{10}$		/	/	/	/	/	/	/	/	24	10	/	/	/	/	/	/	/	/	/	/	/