

Reference materials

Part 4. Guide to general and statistical principles for the certification of reference materials

Committees responsible for this Published Document

The preparation of this Published Document was entrusted by the Quality, Management and Statistics Standards Policy Committee (QMS/-) to Technical Committee QMS/16, upon which the following bodies were represented:

British Coal Corporation
 British Gas plc
 Chemical Industries' Association
 Consumers' Association
 Department of Trade and Industry (Laboratory of the Government Chemist)
 Department of Trade and Industry (National Physical Laboratory)
 Institute of Quality Assurance
 Ministry of Defence
 Newcastle Polytechnic
 Royal Society of Chemistry
 University of London
 Water Research Centre

This British Standard, having been prepared under the direction of the Quality, Management and Statistics Standards Policy Committee, was published under the authority of the Standards Board and comes into effect on 15 May 1992

© BSI 1992

Amendments issued since publication

Amd. No.	Date	Text affected

The following BSI references relate to the work on this Published Document:
 Committee reference QMS/16
 Draft 90/97562
 Special announcement in *BSI News* May 1991

ISBN 0 580 20849 4

Contents

	Page
Committees responsible	Inside front cover
National foreword	ii
Guide	
Introduction	1
1 Scope	1
2 Definitions	2
3 The role of reference materials in measurement science	2
4 Measurement uncertainty	4
5 Homogeneity of materials	8
6 General principles of certification	11
7 Certification by a definitive method	12
8 Certification by interlaboratory testing	14
9 Certification based on a metrological approach	21
Annex	
A Bibliography	32
Tables	
1 ANOVA table	20
2 ANOVA table	21
3 Trace elements in milk	24
4 Results of analyses of olive-tree leaves	24
5 Determination of pesticides in powdered milk spiked with certain compounds	24
Figures	
1 An example of results of measurements by two operators using two instruments on eight different days	5
2 Histogram of the frequency (number of ampoules) versus the ratio of the activity of ^{137}Cs standards to a radium reference standard (RRS20)	10
3 Schematic diagram of the process of preparation and certification of an RM by interlaboratory consensus	15
4 Schematic diagram for statistical evaluation of interlaboratory results for certification of an RM	17
5 Experimental scheme for an interlaboratory programme	19
6 Results of the first and second intercomparison of analyses of carbon monoxide in nitrogen	26
7 Evolution of results in successive intercomparisons for the determination of aflatoxin in milk powder	27
8 Results of individual laboratories for oxygen/nickel ratio	28
9 Determination of cortisol in reconstituted human serum	29
10 a) Successive results for the determination of γ -glutamyltransferase in albumin: First and second intercomparisons	30
b) Successive results for the determination of γ -glutamyltransferase in albumin: Final campaign	31

National foreword

This Part of PD 6532 has been prepared under the direction of the Quality, Management and Statistics Standards Policy Committee. It is identical with ISO Guide 35 : 1989 *Certification of reference materials — General and statistical principles* published by the International Organization for Standardization (ISO).

Three Parts of this Published Document have been published and these may be summarized as follows.

Part 2 *Guide to the contents of certificates of reference materials* (identical with ISO Guide 31-1981). This Part provides guidance on the preparation of clear and concise certificates of reference materials that summarize the information needed by the user of the reference material. The information that should be provided on a certificate is listed under a series of headings. Each heading is further explained and examples are provided where necessary.

Part 3 *Guide to the uses of certified reference materials* (identical with ISO Guide 33 : 1989). This Part provides guidance on the uses of certified reference materials (CRMs). The proper uses, and potential misuses, of CRMs are explained. Recommendations are presented for the assessment of the precision and trueness of a measurement process by the use of CRMs. The use of CRMs for the definition and realization of conventional measurement scales is discussed, including temperature, pH, octane number, soluble solids content (refractometric), turbidity, hardness and the Roga test for the classification of hard coals.

Part 4 *Guide to general and statistical principles for the certification of reference materials* (identical with ISO Guide 35 : 1989). This Part provides guidance on the concepts and practical aspects of the certification of reference materials. The roles of reference materials in quantitative measurements are described and important technical aspects of the certification of reference materials are considered including the measurement uncertainty associated with certified values. Commonly used methods for the certification of reference materials are discussed.

A further Part is in the process of being drafted and this is as follows.

Part 1 *Guide to terms and definitions used in connection with reference materials* (identical with ISO Guide 30 : 1992). This Part will recommend terms and the meanings that should be assigned to them, used in connection with reference materials. Particular attention will be given to terms that are used in reference material certificates and corresponding certification reports.

Cross-references

International standard	Corresponding British Standard
ISO Guide 31 : 1981	PD 6532 <i>Reference materials</i> Part 2 : 1992 <i>Guide to the contents of certificates of reference materials</i> (Identical)
ISO Guide 33 : 1989	PD 6532 <i>Reference materials</i> Part 3 : 1992 <i>Guide to the uses of certified reference materials</i> (Identical)

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Certification of reference materials — General and statistical principles

Introduction

The Committee on reference materials (REMCO) is concerned with guidelines for the preparation, certification and use of reference materials. This Guide is intended to describe the general and statistical principles for the certification of reference materials.

Various sections of this Guide were prepared by different delegates to REMCO. The project was co-ordinated with representatives of ISO/TC 69, *Applications of statistical methods*.

Acknowledgment is given to J. D. Cox (BSI, UK) for preparation of the section on the role of reference materials in measurement systems (clause 3). Much of clauses 4, 5 and 6 is based on material contained in three previously published sources:

- a) CALI, J.P. *et al.* The role of standard reference materials in measurement systems, *NBS Monograph 148*, Washington, DC, National Bureau of Standards, 1975 (especially Chapter III, by H. H. Ku);
- b) URIANO, G. A. and GRAVATT, C. C. The role of reference materials and reference methods in chemical analysis. *Crit. Rev. in Anal. Chem.* **6** 1977: 361;
- c) MARSCHAL, A. *Matériaux de référence*. Bureau National de Métrologie, Laboratoire National d'Essais, Paris.

K. R. Eberhardt (ANSI, USA) prepared clause 7 on the use of a definitive method to certify reference materials. R. Sutarno and H. Steger (SCC, Canada) prepared clause 8 on the use of an interlaboratory testing programme to certify reference materials. H. Marchandise (Community Bureau of Reference, Commission of the European Communities) prepared clause 9 on a metrological approach to certification, included for the first time in the second edition of this Guide. G. Uriano (ANSI, USA) served as editor of the Guide.

Special acknowledgement is given to members of ISO/TC 69/SC 6 and its Secretary K. Petrick (DIN, Germany, F.R.), for their co-operation in preparing those sections of the document concerned with the statistical analysis of data. In particular the

many contributions of Prof. P. T. Wilrich (DIN, Germany, F.R.) and Dr. T. Miyazu (JISC, Japan) of ISO/TC 69/SC 6 to the review and editing of the Guide are gratefully acknowledged.

Earlier Guides^[1-3] prepared by REMCO have dealt with the following aspects of reference materials:

- a) mention of reference materials in International Standards;
- b) terms and definitions used in connection with reference materials;
- c) the contents of certificates of reference materials.

The purpose of this Guide is to provide a basic introduction to concepts and practical aspects related to the certification of reference materials. ISO Guide 33^[29] more fully addresses concepts and practical aspects related to the use of reference materials.

1 Scope

According to the definition given in 2.1, reference materials (RMs) may be used in diverse measurement roles connected with instrument calibration, method assessment and assignment of property values. The purpose of clause 3 is to discuss these measurement roles and to show how traceability¹⁾ of measurement may be secured by use of RMs, thus yielding worldwide compatibility of measurement.

Just as certified reference materials (CRMs) are to be preferred over other classes of RMs in citations in International Standards^[1], so also are CRMs to be preferred over other classes of RMs in measurement science generally, given that CRMs needed for a particular type of measurement exist. Assistance in locating the source(s) of supply of CRMs for various technical fields is afforded by ISO's *Directory of certified reference materials*^[4].

It will be evident that the quality of a measurement based on use of a CRM will depend in part on the effort and care expended by the certifying body on determining the property

1) An internationally agreed definition of "traceability" in measurement science is given in reference [5]:

traceability: The property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons.

value(s) of the candidate CRM. Hence the process of certification^[2] should be carried out using well-characterized measurement methods that have high accuracy as well as precision and provide property values traceable to fundamental units of measurement. Furthermore, the methods should yield values with uncertainties that are appropriate to the expected end-use of the CRM. Clauses 4 and 5 deal with two of the most important technical considerations in the certification of RMs — measurement uncertainties and material homogeneity. Clause 6 provides general principles for RM certification.

Two commonly used general approaches to assuring technically valid RM certification are discussed in clauses 7 and 8. Clause 7 describes the use of a single method of the highest accuracy (i.e. sometimes referred to as a “definitive” or “absolute” method) and usually employed by a single laboratory for RM certification. Clause 8 describes the use of an inter-laboratory testing approach to RM certification, which might involve more than one method.

The metrological approach discussed in clause 9 has as its objective the production of certified values the accuracy and uncertainty of which are demonstrated by experimental evidence.

In summary, the purpose of this Guide is to assist in understanding valid methods for the certification of RMs and also to help potential users to better define their technical requirements. The Guide should be useful in establishing the full potential of CRMs as aids to assuring the accuracy and inter-laboratory compatibility of measurements on a national or international scale.

2 Definitions

Definitions of the basic terms “reference material” and “certified reference material” were first put forward in 1977^[1] and were later amended slightly^[2] to read as follows.

2.1 reference material; RM: A material or substance one or more properties of which are sufficiently well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials.

NOTE — An RM may be in the form of a pure or mixed gas, liquid or solid, or even a simple manufactured object. Some RMs are certified in a batch, any reasonably small part of which should exhibit the property value(s) established for the whole batch within stated uncertainty limits. Other RMs exist as individually manufactured objects which are also certified individually. Numerous RMs have properties which, because they cannot be correlated with an established chemical structure or for other reasons, cannot be measured in mass or amount of substance units or determined by exactly defined physical or chemical measurement methods. Such RMs include certain biological RMs (for example a vaccine to which an international unit has been assigned by

the World Health Organization) and certain technological RMs (for example rubber blocks for the determination of abrasiveness or steel plates for the determination of hardness). It is recognized that the definition of “reference material” given above could involve an overlap with the term “material measure” as defined in the *International Vocabulary of Basic and General terms in Metrology*^[5]; consequently, some materials may be characterized as either reference materials or material measures.

2.2 certified reference material; CRM: A reference material one or more of whose property values are certified by a technically valid procedure, accompanied by or traceable to a certificate or other documentation which is issued by a certifying body.

NOTE — A CRM may consist of units which are each certified individually or which are certified by examination of representative samples from a batch.

3 The role of reference materials in measurement science

Metrology is the field of knowledge concerned with measurement. Metrology or measurement science¹⁾ includes all aspects both theoretical and practical with reference to measurements, whatever their level of accuracy, and in whatever fields of science or technology they occur^[6]. This clause describes the role of reference materials in quantitative measurements.

3.1 The role of reference materials in the storage and transfer of information or property values

By definition (2.1), a reference material has one or more properties, the values of which are well established by measurement. Once the property value(s) of a particular RM have been established, they are “stored” by the RM (up to its expiration date) and are transferred when the RM itself is conveyed from one place to another. To the extent that the property value of an RM can be determined with a well-defined uncertainty, that property value can be used as a reference value for intercomparison or transfer purposes. Hence RMs aid in measurement transfer, in time and space, similar to measuring instruments²⁾ and material measures^[6].

A general scheme for constructing a hierarchical measurement system is illustrated in section 6.5 of the *Vocabulary of Legal Metrology*^[6]. The interlinking of various levels and stations within a measurement system via “reference standards” may, in principle, be effected by either measuring instruments or material measures or RMs.

An RM must be suitable for the exacting role it performs in storing and transferring information on measured property values. The following technical criteria (legal or commercial criteria

1) “Measurement science” is therefore synonymous with “metrology” according to the international definition of the latter term^[6]; it should be noted, however, that current usage generally restricts the term “metrology” to physical measurements at high accuracy. The term “metrology” is, however, being increasingly used in the context of chemical, engineering, biological and medical measurements.

2) Some measuring instruments are not readily movable (by reason of size, mass, fragility, instability or cost), in which case the measurand must be brought to the instrument to effect the measurement transfer. But all RMs and material measures are readily movable and thus can be taken to the measurand.

may be relevant also) apply to the fitness for purpose of RMs in general :

- a) the RM itself and the property value(s) embodied in it should be stable for an acceptable time-span, under realistic conditions of storage, transport and use;
- b) the RM should be sufficiently homogeneous that the property value(s) measured on one portion of the batch should apply to any other portion of the batch within acceptable limits of uncertainty; in cases of inhomogeneity of the large batch, it may be necessary to certify each unit from the batch separately;
- c) the property value(s) of the RM should have been established with a precision and an accuracy sufficient to the end use(s) of the RM;
- d) clear documentation concerning the RM and its established property value(s) should be available. Preferably the property value(s) should have been certified, so the documentation should then include a certificate, prepared in accordance with ISO Guide 31^[3].

The word "accuracy" was advisedly used in c) to indicate that whenever possible, the measurement of a given property value should have been made by a method having negligible systematic error or bias relative to end-use requirements (or where the result has been corrected for a known bias) and by means of measuring instruments or material measures which are traceable to national measurement standards. Subsequent use of an RM with traceable property values ensures that traceability is propagated to the user. Since most national measurement standards are themselves harmonized internationally, it follows that measurement standards in one country should be compatible with similar measurements in another country. In many cases, CRMs are appropriate for the intercomparisons of national measurement standards.

3.2 The role of reference materials in the International System of units (SI)

3.2.1 Dependence of the SI base units on substances and materials

The majority of measurements made in the world today are within the framework of the International System of units^[7]. In its present form, SI recognizes seven base units, namely the units of length (metre, symbol m), mass (kilogram, kg), time (second, s), electric current (ampere, A), thermodynamic temperature (kelvin, K), amount of substance (mole, mol) and luminous intensity (candela, cd). The definitions^[7] of these base units mention the following substances: krypton-86¹⁾ (for defining the metre), platinum-iridium (for fabricating the prototype kilogram), caesium-133 (for defining the second), water (for defining the kelvin) and carbon-12 (for defining the mole). Opinions differ as to whether the substances named fall under the definition of reference material (2.1). The use of these substances in basic metrology is consistent with the use of reference materials in other types of measurement applications.

Certainly such materials have a special status as defined substances on which the SI is based. The dependency strictly applies to definition of the unit, since realization of the units may involve other substances/materials. This is especially true in regard to the realization of the mole^[8] and the kilogram.

3.2.2 The realization of derived SI units with the aid of reference materials

From the seven base units an unlimited number of derived units of the SI are obtainable by combining base units as products and/or quotients. For example, a derived unit of mass concentration is defined as $\text{kg} \cdot \text{m}^{-3}$ and the derived unit of pressure (given the special name pascal, symbol Pa) is defined as $\text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-2}$. Formally speaking, the derived units ultimately depend on the substances on which the base units themselves depend (see 3.2.1). In practice, the derived units are often realized not from base units but from RMs with accepted property values. Thus a variety of substances/materials may be involved in the realization of derived units (examples 1 and 2 below) or even of base units (examples 3 and 4 below).

Example 1: The SI unit of dynamic viscosity, the pascal second ($\text{Pa} \cdot \text{s} = \text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-1}$) may be realized^[9] by taking the value for a well purified sample of water as $0,001\,002\,\text{Pa} \cdot \text{s}$ at $20\,^{\circ}\text{C}$.

Example 2: The SI unit of molar heat capacity, the joule per mole · kelvin ($\text{J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1} = \text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$) may be realized^[10] by taking the value for purified α -alumina as $79,01\,\text{J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ at $25\,^{\circ}\text{C}$.

Example 3: The SI unit of amount of substance, the mole, may be realized by taking $0,069\,72\,\text{kg}$ of highly purified gallium metal^[11].

Example 4: The SI unit of temperature, the kelvin, may be realized at any temperature T_1 ($273,15\,\text{K} < T_1 < 903,89\,\text{K}$) from measurements of the resistance of a highly pure platinum wire at T_1 , at the triple point of purified water, at the freezing point of purified tin and at the freezing point of purified zinc, coupled with use of a specified mathematical relation^[12]. The word "thermodynamic" has been deliberately omitted here to avoid controversy over whether thermodynamic temperatures are, or are not, the same as International Practical Temperatures of 1968: the intention of the International Committee for Weights and Measures was to match the two sorts of temperature exactly, within the framework of knowledge available during 1968-1975.

3.2.3 Connection of analytical chemistry to the International System of units

It will be noted that purified (often called "pure") chemical substances were cited in each of the examples 1 to 4 (3.2.2). The measurement of degree of purity, or more generally of the chemical composition of materials, is within the realm of analytical chemistry. In addition to the dependence of SI on chemical substances, the dependence of analytical chemistry on SI is worthy of examination. Presently, most analytical

1) Recently, the General Conference on Weights and Measures redefined the metre as the distance travelled by light in a vacuum during $1/299\,792\,458$ of a second.

chemists employ units within the SI (all base units except the candela and also many derived units) in their measurements. However, compositional analysis depends on an additional concept, namely that pure chemical species exist to which the chemical compositions of other substances and materials are referred, by invoking the laws of chemical change and stoichiometry.

From one or more pure chemical species, considered to be primary measurement standards, it is feasible to construct measurement hierarchies for analytical chemistry similar to those used in physical measurement^[6]. Examples of such measurement standards are :

- a) the electron, to which other species can be connected by electrochemical analysis^[13];
- b) carbon-12, to which other species can in principle be connected by mass spectrometry, Raoult's law measurements, or volumetric measurements with low-density gases, etc.;
- c) a highly purified element or compound, to which other species can be connected by electrochemical, gravimetric, titrimetric, spectrometric methods, etc.

The "other species" cited in these examples will in many cases be used as RMs. Many substances can fill this role of intermediaries between primary and working analytical standards using the diversity of techniques and chemical reactions that an analyst may employ. The concept of traceability applies to analytical chemistry as much as it does to other branches of measurement science. The quality of the result of a chemical analysis will be enhanced if the result's traceability can be clearly stated in terms of the traceability of the instruments, material measures and RMs employed. In most cases, the traceability will also depend on the values of the relative atomic masses (formerly called "atomic weights") used in the calculations; the source of these should be recorded by the analyst (for example [11]).

3.2.4 The role of reference materials in realizing units outside of the SI

Where the components of a measurement system (for example the Imperial system) can be related exactly to the corresponding components of the SI, it is unnecessary to have independent means for realizing the non-SI measurement system. Where the quantities cannot be related to those of the SI, then independent realization of the non-SI units is in principle necessary. In practice, however, few such systems remain in use and thus are mostly historical curiosities.

3.3 Use of reference materials

REMCO intends to publish a separate guide covering general and statistical principles for the use of reference materials. There are very few published documents that address general problems associated with the use of reference materials. The reader is referred to the documents and recommendations published by IUPAC Commission I.4 on Physico-chemical Reference Materials and Standards, which deal primarily with

the use of reference materials for realization of physical properties. The following IUPAC Commission I.4 publications in *Pure and Applied Chemistry* are concerned with the certification and use of reference materials for physical properties :

Physical property	Volume, date of publication and page number	
Enthalpy	40	1974 : 399
Optical rotation	40	1974 : 451
Optical refraction	40	1974 : 463
Density	45	1976 : 1
Relative molecular mass	48	1976 : 241
Absorbance and wavelength	49	1977 : 661
Reflectance	50	1978 : 1 477
Potentiometric ion activities	50	1978 : 1 485
Viscosity	52	1980 : 2 393
Permittivity	53	1981 : 1 847
Thermal conductivity	53	1981 : 1 863

4 Measurement uncertainty

In discussing measurement uncertainties, the terms "precision", "systematic error or bias", and "accuracy" are usually used. The meanings of these terms are not rigidly fixed, but depend to a large extent on the interpretation and use of the data^[14, 15].

4.1 An illustrative example

If two equally trained operators, A and B, each make four replications of a measurement on a uniform material each day for 4 days on one instrument, and 4 days again on a similar instrument, the results, 16 sets of four measurements, may look like those in figure 1. What can be seen from this plot ?

- a) the spreads among each set of four values are comparable, perhaps slightly smaller for instrument 2 than instrument 1;
- b) there appears to be more variability between daily results than within sets of daily results, particularly for instrument 1;
- c) operator B gives lower results than operator A;
- d) instrument 1 gives lower results than instrument 2.

Figure 1 is constructed for the purpose of demonstration, and actual measurements could be better or worse than shown. However, this plot does show some four types of factors that contributed to the total variability of these measurements :

- 1) factors acting within days;
- 2) factors acting between days;
- 3) factors due to instrument systems;
- 4) factors due to operators.

Appropriate techniques are available for the separate estimation of the effects of these four factors and standard deviations could be computed corresponding to each of them. However, the limited number of operators and instruments prevents the computation of standard deviations as reliably for

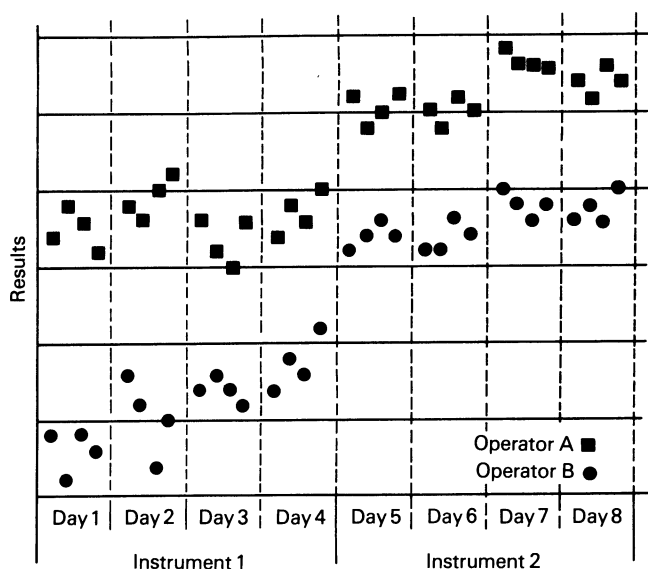


Figure 1 — An example of results of measurements by two operators using two instruments on eight different days

factors 3) and 4) as for factors 1) and 2). The time and work involved certainly impose limits on any efforts to do so.

The failure to allow for factors relating to instruments and operators is one of the main causes for the unreasonable differences usually encountered in interlaboratory, or round-robin, types of tests^[16]. Because instruments vary from time to time and operators change, the result from a laboratory at a given time represents only one of the many results that could be obtained, and the variability caused by these two sources must be considered as part of the precision of the laboratory. The standard deviation computed without regard to these effects would underestimate the true variability.

If, by the proper use of standards and reference methods^[17], these two sources of errors were eliminated, the standard deviation computed from the 16 means of sets of four measurements would be the proper measure of precision. Presumably the grand mean of the 16 mean values would be reported.

The mean of many values is more stable than individual measurements. When extraneous sources of variation, such as instrument and operator effects, are eliminated, the relationship between the standard deviation of individual measurements and the standard deviation of the mean of n such measurements can be expressed as

$$\sigma(\bar{X}_n) = \frac{\sigma(X)}{\sqrt{n}} \quad \dots (1)$$

In other words, the standard deviation of the mean is smaller than the standard deviation of individual measurements by a factor of $1/\sqrt{n}$. One important provision must hold for this relationship to be true, i.e. that the n measurements are independent of each other. "Independence" can be defined in a probability sense, but for present purposes, measurements may be considered independent if they show no trend or pattern. This is certainly not true in figure 1, and to say that the

standard deviation of the mean of all 64 values is $1/8$ ($= 1/\sqrt{64}$) of the standard deviation of individual measurements would seriously underestimate its true variability. Moreover, the relationship in equation (1) is expressed in terms of the true value of the standard deviation, σ , which is usually not known. As the computed standard deviation, s , is itself an estimate of σ from the set of measured values, the standard deviation of the mean in equation (1) is only approximated when s is used in place of σ .

The use of the standard deviation computed from daily averages rather than individual values is preferred because the former properly reflects a component of variability between days, or over time, which is usually present in precision measurement.

4.2 Some basic statistical concepts

The basic information available on the measurement errors is summarized by:

- the number of independent determinations or the number from which a mean was computed and reported;
- an estimate of the standard deviation, s , defined by

$$s = \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}$$

where n measurement results are denoted by $x_1, x_2, \dots, x_n$, and their mean is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

From a) and b) several useful derived statistics can be computed:

- standard deviation of the mean of n measurements

$$s(\bar{x}_n) = \frac{s}{\sqrt{n}}$$

This is sometimes called the standard error of the mean to differentiate it from the standard deviation of individual determinations.

NOTE — As n becomes large, the value of $s(\bar{x}_n)$ becomes very small, showing that the average of a large number of measurements approaches a constant value μ which is usually the objective of the measurement procedure.

- confidence interval for the mean (normal distribution). Each time n measurements are made, a value of the average of the measurements is reported. These averages will differ from time to time within certain limits. Assuming a normal distribution, one interval of the type $\bar{x} \pm \delta$ can be constructed^[18] such that the interval from $\bar{x} - \delta$ to $\bar{x} + \delta$ will

be fairly certain to include the value of μ desired. The interval is computed by:

$$\delta = t \frac{s}{\sqrt{n}} \quad \dots (2)$$

where t is a tabular value of the Student distribution, and depends on the confidence level and the degrees of freedom for s ;

e) 2-sigma (or 2s), 3-sigma (or 3s) limits. These limits describe the distribution of measurement error. If a measurement is made by the user of a CRM having the same precision (i.e. same σ) as that obtained by the certifying laboratory, his measurements should fall (with probability approximately 0,95 to 0,997) within these limits when σ is well-established. Otherwise there is evidence of systematic difference.

4.3 Instrument and operator errors

Instrument and operator types of errors have not yet been treated. An ideal situation would be to eliminate them from the measurement process, or to use more instruments and more operators and then estimate standard deviations associated with these sources. When neither of the above is feasible or practical, the least that can be done is to use two instruments and/or operators. If the confidence intervals for the mean results of the two instruments do not overlap, then there is good evidence of instrument difference.

Using his experience and judgement, a measurement scientist may arrive at reasonable bounds for these types of errors. If the bound is not computed from measurement data, then its validity cannot be supported by statistical analysis. In such cases, these bounds are "guesstimates" and the only recourse is to treat them as limits to systematic errors.

The detection of differences and the separation of the total variability into its identifiable components can be facilitated through careful planning and statistical design of the experiment.

4.4 Differences among measurement methods

Each measurement method purports to measure the desired property of a material, but seldom does a method measure the property directly. In most cases the method actually measures some other property that is related to the property by theory, practice, or tradition, and then converted to the value of the desired property through these relationships. Discrepancies among results of different measurement methods are common, even for measurements leading to the determination of fundamental physical constants^[19].

In the preparation of a CRM, usually two or more measurement methods are employed for each property measured. If these methods are well established by virtue of past experience, the results yielded by these methods usually agree to within the uncertainty assigned to each method.

In a few cases these differences are so large that the results cannot be reconciled, and these results are then reported

separately for each individual method. The RM is either not certified or certified on a method-dependent basis. A historical example of this type of reporting is NBS CRM 1091, Stainless Steel. The nitrogen content was measured by vacuum fusion and pressure bomb-distillation, and gave results of 861 and 945 mg/kg, with standard deviations of 3 and 20 mg/kg, respectively. Clearly one or both methods have a systematic error that is large compared to the variability of material or the measurement uncertainty. A report of the average of the two methods would be highly misleading.

Measurement accuracy in its absolute sense is never realized. In practice, certified values of some reference materials are defined by using a referee method or assigning a value by a well-defined procedure so that at least the same benchmark will be used by everyone in the field. The importance of reference methods to supplement the use of these measurement standards is also being emphasized^[17]. A good example is the reference method for blood haemoglobin and the value assigned as a benchmark to the reference material issued by the International Committee for Standardization in Hematology (ICSH)^[20, 21].

4.5 Uncertainties of certified values

The uncertainty of a CRM value is usually made up of several components, some supported by data and some not:

- a statistical tolerance interval giving bounds to material inhomogeneity based on data and statistical computations;
- a confidence interval for the mean giving bounds to measurement error based on data and statistical computations;
- components of measurement uncertainty due to variation among laboratories and/or operators and measurement methods;
- a combination (addition of absolute values or the square root of the sum of the squares) of estimated bounds to "known" sources of possible systematic error based on experience and judgement (in other words, there are no data, or an insufficient number of data, to make a statistical calculation).

The word "known" is quoted above to contrast with systematic errors that are "unknown" or unsuspected. These unsuspected errors could occur in a number of ways — a component in the physical system, a minor flaw in the theoretical consideration, or the rounding error in a computation. As more homogeneous materials become available, and more precise measurement methods are developed, these types of errors will be detected by design or by chance and hopefully will be eliminated. Improved accuracy in the measurement of a property is basically an expensive iterative process and unwarranted demand for accuracy could mean the waste of resources.

4.6 Statements of uncertainty on CRM certificates

A variety of statements of uncertainty can be found in past and current certificates issued for CRMs around the world. Some of these statements are well formulated and supported by data,

others are not; some of these statements contain a wealth of information that is useful to exacting users, but overwhelming to others; some statements are oversimplified with a resulting loss of information. Because the originator of a CRM has to keep all classes of users in mind, the use of a single form of statement is not usually possible. The intention is that all these statements are unambiguous, meaningful, and contain all the information that is relevant for potential users.

Some commonly used statements, taken from existing certificates, are listed in 4.6.1 to 4.6.4.

4.6.1 Example 1: 95 % confidence limits for the mean

Rubidium chloride

Absolute abundance ratio $2,593 \pm 0,002$

“The indicated uncertainties are overall limits of error based on 95 % confidence limits for the mean and allowances for the effects of known sources of possible systematic error.”

Because the isotopic ratio is a constant for a given batch of material and is not subject to errors of material inhomogeneity, the 95 % confidence limits for the mean refer to measurement error only. This is computed from

$$t \frac{s}{\sqrt{n}}$$

as described in equation (2).

The effects of known sources of possible systematic error are discussed in detail in “Absolute isotopic abundance ratio and atomic weight of terrestrial rubidium” [22].

4.6.2 Example 2: 2-sigma or 3-sigma limits

Glass Filters for Molecular Absorption Spectrometry

Absorbance $0,500 \pm 0,002 \ 5$

“This uncertainty is the sum of the random error of $\pm 0,1$ % relative (2σ limit) and of estimated biases which are $\pm 0,4$ % relative.”

Each glass filter was individually calibrated, and the standard deviation refers to measurement error, including the cleanliness of the surface. As these glass filters will be used time after time, a multiple of the standard deviation is a proper measure of variability.

4.6.3 Example 3: Uncertainty expressed in significant digits

AISI 4340 Steel

Element Mass Fraction

Carbon $3,8_2 \times 10^{-3}$

Manganese $6,6 \times 10^{-3}$

According to the explanation given in the text: “The value listed is not expected to deviate from the true value by more

than ± 1 in the last significant figure reported; for a subscript figure, the deviation is not expected to be more than ± 5 .” Thus, the mass fraction of carbon, expressed as a percentage, is between 0,377 and 0,387; and that for manganese is between 0,65 and 0,67. These uncertainties include material inhomogeneity, measurement imprecision, and possible bias between laboratories and implicit rounding, because these values are “... the present best estimate of the true value based on the results of a co-operative interlaboratory analytical programme.”

When 20 to 30 elements are to be certified for one material, this method gives a concise and convenient summary of the results. As these limits are expressed in units of 5 and 10, some information is unavoidably lost for some of the elements. However, when the certified value is used, it is important to use all of the digits given including the subscripts. The uncertainty stated on this certificate depends heavily on the use of chemical judgement.

4.6.4 Example 4 : Standard deviation, and number of determinations

Method	Oxygen in ferrous metals ($\mu\text{g/g}$)		
	CRM A (Ingot iron)	CRM B (Stainless steel : AISI 431)	CRM C (Vacuum melted steel)
Vacuum fusion			
$\bar{x}$	484	131	28
s	14	8	2
n	216	286	105
Neutron activation			
$\bar{x}$	492	132	28
s	28	7	4
n	6	6	5
Inert gas fusion			
$\bar{x}$	497	129	29
s	13	8	5
n	12	11	20

where

$\bar{x}$ is the mean oxygen value;

s is the standard deviation of an individual determination;

n is the number of determinations.

NOTE — The standard deviation includes error due both to the imprecision of the analytical method and to possible heterogeneity of the material analysed.

One criticism against this mode of presentation is that the user will have to compute the uncertainty based on his own understanding of the relationships.

5 Homogeneity of materials

Most RMs are subjected to a preparation procedure which ultimately includes subdivision into usable units. A subset of individual units from the batch is chosen for measurement according to a statistically valid sampling plan. A measurement uncertainty is derived taking into account material inhomogeneity as well as other factors (see clause 4). Other types of RM are prepared as individual artifacts and the certification is based on separate measurement of each unit rather than on statistical sampling of the complete batch. The second approach is useful when the RM can be measured non-destructively.

5.1 Materials

RMs prepared as solutions or pure compounds are expected to be homogeneous on physical (thermodynamic) grounds. The object of the test for homogeneity is mainly to detect any impurities, interferences or irregularities.

Materials such as mixed powders, ores, alloys, etc. are heterogeneous in composition by nature. RMs prepared from such materials must therefore be tested to assess the degree of homogeneity.

5.2 Concept of homogeneity

In theory, a material is perfectly homogeneous with respect to a given characteristic if there is no difference between the value of this characteristic from one part (unit) to another. However, in practice a material is accepted to be homogeneous with respect to a given characteristic if a difference between the value of this characteristic from one part (or unit) to another cannot be detected experimentally. The practical concept of homogeneity therefore embodies both a specificity to the characteristic and a parameter of measurement (usually the standard deviation) of the measurement method used, including the defined sample size of the test portion.

5.2.1 Characteristic of interest

A material may be sufficiently homogeneous with respect to the characteristic of interest to be useful as an RM even though it is inhomogeneous with respect to other characteristics, provided that this inhomogeneity exerts no detectable influence on the accuracy and precision of the commonly used methods of determination for the characteristic of interest.

5.2.2 Homogeneity measurement method

The degree of homogeneity that a material must have for use as an RM is commensurate with the precision attainable by the best available methods for the determination of the characteristic for which the RM is intended. Therefore, the greater the precision of the measurement method, the higher is the required degree of homogeneity of the material.

The precision attainable by the homogeneity measurement method varies with both the characteristic measured and its value for the RM. An RM intended for more than one characteristic is described by a corresponding number of statements of homogeneity, each of which should be traceable to an experimentally determined precision. The magnitude of the precision can vary widely.

In many cases, the precision attainable by a measurement method is affected by the size of the test portion taken from the RM. The degree of homogeneity of an RM is therefore defined for a given test portion size.

5.2.3 Practice

Ideally, an RM should be characterized with respect to the degree of homogeneity for each characteristic of interest. For RMs intended for a relatively large number of characteristics, the assessment of the degree of homogeneity for all characteristics is both economically and physically burdensome, and in some cases unfeasible. In practice therefore, the degree of homogeneity of such RMs is assessed only for selected characteristics. It is recommended that these characteristics be appropriately selected on the basis of established chemical or physical relationships; for example, an interelement concomitance in the mineral phases of an RM makes reasonable the assumption that the RM also has an acceptable degree of homogeneity for the non-selected elements.

5.3 Experimental design

5.3.1 Objectives

For reference materials that are expected to be homogeneous on physical grounds, the main purpose of homogeneity testing is to detect unexpected problems. Some examples are differential contamination during the final packaging into individual units, or incomplete dissolution or equilibration of an analyte in a solvent (which could lead to steadily changing concentrations from the first vial filled to the last). A statistical trend analysis would be helpful in the latter case. If the material is produced in more than one batch, it is necessary to test the equality of the batches (or to certify the batches separately).

When the nature of a reference material leads one to expect some inhomogeneity, the goal of the testing programme is not simply detection of inhomogeneity, but rather the estimation of its magnitude. This may require a more extensive testing programme than is required for detection.

Inhomogeneity can manifest itself in at least two ways :

- a) different subsamples of an RM unit may differ on the property of interest;
- b) there may be differences between units of the RM.

Differences among subsamples can usually be reduced or controlled to an acceptably low level by making the size of the subsample sufficiently large. Often a study to determine the appropriate subsample size is conducted before the certification experiments are begun. Differences which exist between individual units of the candidate RM must be reflected in the uncertainty statement on the certificate.

In statistical terms, the experimental design must satisfy the following objectives :

- 1) to detect whether the within-unit (short-range) variation is statistically significant in comparison with the known variation of the measurement method;