Eur. J. Clin. Chem. Clin. Biochem.
Vol. 29, 1991, pp. 247–252

Reference Materials[1])

By *D. Stamm*

Munich

(Received February 1, 1991)

In honour of Professor Dr. Dr. Hannes Büttner on the occasion of his sixtieth birthday

Summary: This paper summarizes the practical experience gained in the Federal Republic of Germany in connection with the development and implementation of a national reference system.

The National Reference System is set forth in:

1. the New Calibration Act 1986 (Neufassung des Eichgesetzes vom 22. Februar 1985. Bundesgesetzblatt 1985, Teil I, 410–422);
2. the New Weights and Measures Regulations 1988 (Eichordnung vom 12. August 1988. Bundesgesetzblatt 1988, Teil I, 1657–1674),
3. the Guidelines of the Federal Medical Association (1988) for Quality Assurance in Medical Laboratories (Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien vom 16. Januar und 16. Oktober 1987. Deutsches Ärzteblatt *85* (11) (1988), A-699–A-712).

Introduction

Before the theoretical basis of a given reference material is considered and the desirable characteristics are specified, a decision must be made about the purpose for which the reference material is to be used. There are two different types of reference materials, with basic differences in the desirable characteristics and uses:

1. Calibration materials
2. Control materials

The Basic Requirements for the use and properties of calibration materials and control specimens are:

1. *Calibration materials* and *control specimens* must be used *completely independently* of one another.
2. *Calibration materials* should be as highly purified and well defined as the state of the art allows.
3. The matrix of the *control specimens* should be as much like the matrix of the specimens to be examined as possible so that procedural deficiencies and interference factors due to the matrix of the specimens can be monitored.

This paper deals mainly with control materials, including a discussion of the types of errors that can occur if control materials are used for calibration; such errors can have quite serious consequences.

Over the past decade assigned values have been determined in more than 500 commercial control materials (4) according to a plan specified in detail in regulations that have the status of law. The results show that assigned values obtained for the same quantity in the same control specimen, but with different analytical principles and analytical systems, can differ markedly.

[1]) Based on a lecture given at the Symposium "Reference Methods in Clinical Chemistry – Objectives, Trends, Problems" of the Congress Biochemische Analytik 90, München, May 8, 1990

It could be demonstrated that these differences are of diagnostic and therapeutic relevance, because they can lead to sizeable percentages of false classifications of analytical results from patients. Hence the question arose of what could be done to correct this situation, which was unacceptable for both patient care and clinical research.

Where technically possible, the solution is to use only analytical methods that yield results which are good approximations of the "true value." By definition, the best approximations of the true value are obtained with Definitive Methods, and for the practical problems associated with control specimens, with Reference Methods. These facts served as the basis for development of a New Concept for quality control of clinical laboratory findings.

New Concept

In the New Concept, the principle underlying the basic programme (tab. 1) was retained. However, the following changes were made (3, 5):

1. *Accuracy control* is now performed by comparison of the analytical results with *method-independent reference method values*. These *reference method values* have replaced *method-dependent* assigned values.
2. The *maximum allowable deviations* from the reference method values are specified with reference to the medical requirements.
3. The *maximum allowable imprecision* for precision control is specified in an analogous manner.
4. The *decision limits for the interlaboratory surveys* are specified according to considerations analogous to those for accuracy control.

For implementation of the New Concept the following were necessary:

1. Reference methods
2. Specification of the medical requirements

The reference methods are discussed later in this symposium by experts on this subject.

The *medical requirements* have been formulated as follows.

The maximum allowable imprecision and the maximum allowable inaccuracy of the analyses should be selected in such a way that the resulting *proportion of incorrect classifications* in the medical assessment of analytical results is as small as possible.

Tab. 1. Basic programme acc. to the Guidelines (3)

1. Internal quality control
   a. *Control of precision*
   at the most frequent decision limit, by analysing samples of the same control specimen in every run of analyses.
   b. *Control of accuracy*
   over the whole clinically relevant range of measurement, by analysing an accuracy control specimen in every 4th run of analyses; the control specimen is selected from a relatively large number of different control specimens kept on hand.
2. External quality control
   in the form of short-term interlaboratory surveys, using two control specimens with different concentrations.

In the medical assessment of analytical results the percentage of false classifications resulting from imprecision and inaccuracy is to be kept as small as possible.

An estimate must be made of the percentage of false classifications to be expected.

This estimate is based on the specifications for effective quality control of the analytical procedure.

For the different types of medical assessment — transverse assessment (that is, comparison with a reference interval, a decision limit or a therapeutic range) and longitudinal assessment — a set of rules was developed for specifying the medical requirements.

These medical requirements can be met if routine methods are used that are of proven reliability; some comparisons of reference method values and assigned values obtained with routine methods are provided as documentation of reliability (tab. 2).

An *unsolved problem is the vial-to-vial variability* seen with all lyophilized control specimens, which can vary markedly for different quantities. This vial-to-vial variability must be assessed very carefully in the quality control of the production of control materials, and must be taken into account in the protocol for determining reference method values.

If control materials are used for calibration of analytical methods, considerable uncertainties can be expected (even if the reference method value for the analyte to be measured is used), because the measurement signal can contain sizeable unspecific components from the matrix that do not come from the analyte.

Vial-to-vial variability

In the course of our determinations of reference method values in control specimens we have had some very surprising results.

Tab. 2. Comparison of reference method values and assigned values in lyophilized control specimen reg. No. 52600.

Quantity (Unit)	Reference method value	Assigned value		Method
		Absolute	Fraction of reference method value	
S-calcium (mval/l)	4.43	4.45	1.005	Atomic absorption
		4.35	0.982	Flame emission with lithium-guideline
S-chloride (mval/l)	101.3	100	0.987	Coulometric
S-cholesterol (mg/dl)	123.85	120	0.969	Cholesterol oxidase *p*-aminophenazone reaction
S-glucose (mg/dl)	93.35	95.5	1.023	Hexokinase reaction
		91.5	0.980	Glucose-dehydrogenase reaction
S-uric acid (mg/dl)	5.485	5.3	0.966	Uricase/aldehyde-dehydrogenase reaction
		5.5	1.003	Enzymatic colour test
S-potassium (mval/l)	4.38	4.35	0.993	Flame emission with lithium-guideline
S-creatinine (mg/dl)	1.395	1.4	1.004	Enzymatic *p*-aminophenazone reaction
		1.5	1.075	*Jaffé* reaction, kinetic
S-magnesium (mval/l)	1.811	1.825	1.008	Atomic absorption
S-sodium (mval/l)	143.8	143.5	0.998	Flame emission with lithium-guideline
S-total protein (g/dl)	5.895	5.95	1.009	Biuret method *without* consideration of sample blank
		5.8	0.984	*with* consideration of sample blank
S-total glycerol for triacylglycerols (mg/dl)	11.52	10.8	0.938	Enzymatic after KOH saponification
		11.3	0.981	Enzymatic, *p*-aminophenazone reaction

For example, three reference laboratories determined serum glucose with the hexokinase/glucose-6-phosphate dehydrogenase procedure, a National Glucose Reference Method (6) established by the CDC, Atlanta, USA in 1976. Each laboratory made 9 determinations for each of 10 vials, and each laboratory followed the protocol very carefully. Even the results from a single laboratory (tab. 3) are surprising. Thus,

1. the mean values for the vials range from 103.05 to 106.84 mg/dl;
2. analytical imprecision, expressed as the coefficient of variation, is quite different for the different vials, ranging from 0.55 to 0.96%.

Tab. 3. Vial-to-vial variability
S-glucose (mg/dl): 9 analyses per vial with National Reference Method, USA.

Vial No.	Mean (mg/dl)	CV_A (%)
1	103.05	0.73
2	104.90	0.96
3	105.43	0.88
4	103.79	0.95
5	105.35	0.86
6	106.54	0.86
7	106.32	0.55
8	106.84	0.82
9	105.42	0.62
10	106.10	0.69

These results were a compelling argument for us to undertake a systematic study of the vial-to-vial variability of the reference method values in at least two lots of control specimens, and the possible causes of this variability. The results of these investigations were then to serve as the basis for an experimental design to determine reference method values that would be representative of an entire lot of a control specimen.

As possible causes of vial-to-vial variability, we looked at the variability in the amount of material put into each vial and at changes in the quantities dependent on the time of filling and the position in the lyophilizer (tab. 4).

Tab. 4. Factors studied
as possible causes of vial-to-vial variability, and suitable quantities (in brackets).

1. Variability in the amount filled [S-sodium]
2. Change in quantity to be measured [S-glucose, S-bilirubin, S-creatine kinase, S-aspartate aminotransferase (S-glutamate oxaloacetate transaminase)] related to
   2.1 Time of filling
   2.2 Position in lyophilizer

Quantities suitable for assessment of these effects are shown in square brackets.

As the sampling plan shows (tab. 5), 6 vials each were selected from 36 locations and clusters of 20 vials each from 2 locations.

The protocol for analysis can only be outlined here (tab. 6).

The results were evaluated with respect to both the order of the analysis numbers and the order of the sequence numbers, i. e. of time of filling and position in the lyophilizer (tab. 7).

The following is a summary of the results.

1. There is a *vial-to-vial variability* which is different for different analytes and which cannot be explained in terms of imprecision of measurement.
2. For a given quantity this *vial-to-vial variability* is independent of filling time and position in the lyophilizer.
3. Therefore two clusters of 20 vials each yield the same information on vial-to-vial variability.

Tab. 5. Sampling plan.

1. 36 locations, 6 vials each = 216 vials
2. 2 locations, a cluster of 20 vials each = 40 vials
Associated with each sample and cluster is
– a specific filling time
– a specific position in the lyophilizer.
Each vial has
– a *sequence number* (reflecting order of sampling)
– an *analysis number* (assigned at random and indicating order of analysis).

Tab. 6. Protocol for analysis.
Analyses are performed in the order of the analysis numbers.

This protocol excludes:
1. Errors occurring during reconstitution
2. Bacterial or other contamination
3. Changes before or during the analysis
5 analyses per vial
Calculation of mean and standard deviation for each vial
Precision control with pools of 5 vials of lyophilized material; S-sodium and S-glucose with additional liquid control material

Tab. 7. Evaluation of the analytical results.
The characteristics listed below were examined.

1. *In the order of analysis*
1.1 Vial means
1.2 Standard deviation for each vial
1.3 Analysis of outliers
2. *In the order of filling*
2.1 Distribution of vial means
2.1.1 in relation to position *and* filling time
2.1.2 in relation to filling time
2.1.3 in relation to position
2.1.4 for the clusters
2.1.5 for all samples
2.1.6 and comparison with clusters

Table 8 shows the extent of the vial-to-vial variability for 6 lots of specimens; 4 of these are control specimens and 2 are very commonly used calibration specimens from highly regarded manufacturers, but purchased from a retail supplier, not directly from the manufacturer.

After a look at these figures, the question arises of the range of mean values that can be expected, expressed as a percentage of the location parameter, for each of the vial-to-vial variabilities found (tab. 9).

We then estimated the chances that a participant in an interlaboratory survey would not get a certificate for a given quantity because of vial-to-vial variability. We made our estimate, assuming the maximum allowable imprecision as specified after consideration of the medical requirements.

As a rule, the figure is no more than 0.3%. A publication on the model used and some realistic estimates is in preparation.

Problems associated with the use of calibration specimens

Two types of problems arise.

1. *Influence of vial-to-vial variability*
   Let us assume that a completely specific analytical method is used. Even then, as a result of the vial-to-vial variability that we found in calibration specimens from highly regarded manufacturers, and depending on the size of the lot, the range of concentrations around the vial mean can vary. In a lot of 100 000 vials, for example, serum glucose values can be expected to have a range of 9%.
2. *Effect of the matrix on the "total measurement signal"*
   The matrix component of the measurement signal (fig. 1) can have an additive effect, yielding a result greater than the "true value", or quench phenomena or cross reactions can result in a signal less than the "true value." In other words, if an unspecific analytical method is used, the measurement is inexact; in addition, the measured property is not characterized unambiguously.

Tab. 8. Coefficients of variation for the vial means (vial-to-vial variability in %).

Material	Analyte				
	Sodium	Glucose	Bilirubin	Creatine kinase	Aspartate aminotransferase (Glutamate oxaloacetate transaminase)
Control specimen					
1	0.64	2.23	1.18	1.83	2.41
2	1.298	1.024	2.47	3.9	1.419
3	0.73	0.68	0.92	4.77	2.07
4	0.83	0.975	0.78	3.17	1.70
Calibration specimen					
A	0.96	1.22	0.92	2.36	1.10
B	0.82	0.885	1.11	1.62	1.30

Tab. 9. Estimate of the range of the vial means based on the vial-to-vial (vtv) variability for lots of different sizes.

S-glucose		
Location parameter	125.0 mg/dl	
s_{vtv}	1.2825 mg/dl	
CV_{vtv}	1.026 %	
95% range	122.486–127.514 mg/dl (±2.01%)	
Range for		
100 vials	121.78–128.22	(±2.57%)
1 000 vials	120.84–129.16	(±3.32%)
40 000 vials	119.64–130.36	(±4.29%)
80 000 vials	119.37–130.57	(±4.45%)
100 000 vials	119.40–130.60	(±4.48%)

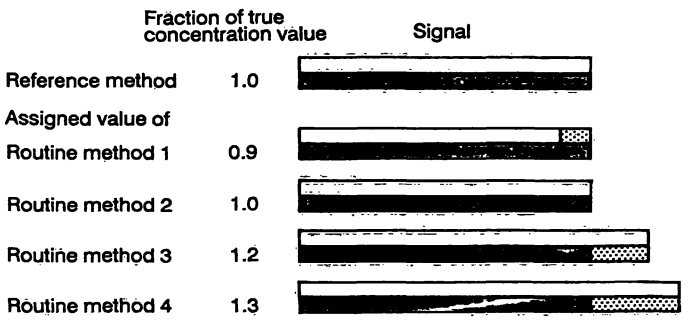

Fig. 1. Components of the signal in a measuring procedure. □ signal fraction resulting from the value of true concentration (■) as influenced by signals due to unspecific constituents (▨)

Consequently, if the reference method value is used in calibration, the following errors result:

$$A_{total} = A_{TV} \pm A_{matrix}$$

where

A_{total} : Total signal

A_{TV} : Signal corresponding to true value

A_{matrix} : Part of signal from unspecific components in the matrix (different from lot to lot)

$$c_P = \frac{c_{RMV\ cal}}{A_{TV\ cal} \pm A_{matrix\ cal}} \times A_{total\ specimen}$$

where

c_{RMV} : Concentration corresponding to reference method value

c_P : Concentration measured in the specimen analysed, dependent on calibrator and analytical method

Example:

If $A_{matrix\ cal} = 0.1 \times A_{TV\ cal}$,
then c_P is roughly 10% too low or 10% too high.

Awareness of these problems and of the extent to which the analytical results can deviate from the reference method value should be enough to prevent critical clinical chemists from using calibration specimens for unspecific analytical methods.

The cholesterol story, which *Gerald Cooper* tells us about elsewhere in this issue (7), illustrates well how precise and accurate analytical results in clinical chemistry must be if they are to serve as the basis for valid clinical laboratory findings, leading to responsible medical action.

References

1. Neufassung des Eichgesetzes vom 22. Februar 1985. Bundesgesetzblatt 1985, Teil I, 410–422.
2. Eichordnung vom 12. August 1988. Bundesgesetzblatt 1988, Teil I, 1657–1674.
3. Richtlinien der Bundesärztekammer zur Qualitätssicherung in medizinischen Laboratorien vom 16. Januar und 16. Oktober 1987. Deutsches Ärzteblatt *85* (11) (1988), A-699 to A-712.
4. Hansert, E. & Stamm, D. (1980) Determination of assigned values in control specimens for internal accuracy control and for interlaboratory surveys. J. Clin. Chem. Clin. Biochem. *18*, 461–490.
5. Stamm, D. (1982) A new concept for quality control of clinical laboratory investigations in the light of clinical requirements and based on reference method values. J. Clin. Chem. Clin. Biochem. *20*, 817–824.
6. Neese, J. W., Duncan, P., Bayse, D., Robinson, M., Cooper, T. & Stewart, C. (1976) Development and Evaluation of a Hexokinase/Glucose-6-Phosphate Dehydrogenase Procedure for Use as a National Glucose Reference Method. Department of Health, Education, and Welfare, Washington, D. C. HEW Publication No. (CDC) 77–8330.
7. Cooper, G. R., Myers, G. L. & Henderson, L. O. (1991) Establishment of reference methods for lipids, lipoproteins end apolipoproteins. Eur. J. Clin. Chem. Clin. Biochem. *29*, 269–275.

Prof. Dr. Dr. D. Stamm
Friedrich-Rein-Weg 21
W-8000 München 60
Bundesrepublik Deutschland