SIMPOSIO INTERNAZIONALE SULLA STANDARDIZZAZIONE E SUL CONTROLLO DI QUALITÀ NEI LABORATORI CHIMICO-CLINICI

International nomenclature for clinical chemical quantities

R. DYBKAER

Institute of Medical Microbiology, University of Copenhagen and De Gamles By, Geriatric Unit, Department of Clinical Chemistry, Copenhagen, Denmark.

Introduction

It may seem strange, to start a symposium on quality control with a talk on quantity names and units. It is evident however, that we cannot hope to control quality, if we are not able to specify what we are trying to control. With increasing sophistication of clinical chemistry this specification has become more difficult and also more necessary - especially since quality control is now of international scope and involves specialists from many disciplines.

To this date the way of presenting clinical chemical data has been incredibly diverse, inconsistent, and often incomprehensible to outside colleagues. This, sometimes, leads to dangerous mistakes.

I have met with an instance where a clinician suggested bleeding in a patient having a blood hemoglobin mass concentration of 135 grammes per litre because he thought that the result was given as per cent of a normal mean.

Additionally, the conventions have been contrary to internationally accepted terminology used in other branches of science, notably chemistry, biochemistry, and physics, with which clinical chemistry must communicate.

Recognizing the chaos in the clinical chemical language, international efforts in 1966 created a first *Recommendation on Quantities and Units* with the approval of the International Union of Pure and Applied Chemistry and of the International Federation of Clinical Chemistry ([1-3]).

Recommendation 1966

The subject matter of this recommendation may be divided in three parts: The naming principle of the parameters that are measured, the preferred way in which to express them, the units to be used.

Concerning names three parts are necessary: system, component, and kind of quantity. The system may be the blood serum of a stated patient at a stated moment; the component could be the sodium ions of the serum, and the kind of quantity would be the way in which the component should be related to the system, e. g. the mass concentration or the molar concentration of the sodium ions in the serum. The full set of information is called the quantity name.

The preferred way of relating component and system, when a choice is possible, is that giving the best information in a given situation. Is it better to express the concentration of sodium ion in serum as mass concentration or molar concentration? Should the concentration of hemoglobin in blood be given as molar concentration or mass concentration or as per cent of normal? Here, it seems evident that the « molecular » concepts possess inherent advantages which decide the issue. Mass and mass concentration very seldom correlate component and system in a biologically useful way, whereas normal and pathological biochemical processes are governed by laws that are molecular in nature. Thus, « molecular concepts » clarify functional relationships, « mass concepts » obscure them. This fact is mostly reflected today by the use of molar concentration or « equivalent concentration » in the field of inorganic electrolyte concentrations in serum, i. e. for the components sodium-ion, potassium-ion, chloride, and hydrogen carbonate. A short list of other groups of interrelated components should show the advantages of extending this usage:

acetoacetate – acetone – β-hydroxybutyrate;
adrenalinium – noradrenalinium – 4-hydroxy-3-methoxymandelate;
« base excess » – lactate – glucose;
bilirubin – bilirubin conjugates – albumin;
calcium(II) – phosphate;
chloride – bromide;
cholesterol – cholesterol esters;
glycerol – triglycerides;
hemoglobin(Fe) – dioxygen(O_2) – iron(II) – transferrin.

Consequently the International Federation of Clinical Chemistry (IFCC) recommends the use of quantities of a « molecular nature » whenever possible.

As far as units are concerned a set of seven or eight kinds of quantities and corresponding units (Table 1) is regarded as basic, whereas all other quantities and units are derived from this set by simple equations, e. g. volume is length to the third power. When the base units have an inconvenient size, subunits are created by the use of a list of internationally approved factors having step « heights » of one thousand (Table 2). This means that the volume denominator often used for concentration units « one hundred millilitres » is abandoned.

TABLE 1

International basic kinds of quantities and corresponding base units

BASIC KIND OF QUANTITY		BASE UNIT	
Name	Symbol	Name	Symbol
length	*l*	metre	m
mass	*m*	kilogramme	kg
time	*t*	second	s
electric current	*I*	ampere	A
thermodynamic (absolute) temperature	*T*	kelvin	K
luminous intensity	*I*	candela	cd
amount of substance	*n*	mole	mol
« amount of enzyme » *	—	enzyme unit	U

* Cf. « Added in proof ».

TABLE 2

Names and symbols of factors placed before unmultiplied units

PREFIXES SYMBOLIZING FACTORS					
Factor	Name	Symbol	Factor	Name	Symbol
			10^{-3}	milli-	m
			10^{-6}	micro-	μ
10^{12}	tera-	T	10^{-9}	nano-	n
19^{9}	giga-	G	10^{-12}	pico-	p
10^{6}	mega-	M	10^{-15}	femto-	f
10^{3}	kilo-	k	10^{-18}	atto-	a
10^{2}	hecto-	h	10^{-1}	deci-	d
10^{1}	deca-	da	10^{-2}	centi-	c

The recommended system may be illustrated by mentioning a few kinds of quantities and examples of their use.

For *volume* the simplest, so-called coherent unit is the cubic metre, but IFCC has decided — for the time being — to prefer the litre and its subunits, e. g.:

Patient--Urine, volume = 1.20 l

The kind of quantity *mass* should not be confused with « weight », e. g.:

Patient--Body, mass = 70.0 kg

Amount of substance is the new « chemical » basic kind of quantity with the base unit *mole* (defined as the amount of substance of a given component

which contains as many formula units as there are atoms in exactly 0.012 kg of the pure carbon nuclide ^{12}C), e. g.:

24 hours Urine--Calcium(II), amount of substance = 4.3 mmol/l

Mass concentration is used only when « molecular » kinds of quantities will not serve, e. g.:

Serum--Lipid(total), mass concentration = 7.0 g/l

Amount of substance concentration (molar concentration) should be used whenever possible, e. g.:

Blood--Hemoglobin(Fe), molar concentration = 8.9 mmol/l

Particle concentration, much used in haematology, should not employ the microlitre as volume denominator.

Blood--Leukocytes, particle concentration = 6.5×10^9/l

Time does not permit discussing the many other kinds of quantities used in clinical chemistry, only the most important have been touched upon.

Implementation.

A few words should be spared on the implementation of the principles of the Recommendation.

Evidently, the use in practice of the system requires thorough education of clinicians, nursing staff, and laboratory personnel. Informatory articles, lectures, and discussions are necessary for, perhaps, half a year preceding the change. A list of names and units as well as conversion factors from former to new values should be prepared in handy format. It would be advantageous if all the laboratories in a country change at the same time, but not a prerequisite. I prefer a voluntary act, decided by a conviction that the advantages of the recommended system outweigh the problems of transition. Personally, I prefer a sudden — rather than a stepwise — change for all quantities measured. A prolonged period of transition only drags out the inevitable pain of rethinking. Incidentally, from experience I know that this pain is bearable since only about ten of the more used quantities alter the values of the results. For the rest, the users have to consult the normal ranges anyway.

The advantages of the recommended system are prominent: international and national unification in presentation of results, a common language with other scientific disciplines, biological insight and, ultimately, fewer misunderstandings.

The American Association of Clinical Chemists already adopted the Recommendation 1966 in principle. The Netherlands, Finland, Norway, and Denmark decided to change during this year. Great Britain will change stepwise. The periodicals Clinical Chemistry, Clinica Chimica Acta, and Scandinavian Journal of Clinical and Laboratory Investigation will recommend the system to the authors.

Why don't you join the club? I think it would be extremely valuable if the Italian clinical biochemists at their first National Congress decided to adopt the principles of Recommendation 1966 of IUPAC and IFCC.

Summary. — The problem of terminology in the field of clinical chemistry was faced by two International Bodies: the Section on Clinical Chemistry of the IUPAC and the International Federation of Clinical Chemistry, in order to achieve the following:

1) To reduce the number of the ways of presentation of clinical chemical results.

2) To unify the terminology used in clinical chemistry with that used in related fields.

3) To achieve increased biological insight through the preferred use of kinds of quantities of a « molecular » nature.

Therefore, a « Recommendation on Quantities and Units» was prepared in 1966, based on the recommendations of the IUPAC, IUB, IUPAP and ISO, and concerning the basic and derived quantities and the corresponding units of major importance for the clinical chemist, and particularly for his communication with clinicians.

The principles of this Recommendation are illustrated with practical examples. The adoption of a unified nomenclature on a national and international level is strongly encouraged. The advantages for both scientists and patients of reducing the possibilities of errors and misunderstandings are emphasized.

Riassunto (*Terminologia internazionale unificata per i risultati quantitativi delle analisi chimico-cliniche*). — Il problema della terminologia nel campo della Chimica Clinica è stato affrontato da due Enti Internazionali: la Sezione di Chimica Clinica della IUPAC, e la Federazione Internazionale di Chimica Clinica, con i seguenti obiettivi:

1) ridurre il numero delle modalità di presentazione dei risultati chimico-clinici;

2) facilitare il coordinamento della nomenclatura nel campo della chimica clinica e nei campi affini.

3) conseguire una più profonda comprensione biologica attraverso l'impiego di unità di natura « molecolare ».

È stata quindi approntata nel 1966, in base alle raccomandazioni della IUPAC, IUB, IUPAP e ISO, una « Recommendation on Quantities and Units» che riguarda le grandezze di base, quelle da esse derivate e le corrispondenti unità di misura di maggiore importanza per il chimico clinico, specie nei suoi rapporti con i clinici. I principi in essa contenuti vengono illustrati mediante una serie di esempi pratici.

Viene rivolto un invito ad adottare a livello nazionale la nomenclatura internazionale unificata sottolineando il vantaggio derivante, in ultima analisi, non soltanto per lo studioso ma anche per il malato, in quanto le possibilità di equivoci saranno ridotte.

Added in proof.

At its meeting in Menton, France, in May 1971, the Commission on Biochemical Nomenclature (of IUPAC/IUB) decided to recommend the kind of quantity « catalytic amount » with the unit « katal » symbolized « kat » as preferable to amount of enzyme and enzyme unit (U). The katal may be defined as the catalytic amount of any catalyst which catalyses as many reaction cycles per second as there are carbon atoms in 0.012 kg (exactly) of the pure nuclide ^{12}C. In other words the unit is measured by a rate of reaction (in mol/s) of a defined catalysed reaction. The new definition is in accordance with the Système International d'Unités (SI). For conversion 1U $\triangleq$ 16.67 nkat.

REFERENCES

(1) DYBKAER, R. & K. JØRGENSEN. *Quantities and Units in Clinical Chemistry, including Recommendation 1966 of the Commission on Clinical Chemistry of the International Union of Pure and Applied Chemistry and of the International Federation for Clinical Chemistry*. Munksgaard, Copenhagen 1967.

(2) DYBKAER, R. *Clin. Biochem.*, **2**, 227 (1969).

(3) DYBKAER, R. Nomenclature for quantities and units. In: *Standard Methods of Clinical Chemistry*. Acad. Press, Inc., New York 1970. Vol. 6, p. 223.

Statistical considerations on the variability of analytical results

G. AGNESE

Istituto di Statistica Medica e Biometria, Università di Genova, Italy

Laboratory results are estimates of analytical quantities in biological specimens: each result presents some degree of variability, owing to random errors. Different kinds of frequency distribution may be used by the analyst to estimate this variability:

a) binomial distribution may be used to estimate the variability of percentages, such as those of the different kinds of white blood cells in leucocyte formulas;

b) Poisson's distribution may be used to estimate the variability of quantitative discontinuous observations such as cell counts, for instance those of RBC, WBC, platelets, bacteria, etc.;

c) normal (or gaussian) distribution may be used to estimate the variability of quantitative « continuous» observations, that is, to measure the intensity of a parameter that can assume any value within a given range. This applies to biochemical and to clinical chemical assays.

— We can estimate the standard deviation and confidence limits (*CL*) of the percentages by consulting precalculated tables prepared for binomial distribution.

— We can estimate the standard deviation and confidence limits of cell counts by means of Poisson's distribution, from the total number of cells that were counted (N) (*):

$$\text{Standard deviation} = \sqrt{N} \qquad [1]$$

$$CL = \text{Confidence limits at the 95 \% level} = N \pm 2\sqrt{N} \qquad [1']$$

(*) When N is large. For small values of N one must consult precalculated tables.

— We can estimate the standard deviation and the confidence limits of clinical chemical measurements by a gaussian distribution from an adequate number ($n \geqq 30$) of replicate analyses on the same biological specimen using the well-known formulas:

$$\sigma = \sqrt{\frac{\Sigma (m - x)^2}{n - 1}}$$

$$CL = x \pm 2\,\sigma \qquad [2]$$

If analysis of a unknown specimen is repeated « n » times, the confidence limits will be:

$$CL = m \pm 2\,\sigma / \sqrt{n} \qquad [3]$$

and if the analyses belong to a single series:

$$CL = m \pm t_{0.05}\, s / \sqrt{n} \qquad [3']$$

$t_{0.05}$ must be found in the appropriate table, for $n - 1$ degrees of freedom.

With the above formulas we can estimate random variability, which is the source of analytical imprecision: however, we must also consider systematic errors (so-called « bias » of the analyses) which may lead to inaccurate results.

Systematic variability may affect leucocyte formulas as a consequence of poor staining, unequal cell distribution on the glass slide, and the like; it may affect cell counts as a consequence of faulty calibration of the micropipets or counting chambers, insufficient mixing of blood specimens, etc. This variability is not likely to occur in well-organized laboratories.

Systematic variability is typical, and more difficult to avoid, in clinical chemical assays:

a) first, because many non-specific methods are in current use, and this may lead to different levels of inaccuracy, due to the presence of interfering substances in the biological specimens;

b) second, because the results of quantitative assays are usually obtained by comparing the results given by the unkonown biological specimen with the results given by a reference solution of a chemical standard, according to the well-known formula:

$$C_x = C_{st} \cdot \left(\frac{OD_x}{OD_{st}} \right) \qquad [4]$$

The composition of standards, and the way they are used, should therefore be regarded as potential sources of error.

Therefore, the overall variability of results is the sum of three main components:

1) Within-run variability = random variability in a single analytical series. This is related to the features of the analytical method, to small sample-to-sample changes in the performance of the analytical instruments and of the operator, etc.

2) Day-to-day variability. This is related to small day-to-day changes of analytical conditions: reaction times and temperatures, small changes in the volume or reactivity of the reference solution, etc. (so-called day bias).

3) Variability « between laboratories». This is related mainly to use of different methods, and to different reference solutions.

Recent interlaboratory surveys (1) have shown that, if the coefficient of variation (CV) of a given method is usually $\pm 5\%$ in the hands of its author, it may become $\pm$ 5–10% in the routine analytical laboratory, and $\pm$ 15–25%, or even more, in different laboratories.

I shall attempt to analyse the factors contributing to this situation, so that we may understand how to improve it.

Use of chemical standards

These standards are required and commonly used to check the correct functioning of analytical methods. However, as I said before, these may themselves give rise to increased variability.

Let us consider a method with a theoretical CV of $\pm 5\%$, and a practical CV of $\pm 8\%$ in a given laboratory. If we apply formula [3], since OD_x and OD_{st} are both subject to random errors, the overall standard deviation will be:

$$s\left(\frac{OD_x}{OD_{st}}\right) = \frac{1}{(OD_{st})^2} \cdot \sqrt{s_2^2 \cdot (OD_x)^2 + s_1^2 (OD_{st})^2} \qquad [5]$$

where s_1 is the SD of OD_x , and s_2 is the SD of OD_{st}.

If OD_x and OD_{st} are in the same range, the overall coefficient of variation will be:

in the case of a single standard: $CV = 8\% \cdot \sqrt{1 + \frac{1}{1}} = 11.3\%$

in the case of a two standards: $CV = 8\% \cdot \sqrt{1 + \frac{1}{2}} = 9.8\%$

This means that the variability due to the standard increases the « imprecision» of about 40 % if a single standard is assayed, and of about 20 % if the standard is assayed in duplicate and the results are averaged.

The analyst, however, should not rely too much on formula [5] for the assessment of analytical variability, since the chemical standards do not always afford reliable day-to-day (or batch-to-batch) estimates of variability in biological specimens. In fact, several analytical steps required for the analysis of biological specimens are likely to increase the variability of results « between batches» (and also « between laboratories»). Therefore to assess this variability, we must subject a biological control material (for instance a control serum, or a pool of sera) to replicate analysis in subsequent days under routine conditions.

Also, the purity of chemical standards may affect the results: measurement errors, impurity of reagents, and instability of solutions may give rise to systematic differences « between laboratories». This variability can be avoided by using the same reference standard.

Quality control

Quality control [2-5] may be defined as a statistical procedure for checking whether the quality of serially produced items (for instance, analytical results) corresponds to given parameters. These parameters may concern the presence or absence of technical faults: in this case, a given (maximum) frequency of faults should not be exceeded. Such a control may be useful in the fields of hematology and histology, if we want to make sure that errors of cell identification do not exceed a given level. In this case we may apply the formula:

$$P_0 = e^{-a} \; ; \; P_1 = a \cdot e^{-a} \; ; \; P_2 = \frac{a^2}{2!} \cdot e^{-a} \; ; \; P_3 = \frac{a^3}{3!} e^{-a} \; ; \text{ etc.} \qquad [6]$$

where P_0, P_1, P_2, P_3, etc. indicate the probability of 0, 1, 2, 3 ... observational errors, and a is the average frequency of errors in normal conditions. By using formula [6], which is related to Poisson's distribution, we can estimate the maximum allowable number of errors, with reference to a given level of probability ($P = 0.10$–0.05).

In the case of clinical chemical analyses, as a rule, we should apply the formulas of gaussian distribution, and estimate the average value (μ) and the standard deviation (σ) of the results given in subsequent days by a control specimen, for instance pooled sera. Thereafter we should analyze the same control specimen with each new analytical series; and we should tolerate deviations from the expected value (μ) only if they do not exceed the limits of gaussian distribution, according to which:

99.8 % of the individual observations should be comprised within $\pm 3.00\,\sigma$;
99.0 % of the individual observations should be comprised within $\pm 2.57\,\sigma$;
95.0 % of the individual observations should be comprised within $\pm 1.96\,\sigma$.

If each analysis is done more than once, σ should be replaced by $\sigma_m = \sigma / \sqrt{n}$, where n is the number of analyses done on each biological sample, as well as on the control specimens.

The daily results given by the control specimens should be plotted on control charts, as shown in Fig. 1: the middle line corresponds to the theoretical average, and the outer lines to the confidence limits (usually at the 95 % level). Results outside these limits should appear on the chart with an average frequency not exceeding 1/20 (or 1/100, or 1/500, respectively) of the total number of observations; if this frequency is exceeded the analyst should look for possible sources of error.

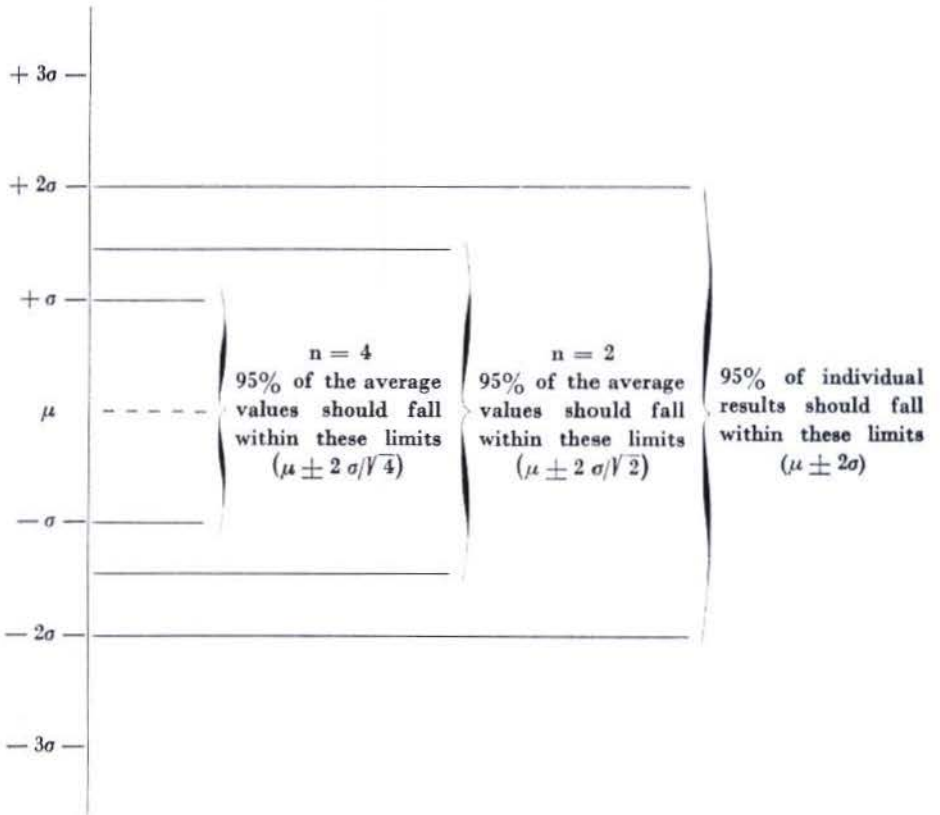

Fig. 1. — Quality control chart for average and individual daily values: limits for 95 % of the observed values.

If two or more specimens of the control solution are assayed daily, this will reduce the scatter and improve the « normality » of the frequency distribution. In case of duplicate analyses the results should be plotted on a

control chart like that shown in Fig. 2; average values are recorded in the upper diagram, and the daily differences between duplicates in the lower diagram. The difference chart affords monitoring of the within-run precision of the analyses, while the upper diagram monitors not only the within-run precision, but also the variability « between batches»: the confidence limits at the 95 % level in the upper diagram will be $\mu \pm 1.20\ \sigma$.

By assaying the control sera every day at least in duplicate, the analyst can monitor the random variability as well as the day-to-day variability of all routine methods (Fig. 2). This affords an answer, at least in terms of probability, to the following questions:

a) does random variability exceed the limits expected on the basis of the initial estimate for σ ? (estimate of imprecision);

b) is there a significant difference between observed average value of the day (m) and expected value μ ? (control of « day-to-day» variability).

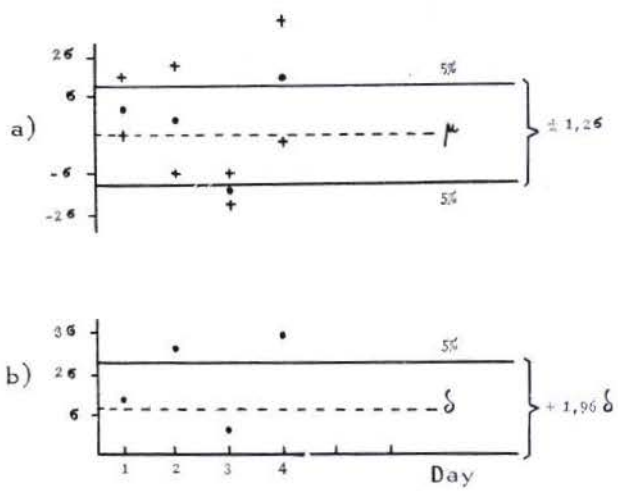

Fig. 2. — Daily chart for the control solution (duplicate analysis):

a) Chart of average values (m) for $n = 2$.

b) Chart of differences (d), in units of σ.

To answer the first question we may record the *daily range d* of results given by the control material: that is, the difference between the maximum and minimum values of the day.

Next, we can estimate the *daily value* of s by one of the following formulas:

for $n = 2$ $s = 0.886\ d$	for $n = 4$ $s = 0.486\ d$
for $n = 3$ $s = 0.591\ d$	for $n = 5$ $s = 0.430\ d$ etc.

Even better, we can estimate $\delta = \sigma/k$, where k is one of the above factors: for instance, in the case of duplicate analyses we shall find $\delta = \sigma/0.886 = 1.136\ \sigma$. We may use the value of δ to set the confidence limits for the daily range d in the chart, as shown in Fig. 2. The value of d may either be within the limits, or out of them when it exceeds $1.96\ \delta$: in this

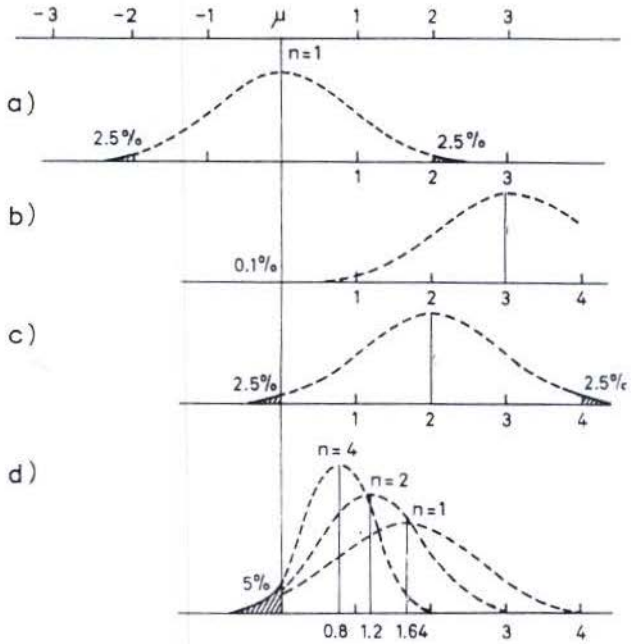

Fig. 3. — Quality control-chart of daily individual and average values.

Variability of daily values (m) around the expected value (μ), expressed in m/σ units. Significance of difference $m - \mu$.

event the analyst must suspect an increase of normal variability, requiring a careful evaluation of results and often a repetition of the entire batch of tests.

We shall now consider the variability of the difference $m - \mu$ between the daily result (or daily average) given by the control material (m) and the reference value (μ). As shown in Fig. 3, in the absence of systematic bias, the difference will vary according to a gaussian curve centered on zero.

If the difference exceeds a given limit (day n. 3 in Fig. 2), a systematic error must be suspected.

If we accept as reference the expected value μ of the control material, in the case of a single control specimen we must admit a systematic deviation

of analytical values (with a probability $\geq 95\%$) when $m - \mu$ exceeds $1.64\,\sigma$: we must adopt this limit instead of $1.96\,\sigma$ because we are comparing a single variable result with a fixed reference (one-tailed test of significance). As shown in Fig. 3 *d*, the corresponding limit in the case of daily averages of duplicate or quadruplicate analyses of the control material will be about $1.2\,\sigma$ and $0.8\,\sigma$, respectively. An increase in the number of control specimens will not only increase the probability of detecting a systematic bias, but also afford an independent check of the daily analytical scatter (Fig. 2).

If the difference $m - \mu$ falls outside the confidence limits in coincidence with a high value of d (day n. 4 in Fig. 2), this may reflect an abnormal scatter of results without systematic error.

If the difference $m - \mu$ exceeds the confidence limits, while the daily range d is within limits, we may infer the existence of a day-to-day variability in excess of random variability, leading to increased variability « between batches». An evaluation of all possible sources of error may be useful, but as we said, chemical standards are insensitive to certain sources of variability. Besides, until the source of variability is eliminated, there is the risk of releasing biased results, unless whole batches of tests are repeated.

In such a case we may use the value of m to correct the results of the whole batch. By doing so, we can avoid major systematic errors in the results even though the exact concentration of the substance in the control serum remains unknown. On the other hand, with this procedure we cannot detect or correct differences « between batches» below $1.6\,\sigma$; $1.2\,\sigma$, and $0.8\,\sigma$ respectively for 1,2, and 4 replicate analyses of the control serum.

Alternatively, we could use in formula [4] the daily value of the optical density and the known concentration of the control serum as daily reference, instead of the corresponding values of chemical standards; by doing so we can control those variations (due to factors that cannot be detected by the chemical standard) which may result in increased variability « between batches» and also « between laboratories».

Today, however, we do not recommend the use of control sera as a substitute for chemical standards: this would favor error, since commercial sera supplied by different manufacturers do not always give consistent results. However, we strongly recommend the future preparation of control sera containing the main groups of substances in well controlled amounts: these sera should be analyzed by carefully selected reference methods. If officially controlled reference materials of this kind will become available to laboratories on a regional, national or international scale, this will contribute to eliminate an important component of the overall variability, allowing greater reliability of analytical results.

A last important factor of variability is the great number of methods used in different countries and in different laboratories for the same assay, and the frequent use of imprecise and non-specific methods. Until now, in spite of many efforts, little progress has been made in the selection of standard clinical chemical methods; the situation is far worse in this field than in other fields of chemical analysis ([1,6-8]).

If the analysts want to improve their work, they must establish national and international committees for the selection of accurate and precise standard methods, and they must encourage systematic adoption of such methods in the clinical laboratories. In the meantime, the use of control sera as biological reference will sometimes be useful to facilitate comparison of results: the values assigned by selected reference methods to a control serum may be used, at least in some cases, to convert the results in terms of a selected standard method.

Normal values

An improved analytical performance will certainly lead, among other things, to a more reliable assessment of normal values and ranges by different methods ([9-13]). This problem is beyond the scope of this paper: I shall only stress the need for sampling normal subjects from a truly homogeneous (unimodal) reference population since this requirement is essential for assessing normal values.

However, such a requirement is not easily met because within a so-called « normal» population there may be groups with a hidden, subclinical pathological condition (abnormal subjects). As a consequence, abnormal values may inadvertently be included in the « normal» range.

In case of a bimodal distribution, a minor (abnormal) component may remain hidden in the right-hand tail of the distribution diagram: this is shown in Fig. 4*a*; other distributions with multiple components are illustrated in Figs. 4*b* and 4*c*. One can easily see that the presence of abnormal com-

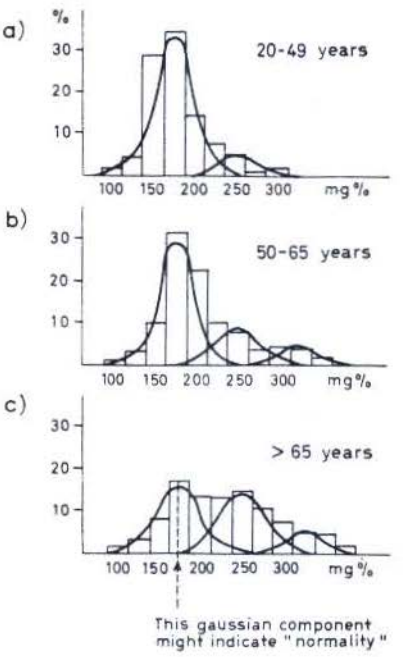

Fig. 4. — Serum cholesterol levels: frequency distribution in different age groups, with multiple gaussian components.

ponents within the reference population leads to abnormally broad « normal » limits; as shown in the diagram of Fig. 5, this may greatly reduce the discriminating power of a given biochemical parameter. If the reference population includes subclinical pathological members (abnormal), the total frequency

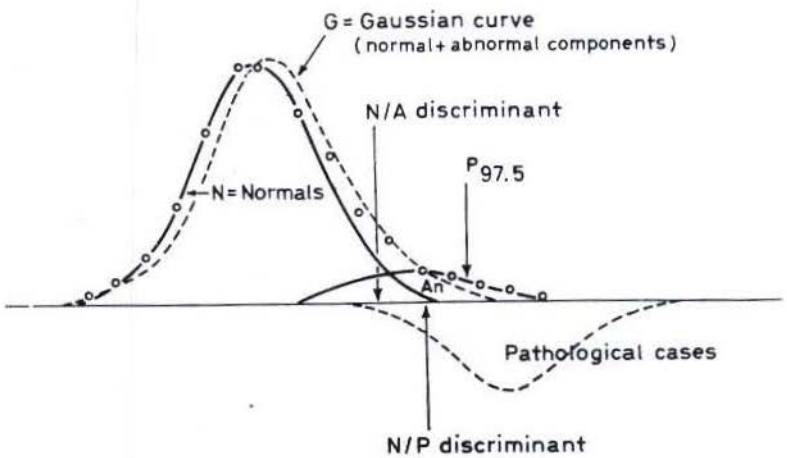

Fig. 5. — Interference of abnormal components in a « normal » reference population.

o = Observed frequency distribution.

An = Abnormal cases = minor gaussian component.

A change in the size of the abnormal component « An » will lead to a shift in the position of N/P and of $P_{97.5}$, while the position of N/A will remain the same.

curve will be an artifact, and it would be misleading to describe it in terms of « lognormal distribution » of normal values, or to cut off percentiles from the assembled data.

As shown by some investigators ([14,15]) we may often split by statistical analysis a mixed series into gaussian components: one of these components indicates the truly « normal » subjects and gives narrower limits for the normal range. This may allow a better discrimination between normal and abnormal results, and increase the diagnostic value of the analyses.

A general outline of the above mentioned components of the variability of laboratory results is reported in Table 1.

To sum up, the analyst can take the following steps to improve the performance of his laboratory:

1) Assaying the daily reference standards at least in duplicate;

2) Adopting a standard analytical method, characterized by an adequate precision, stability, and accuracy: or at least, comparing from time

TABLE 1

General outline of components of analytical variability

VARIABILITY (variance)	NATURE	SOURCES OF ERRORS	CONTROLS	RECOMMENDED STEPS FOR ANALYTICAL IMPROVEMENT
σ^2	Random variance (theoretical value)	Characteristics of analytical method		
$s_1^2 +$	« Within-run » random variability	Analytical technique (precision)	Difference control chart (*d*)	Improve technique (or select another method)
$s_2^2 +$	Day bias (systematic)	Analytical technique (day-to-day changes of experimental conditions)	Control chart of the average values (*m*)	Improve technique Use replicate biological standards
$s_3^2 +$	Interlaboratory bias due to different standards	Use of different (or unreliable) reference standards		Use single or multiple reference standards of known composition
s_4^2	Interlaboratory bias due to different methods	Use of methods of different accuracy		Use reliable standard methods
TOTAL $= S^2$	Random and systematic variance	Multiple sources	Quality control	Use standard methods and controlled biological standard preparations Adopt statistical control Check analytical technique

to time the analytical results given by the method in current use with those of a standard analytical method;

3) Estimating the confidence limits of new methods, using the values of imprecision (CV within runs) increased by 20–40% according to the number of assays of reference standards;

4) Assessing the imprecision and inaccuracy of analyses by assaying a biological control specimen in duplicate every day, and by recording the analytical differences, as well as the daily average values, on a chart.

I also hope that in the near future the preparation and use of officially controlled biological reference materials will assist the analyst in his daily effort to reduce analytical errors and to decrease the variability of results « between days» and « between laboratories».

Summary. — After a brief, general introduction on the variability in the various types of laboratory tests and on the statistical frequency distributions to be utilized for its evaluation, the author emphasizes the fundamental distinction between casual and systematic errors. The attention is focused on the situation existing in the field of clinical chemical analyses.

The possible sources of casual and systematic errors (that is imprecision and inaccuracy, respectively) are briefly examined. In particular, on the basis of the real magnitude of the whole variability resulting from the analytical inquiries, the following problems are considered from a statistical standpoint:

a) Clear evaluation of the incidence of casual and systematic errors, in the whole, and within each laboratory.

b) Control of the sources of error, that is evaluation and correction of their effects.

The statistical basis of « quality control» is illustrated, and methods for the preparation of « control charts» are discussed.

Finally, the problem of evaluation and correction of variability among different laboratories is examined, considering such problems as methods of analysis, choice of reference standards and definition of « normal values».

Riassunto (*Considerazioni statistiche sulla variabilità dei risultati analitici*). — Dopo una breve premessa generale sulla variabilità nei vari tipi di esami di laboratorio e sulle distribuzioni statistiche di frequenza cui si deve fare riferimento per valutarla, si sottolinea la fondamentale distinzione fra errori casuali e sistematici, e, per questi ultimi, si considera in particolare la situazione nel campo delle analisi chimico-cliniche.

Si esaminano brevemente le possibili fonti di errori casuali e sistematici (ovvero di imprecisione ed inaccuratezza). In particolare, sulla base della

effettiva consistenza della variabilità globale quale è risultata dalle inchieste analitiche, si considera sotto l'aspetto statistico il problema di:

a) valutare distintamente, nel complesso e per ciascun laboratorio, l'entità degli errori casuali e sistematici;

b) controllare le fonti di errore, ovvero valutarne e correggerne gli effetti.

Si illustrano i presupposti statistici sui quali si basa il « controllo di qualità » e si discutono le modalità pratiche per la messa a punto delle « carte di controllo ».

Si esamina infine il problema della stima e della correzione della variabilità tra laboratori diversi, in relazione ai metodi d'analisi, alla scelta degli standard di riferimento ed alla definizione dei « valori normali ».

REFERENCES

(1) Saracci, R. *Appl. Bio-Med. Calcolo Elettron.*, **3**, 169 (1968).

(2) Agnese, G., G. Lamedica & U. De Leoni. *Giorn. Igiene Med. Preventiva*, **4**, 269 (1963).

(3) Barnett, R. N. *Ann. N. Y. Acad. Sci.*, **161**, 477 (1969).

(4) Büttner, H. *Z. Klin. Chem. Klin. Biochem.*, **5**, 41, (1967).

(5) Whitby, L. G., F. L. Mitchell & W. Moss. *Advan. Clin. Chem.*, **10**, 65 (1967).

(6) Tonks, D. B. *Clin. Chem.*, **9**, 217 (1963).

(7) Vanzetti, G. *Minerva Med.*, **58**, 2724 (1967).

(8) Wootton, I. D. P. & E. J. King. *Lancet*, **i**, 470 (1953).

(9) Gabrieli, E. R. *Ann. N. Y. Acad. Sci.*, **161**, 460 (1969).

(10) Herrera, L. *J. Lab. Clin. Med.*, **52**, 34 (1958).

(11) Mainland, D. *Ann. N. Y. Acad. Sci.*, **161**, 527 (1969).

(12) Murphy, E. A. *Perspectives Biol. Med.*, **9**, 333 (1967).

(13) Sunderman, F. W. Jr. *Ann. N. Y. Acad. Sci*, **161**, 549 (1969).

(14) Agnese, G. *Minerva Med.*, **62**, 963 (1971).

(15) Bo, G. *Giorn. Gerontol.*, **17**, 535 (1969).

Confidence limits of laboratory data and normal values

D. B. TONKS

Division of Clinical Chemistry, Montreal General Hospital and Faculty of Medicine, McGill University, Montreal, Canada

There are two types of « confidence limits» to which I wish to refer in this discussion. One of these is not a confidence limit in the true sense, however; this type I call « allowable limits of error», by which I mean desirable error limits within which we should try to operate in our analytical procedures. The other is the usual type of statistical limits based on standard deviations calculated from laboratory data. Rather than using the term standard deviation, however, I prefer the companion term coefficient of variation since it is preferable to express the data as percentages rather than in concentration units.

The quality of clinical laboratory data varies widely from laboratory to laboratory. Three of the principal reasons for this variability are the following:

(*a*) Only a relatively few directors insist on great accuracy and precision, and collect proper data for their evaluation. Most directors do not give adeguate consideration to the measurement and control of these factors.

(*b*) Methods of analysis of course vary greatly, and each one has a different precision and accuracy.

(*c*) A generally-accepted list of allowable limits of error, based on clinical considerations, is not available. No recognized group of physicians or clinical chemists has decided, for example, that for a serum amylase estimation to be clinically useful, it must be within 10 % or 20 %, say, of the true value.

Some clinical chemists, e. g. Campbell and Annan (1), have stated that it may be too early to define « allowable limits of error» in clinical chemistry, suggesting that such limits cannot be properly delineated in the present imperfect state of medical knowledge. This seems to me to be too pessimistic a view. In any case, I doubt whether we will ever reach a « perfect state» of knowledge and I feel strongly that we should try to solve this problem now.

Most laboratories to-day have an internal quality control system, but in many cases this is not being applied properly because performance standards or specifications have not been established beforehand. Regardless

of the width of the control limits obtained by estimating the standard deviation (2 or 3 standard deviations are usually used as control limits), the results are considered to be fully acceptable as long as the method is « in control», that is, operates within these locally-established limits. But these limits may be far too wide for the test to be clinically useful. Unless they fall within previously-selected « allowable limits of error», quality control cannot be considered to be effective or adequate.

In my laboratories, for all methods controlled by our quality control system, the basic operating rule is as follows:

« A test must be performed in such a way that the working control limits of $\pm$ 2 $\times$ the coefficient of variation ($\pm$ 2 C.V.) for a method should not be wider than allowable limits of error previously decided upon as being acceptable for the method».

The total error of each analytical method of course includes the component errors of both accuracy and precision. In this discussion we will be concerned primarily with precision, and will assume that the methods chosen are accurate enough for their intended purpose, and that any method bias is compensated for by the establishment of a proper normal range.

Since the primary purpose of a clinical laboratory test is to serve as a diagnostic aid, it must be sufficiently precise to enable us to differentiate between normal and abnormal values in patients. Often the physician must take a decision about a value which is at a borderline of the normal range. Is it truly an abnormal result or is it simply a high or low normal value? This type of decision is required more frequently to-day because of mass screening programs.

If errors of $\pm$ 10 % occur quite often when estimating serum calcium (normal range 9–11 mg/100 ml), it is obvious that one cannot hope to differentiate consistently between a value of 10.5, which is well inside the normal range, and a value of 11.2 which is significantly above normal, since the 10 % error ranges would be 9.45 to 11.55, and 10.1 to 12.3, respectivelly. There is considerable overlapping here, and this situation is illustrated by Fig. 1, in which the triangles are used to simulate the distribution curves of the values which would be obtained by repeated estimations of calcium in the two samples. Diagram No. 1 in Fig. 1 shows the extensive overlapping which occurs when the method is operating within $\pm$ 10 % limits. More appropriate limits, and those usually specified for calcium estimations, are $\pm$ 5 %. Diagram No. 2 shows that in this case there is still some overlapping but much less than with the $\pm$ 10 % limits. In order to eliminate overlapping entirely the method would have to operate within $\pm$ 3 % limits, as illustrated in Diagram No. 3 of Fig. 1. It is obvious that when deciding upon allowable limits of error (A.L.E.) for a method, particular consideration must be given to its range of normal values.

The type of diagram presented in Fig. 1 could be extended to cover the entire range of usefulness of a test. By using a computer to plot accurately the many distribution curves involved, the relative sizes of the common or overlapping areas could be estimated exactly for various concentration levels and A.L.E. values. This information would assist us greatly when trying to decide upon appropriate allowable limits of error for an analytical method.

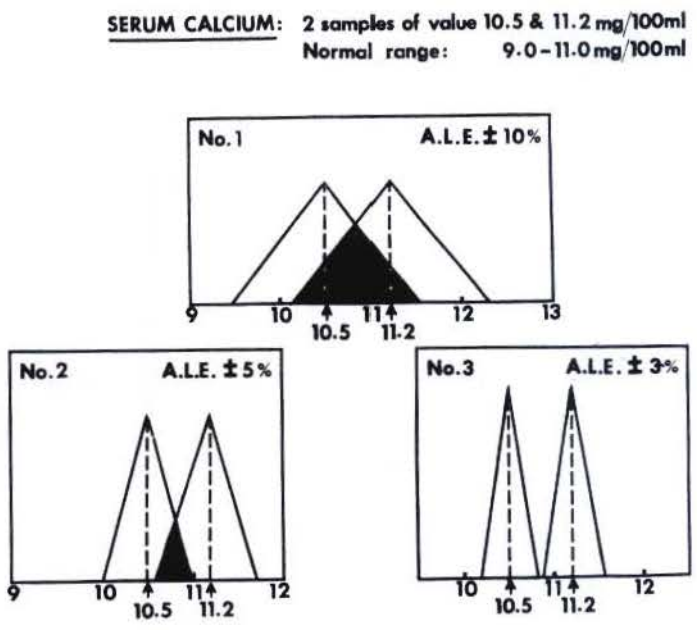

Fig. 1. — Diagrams showing overlapping of results which can occur with various allowable limits of error.

In 1958 the author proposed [2] a general rule or formula for establishing allowable limits of error for clinical chemistry estimations, to be used initially for a proficiency study of 170 Canadian laboratories [3]. It was based on the proposition that in order to distinguish between normal and abnormal values at the borderlines of the normal range, the allowable limits of error in concentration units must not be greater than $\frac{1}{4}$ of this range. On this basis, the following empirical formula was established for calculating maximum allowable limits of error:

$$\text{A.L.E. (in \%)} = \pm \frac{\frac{1}{4}\,(\text{normal range})}{\text{mean of normal range}} \times 100.$$

For example for *serum chloride*, normal range 98–108 mEg/l,

$$\text{A.L.E.} = \pm \frac{\frac{1}{4}\,(108-98)}{103} \times 100 = \pm\, 2.4\ \%$$

The factor of $\frac{1}{4}$ was chosen by taking into consideration the desired relationship between the widths of the normal ranges and the required clinical sensitivities of a number of common estimations. (Here by sensitivity I mean the ability of a method to differentiate between two concentrations which are quite close to each other but nevertheless have different clinical significances).

Other numerical factors of course can be used in the formula, which could be thus adapted for different estimations. The effect of changing the factor, and thus the specified allowable limits of error, is illustrated in Fig. 2. With a factor of $\frac{1}{2}$ the limits of $\pm$ 25 % are actually wider than the

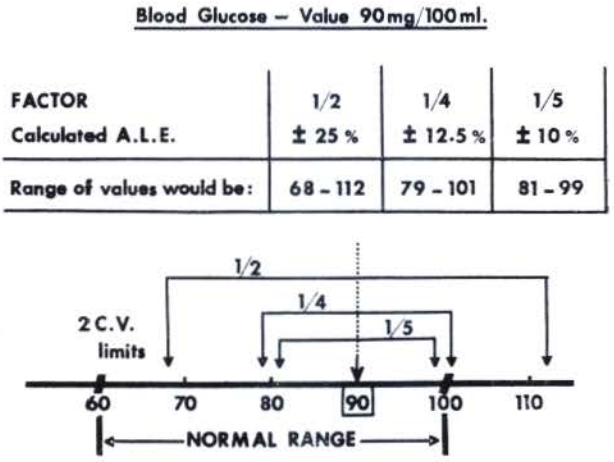

Fig. 2. — Effect of using different factors in the Tonks formula.

normal range — this of course would be entirely unacceptable. With a factor of 1/4 or 1/5, the limits of $\pm$ 12.5 % and $\pm$ 10 %, respectively, encompass about half of the normal range, making it possible to differentiate between some normal and abnormal values. One should realize of course that the distribution of values would be heaviest at the center (that is, at value 90) following the usual distribution curve patterns.

For the proficiency study of Canadian clinical laboratories referred to above, it was specified that the A.L.E. method should not exceed $\pm$ 10 %. A list of allowable limits of error was thus established, as listed in Table 1. The values obtained in the survey were classified as acceptable or unacceptable, depending upon whether they fell within or without the acceptable range: target value $\pm$ A.L.E. Table 1 gives the percentage of unacceptable results for each estimation and sample. No serious objections were ever made to the criteria of acceptability used for this survey even though the study has been referred to many times. In fact, very similar limits have been used for several other proficiency studies.

TABLE 1

1960 survey of Canadian laboratories
Percentage of unacceptable values for the various determinations

COMPONENT	A. L. E.	TARGET VALUES		% UNACCEPTABLE	
		Spec. A	Spec. B	Spec. A	Spec. B
Glucose (tot. red. subs.) . .	± 10%	100	220	27.2	19.1
Glucose (true)	± 10%	86	197	34.2	20.0
Inorg. P.	± 10%	3.7	8.0	22.0	22.4
Protein	± 7.0%	6.9	4.4	40.0	71.5
Sodium	± 2.0%	139	126	53.5	53.7
Chloride	± 1.8%	102	89	41.8	75.0
Urea N.	± 10%	12.1	29.5	55.0	47.6
NPN	± 10%	27	47	74.5	47.1
AVERAGE . . .	—	—	—	43.6%	44.6%
Cholesterol (S & S method)	± 10%	—	90	—	39.0
($FeCl_3$ method)	± 10%	—	103	—	53.9
(Bloor method)	± 10%	—	106	—	80.5
(Direct L–B)	± 10%	—	106	—	83.5
AVERAGE . . .	—	—	—	—	61.7%

The formula is now used in my laboratories to calculate allowable limits of error for use as preliminary control limits for methods being brought into our quality control program. Only after a method has been operating within these preliminary control limits for some time, with the control values plotted on the control chart being evenly distributed about the zero line, and 31–40 valid, consecutive control values have been obtained, is the coefficient of variation calculated. From this point on, ± 2 C.V. limits are used to control the method, with the proviso, however, that they must fall within the allowable limits of error previously selected. Considerable

time, even several months, is sometimes required before the statistical limits can be properly estimated. During this time, the A.L.E. for the method can serve usefully as operating control limits for quality control purposes.

After experience has been gained with the method, and several bimonthly calculations of the coefficient of variation have been made, the allowable limits of error may be adjusted to a more practical or desirable level. Adjustments have in fact been made to a considerable number of our A.L.E.'s:

(*a*) It is now realized that with a number of tests, chiefly the enzyme procedures, a precision of $\pm$ 10 % is attainable in only a very few laboratories. Therefore, it is now recommended by the author ([4]) that the maximum A.L.E. for any test be $\pm$ 20 %, rather than $\pm$ 10 % as originally proposed.

(*b*) Comparisons between the allowable limits calculated by the formula and $\pm$ 2 C.V. values estimated from laboratory data, have indicated that sodium, chloride, and calcium analyses are seldom performed well enough to operate within the former limits (i. e. the A.L.E. calculated by formula, namely $\pm$ 2.3 %, $\pm$ 2.4 %, and $\pm$ 5 %, respectively). These have therefore been adjusted upwards to $\pm$ 4 %, $\pm$ 4 %, and $\pm$ 6 % respectively, so that they conform more closely to the actual working limits used in our laboratories and others. However, it is apparent that these adjusted limits are too wide for the tests to have a true clinical significance when single analyses only are performed.

(*c*) It has been possible to lower the allowable limits of error for a number of tests. For example, the calculated A.L.E. of $\pm$ 17 % which was used for a considerable length of time for urea estimations, has now been adjusted to $\pm$ 12 %.

Table 2 lists our presently-recommended allowable limits of error for a number of common estimations, and compares them with those calculated by the formula. It is not the comparisons, however, that I wish to emphasize, but rather the recommended A.L.E. values themselves, which are based on two years of precision measurements from our quality control program. The data for these estimations were all obtained from our quality control program, using multiple unknown control samples prepared by our quality control chemist. The analyses were performed by several technologists by the regular methods, singly, one per day, on many days during a period of at least 12 months. All values obtained were included except those grossly in error (exceeding 3 C.V.). The concentration levels of the control samples were usually within or close to the normal range (except for bilirubin, acid phosphatase, and creatinine estimations where higher levels were used).

TABLE 2

Allowable limits of error for a number of common estimations

SERUM TESTS	A. L. E. % calculated by Formula	Adjusted A. L. E. %
Class I		
Bicarbonate	± 8.8	
P.B.I.	± 17.0	No
Potassium	± 8.8	adjustment
Protein, total	± 7.0	
Class II		(Raised)
Calcium	± 5.0	6.0
Chloride	± 2.4	4.0
Sodium	± 2.3	4.0
Class III		(Lowered)
Bilirubin (A.A.) *	± 33	± 10
Cholesterol (A.A.)	± 17	± 10
Creatinine (A.A.)	± 25	± 10
Glucose	± 12	± 10
Phosphorus	± 14	± 10
Uric acid	± 20	± 10
Urea N	± 17.0	± 12
Amylase (van Loon *et al.*)	± 26	± 20
L.D.H. (A.A.)	± 21	± 20
Phosphatase, acid	± 30	± 20
Phosphatase, alk. (K.A.)	± 31	± 20
SGOT (A.A.)	± 30	± 20

* AutoAnalyser. **Registered trademark** of Technicon.

A brief description (*) of our quality control program will help to clarify some of the above points. In this program, two separate control systems are actually run in parallel. For one system there is used a single, very large lot of control serum having values known to the analysts (usually called a « bench» control). This system is administered by the technicians themselves. For the other, at least 4 small lots (at a time) of control serums whose values are not known to the analysts, are used. The latter system is administered by a special quality control chemist under the author's direction.

(*) Full details are provided in another publication of the author (5).

Each method placed under the program is treated according to an established plan, which is as follows:

1. First the method itself is studied in the laboratory and by referring to pertinent articles concerning it. An official technique is decided upon, and a detailed description prepared. A preliminary test of precision is made by running a few samples in duplicate or triplicate on three days; and if this is satisfactory, the accuracy of the method and the normal range are verified. If there are no obvious faults in the method, it will now be introduced into the laboratory for actual use. These studies of the method are considered to be part of the quality control program.

2. Next, the necessary preliminary preparations are made for placing the method under quality control:

Suitable control samples, « known » and « unknown », are obtained or prepared, and a schedule for their daily use is established. The necessary forms and control charts are readied. Values for the control samples are verified or established in the first 10 days of regular operation, but these values may be changed somewhat later when the technicians have gained more experience with the method.

At this initial stage, Type A control charts (see Fig. 3) are used for both types of controls. The % error of each control value is plotted on the

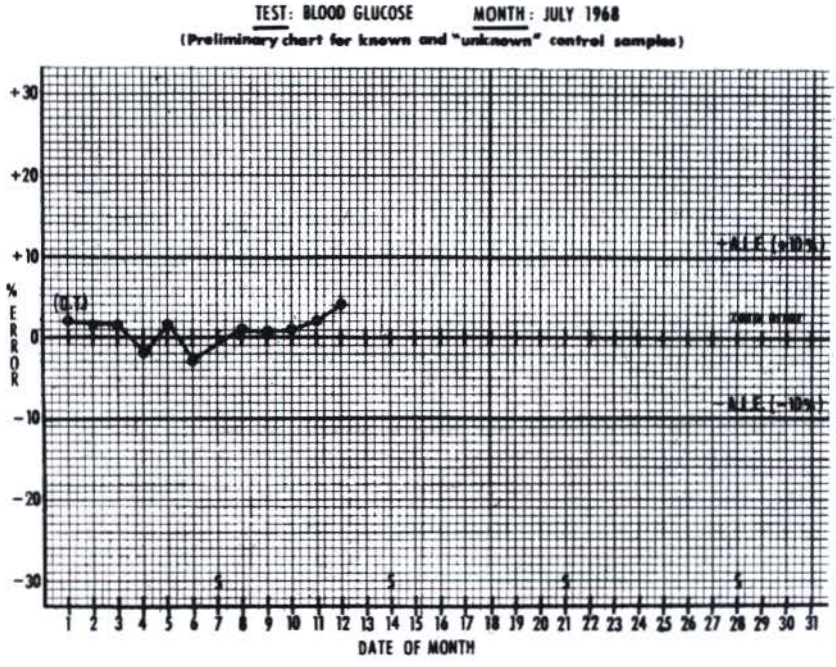

Fig. 3. — Quality control chart Type A.

Q.C. FORM 1 — DAILY Q.C. RECORD FORM — Room No. 720 — Sheet #1 — DATE: June 23/69

TEST	CONTROL TYPE	SAMPLE NAME	TARGET VALUE	VALUE OBTAINED	% ERROR	CONTROL LIMITS: FULLY ACCEPTABLE RANGE	CONTROL LIMITS: WARNING RANGES	CONTROL LIMITS: ALE %	RATING	COMMENTS	ANALYST
Amylase	Bench	Pool	98	108		88-108	78-88 / 108-118	—	✓		J.H.
	Unknown	Versatol E-EH Mix	190	197	+3.6	±10%	±(10-20)%	±20	✓		
Bicarbonate AA	Bench	POOL	18	(21)		17-19	16-17 / 19-20	—	(U)	check!	H.H.
	Unknown	Clinton Mix	20	21	+5.0	±4.4%	±(14-18)%	±9.8	✓		
Bilirubin	Bench	POOL	1.3	1.4		1.25-1.35	1.20-1.25 / 1.35-1.40	—	✓		L.U.
	Unknown	"Calibrate" Mix	5.1	5.3	+3.9	±4.7%	±(4.7-9.4)%	±10	✓		
Calcium	Bench	Versatol	9.8	9.8		9.5-10.1	9.2-9.5 / 10.1-10.4	—	✓		R.W.
	Unknown	"Calibrate" Mix	10.1	10.2	+1.0	±3.0%	±(3.0-6.0)%	±6	✓		
Chloride AA	Bench	POOL	115	(120)		113-117	111-113 / 117-119	—	(U)	check!	H.H.
	Unknown	Clinton Mix	105	108	+2.9	±1.9%	±(1.9-3.8)%	±4	✓		
Cholesterol	Bench	SERACHOL	350	345		340-360	330-340 / 360-370	—	✓		L.U.
	Unknown	Pure solutions	270	255	-5.5	±2.8%	±(2.8-5.6)%	±10	✓	?	
Creatinine	Bench	POOL	1.8	1.9		1.7-1.9	1.6-1.7 / 1.9-2.0	—	✓		L.B.
	Unknown	"Calibrate" Mix	4.5	4.7	+4.4	±3.7%	±(3.7-7.4)%	±10	✓		
Glucose	Bench	POOL	127	126		124-130	121-124 / 130-133	—	✓		A.R.
	Unknown	Pure solution	95	94	-1.1	±2.6%	±(2.6-5.2)%	±10	✓		
LDH	Bench	Versatol EH	313	350		286-340	267-286 / 340-367	—	✓	watch!	N.E
	Unknown	Versatol Mix.	215	247	+15	±8.7%	±(8.7-17)%	±20	✓	?	
Phosphorus	Bench	Versatol	4.0	4.1		3.8-4.2	3.6-3.8 / 4.2-4.4	—	✓		R.W.
	Unknown	"Calibrate" Mix	4.5	4.4	-2.2	±4.1%	±(4.1-8.2)%	±10	✓		

DRS.

Fig. 4. — A daily record form for quality control.

appropriate chart. The allowable limits of error (A.L.E.), calculated by the above formula if the method is a new type, are used as control limits. But only the chart for « known» controls is posted in the laboratory since plotting the values for the « unknown» controls would make it possible for the technician to calculate backwards from the plotted points to obtain their values.

3. Quality control is now started. The control samples are analysed each day along with the patients' specimens, and the % error of each control result is plotted on the appropriate chart. If the control values fall outside of the limits, the necessary corrective action is taken. When the method is operating regularly within the A.L.E., or is working as well as possible; and at least 31 valid control results have been obtained for the unknown controls, the coefficient of variation is calculated from these values.

4. Now, for the « unknown» controls, an acceptable range ($\pm$ 1 C.V.), and warning ranges ($\pm$ 1 C.V. to $\pm$ 2 C.V.) are calculated in terms of concentration units, and are typed on the appropriate form shown in Fig. 4. One of these forms is posted each day on the bulletin board in the

laboratory and the daily control values are marked on it by the analysts as soon as obtained.

For the « unknown» controls, a Type B control chart, shown in Fig. 5, is now posted in the laboratory. The % error of the daily control value is plotted each day by the quality control chemist. No units are marked on this chart — only lines representing the C.V. limits. (These lines are placed on the chart when it is printed). The % error values are plotted approximately « by eye». This one type of printed chart, which is designed specifically to preserve the anonymity of the unknown control values, can be used for all tests regardless of the concentration levels of the control and the width of the control limits. For most methods, we use the $\pm$ 2 C.V. lines as the « action» control limits. Any point outside of these lines is considered to be « outside-limits».

5. Every two months during the first year at least, the coefficient of variation is re-calculated from the past 2 months data for the unknown

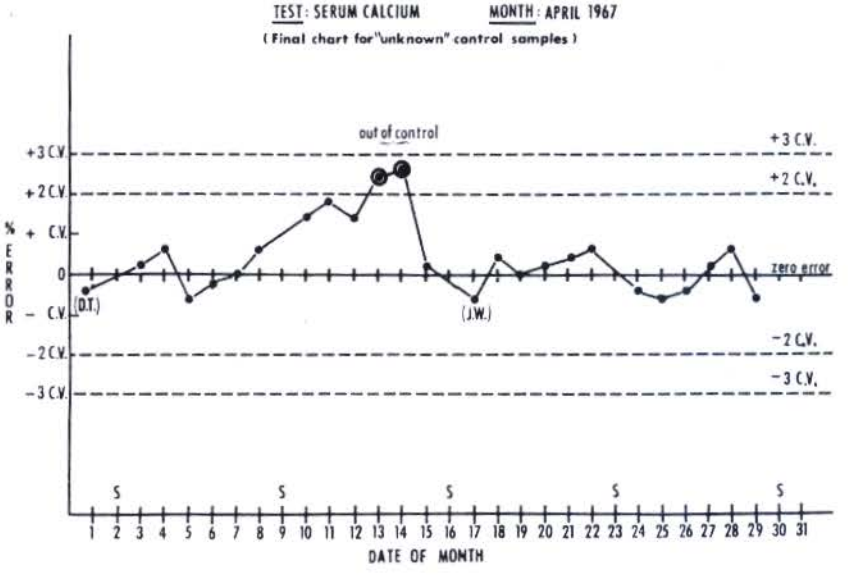

Fig. 5. — Quality control chart Type B.

control samples. A comparison is made each time, using the printed form shown in Fig. 6, between the newly-calculated 2 C. V. value and the original allowable limits of error (A.L.E.). Some figures in the 2 C. V. column of the table are circled to indicate that the method had not been sufficiently precise or accurate in the period in question.

A 6 months' average of the coefficient of variations is used to adjust the working control limits, that is, the 2 C.V. limits on the control chart.

A twelve months' average is used as a basis for changing the original allowable limits of error, if this seems to be warranted from practical considerations and provided the method will retain its clinical usefulness when operating within the proposed new limits.

Q.C. FORM 4

BI-MONTHLY SUMMARY SHEET- FOR ALL TESTS

Month(s): Sept. – Oct.

Year: 1969

TEST	SPECIMEN	TYPE OF CONTROL	MEAN VALUE	RANGE OF% ERRORS	MEAN ERROR %	2 C.V.	A.L.E.	RATING, COMMENTS	TECH.
			conc. units	%					
AMYLASE	SERUM	COMM.	120	-9 ↔ 19		(8.2)	±20	U	N.R
BICARBONATE	"	Pure subs.	25	-7 ↔ 32		(20.0)	±8.0	U	J.M.
BILIRUBIN	"	B.I.-Alb.	50	-19 ↔ 20		(11.6)	±10	U	P.R.
CALCIUM	"	Pure subs.	10.0	-9 ↔ 2		4.6	±6	✓	J.K.
CHLORIDE	"	COMM.	100	-1 ↔ 4		3.2	±4	✓	L.D.
CHOLESTEROL	"	"	320	-2 ↔ 9		4.6	±10	✓	D.G.
CREATININE	"	Pure subs.	2.5	-11 ↔ 8		8.0	±10	✓	G.W.
GLUCOSE	"	"	105	-7 ↔ 9		7.6	±10	✓	J.E.
PHOSPHORUS	"	"	4.0	-11 ↔ 6		6.0	±10	✓	J.K
PROTEIN, TOT	"	COMM.	7.0	-3 ↔ 3		3.4	±7	✓	D.G
SODIUM	"	COMM.	136	-9 ↔ 5		6.4	±20	✓	J.E.

Fig. 6. — A bi-monthly summary record form for quality control.

Thus, two kinds of « confidence » limits are finally arrived at for each method:

(*a*) Allowable limits of error, which have now become long-term, desirable limits, and

(*b*) 2 C. V. limits, which are the regular, working, statistical control limits for the method.

In Table 3 these two types of limits are compared with data provided by Straumfjord and Copeland (6) in a survey of U. S. University Hospitals, and with confidence limits used at the University of Minnesota Hospital in 1966 (7). The values for the two hospitals, Montreal General Hospital and University of Minnesota, were compared with the other data in the tables and classified as Satisfactory (S), Unsatisfactory (U), or Good (G). The former hospital has been given two « U » ratings for amylase and sodium; the latter one « U » rating for creatinine.

It is pertinent to this discussion to mention the allowable limits calculated from individual normal ranges by the Japanese clinical chemist, M. Kitamura (8). He studied, for a number of serum components, the physiological variations which occurred in 200 individuals, and thus obtained individual normal ranges. From these individual normals, which have about half the range of the general population normals, he calculated allowable limits of error by the author's formula. These are given in Table 4 and are

TABLE 3

Method performance comparisons using the coefficient of variation

TEST	A.L.E. (Tonks) adjusted	MONTREAL GENERAL HOSPITAL (1967)		UNIV. OF MINNESOTA HOSPITAL (1966)		U.S. UNIV. HOSPS. (*)
		± 2 C.V.	Rating *	± 2 C.V.	Rating *	± 2 C.V.
	± %	± %		± %		± %
Amylase	20	31.3	U	16.7	S	16.2
Bilirubin	10	9.3	S	8.9	G	15.2
Calcium	6.0	6.1	S	5.0	S	6.6
Chloride	4.0	3.8	S	2.2	S	2.0
Cholesterol	10	5.6	S	5.0	S	10.2
Creatinine	10	7.4	G	11.2	U	15.6
Glucose	10	5.2	S	6.0	S	6.2
LDH	20	17.4	S	6.0	G	—
Phosphatase, acid	20	14.8	S	10.2	S	—
Phosphatase, alk.	20	18.0	S	15.4	S	21.2
Phosphorus, inorg.	10	8.1	S	5.6	G	9.8
Potassium (A. A.)	8.8	6.7	S	4.0	S	5.0
Potassium (Man)	8.8	4.2	S	—	—	—
Protein (TP)	7.0	4.2	S	4.5	S	5.2
Sodium (A. A.)	4.0	4.1	U	2.4	S	2.2
Sodium (Man)	4.0	2.8	S	—	—	—
Urea N	12	11.4	S	8.2	G	11.6
Uric acid	10	9.6	S	6.1	S	5.5
SGOT	20	15.9	S	—	—	23.6

* S = Satisfactory; U = Unsatisfactory; G = Good.

TABLE 4

Kitamura's limits for individual normals

CONSTITUENT	A.L.E. (2 C.V.)	CONSTITUENT	A.L.E. (2 C.V.)
Sodium	± 1.6%	Urea N	± 11.6%
Potassium	± 5.0%	Uric acid	± 6.4%
Calcium	± 3.4%	Cholesterol (tot.)	± 5.0%
Chloride	± 1.8%	Phosphatase (alk.)	± 9.2%
Phosphorus, inorg.	± 10.4%	SGOT	± 15.2%
Protein, total	± 3.6%	Amylase	9.0%

of course considerably narrower than those calculated from general population normals. Kitamura has thus pointed out very well that the precision of many of our methods will have to be improved if we are to be able to use individual normal ranges as a basis for the diagnosis of disease.

Summary. — One of the essentials for clinical laboratory data is that they be sufficiently precise to differentiate between concentrations which are in the normal range and those which fall just outside of this range. This obviously cannot be done if the coefficient of variation is very large in relation to the width of the normal range. A useful rule-of-thumb is that allowable limits of error should not exceed ¼ of the normal range. The application of this rule to a quality control system is discussed; and a comparison is made between limits established from the rule and by statistical calculations from actual laboratory data. The quality control system used by the author is described briefly.

Riassunto (*Limiti fiduciari delle analisi di laboratorio e loro rapporto con i valori normali*). — I dati del laboratorio clinico debbono avere una precisione sufficiente per differenziare le concentrazioni normali dalle concentrazioni appena al di fuori della norma. Ovviamente, ciò si può ottenere solo se il coefficiente di variazione del metodo non è eccessivo in rapporto all'ampiezza dell'intervallo normale. Un'utile regola approssimativa è che i limiti di errore non oltrepassino ¼ dell'intervallo normale. Si discute l'applicazione di questa regola ad un sistema di controllo di qualità; si confrontano inoltre i limiti stabiliti mediante questa regola con quelli calcolati statisticamente in base ai dati reali ottenuti in laboratorio. Si descrive infine brevemente il sistema di controllo di qualità impiegato nel laboratorio dell'autore.

REFERENCES

(1) CAMPBELL, D. G. & W. ANNAN. *J. Clin. Pathol.*, **19**, 513 (1966).

(2) TONKS, D. B. Quality control in clinical chemistry laboratories. Prepared for Quality Control Committee of the Canadian Society of Clinical Chemists. 1958.

(3) TONKS, D. B. *Clin. Chem.*, **9**, 217 (1963).

(4) TONKS, D. B. *Z. Anal. Chem.*, **243**, 760 (1968).

(5) TONKS, D. B. *Quality control in clinical laboratories*, 1970 Edition. Published by Diagnostics Reagents Division, Warner-Chilcott Laboratories Co. Limited, 2200 Eglington Avenue East, Scarborough, Ontario, Canada.

(6) STRAUMFJORD, J. V. & B. E. COPELAND. *Am. J. Clin. Pathol.*, **44**, 452 (1965).

(7) HENDRICKS, E. J. & E. F. FRIER. *Postgrad. Med.*, **40**, A–70 (1966).

(8) KITAMURA, M. Director Department of Clinical Chemistry, Toranomon Hospital, Akasaka, Minato-Ku, Tokyo, Japan. Personal communication, Nov. 1, 1969.