INTERNATIONAL STANDARD

ISO 13528

First edition
2005-09-01

Statistical methods for use in proficiency testing by interlaboratory comparisons

Méthodes statistiques utilisées dans les essais d'aptitude par comparaisons interlaboratoires

Reference number
ISO 13528:2005(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 13528 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

0 Introduction

0.1 The aims of proficiency testing

Proficiency testing by interlaboratory comparisons is used to determine the performance of individual laboratories for specific tests or measurements, and to monitor the continuing performance of laboratories. The Introduction to ISO/IEC Guide 43-1:1997 should be consulted for a full exposition of the purposes of proficiency testing. In statistical language, the performance of laboratories can be described by three properties: laboratory bias, stability and repeatability. Laboratory bias and repeatability are defined in ISO 3534-1, ISO 3534-2 and ISO 5725-1. The stability of a laboratory's results is measured by intermediate precision as defined in ISO 5725-3.

Laboratory bias may be assessed by tests on reference materials, when these are available, using the procedure described in ISO 5725-4. Otherwise, proficiency testing by interlaboratory comparisons provides a generally available means of obtaining information about laboratory bias, and the use of data from proficiency tests to obtain estimates of laboratory bias is an important aspect of the analysis of such data. However, stability and repeatability will affect data obtained in proficiency tests, so that it is possible for a laboratory to obtain data in a round of a proficiency test which indicate bias that is actually caused by poor stability or poor repeatability. It is therefore important that these aspects of laboratory performance are assessed regularly.

Stability may be assessed by re-testing of retained samples, or by making regular measurements on a reference material or an in-house reference material (a stock of material established by a laboratory to use as private reference material). These techniques are described in ISO 5725-3. Stability may also be assessed by plotting estimates of laboratory bias derived from proficiency tests in control charts. This can provide information about laboratory performance that is not apparent from the examination of the results of individual rounds of proficiency testing schemes, and is another important aspect of the analysis of such data.

Data suitable for assessing repeatability may be generated by tests carried out in the normal course of the work of a laboratory, or by extra tests carried out within a laboratory specifically to assess repeatability. Consequently, the assessment of repeatability is not necessarily an important aspect of proficiency testing, although it is important that laboratories monitor their repeatability in some way. Repeatability may be assessed by plotting ranges of duplicate measurements on a control chart as described in ISO 5725-6.

The flowchart (Figure 1) illustrates how the techniques described in this International Standard are to be applied.

0.2 ISO/IEC Guide 43

ISO/IEC Guide 43-1 describes different types of proficiency testing schemes and gives guidance on the organization and design of proficiency testing schemes. ISO/IEC Guide 43-2 gives guidance on the selection and use of proficiency testing schemes by laboratory accreditation bodies. Those documents should be consulted for detailed information in those areas (the information is not duplicated here). ISO/IEC Guide 43-1 contains an annex that briefly describes the statistical methods that are used in proficiency testing schemes.

This International Standard is complementary to ISO/IEC Guide 43, providing detailed guidance that is lacking in that document on the use of statistical methods in proficiency testing. ISO 13528 is to a large extent based on a harmonized protocol for the proficiency testing of analytical laboratories [1], but is intended for use with all measurement methods.

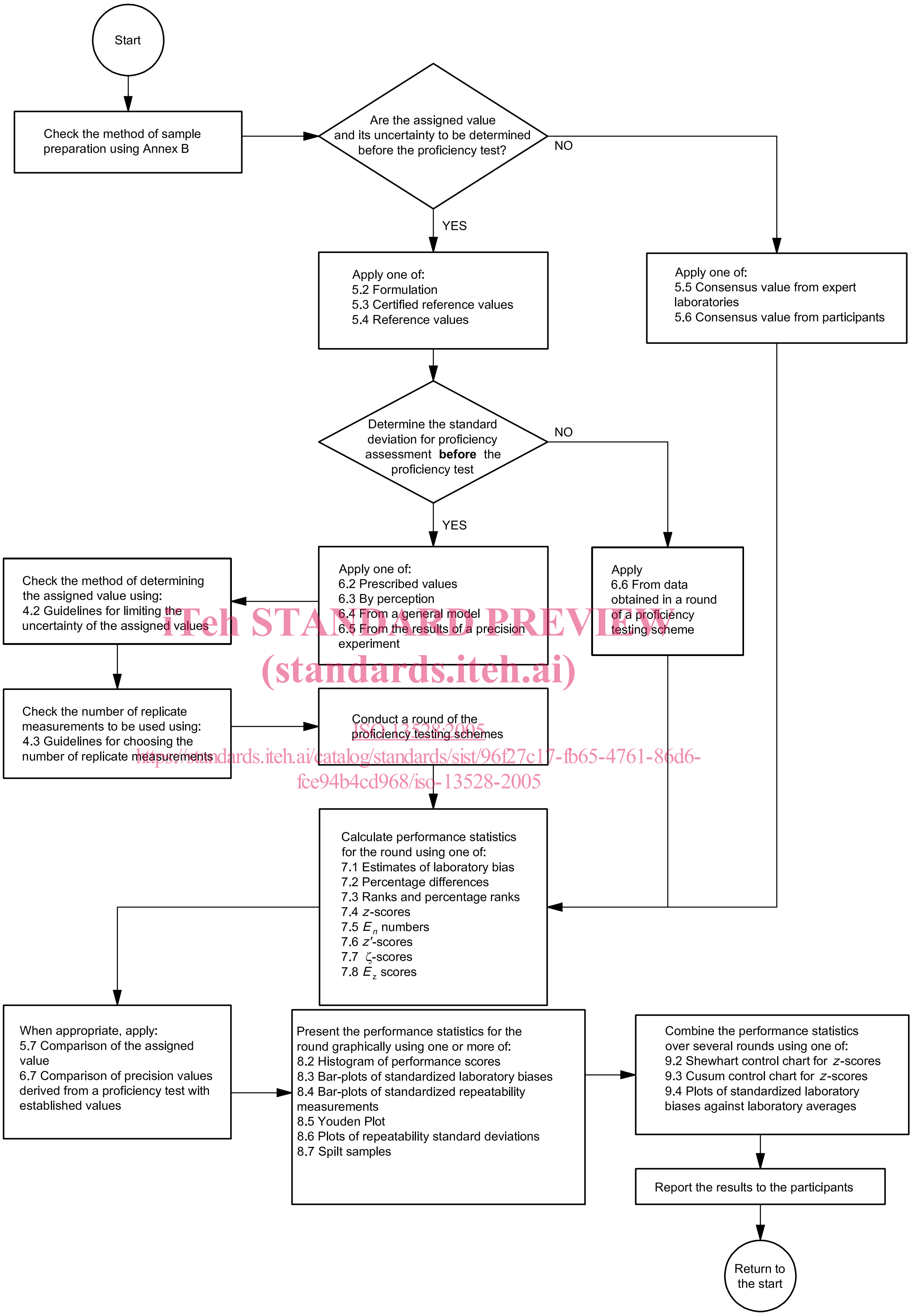

Figure 1 — Flowchart showing the activities requiring the use of statistical methods when operating a proficiency testing scheme

Statistical methods for use in proficiency testing by interlaboratory comparisons

1 Scope

This International Standard complements ISO Guide 43 (all parts) by providing detailed descriptions of sound statistical methods for organizers to use to analyse the data obtained from proficiency testing schemes, and by giving recommendations on their use in practice by participants in such schemes and by accreditation bodies.

This International Standard can be applied to demonstrate that the measurement results obtained by laboratories do not exhibit evidence of an unacceptable level of bias.

It is applicable to quantitative data but not to qualitative data.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

ISO 3534-2:—[1], *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO/IEC Guide 43-1:1997, *Proficiency testing by interlaboratory comparisons — Part 1: Development and operation of proficiency testing schemes*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 5725-1 and the following apply.

3.1
interlaboratory comparison
organization, performance and evaluation of tests or measurements on the same or similar test items by two or more laboratories in accordance with predetermined conditions

NOTE Adapted from ISO/IEC Guide 43-1.

3.2
proficiency testing
determination of laboratory testing performance by means of interlaboratory comparisons

1) To be published.

3.3
assigned value
value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose

3.4
standard deviation for proficiency assessment
measure of dispersion used in the assessment of proficiency, based on the available information

3.5
***z*-score**
standardized measure of laboratory bias, calculated using the assigned value and the standard deviation for proficiency assessment

3.6
coordinator
organization (or person) with responsibility for coordinating all of the activities involved in the operation of a proficiency testing scheme

4 Statistical guidelines for the design and interpretation of proficiency tests
(see ISO/IEC Guide 43-1:1997, 5.4.2.)

4.1 Action and warning signals

4.1.1 This International Standard describes some simple numerical or graphical criteria that should be applied to the data obtained in a proficiency test to see if they give rise to action or warning signals. Even in a well-run laboratory, with experienced staff, anomalous results may sometimes be obtained. Also, it is possible that a standardized measurement method, even though it has been validated by a precision experiment, may contain faults that become apparent only after several rounds of a proficiency testing scheme. The proficiency scheme itself may contain faults. For these reasons, the criteria given here shall not be used to condemn laboratories, as being unfit to perform the measurement method under examination. If proficiency testing is used to condemn laboratories, then it shall be necessary to devise appropriate criteria for that purpose.

4.1.2 The criteria given here are designed so that, when the standard deviation for proficiency assessment is based on observed performance (using one of the methods described in 6.4 to 6.6), the criteria give action signals when results are so exceptional as to merit investigation and corrective action.

4.1.3 The coordinator should have an understanding of the major sources of variability that can be anticipated in proficiency test data for the measurement in question. The first step in any analysis should be to examine the distribution of results for evidence of unanticipated sources of variability. For example, a bimodal distribution might be evidence of a mixed population of results caused by different methods, contaminated samples or poorly worded instructions. In this situation, the concern should be resolved before proceeding with analysis or evaluation. Accrediting bodies shall have policies for response to unacceptable performance in proficiency testing. Follow-up actions are determined by that policy or by the laboratory's quality procedures. However, there are generally recommended actions when a laboratory produces an unacceptable result in a proficiency test. Guidance for actions by laboratories in response to unsuccessful performance on a proficiency test is given in 4.1.4.

4.1.4 In schemes where the standard deviation for proficiency assessment is based on observed performance, when a result gives an action signal, the laboratory shall decide what investigations and corrective actions are appropriate, in consultation with the coordinator or an accreditation body if necessary. Unless there is a valid reason not to do so, the laboratory shall examine its procedures and identify one or more corrective actions that, according to staff in the laboratory, are likely to prevent the recurrence of such results. The laboratory may ask the coordinator for advice on possible causes of its problem, or may ask the coordinator to consult other experts. The laboratory shall take part in further rounds of the proficiency testing scheme to assess the effectiveness of the corrective actions. Appropriate corrective actions may be one of the following:

a) checking that staff understand and follow the measurement procedure;

b) checking that all details of the measurement procedure are correct;

c) checking the calibration of equipment and the composition of reagents;

d) replacing suspect equipment or reagents;

e) comparative tests of staff, equipment and/or reagents with another laboratory.

The use of the results of proficiency tests by laboratory accreditation bodies is described in ISO/IEC Guide 43-2:1997, Clause 6.

4.2 Guidelines for limiting the uncertainty of the assigned value

The assigned value X has a standard uncertainty u_X that depends on the method that is used to derive it, and also, when it is derived from tests in several laboratories, on the number of laboratories and, perhaps, on other factors. Methods for calculating the standard uncertainty of the assigned value are given in Clause 5.

The standard deviation for proficiency testing $\hat{\sigma}$ is used to assess the size of estimates of laboratory bias found in a proficiency test. Methods for obtaining the standard deviation for proficiency testing are given in Clause 6 and criteria that compare it with estimates of laboratory bias are given in Clause 7.

If the standard uncertainty u_X of the assigned value is too large in comparison with the standard deviation for proficiency testing $\hat{\sigma}$, then there is a risk that some laboratories will receive action and warning signals because of inaccuracy in the determination of the assigned value, not because of any cause within the laboratories. For this reason, the standard uncertainty of the assigned value shall be established and shall be reported to laboratories participating in proficiency testing schemes (see ISO/IEC Guide 43-1:1997, A.1.4 and A.1.6).

If

$$u_X \leqslant 0{,}3\hat{\sigma} \tag{1}$$

then the uncertainty of the assigned value is negligible and need not be included in the interpretation of the results of the proficiency test.

If these guidelines are not met, then the coordinator shall consider the following.

a) Look for a method for determining the assigned value such that its uncertainty meets the above guideline.

b) Use the uncertainty of the assigned value in the interpretation of the results of the proficiency test (see 7.5 on E_n numbers or 7.6 on the z'-score).

c) Inform the participants in the proficiency test that the uncertainty of the assigned value is not negligible.

EXAMPLE Suppose that the assigned value X is determined as the average $\bar{x}$ of the results of tests in 11 laboratories, and that the standard deviation for proficiency testing is determined as the standard deviation s of these same 11 results, so $\hat{\sigma} = s$. As a first approximation, the standard uncertainty of the assigned value in this situation may be estimated by $u_X = s/\sqrt{11} = 0{,}3\,s$, so that the requirement appears to be met. However, the requirement cannot be met in this situation with fewer than 11 laboratories. Further, the uncertainty of the assigned value will be larger than $s/\sqrt{11}$ if the samples suffer from non-homogeneity or instability, or if there is a factor that causes a common bias in the results of the laboratories (e.g. if they all use the same reference standard).

4.3 Guidelines for choosing the number of replicate measurements

Repeatability variation contributes to the variation between the laboratory biases in a proficiency test. If the repeatability variation is too large in comparison with the standard deviation for proficiency testing, then there is a risk that repeatability variation will cause the results of the proficiency test to be erratic. In this situation, a laboratory could have large bias in one round, but not the next, and they will have difficulty identifying the cause.

For this reason, when it is considered desirable to limit the influence of repeatability variation, the number of replicate measurements n made by each laboratory in a proficiency test shall be chosen so that:

$$\sigma_r / \sqrt{n} \leqslant 0{,}3\hat{\sigma} \tag{2}$$

where σ_r is the repeatability standard deviation that has been established in a previous interlaboratory experiment.

The justification for the factor of 0,3 is that when this criterion is met, the repeatability standard deviation contributes no more than about 10 % of the standard deviation for proficiency testing.

Further, all laboratories shall carry out the same number of replicate measurements. (The methods of analysis given later in this International Standard assume that this requirement is met.) If the requirement inequality of (2) is not met, then the number of replicate measurements shall be increased, or the results of the proficiency test shall be interpreted with caution.

This approach assumes that laboratories have generally similar repeatability. Cases can arise where this is not so. In such cases, for the methods described in this International Standard to be applied, the following device may be used. The coordinator should fix the number of replicate measurements n, using a typical value for the repeatability standard deviation. Then, each laboratory should check that it satisfies Inequality (2) with its own repeatability standard deviation. If it does not, then it should modify its measurement procedure so that it obtains a test result as the average of some number of determinations chosen so that Inequality (2) is satisfied.

4.4 Homogeneity and stability of samples (see ISO/IEC Guide 43-1:1997, 5.6.2 and 5.6.3)

Methods are given in Annex B for checking that the samples to be used in a proficiency test are adequately homogeneous and stable.

When a method of sample preparation is used such that the homogeneity criterion in Annex B is not met, then replicate samples shall be tested by the participants, or the standard deviation for proficiency testing shall include an allowance for the heterogeneity of the samples, as described in Annex B.

4.5 Operationally defined measurement methods

With an operationally defined measurement method, the measurement result is defined by the measurement procedure. For example, the size distribution of a particulate material may be determined using either square-holed or round-holed sieves. There may be no good reason why one type of sieve should be preferred, but unless the type of sieve is specified, laboratories that use different types of sieves may obtain differing results. If a participant uses a different method from that used to establish the assigned value, then their results may show a bias when no fault in execution is present.

If participants are free to choose between operationally defined methods, no valid consensus may be evident amongst them. Two recourses are available to overcome this problem:

a) when a standardized method is in routine use by the participants, it is used to establish the assigned value, and participants are instructed to use it in the proficiency test;

b) a separate value of the assigned value is produced for each method used.

A similar situation arises when the measurand is specified, but not the procedure, and the same choice has to be made.

4.6 Reporting of data (see ISO/IEC Guide 43-1:1997, 6.2.3)

For the purposes of the calculations required in proficiency testing, it is recommended that individual measurement results should not be rounded by more than $\sigma_r / 2$.

Participants shall be asked to report the actual values of their measurement results. Measurement results shall not be truncated (i.e. results shall not be reported in the form "<0,1" or "less than the detection limit"). Likewise, when a negative result is observed, the actual negative value shall be reported (i.e. results shall not be reported as zero even when logically the measurement result cannot be negative). Participants shall be informed that if they report truncated results on a sample, or zeros when results are negative, then all the data for that sample will be excluded from the analysis. If necessary, the form used to report results may contain a box to allow a participant to indicate that a result is below the detection limit.

4.7 Period of validity of the results of proficiency tests

The period of validity of the result obtained by a laboratory in a single round of a proficiency testing scheme is limited to the time that the laboratory performed the test. Thus, if a laboratory achieves a satisfactory result in a single round, the result shall not be used to support a claim that the laboratory obtained reliable data on any other occasion.

A laboratory that operates a quality system and achieves a history of satisfactory results in many rounds of a proficiency testing scheme shall be entitled to use the results as evidence that it is able to obtain consistently reliable data.

5 Determination of the assigned value and its standard uncertainty

5.1 Choice of method of determining the assigned value

Five ways of determining the assigned value X are described in 5.2 to 5.6. The choice between these methods shall be the responsibility of the coordinator following consultation with technical experts as described in ISO/IEC Guide 43-1. The methods described in 5.5 and 5.6 are unlikely to be applicable when the number of laboratories participating in the scheme is small. The methods of calculating the standard uncertainty, u_X, of the assigned value given in this clause will usually be adequate for the applications for which they are used in this International Standard. Alternative methods may be used provided that they have a sound statistical basis and that the method used is described in the documented plan for the scheme.

The determination of the assigned value shall be the responsibility of the coordinator. The assigned value shall not be disclosed to the participants until they have reported their results to the coordinator. The coordinator shall prepare a report giving details of how the assigned value was obtained, the identities of laboratories involved in its determination, and statements of the traceability and measurement uncertainty of the assigned value.

The *Guide to the expression of uncertainty in measurement* gives guidance on the evaluation of measurement uncertainties.

This International Standard recommends the use of robust statistical methods when it is considered that they are the most appropriate methods to use (for example, as in 5.5 and 5.6). Alternatively, procedures that involve the detection and removal of outliers may be used provided that they have a sound statistical basis and the method that is used is reported. Guidance on the use of tests for outliers is given in ISO 5725-2.

5.2 Formulation [see ISO/IEC Guide 43-1:1997, A.1.1, item a)]

5.2.1 General

The test material may be prepared by mixing constituents in specified proportions, or by adding a specified proportion of a constituent to a base material. In this case, the assigned value X is derived by calculation from the masses used.

The approach is especially valuable when individual samples may be prepared in this way, and it is the proportion of the constituents or of the addition that is to be determined: there is then no need to prepare a bulk quantity and ensure that it is homogeneous. However, when formulation gives samples in which the addition is more loosely bonded than in typical materials, or in a different form, it may be preferable to use another approach.

5.2.2 Standard uncertainty u_X of the assigned value

When the assigned value is calculated from the formulation of the test material, the standard uncertainty is estimated by combination of uncertainties using the approach described in the *Guide to the expression of uncertainty in measurement*. For example, in chemical analyses the uncertainties will usually be those associated with gravimetric and volumetric measurements.

The limitation of this method (in chemical analysis) is that care is needed to ensure that:

a) the base material is effectively free from the added constituent, or that the proportion of the added constituent in the base material is accurately known;

b) the constituents are mixed together homogeneously (where this is required);

c) all sources of error are identified (e.g. it is not always realized that glass absorbs mercury compounds, so that the concentration of an aqueous solution of a mercury compound can be altered by its container);

d) there is no interaction between the constituents and the matrix.

5.2.3 Example: Determination of the cement content of hardened concrete

In this case, concrete specimens may be prepared by weighing out quantities of the constituents (cement, aggregates and water) and mixing them together to form each concrete sample. The approach is satisfactory because the accuracy with the specimens can be prepared is far superior to that of the analytical method used to determine the cement content.

5.3 Certified reference values [see ISO/IEC Guide 43-1:1997, A.1.1 item b)]

5.3.1 General

When the material used in a proficiency test is a certified reference material (CRM), its certified reference value is used as the assigned value X.

5.3.2 Standard uncertainty u_X of the assigned value

When a certified reference material is used as the test material, the standard uncertainty of the assigned value is derived from the information on uncertainty provided on the certificate.

The limitation of this approach is that it can be expensive to provide every participant in a proficiency test with a sample of a certified reference material.

5.3.3 Example: Los Angeles value of aggregates

The “Los Angeles value” is a measure of the mechanical strength of aggregates that are used for road construction, and results of the test are measured in “LA units”. In an exercise to certify a reference material, a large number of samples of a particular aggregate were prepared, and some of these samples were used in an interlaboratory experiment involving 28 laboratories, allowing an assigned value of X_{CRM} = 21,62 LA units to be established with a standard uncertainty of $u_{X;\text{CRM}}$ = 0,26 LA units. The remaining samples of this aggregate could be used in proficiency tests.

5.4 Reference values [see ISO/IEC Guide 43-1:1997, A.1.1 item c)]

5.4.1 General

In this approach, samples of the test material that is to be the reference material (RM) are prepared first, ready for distribution to the participants. A number of the samples are then selected at random and tested along with certified reference materials, in one laboratory, using a suitable measurement method, and under repeatability conditions (as defined in ISO 3534-2). The assigned value X_{RM} of the test material is then derived from a calibration against the certified reference values of the CRMs.

5.4.2 Standard uncertainty u_X of the assigned value

When the assigned value of a test material is derived from the results of a series of tests on that material and on CRM, the standard uncertainty of the assigned value is derived from the test results, and the uncertainties of the certified reference values of the CRM. If the test material and the CRM are not similar (in matrix, composition and level of results), then the uncertainty arising from this is also to be included.

This method allows the assigned value to be established in a manner that is traceable to the certified values of the CRMs, with a standard uncertainty that can be calculated, and avoids the cost of distributing the CRM to all the participants. These are good reasons for preferring it to other methods. However, the method assumes that there are no interactions between the materials used and the test conditions.

The example in 5.4.3 illustrates how the required uncertainty may be calculated in the simple case when the assigned value of a test material is established by direct comparison with a single CRM.

5.4.3 Example: Los Angeles value of aggregates

The CRM described in the example in 5.3 may be used to determine the assigned value for an RM which is another, similar, aggregate. This determination requires a series of tests to be carried out, in one laboratory, on samples of the two aggregates, using the same measurement method, and under repeatability conditions. Let

X_{CRM} is the assigned value for the CRM

X_{RM} is the assigned value for the RM

D_i is the difference (RM − CRM) between the average results for the RM and the CRM on the i^{th} samples

$\bar{D}$ is the average of the differences D_i

Then

$$X_{RM} = X_{CRM} + \bar{D} \tag{3}$$

The standard uncertainty of the assigned value of the RM may be calculated as:

$$u_{X;RM} = \sqrt{u_{X;CRM}^2 + u_D^2} \tag{4}$$

Table 1 gives an example of data that might be obtained in such a series of tests, and shows how the standard uncertainty u_D of the difference is calculated.

With these results,

$$X_{RM} = 21{,}62 + 1{,}73 = 23{,}35 \text{ LA units} \tag{5}$$