



GUIDE 33

Uses of certified reference materials

First edition 1989



Contents

	Page
Foreword	ii
Introduction	iii
Section 1: General	1
1.1 Scope	1
1.2 Definitions	1
1.3 Statistical considerations	2
Section 2: Assessment of a measurement process ..	3
2.1 The cases to be considered	3
2.2 Requirements of limits	3
2.3 Choices of CRM	3
2.4 Carrying out the experiment	3
2.5 General remarks	8
Section 3: Defining and realizing conventional scales	10
3.1 General principles	10
3.2 The International Practical Temperature Scale	10
3.3 The pH scale	11
3.4 The octane-number scale	11
3.5 Soluble solids content by the refractive index method	11
3.6 Unit of turbidity	11
3.7 Hardness scales	11
3.8 Caking power of coals by the Roga method	11
Annex A: Bibliography	12

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

ISO guides are intended essentially for internal use in ISO committees or in some cases for the guidance of member bodies when dealing with matters that would not normally be the subject of an International Standard.

ISO Guide 33 was drawn up by the ISO Committee on reference materials (REMCO) in collaboration with Technical Committee ISO/TC 69, *Applications of statistical methods*, and was submitted directly to ISO Council for acceptance.

© ISO 1989

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

Introduction : “Proper use” and “misuse” of CRMs

Today’s world of modern technology requires a vast number of certified reference materials (CRMs) in widely diverse fields and the demand for such materials is expected to increase. The preparation of a CRM is a time-consuming, meticulous and expensive endeavour and consequently it has not always been, and will continue not to be, possible to satisfy the demand for all types and quantities of CRMs. For this reason, CRMs must be used properly, i.e. effectively, efficiently and economically.

Certified reference materials must be used consistently to ensure reliable measurements. However, in doing so, the magnitude of the supply of that RM, its relative cost, its availability (accessibility) and the measurement technique, be it destructive or non-destructive, must be considered. Also important to the user is the fact that the misuse of an RM may not provide the intended information.

Misuse of CRMs differs from incorrect use. The user of a CRM is expected to be familiar with all information pertinent to the use of the CRM as specified by the producer. He should comply with such factors as the period of validity of the CRM, the prescribed conditions for storage of the CRM, instructions for the use of the CRM, and specifications for validity of the certified properties of the CRM. A CRM should not be used for a purpose other than that for which it was intended. Nevertheless, from time to time, when a user must resort to applying a CRM in an incorrect manner because of the unavailability of a suitable CRM, he must be fully cognizant of the potential pitfalls and therefore assess his measurement output accordingly.

There are many measurement processes where CRMs are in general use but are replaceable by a host of working standards such as homogeneous materials, previously analysed materials, pure compounds, solutions of pure elements, etc. Some examples are where only a “rough” estimate of the trueness or precision of a method is sought, where “blind” unknown check samples are used routinely in quality control programmes, and where only the variation in trueness or precision of a method with some parameter such as time, analyst, instrument, etc., is being evaluated. The first example illustrates the use of a CRM where the well-defined certified value and uncertainty of the CRM is under-utilized. The others illustrate the case where a series of “one-time” trueness and precision assessments are compared with one another. There is no need to base that comparison on a well-defined certified value and uncertainty of a CRM. The advantages in using CRMs are that the user has the means to assess the trueness and precision of his measurement method and establishes metrological traceability for his results.

Whether the use of CRMs in these procedures is in fact “misuse” depends largely on the availability and relative cost of the CRMs. Where CRMs are in short supply or very expensive, their use would indeed be misuse. However, for CRMs in ample supply or where similar CRMs are available from one or more producers, it is strongly recommended that CRMs always be used instead of working standards because of the resultant enhanced confidence in the measurement output.

It is important that users remain aware that the preparation of working standards for use instead of CRMs has an associated cost based on factors such as material cost, facility usage charges, personnel labour rates, etc., in which the material cost is in general the lowest. For some CRMs such as the complex compositional materials certified for chemical composition, the cost of preparing working standards to match the composition of real samples can exceed that of available CRMs. In these cases, the use of CRMs is recommended.

The user should be aware of the potential misuse of CRMs as “blind” unknown check samples in quality control programmes. Where there are only a few CRMs in an area of expertise, they are easily recognized and they may therefore not satisfy the intended purpose. Moreover, the same CRMs should never be used for both calibration purposes and as “blind” unknown check samples in a measurement process.

The misuse of CRMs can occur also when the user does not fully take into account the uncertainty in the certified property. The overall uncertainty of a certified property of a CRM can have contributions from the inhomogeneity of the material, the within-laboratory uncertainty and, where applicable, the between-laboratories uncertainty. The level of homogeneity defined for a CRM by the producer is dependent on both the statistical design used to evaluate it and the repeatability of the method of measurement. For certain CRMs, the level of homogeneity is valid for a test portion defined by mass, physical dimension, time of measurement, etc. The user must be aware that the use of a test portion that does not meet or exceed that specification could severely increase the contribution of the inhomogeneity of the CRM to the uncertainty of the certified property to the point where the statistical parameters of certification are no longer valid.

The variation in the repeatability of different methods has another implication for the user. Since the degree of inhomogeneity of a CRM is dependent on the repeatability of the method of measurement, it is possible that a user, in applying a

method capable of better repeatability, could detect inhomogeneity in that CRM. In such cases, the observed inhomogeneity is already accounted for in the statistical parameters for the certified property and therefore the statistical tests presented in this Guide remain valid but the scientific basis for using that particular CRM to give a true assessment of the user's method must again be questioned.

It is well known that the methods of measurement of a property are not capable of equal repeatability. Accordingly there could arise instances where the user may wish to assess a method that has greater repeatability than that or those used in the certification of the CRM. In such cases, the statistical tests presented in this Guide remain valid but the scientific basis for using that particular CRM to give a true assessment of the precision (and possibly the trueness) normally expected from the user's method must be questioned. It is recommended that the user resort to a CRM of lesser uncertainty, if available.

For RMs certified by a "definitive" method, the user should not assume that his method is capable of matching the precision and trueness reported for the CRM. It is unreasonable therefore to apply the statistical procedures in this Guide for assessing the trueness and precision of a method by application to a CRM using the certification parameters for a property reported by the producer. The user, as a consequence, must either experimentally establish or make estimates based on available information for those parameters that are more appropriate. Similarly, where a user applies a method to an RM that has been certified by a single different method, the user must not assume that the certification parameters for the certified property are applicable to his method except in cases where the trueness and precision capable by both methods are known to be comparable.

One of the important considerations in selecting a CRM for use either in assessing the trueness and precision of a method or in the calibration of instruments in a method is the level of uncertainty required by the end-use of the method. Obviously the user should not apply a CRM of greater uncertainty than permitted by the end-use. This Guide states that the uncertainty of the replicated measurement of a CRM is twice the between-laboratories standard deviation of the certification programme, when this parameter is known, or four times the repeatability of the method on the CRM. These values represent the lowest level of uncertainty achievable with this CRM and may be used as a guide in determining whether it satisfies the uncertainty requirement of the end-use.

The selection of CRMs must take into account not only the level of uncertainty required for the intended purpose but also their availability, cost, and chemical and physical suitability for the intended purpose. For example, the unavailability or high cost of one CRM could force a user to resort to using another CRM of greater uncertainty than the preferred one. Also, in chemical analysis, a CRM of greater, but still acceptable, uncertainty in the certified property may be preferred over another CRM because of better matching with the composition of real samples. This could result in minimizing "matrix" or chemical effects in the measurement process which are capable of causing errors far greater than the difference between the uncertainties of the CRMs.

In conclusion, CRMs are meant to fulfil many purposes. Accordingly, a CRM used properly for one purpose in one laboratory may be misused for another purpose in another laboratory. It is recommended to the user that he consider the suitability of a CRM for his intended purpose on a case-by-case basis.

Uses of certified reference materials

Section 1 : General

1.1 Scope

Section 1 of this Guide presents definitions (with indication of their sources) of terms used, and sets out the statistical considerations on which the Guide is based.

Section 2 presents recommendations for developing criteria for the assessment of the precision and trueness of a measurement process by the use of CRMs. It pertains only to CRMs characterized to be homogeneous as described in ISO Guide 35^[1]. The use of CRMs is essential for assessment of trueness and optional for assessment of precision.

Section 3 discusses the use of CRMs for the definition and realization of conventional measurement scales.

1.2 Definitions

1.2.1 measurement process : All the information, equipment and operations relevant to a given measurement.

NOTE — This concept embraces all aspects relating to the performance and quality of the measurement; it includes, for example, the principle, method, procedure, values of the influence quantities and the measurement standards.

[VIM : 1984^[2]]

1.2.2 influence quantity : A quantity which is not the subject of the measurement but which influences the value of the measurand or the indication of the measuring instrument.

EXAMPLES — ambient temperature; frequency of an alternating measured voltage.

[VIM : 1984^[2]]

1.2.3 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.

[ISO Guide 30 : 1981^[3]]

1.2.4 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.

[ISO Guide 30 : 1981^[3]]

1.2.5 precision : The closeness of agreement between independent test results obtained under prescribed conditions.

[ISO 3534-1^[4]]

1.2.6 repeatability : Precision under repeatability conditions.

[ISO 3534-1^[4]]

1.2.7 repeatability conditions : Conditions where independent test results are obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within short intervals of time.

[ISO 3534-1^[4]]

1.2.8 repeatability standard deviation : The standard deviation of test results obtained under repeatability conditions.

NOTE — It is a measure of the dispersion of the distribution of test results under repeatability conditions.

[ISO 3534-1^[4]]

1.2.9 repeatability limit, r : The value less than or equal to which the absolute difference between two single test results obtained under repeatability conditions is expected to be with a probability of 95 %.

[ISO 3534-1^[4]]

1.2.10 reproducibility : Precision under reproducibility conditions.

[ISO 3534-1^[4]]

1.2.11 reproducibility conditions : Conditions where test results are obtained with the same method on identical material in different laboratories by different operators using different equipment.

[ISO 3534-1^[4]]

1.2.12 reproducibility limit, R : The value less than or equal to which the absolute difference between two single test results obtained under reproducibility conditions is expected to be with a probability of 95 %.

[ISO 3534-1^[4]]

1.2.13 bias : The difference between the expectation of the test results and an accepted reference value.

NOTE — Bias is a systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

[ISO 3534-1^[4]]

1.2.14 trueness : The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

NOTE — The measure of trueness is usually expressed in terms of bias.

[ISO 3534-1^[4]]

1.2.15 estimation, estimate, estimator (of parameters)

1.2.15.1 estimation : The operation of assigning, from the test result in a sample, numerical values to the parameters of a distribution chosen as the statistical model of the population from which this sample is taken.

[ISO 3534-1^[4]]

1.2.15.2 estimate : The result of an estimation.

[ISO 3534-1^[4]]

1.2.15.3 estimator : A statistic used to estimate a population parameter.

[ISO 3534-1^[4]]

1.3 Statistical considerations

1.3.1 Basic assumptions

All statistical methods used in this Guide are based on the following assumptions.

- a) The certified value is the best estimate of the true value of the property of the CRM.
- b) All variation, be it associated with the material (i.e. homogeneity) or the measurement process, is random and follows a normal probability distribution. The values of probabilities stated in this Guide assume normality. They may be different if there is deviation from normality.

1.3.2 Decision errors

The assessment of a measurement process on the basis of precision and trueness is always subject to rendering an incorrect conclusion because of

- a) the uncertainty of measurement results and
- b) the limited number of replicate results usually performed.

Increasing the number of measurements decreases the chance of an incorrect conclusion but, in many instances, the risk of making a wrong conclusion has to be balanced in economic terms against the cost of increasing the number of measurements. Accordingly, the rigour of the criteria developed for assessing a measurement process must take into account the level of precision and trueness requisite for the end-use.

For the purposes of this Guide, it is necessary to define the term “null hypothesis”.

The null hypothesis is the hypothesis to be accepted or rejected based on the outcome of the measurement. In this case the null hypothesis is that the measurement process has bias no greater than the limit chosen by the experimenter and variances no greater than the predetermined value; the alternative hypothesis is the hypothesis which is opposed to the null hypothesis (see also ISO 3534-1^[4]).

There are two types of possible error in accepting or rejecting the null hypothesis :

- a) **error type I** : The error committed in rejecting the null hypothesis when in reality the null hypothesis is true.

type I risk : The probability of committing error type I. Its value varies according to the real situation. The maximum value is called the significance level.

significance level : The given value, usually designated by α , which limits the probability of committing error type I.

- b) **error type II** : The error committed in failing to reject the null hypothesis when in reality the alternative hypothesis is true.

type II risk : The probability, usually designated by β , of committing error type II. Its value depends on the real situation and can be calculated only if the alternative hypothesis is adequately specified.

power of test : The probability of not committing error type II, usually designated by $(1 - \beta)$. It is the probability of rejecting the null hypothesis when in reality the alternative hypothesis is true.

The choice of the values of both α and β is usually based on economic considerations dictated by the importance of the consequences of the decision. These values as well as the alternative hypothesis should be chosen before the start of the measurement process.