INTERNATIONAL STANDARD

ISO
21748

First edition
2010-11-01

Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation

Lignes directrices relatives à l'utilisation d'estimations de la répétabilité, de la reproductibilité et de la justesse dans l'évaluation de l'incertitude de mesure

---

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.

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 21748 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

This first edition cancels and replaces ISO/TS 21748:2004, which has been technically revised.

Introduction

Knowledge of the uncertainty associated with measurement results is essential to the interpretation of the results. Without quantitative assessments of uncertainty, it is impossible to decide whether observed differences between results reflect more than experimental variability, whether test items comply with specifications, or whether laws based on limits have been broken. Without information on uncertainty, there is a risk of misinterpretation of results. Incorrect decisions taken on such a basis may result in unnecessary expenditure in industry, incorrect prosecution in law, or adverse health or social consequences.

Laboratories operating under ISO/IEC 17025 accreditation and related systems are accordingly required to evaluate measurement uncertainty for measurement and test results and report the uncertainty where relevant. The *Guide to the expression of uncertainty in measurement* (GUM), published by ISO/IEC as ISO/IEC Guide 98-3:2008, is a widely adopted standard approach. However, it applies to situations where a model of the measurement process is available. A very wide range of standard test methods is, however, subjected to collaborative study in accordance with ISO 5725-2:1994. This International Standard provides an appropriate and economic methodology for estimating uncertainty associated with the results of these methods, which complies fully with the relevant principles of the GUM, whilst taking account of method performance data obtained by collaborative study.

The general approach used in this International Standard requires that

— estimates of the repeatability, reproducibility and trueness of the method in use, obtained by collaborative study as described in ISO 5725-2:1994, be available from published information about the test method in use. These provide estimates of the intra- and inter-laboratory components of variance, together with an estimate of uncertainty associated with the trueness of the method;

— the laboratory confirms that its implementation of the test method is consistent with the established performance of the test method by checking its own bias and precision. This confirms that the published data are applicable to the results obtained by the laboratory;

— any influences on the measurement results that were not adequately covered by the collaborative study be identified and the variance associated with the results that could arise from these effects be quantified.

An uncertainty estimate is made by combining the relevant variance estimates in the manner prescribed by the GUM.

The general principle of using reproducibility data in uncertainty evaluation is sometimes called a "top-down" approach.

The dispersion of results obtained in a collaborative study is often also usefully compared with measurement uncertainty estimates obtained using GUM procedures as a test of full understanding of the method. Such comparisons will be more effective given a consistent methodology for estimating the same parameter using collaborative study data.

Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation

1 Scope

The International Standard gives guidance for

— evaluation of measurement uncertainties using data obtained from studies conducted in accordance with ISO 5725-2:1994;

— comparison of collaborative study results with measurement uncertainty (MU) obtained using formal principles of uncertainty propagation (see Clause 13).

ISO 5725-3:1994 provides additional models for studies of intermediate precision. However, while the same general approach may be applied to the use of such extended models, uncertainty evaluation using these models is not incorporated in the present International Standard.

This International Standard is applicable in all measurement and test fields where an uncertainty associated with a result has to be determined.

This International Standard does not describe the application of repeatability data in the absence of reproducibility data.

This International Standard assumes that recognized, non-negligible systematic effects are corrected, either by applying a numerical correction as part of the method of measurement, or by investigation and removal of the cause of the effect.

The recommendations in this International Standard are primarily for guidance. It is recognized that while the recommendations presented do form a valid approach to the evaluation of uncertainty for many purposes, it is also possible to adopt other suitable approaches.

In general, references to measurement results, methods and processes in this International Standard are normally understood to apply also to testing results, methods and processes.

2 Terms and definitions

For the purposes of this document, the following terms and definitions apply. In addition, reference is made to "intermediate precision conditions", which are discussed in detail in ISO 5725-3:1994.

2.1
bias
difference between the expectation of a test result or measurement result and a true value

NOTE 1   Bias is the total 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 true value is reflected by a larger bias value.

NOTE 2   The bias of a measuring instrument is normally estimated by averaging the error of indication over an appropriate number of repeated measurements. The error of indication is the "indication of a measuring instrument minus a true value of the corresponding input quantity".

NOTE 3   In practice, the accepted reference value is substituted for the true value.

[ISO 3534-2:2006, definition 3.3.2]

1

2.2
combined standard uncertainty
$u(y)$
standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities

[ISO/IEC Guide 98-3:2008, definition 2.3.4]

2.3
coverage factor
k
numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty

NOTE   A coverage factor, k, is typically in the range 2 to 3.

[ISO/IEC Guide 98-3:2008, definition 2.3.6]

2.4
expanded uncertainty
U
quantity defining an interval about a result of a measurement expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand

NOTE 1   The fraction may be regarded as the coverage probability or level of confidence of the interval.

NOTE 2   To associate a specific level of confidence with the interval defined by the expanded uncertainty requires explicit or implicit assumptions regarding the probability distribution characterized by the measurement result and its combined standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions can be justified.

NOTE 3   Expanded uncertainty is termed overall uncertainty in paragraph 5 of Recommendation INC-1 (1980).

[ISO/IEC Guide 98-3:2008, definition 2.3.5]

2.5
precision
closeness of agreement between independent test/measurement results obtained under stipulated conditions

NOTE 1   Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.

NOTE 2   The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results or measurement results. Less precision is reflected by a larger standard deviation.

NOTE 3   Quantitative measures of precision depend critically on the stipulated conditions. Repeatability conditions and reproducibility conditions are particular sets of extreme stipulated conditions.

[ISO 3534-2:2006, definition 3.3.4]

2.6
repeatability
precision under repeatability conditions

NOTE   Repeatability can be expressed quantitatively in terms of the dispersion characteristics of the results.

[ISO 3534-2:2006, definition 3.3.5]

2.7
repeatability conditions
observation conditions where independent test/measurement results are obtained with the same method on identical test/measurement items in the same test or measuring facility by the same operator using the same equipment within short intervals of time

NOTE   Repeatability conditions include:

— the same measurement procedure or test procedure;

— the same operator;

— the same measuring or test equipment used under the same conditions;

— the same location;

— repetition over a short period of time.

[ISO 3534-2:2006, definition 3.3.6]

2.8
repeatability standard deviation
standard deviation of test results or measurement results obtained under repeatability conditions

NOTE 1   It is a measure of the dispersion of the distribution of test or measurement results under repeatability conditions.

NOTE 2   Similarly, "repeatability variance" and "repeatability coefficient of variation" can be defined and used as measures of the dispersion of test or measurement results under repeatability conditions.

[ISO 3534-2:2006, definition 3.3.7]

2.9
reproducibility
precision under reproducibility conditions

NOTE 1   Reproducibility can be expressed quantitatively in terms of the dispersion characteristics of the results.

NOTE 2   Results are usually understood to be corrected results.

[ISO 3534-2:2006, definition 3.3.10]

2.10
reproducibility conditions
observation conditions where independent test/measurement results are obtained with the same method on identical test/measurement items in different test or measurement facilities with different operators using different equipment

[ISO 3534-2:2006, definition 3.3.11]

2.11
reproducibility standard deviation
standard deviation of test results or measurement results obtained under reproducibility conditions

NOTE 1   It is a measure of the dispersion of the distribution of test or measurement results under reproducibility conditions.

NOTE 2   Similarly, "reproducibility variance" and "reproducibility coefficient of variation" can be defined and used as measures of the dispersion of test or measurement results under reproducibility conditions.

[ISO 3534-2:2006, definition 3.3.12]

2.12
standard uncertainty
$u(x_i)$
uncertainty of the result of a measurement expressed as a standard deviation

[ISO/IEC Guide 98-3:2008, definition 2.3.1]

2.13
trueness
closeness of agreement between the expectation of a test result or a measurement result and a true value

NOTE 1   The measure of trueness is usually expressed in terms of bias.

NOTE 2   Trueness is sometimes referred to as "accuracy of the mean". This usage is not recommended.

NOTE 3   In practice, the accepted reference value is substituted for the true value.

[ISO 3534-2:2006, definition 3.3.3]

2.14
uncertainty
⟨measurement⟩ parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand

NOTE 1   The parameter may be, for example, a standard deviation (or a given multiple of it), or the half-width of an interval having a stated level of confidence.

NOTE 2   Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of a series of measurements and can be characterized by experimental standard deviations. Other components, which also can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information.

NOTE 3   It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects such as components associated with corrections and reference standards, contribute to the dispersion.

[ISO/IEC Guide 98-3:2008, definition 2.2.3]

2.15
uncertainty budget
list of sources of uncertainty and their associated standard uncertainties, compiled with a view to evaluating a combined standard uncertainty associated with a measurement result

NOTE   The list often includes additional information such as sensitivity coefficients (change of result with change in a quantity affecting the result), degrees of freedom for each standard uncertainty, and an identification of the means of evaluating each standard uncertainty in terms of a Type A or Type B evaluation (see ISO/IEC Guide 98-3:2008).

3   Symbols

a coefficient indicating an intercept in the empirical relationship $\hat{s}_R = a + bm$

B laboratory component of bias

b coefficient indicating a slope in the empirical relationship $\hat{s}_R = a + bm$

c coefficient in the empirical relationship $\hat{s}_R = cm^d$

c_i sensitivity coefficient $\partial y / \partial x_i$

d coefficient indicating an exponent in the empirical relationship $\hat{s}_R = cm^d$

e random error under repeatability conditions

k numerical factor used as a multiplier of the combined standard uncertainty u in order to obtain an expanded uncertainty U

l laboratory number

m mean value of the measurements

N number of contributions included in combined uncertainty calculations

n' number of contributions incorporated in combined uncertainty calculations in addition to collaborative study data

n_l number of replicates by laboratory l in the study of a certified reference material

n_r number of replicate measurements

p number of laboratories

Q number of test items from a larger batch

q number of assigned values by consensus during a collaborative study

r_{ij} correlation coefficient between x_i and x_j, in the interval -1 to $+1$

s_b between-group component of variance expressed as a standard deviation

s_b^2 between-group component of variance

s_D estimated, or experimental, standard deviation of results obtained by repeated measurement on a reference material used for checking control of bias

s_{inh} uncertainty associated with the inhomogeneity of the sample

s_{inh}^2 component of variance associated with the inhomogeneity of the sample

s_l estimated repeatability standard deviation with ν_l degrees of freedom for laboratory l during verification of repeatability

s_L experimental or estimated inter-laboratory standard deviation

$\hat{s}_L$ adjusted estimate of standard deviation associated with B where s_L is dependent on the response

s_L^2 estimated variance of B

s_r estimate of intra-laboratory standard deviation; the estimated standard deviation for e

s_r' adjusted estimate of intra-laboratory standard deviation, where the contribution is dependent on the response

s_r^2 estimated variance of e

s_R estimated reproducibility standard deviation

s_R' estimate of the reproducibility standard deviation adjusted for laboratory estimate of repeatability standard deviation

$\hat{s}_R$ adjusted estimate of reproducibility standard deviation calculated from an empirical model, where the contributions are dependent on the response

s_w estimate of intra-laboratory standard deviation derived from replicates or other repeatability studies

s_w^2 estimated intra-group component of variance (often an intra-laboratory component of variance)

$s_{\hat{\delta}}$ estimated standard deviation of bias $\hat{\delta}$ measured in a collaborative study

$s(\Delta_y)$ laboratory standard deviation of differences during a comparison of a routine method with a definitive method or with values assigned by consensus

$u(\hat{\delta})$ uncertainty associated with δ due to the uncertainty of estimating δ by measuring a reference measurement standard or reference material with certified value $\hat{\mu}$

$u(\hat{\mu})$ uncertainty associated with the certified value $\hat{\mu}$

$u(x_i)$ uncertainty associated with the input value x_i; also uncertainty associated with x_i' where x_i and x_i' differ only by a constant

$u(y)$ combined standard uncertainty associated with y where $u(y) = \sqrt{\sum_{i=1,n} c_i^2 u^2(x_i)}$

$u_i(y)$ contribution to combined uncertainty in y associated with the value x_i. In terms of the definition of $u(y)$ above, $u_i(y) = c_i u(x_i)$

$u(y_i)$ combined standard uncertainty associated with result or assigned value y_i

$u(Y)$ combined uncertainty for the result $Y = f(y_1, y_2, ...)$ where $u(Y) = \sqrt{\sum_i [c_i u(y_i)]^2}$

$u^2(y)$ combined standard uncertainty associated with y, expressed as a variance

u_{inh} uncertainty associated with sample inhomogeneity

U expanded uncertainty, equal to k times the standard uncertainty u

$U(y)$ expanded uncertainty in y where $U(y) = ku(y)$, where k is a coverage factor

x_i value of the ith input quantity in the determination of a result

x_i' deviation of the ith input value from the nominal value of x

Y combined result formed as a function of other results y_i

y_i result for test item i from the definitive method during a comparison of methods or assigned value in a comparison with values assigned by consensus

$\hat{y}_i$ result for test item i from the routine test method during a comparison of methods

y_0 assigned value for proficiency testing

Δ laboratory bias

Δ_l estimate of bias of laboratory l, equal to the laboratory mean, m, minus the certified value, $\hat{\mu}$

$\bar{\Delta}_y$ mean laboratory bias during a comparison of a routine method with a definitive method or with values assigned by consensus

δ bias intrinsic to the measurement method in use

$\hat{\delta}$ estimated or measured bias

μ unknown expectation of the ideal result

$\hat{\mu}$ certified value of a reference material

σ_0 standard deviation for proficiency testing

σ_D true value of the standard deviation of results obtained by repeated measurement on a reference material used for checking control of bias

σ_L inter-laboratory standard deviation; standard deviation of B

σ_L^2 variance of B; inter-laboratory variance

σ_r intra-laboratory standard deviation; standard deviation of e

σ_r^2 variance of e; intra-laboratory variance

σ_w within-group standard deviation

σ_{w0} standard deviation required for adequate performance (see ISO Guide 33)

ν_{eff} effective degrees of freedom for the standard deviation of, or uncertainty associated with, a result y_i

ν_i degrees of freedom associated with the ith contribution to uncertainty

ν_l degrees of freedom associated with an estimate s_l of the standard deviation for laboratory l during verification of repeatability

4 Principles

4.1 Individual results and measurement process performance

4.1.1 Measurement uncertainty relates to individual results. Repeatability, reproducibility and bias, by contrast, relate to the performance of a measurement or testing *process*. For studies under all parts of ISO 5725, the measurement or testing process will be a single measurement method, used by all laboratories taking part in the study. Note that for the purposes of this International Standard, the measurement method is assumed to be implemented in the form of a single detailed measurement procedure (as defined in ISO/IEC Guide 99:2007, 2.6). It is implicit in this International Standard that process-performance figures derived from method-performance studies are relevant to all individual measurement results produced by the process. It will be seen that this assumption requires supporting evidence in the form of appropriate quality control and assurance data for the measurement process (Clause 6).

4.1.2 It will be seen below that differences between individual test items may additionally need to be taken into account, but, with that caveat, it is unnecessary to undertake individual and detailed uncertainty studies for every test item for a well-characterized and stable measurement process.

4.2 Applicability of reproducibility data

The application of this International Standard is based on two principles.

— First, the reproducibility standard deviation obtained in a collaborative study is a valid basis for measurement uncertainty evaluation (see A.2.1).

— Second, effects not observed within the context of the collaborative study must be demonstrably negligible or explicitly allowed for. The latter principle is implemented by an extension of the basic model used for collaborative study (see A.2.3).

     7

4.3 Basic equations for the statistical model

4.3.1   The statistical model on which this International Standard is based is formulated as in Equation (1):

$$y = \mu + \delta + B + \sum c_i x_i' + e \tag{1}$$

where

  y is the measurement result, assumed to be calculated from an appropriate function;

  μ is the (unknown) expectation of ideal results;

  δ is a term representing bias intrinsic to the measurement method;

  B is the laboratory component of bias;

  x_i' is the deviation from the nominal value of x_i;

  c_i is the sensitivity coefficient, equal to $\partial y / \partial x_i$;

  e is the random error term under repeatability conditions.

B and e are assumed to be normally distributed, with variances of σ_L^2 and σ_r^2, respectively. These terms form the model used in ISO 5725-2:1994 for the analysis of collaborative study data.

Since the observed standard deviations of method bias, δ, laboratory bias, B, and random error, e, are overall measures of dispersion under the conditions of the collaborative study, the summation $\sum c_i x_i'$ is over those effects subject to deviations *other* than those incorporated in δ, B, or e, and the summation accordingly provides a method for incorporating effects of operations that are not carried out in the course of a collaborative study.

Examples of such operations include the following:

a)   preparation of test item carried out in practice for each test item, but carried out prior to circulation in the case of the collaborative study;

b)   effects of sub-sampling in practice when test items subjected to collaborative study were, as is common, homogenized prior to the study. The x_i' are assumed to be normally distributed with expectation zero and variance $u^2(x_i)$.

The rationale for this model is presented in detail in Annex A for information.

NOTE   Error is generally defined as the difference between a reference value and a result. In the GUM, "error" (a value) is clearly differentiated from "uncertainty" (a dispersion of values). In uncertainty estimation, however, it is important to characterize the dispersion due to random effects and to include them in an explicit model. For the present purpose, this is achieved by including "error terms" with zero expectation as in Equation (1) above.

4.3.2   Given the model described by Equation (1), the uncertainty $u(y)$ associated with an observation can be estimated using Equation (2).

$$u^2(y) = u^2(\hat{\delta}) + s_L^2 + \sum c_i^2 u^2(x_i) + s_r^2 \tag{2}$$

where

  s_L^2 is the estimated variance of B;

  s_r^2 is the estimated variance of e;

$u(\hat{\delta})$ is the uncertainty associated with δ due to the uncertainty of estimating δ by measuring a reference measurement standard or reference material with certified value $\hat{\mu}$;

$u(x_i)$ is the uncertainty associated with x'_i.

Given that the reproducibility standard deviation s_R is given by $s_R^2 = s_L^2 + s_r^2$, s_R^2 can be substituted for $s_L^2 + s_r^2$ and Equation (2) reduces to Equation (3):

$$u^2(y) = u^2(\hat{\delta}) + s_R^2 + \sum c_i^2 u^2(x_i) \tag{3}$$

4.4 Repeatability data

It will be seen that repeatability data are used in this International Standard primarily as a check of precision, which, in conjunction with other tests, confirms that a particular laboratory may apply reproducibility and trueness data in its estimates of uncertainty. Repeatability data are also employed in the calculation of the reproducibility component of uncertainty (see 6.3 and Clause 10).

5 Evaluating uncertainty using repeatability, reproducibility and trueness estimates

5.1 Procedure for evaluating measurement uncertainty

The principles on which this International Standard is based (see 4.1) lead to the following procedure for evaluating measurement uncertainty.

a) Obtain estimates of the repeatability, reproducibility and trueness of the method in use from published information about the method.

b) Establish whether the laboratory bias for the measurements is within that expected on the basis of the data obtained in 5.1 a).

c) Establish whether the precision attained by current measurements is within that expected on the basis of the repeatability and reproducibility estimates obtained in 5.1 a).

d) Identify any influences on the measurement that were not adequately covered in the studies referenced in 5.1 a), and quantify the variance that could arise from these effects, taking into account the sensitivity coefficients and the uncertainties for each influence.

e) Where the bias and precision are under control, as demonstrated in 5.1 b) and c), combine the reproducibility estimate [5.1 a)] with the uncertainty associated with trueness [5.1 a) and b)] and the effects of additional influences [5.1 d)] to form a combined uncertainty estimate.

These different steps are described in more detail in Clauses 6 to 10.

NOTE   This International Standard assumes that where bias is not under control, corrective action is being taken to bring the process under such control.

5.2 Differences between expected and actual precision

Where the precision differs in practice from that expected from the studies in 5.1 a), the associated contributions to uncertainty should be adjusted. Subclause 7.5 describes adjustments to reproducibility estimates for the common case where the precision is approximately proportional to level of response.

6 Establishing the relevance of method performance data to measurement results from a particular measurement process

6.1 General

The results of collaborative study yield performance indicators (s_R, s_r) and, in some circumstances, a method bias estimate, which form a "specification" for the method performance. In adopting the method for its specified purpose, a laboratory is normally expected to demonstrate that it is meeting this "specification". In most cases, this is achieved by studies intended to verify control of repeatability (see 6.3) and of the laboratory component of bias (see 6.2), and by continued performance checks [quality control and assurance (see 6.4)].

6.2 Demonstrating control of the laboratory component of bias

6.2.1 General requirements

6.2.1.1 A laboratory should demonstrate, in its implementation of a method, that bias is under control, that is, the laboratory component of bias is within the range expected from the collaborative study. In the following descriptions, it is assumed that bias checks are performed on materials with reference values closely similar to the items actually under routine test. Where the materials used for bias checks do not have reference values close to those of the materials routinely tested, the resulting uncertainty contributions should be amended in accordance with the provisions of 7.4 and 7.5.

6.2.1.2 In general, a check on the laboratory component of bias constitutes a comparison between laboratory results and some reference value(s), and constitutes an estimate of B. Equation (2) shows that the uncertainty associated with variations in B is represented by s_L, itself included within s_R. However, because the bias check is itself uncertain, the uncertainty of the comparison in principle increases the uncertainty of results obtained in future applications of the method. For this reason, it is important to ensure that the uncertainty associated with the bias check is small compared to s_R (ideally less than $0,2\ s_R$) and the following guidance accordingly assumes negligible uncertainties associated with the bias check. Where this is the case, and no evidence of an excessive laboratory component of bias is found, Equation (3) applies without change. Where the uncertainties associated with the bias check are large, it is prudent to increase the uncertainty estimated on the basis of Equation (3), for example by including additional terms in the uncertainty budget (2.15).

Where the method is known from collaborative trueness studies to have non-negligible bias, the known bias of the method should be taken into account in assessing laboratory bias, for example by correcting the results for known method bias.

6.2.2 Methods of demonstrating control of the laboratory component of bias

6.2.2.1 General

Bias control may be demonstrated, for example, by any of the following methods. For consistency, the same general criteria are used for all tests for bias in this International Standard. More stringent tests may be used.

6.2.2.2 Study of a certified reference material or measurement standard

A laboratory l should perform n_l replicate measurements on the reference standard under repeatability conditions, to form an estimate Δ_l (equal to the laboratory mean, m, minus the certified value, $\hat{\mu}$) of bias on this material. Where practical, n_l should be chosen such that the uncertainty $\sqrt{s_w^2/n_l} < 0,2\ s_R$. Note that this reference standard is not, in general, the same measurement standard as that used in assessing trueness for the method. Further, Δ_l is generally not equal to B. Following ISO Guide 33 (see Bibliography) with appropriate changes of symbols, the measurement process is considered to be performing adequately if

$$|\Delta_l| < 2\sigma_D \tag{4}$$

σ_D in Equation (4) is estimated by s_D, given by Equation (5):

$$s_D^2 = s_L^2 + \frac{s_w^2}{n_l} \tag{5}$$

where

n_l is the number of replicates by laboratory l;

s_w is the intra-laboratory standard deviation for the n_l replicates or derived from other repeatability studies;

s_L is the inter-laboratory standard deviation derived from collaborative study.

Compliance with the criterion in Equation (4) is taken to be confirmation that the laboratory component of bias B is within the population of values represented in the collaborative study. Note that the reference material or standard is used here as an independent check, or control material, and not as a calibrant.

NOTE 1   A laboratory is free to adopt a criterion more stringent than Equation (4), either by using a factor smaller than 2 or by implementing an alternative and more sensitive test for bias.

NOTE 2   This procedure assumes that the uncertainty associated with the reference value is small compared to σ_D.

6.2.2.3   Comparison with a definitive test method of known uncertainty

A laboratory l should test a suitable number n_l of test items using both the definitive method and the test method in use in the laboratory, to generate n_l pairs of values $(y_i,\ \hat{y}_i)$, where y_i is the result from the definitive method for test item "i", and $\hat{y}_i$ is the value obtained from the routine test method for test item "i." The laboratory should then calculate its mean bias $\overline{\Delta}_y$ using Equation (6) and the standard deviation $s(\Delta_y)$ of the differences as in Equation (7).

$$\overline{\Delta}_y = \frac{1}{n_l}\sum_{i=1}^{n_l}(\hat{y}_i - y_i) \tag{6}$$

$$s(\Delta_y) = \sqrt{\frac{1}{n_l - 1}\sum_{i=1}^{n_l}(\Delta_{y_i} - \overline{\Delta}_y)^2} \tag{7}$$

where $\Delta_{y_i} = \hat{y}_i - y_i$.

Where practical, n_l should be chosen so that the standard deviation $\sqrt{s^2(\Delta_y)/n_l} < 0{,}2\, s_R$. By analogy with Equations (4) and (5), the measurement process is considered to be performing adequately if $|\overline{\Delta}_y| < 2\, s_D$ where $s_D^2 = s_L^2 + s^2(\Delta_y)/n_l$. In this case, Equation (3) is used without change.

NOTE 1   A laboratory is free to adopt a more stringent criterion than $|\overline{\Delta}_y| < 2\, s_D$, either by using a coverage factor smaller than 2 or by implementing an alternative and more sensitive test for bias.

NOTE 2   This procedure assumes that the standard uncertainty associated with the reference method is small compared to σ_D and that the deviations $\Delta_{y_i} = \hat{y}_i - y_i$ can be assumed to arise from a population with approximately constant variance.

6.2.2.4 Comparison with other laboratories using the same method

If a testing laboratory l participates in additional collaborative exercises (for example, proficiency testing as defined in ISO/IEC 17043) from which it may estimate a bias, the data may be used to verify control of bias. There are two likely scenarios.

a) The exercise involves testing a measurement standard or reference material with an independently assigned value and uncertainty. The procedure of 6.2.2.2 then applies exactly.

b) The comparison generates q ($\geqslant 1$) assigned values $y_1, y_2, ..., y_q$ by consensus. The testing laboratory, whose results are represented by $\hat{y}_1, \hat{y}_2, ..., \hat{y}_q$, should then calculate its mean bias $\bar{\Delta}_y$ in accordance with Equation (8) and the standard deviation $s(\Delta_y)$ with respect to the consensus means as in Equation (9).

$$\bar{\Delta}_y = \frac{1}{q}\sum_{i=1}^{q}(\hat{y}_i - y_i) \tag{8}$$

$$s(\Delta_y) = \sqrt{\frac{1}{q-1}\sum_{i=1}^{q}(\Delta_{y_i} - \bar{\Delta}_y)^2} \tag{9}$$

where $\Delta_{y_i} = \hat{y}_i - y_i$.

The measurement process is considered to be performing adequately if $\left|\bar{\Delta}_y\right| < 2\,s_\text{D}$, where $s_\text{D}^2 = s_\text{L}^2 + s^2(\Delta_y)/q$. In this case, Equation (3) is used without change.

NOTE 1   This procedure assumes that the consensus value is based on a number of results that is large compared to q, leading to a negligible uncertainty associated with the assigned value, and that the deviations Δ_{y_i} can be considered to be drawn from a population with approximately constant variance.

NOTE 2   In some proficiency schemes, all returned results $\hat{y}_i$ are converted to z-scores, $z_i = (\hat{y}_i - y_i)/\sigma_0$, by subtracting the assigned value y_i and dividing by the standard deviation σ_0 for proficiency testing (ISO/IEC 17043). Where this is the case, and the standard deviation for proficiency testing is less than or equal to s_R for the method, a mean z-score between $\pm 2/\sqrt{q}$ for q assigned values provides sufficient evidence of bias control. This is convenient to calculate, and is less sensitive to the assumption of constant variance in Note 1, but it should be noted that it is usually a more stringent criterion than described in 6.2.2.4. The laboratory is free to use a more stringent criterion (see Note 3), but the calculation described in 6.2.2.4 is necessary for exact equivalence.

NOTE 3   A laboratory is free to use a more stringent criterion than that described in 6.2.2.4.

6.2.3 Detection of significant laboratory component of bias

As noted in the Scope, this International Standard is applicable only where the laboratory component of bias is demonstrably under control. Where excessive bias is detected, it is assumed that action will be taken to bring the bias within the required range before proceeding with measurements. Such action will typically involve investigation and elimination of the cause of the bias.

6.3 Verification of repeatability

6.3.1   The test laboratory l should show that its repeatability is consistent with the repeatability standard deviation obtained in the course of the collaborative exercise. The demonstration of consistency should be achieved by replicate analysis of one or more suitable test materials, to obtain (by pooling results, if necessary) a repeatability standard deviation s_l with v_l degrees of freedom. The values of s_l should be compared, using an F-test at the 95 % level of confidence, if necessary, with the repeatability standard deviation s_r derived from the collaborative study. Where practical, sufficient replicates should be taken to obtain $v_l \geqslant 15$.

6.3.2 If s_l is found to be significantly greater than s_r, the laboratory concerned should either identify and correct the causes or use s_l in place of s_r in all uncertainty estimates calculated using this International Standard. Note particularly that this will involve an increase in the estimated value of the reproducibility standard deviation s_R, as $s_R = \sqrt{s_L^2 + s_r^2}$ is replaced by $s'_R = \sqrt{s_L^2 + s_l^2}$, where s'_R is the adjusted estimate of the reproducibility standard deviation. Conversely, where s_l is significantly smaller than s_r, the laboratory may also use s_l in place of s_r, giving a smaller estimate of uncertainty.

In all precision studies, it is important to confirm that the data are free from unexpected trends and to check whether the standard deviation s_w is constant for different test items. Where the standard deviation s_w is not constant, it may be appropriate to assess precision separately for each different class of items, or to derive a general model (such as in 7.5) for the dependence.

NOTE   Where a specific value of precision is required, ISO Guide 33 provides details of a test based on

$$\chi_c^2 = \left(\frac{s_w}{\sigma_{w0}} \right)^2$$ with σ_{w0} set to the required precision value.

6.4   Continued verification of performance

In addition to preliminary estimation of bias and precision, the laboratory should take due measures to ensure that the measurement procedure remains in a state of statistical control. In particular, this will involve the following:

— appropriate quality control, including regular checks on bias and precision. These checks may use any relevant stable, homogeneous test item or material. Use of quality control charts is strongly recommended (see References [8] and [9]);

— quality assurance measures, including the use of appropriately trained and qualified staff operating within a suitable quality system.

NOTE   Where control charts are in use, the standard deviation for quality control observations over a period of time should normally be less than the value of s'_R calculated in 6.3.2 if precision and bias are under adequate control.

7   Establishing relevance to the test item

7.1   General

In a collaborative study or an estimation of intermediate measures of precision under ISO 5725-2:1994 and ISO 5725-3:1994, it is normal to measure values on homogeneous materials or test items of a small number of types. It is also common practice to distribute prepared materials. Routine test items, on the other hand, may vary widely, and may require additional treatment prior to testing. For example, environmental test samples are frequently supplied dried, finely powdered and homogenized for collaborative study purposes; routine samples are wet, inhomogeneous and coarsely divided. It is accordingly necessary to investigate, and if necessary allow for, these differences.

7.2   Sampling

7.2.1   Inclusion of sampling process

Collaborative studies rarely include a sampling step; if the method used in-house involves sub-sampling, or the procedure as used routinely is estimating a bulk property from a small sample, then the effects of sampling should be investigated. It may be helpful to refer to sampling documentation such as ISO 11648-1 or other standards for specific purposes.

7.2.2   Inhomogeneity

Inhomogeneity is typically investigated experimentally via homogeneity studies that can yield a variance estimate, usually from an analysis of variance (ANOVA) of replicate results on several test items, in which the inter-item component of variance s_{inh}^2 represents the effect of inhomogeneity. Where test materials are found to be significantly inhomogeneous (after any prescribed homogenization), this variance estimate should be converted directly to a standard uncertainty (i.e. $u_{inh} = s_{inh}$). In some circumstances, particularly when the inhomogeneity standard deviation found from a sample of Q test items from a larger batch and the mean result will be applied to other items in the batch, the uncertainty contribution is based on the prediction interval (i.e. $u_{inh} = s_{inh}\sqrt{(Q+1)/Q}$). It is also possible to estimate inhomogeneity effects theoretically, using a knowledge of the sampling process and appropriate assumptions about the sampling distribution.

7.3   Sample preparation and pre-treatment

In most studies, samples are homogenized, and may additionally be stabilized, before distribution. It may be necessary to investigate and allow for the effects of the particular pre-treatment procedures applied in-house. Typically, such investigations establish the effect of the procedure on the measurement result by studies on materials with approximately or accurately established properties. The effect may be a change in dispersion, or a systematic effect. Significant changes in dispersion should be accommodated by adding an appropriate term to the uncertainty budget (assuming the effect is to increase the dispersion). Where a significant systematic effect is found, it is most convenient to establish an upper limit for the effect. Following the recommendations of the GUM, this may be treated as a limit of a rectangular or other appropriate finite symmetric distribution, and a standard uncertainty estimated by division of the half-width of the distribution by the appropriate factor.

7.4   Changes in test-item type

The uncertainty arising from changes in type or composition of test items compared to those used in the collaborative study should, where relevant, be investigated. Typically, such effects should either be predicted on the basis of established effects arising from bulk properties (which then lead to uncertainties estimated using the basic approach in the GUM) or investigated by systematic or random change in test-item type or composition (see Annex B).

7.5   Variation of uncertainty with level of response

7.5.1   Adjusting s_R

It is common to find that some or most contributions to uncertainty for a given measurement are dependent on the value of the measurand. ISO 5725-2:1994 considers three simple cases where the reproducibility standard deviation for a particular positive value m is approximately described by one of the models

$$\hat{s}_R = bm \tag{10}$$

$$\hat{s}_R = a + bm \tag{11}$$

$$\hat{s}_R = cm^d \tag{12}$$

where

$\hat{s}_R$ is the adjusted reproducibility standard deviation calculated from the approximate model;

a, b, c and d are empirical coefficients derived from a study of five or more different test items with different mean responses m (a, b and c are positive).

Where one of the Equations (10) to (12) applies, the uncertainty should be based on a reproducibility estimate calculated using the appropriate model.

Where the provisions of 6.3 apply, $\hat{s}_R$ should also reflect the changed contribution of the repeatability term s_r. For most purposes, a simple proportional change in $\hat{s}_R$ should suffice, that is

$$s'_R = (a + bm)\frac{\sqrt{s_L^2 + s_l^2}}{\sqrt{s_L^2 + s_w^2}} \tag{13}$$

where s'_R has the same meaning as in 6.3.

7.5.2 Changes in other contributions to uncertainty

In general, where any contribution to uncertainty changes with measured response in a predictable manner, the relevant standard uncertainty in y should be adjusted accordingly.

NOTE   Where many contributions to uncertainty are strictly proportional to y, it is often convenient to express all significant effects in terms of multiplicative effects on y and all uncertainties in the form of relative standard deviations.

8   Additional factors

Clause 7 considers the main factors that are likely to change between collaborative study and routine testing. It is possible that other effects may operate in particular instances, either because the controlling variables were fortuitously or deliberately constant during the collaborative exercise, or because the full range of conditions attainable in routine practice was not adequately covered within the selection during the collaborative study.

The effects of factors which are held constant or which vary insufficiently during collaborative studies should be estimated separately, either from experimental variation or by prediction from established theory. Where these effects are not negligible, the uncertainty associated with such factors should be estimated, recorded and combined with other contributions in the normal way [i.e. following the summation principle in Equation (3)].

9   General expression for combined standard uncertainty

Equation (3), taking into account the need to use the adjusted estimate $\hat{s}_R^2$ instead of s_R^2 to allow for factors discussed in Clause 7, leads to the general expression in Equation (14) for the estimation of the combined standard uncertainty $u(y)$ associated with a result y:

$$u^2(y) = \hat{s}_R^2 + u^2(\hat{\delta}) + \sum_{i=1}^{n'}\left[c_i^2 u^2(x_i)\right] \tag{14}$$

where $u(\hat{\delta})$ is calculated as specified in Equation (15); see also Equation (A.8):

$$u(\hat{\delta}) = \sqrt{s_{\hat{\delta}}^2 + u^2(\hat{\mu})} = \sqrt{\frac{s_R^2 - (1 - 1/n)s_r^2}{p} + u^2(\hat{\mu})} \tag{15}$$

where

p is the number of laboratories;

n is the number of replicates in each laboratory;

$u(\hat{\mu})$ is the uncertainty associated with the certified value $\hat{\mu}$ used to estimate the bias in the collaborative study.

The variable $u(B)$ does not appear in Equation (14) because s_L, the uncertainty associated with B, is already included in $\hat{s}_R^2$. The subscript "i" covers effects identified in Clauses 7 and 8 (assuming these have indices running contiguously from 1 to n'). Clearly, where any effects and uncertainties are small compared to s_R, they may, for most practical purposes, be neglected. For example, uncertainties less than $0{,}2\,s_R$ lead to changes of under $0{,}02\,s_R$ in the overall uncertainty estimate.

NOTE   Where all uncertainty contributions are expressed in the form of relative standard deviations or percentages as suggested in the Note to 7.5.2, Equations (14) and (15) can be applied directly to the relative values and the resulting uncertainty $u(y)$ will be obtained in the form of a relative standard deviation or percentage.

10 Uncertainty budgets based on collaborative study data

This International Standard assumes essentially only one model for the results of a measurement or test: that given in Equation (3). The evidence required to support continued reliance on the model may come from a variety of sources, but where the uncertainties associated with the tests involved remain negligible, Equation (3) is used. However, there are some different situations for which the form of Equation (3) changes slightly, particularly where the reproducibility or repeatability terms depend on the response. The uncertainty budget where the uncertainty is essentially independent of the response over the range of interest is summarized in Table 1, and where the uncertainty depends on the response, in Table 2.

Table 1 — Uncertainty contributions independent of response

Effect	Standard uncertainty[a] associated with y	Comment
δ	$u(\hat{\delta})$	Only included if the collaborative study incorporates a correction for bias and the uncertainty is non-negligible.
B	s_L	See Table 2.
e	s_r	If an average of n_r complete replicates of the method[b] are used in practice on a test item, the uncertainty associated with e becomes $\hat{s}_r / \sqrt{n_r}$
x_i	$\|c_i\|\,u(x_i)$	See Clause 7 and Annex B.
a	These standard uncertainties have the same units as y. They may also be expressed in relative terms (see Note to Clause 9).	
b	The method may itself mandate replication; n_r relates to repetition of the whole method including any such replication.	

Table 2 — Uncertainty contributions dependent on response

Effect	Standard uncertainty[a,b] associated with y	Comment
δ	$\left\|\dfrac{\partial y}{\partial \hat{\delta}}\right\| u\left(\hat{\delta}\right)$	Only included if the collaborative study incorporates a correction for bias and the uncertainty is non-negligible. (The differential is included to cover cases where the correction is not a simple addition or subtraction.)
B	$\hat{s}_L = a_L + b_L m$	a_L and b_L are the coefficients of a presumed linear relationship between s_L and the mean response m, analogous to Equation (11). This form is applicable only when the dependence of s_L on m has been established. Where it has not, use the combined estimate associated with B and e in Table 1.
e	$\hat{s}_r = a_r + b_r m$	a_r and b_r are the coefficients of a presumed linear relationship between s_r and the mean response m, analogous to Equation (11). If an average of n_r complete replicates of the method[c] is used in practice on a test item, the uncertainty associated with e becomes $\hat{s}_r / \sqrt{n_r}$ This form is applicable only when the dependence of s_r on m has been established. Where it has not, use the combined estimate associated with B and e in Table 1.
B, e	$\hat{s}_R = bm$ or $\hat{s}_R = a + bm$ or $\hat{s}_R = cm^d$	a and b are the coefficients of the appropriate established relationship between s_R and the mean response m, as specified in Equations (10), (11) or (12). This combined estimate should be used *instead of* the separate estimates associated with B and e (see Table 1) when the separate dependencies of s_L and s_r on m have not been established.
x_i	$\|c_i\| u(x_i)$	See Clause 7 and Annex B.

a   These standard uncertainties have the same units as y. They may also be expressed in relative terms (see Note to Clause 9).

b   The following assumes a simple linear dependence of the form in Equation (11).

c   The method may itself mandate replication; n_r relates to repetition of the whole method, including any such replication.

11 Evaluation of uncertainty for a combined result

11.1   A "combined result" Y is formed from the results y_i of a number of different tests, each characterized by collaborative study. For example, a calculation for "meat content" would typically combine a protein content, calculated from a nitrogen determination, with a fat and a moisture content, each determined by different standard methods.

11.2   Uncertainties $u(y_i)$ for each contributing result y_i may be obtained by using the principles specified in this International Standard, or directly by using Equation (A.1) or (A.2), as appropriate. Where, as is often the case, the input values y_i are independent, the combined uncertainty $u(Y)$ for the result $Y = g(y_1, y_2, ...)$ is given by Equation (16).

$$u(Y) = \sqrt{\sum_i \left[c_i u(y_i) \right]^2} \tag{16}$$

Where the results y_i are not independent, due allowance should be made for correlation by reference to the GUM [which uses Equation (A.2)].

12 Expression of uncertainty information

12.1 General expression

Uncertainties may be expressed as standard uncertainties $u(y)$ or as expanded uncertainties, $U(y) = ku(y)$, where k is a coverage factor (see 12.2), following the principles of the GUM. It may also be convenient to express uncertainties in relative terms, for example as a coefficient of variation or an expanded uncertainty expressed as a percentage of the reported result.

12.2 Choice of coverage factor

12.2.1 General

In estimating combined expanded uncertainty, the following considerations are relevant in choosing the coverage factor, k.

12.2.2 Level of confidence desired

For most practical purposes, combined expanded uncertainties should be quoted to correspond approximately to a level of confidence of 95 %. However, the choice of level of confidence is influenced by a range of factors, including the criticality of application, and the consequences of incorrect results. These factors, together with any guidance or legal requirement relating to the application, should be given due consideration when choosing k.

12.2.3 Degrees of freedom associated with the estimate

12.2.3.1   For most practical purposes, when approximately 95 % confidence is required and the degrees of freedom in the dominant contributions to uncertainty is large (> 10), the choice of $k = 2$ provides a sufficiently reliable indication of the likely range of values. However, there are circumstances in which this might lead to significant underestimation, notably where one or more significant term(s) in Equation (14) is/are estimated with fewer than 7 degrees of freedom.

12.2.3.2   Where one such term $u_i(y)$ with v_i degrees of freedom is dominant [an indicative level is $u_i(y) \geqslant 0{,}7\ u(y)$], it is normally sufficient to take the effective degrees of freedom v_{eff} associated with $u(y)$ as v_i.

12.2.3.3   Where several significant terms are of approximately equal size and all have limited degrees of freedom (i.e. $v_i \ll 10$), apply the Welch-Satterthwaite equation [Equation (17)] to obtain the effective degrees of freedom v_{eff}.

$$\frac{u^4(y)}{v_{\text{eff}}} = \sum_{i=1}^{N} \frac{u_i^4(y)}{v_i} \tag{17}$$

The value of k is then chosen from v_{eff} by using the appropriate two-tailed value of Student's t for the level of confidence required and v_{eff} degrees of freedom. It is generally safest to round non-integer values of v_{eff} downward to the next lower integer value.

NOTE   In many fields of measurement and testing, the frequency of statistical outliers is sufficiently high compared to the expectation from the normal distribution to warrant extreme caution in extrapolating to high levels of confidence (> 95 %) without good knowledge of the distribution concerned.

13 Comparison of method performance figures and uncertainty data

13.1 Basic assumptions for comparison

Evaluation of measurement uncertainty in accordance with this International Standard will provide a standard uncertainty which, while based primarily on reproducibility or intermediate precision estimates, makes due

allowance for factors that do not vary during the study on which these precision estimates are based. In principle, the resulting standard uncertainty $u(y)$ should be identical to that formed from a detailed mathematical model of the measurement process. A comparison between the two separate estimates, if available, forms a useful test of the reliability of either estimate. The test procedure in 13.2 is recommended.

Note, however, that the procedure is based on two important assumptions.

— First, however a standard uncertainty $u(y)$ with ν_{eff} effective degrees of freedom is estimated, it follows the usual distribution for a standard deviation s with $n - 1$ degrees of freedom [i.e. $(n - 1)(s^2/\sigma^2)$ is distributed as χ^2 with $n - 1$ degrees of freedom]. This assumption permits the use of an ordinary F-test. However, because combined uncertainties may include uncertainties associated with terms from a variety of distributions, and also terms with different variances, the test should be treated as indicative and the level of confidence implied should be viewed with due caution.

— Second, the two uncertainty estimates to be compared are entirely independent. This is also unlikely in practice, as some factors may be common to both estimates. (A more subtle effect is the tendency for judgements about uncertainties to be influenced by known inter-laboratory performance; it is assumed that due care is taken to avoid this effect.) Where significant factors are common to two estimates of uncertainty, the two estimates will clearly be similar far more often than chance alone would dictate. In such cases, where the following test fails to find a significant difference, the result should not be taken as strong evidence for measurement model reliability.

13.2 Comparison procedure

Compare the two estimates $u(y)_1$ and $u(y)_2$, chosen such that $u(y)_1$ is the larger of the two, with effective degrees of freedom ν_1 and ν_2, respectively, using a level of confidence α (e.g. for 95 % confidence, $\alpha = 0,05$), as follows.

a) Calculate $F = [u(y)_1/u(y)_2]^2$.

b) Look up, or obtain from software, the one-sided upper critical value $F_{crit} = F(\alpha/2, \nu_1, \nu_2)$. Where an upper and a lower value are given, take the upper value, which is always greater than 1.

c) If $F > F_{crit}$, $u(y)_1$ should be considered significantly greater than $u(y)_2$.

13.3 Reasons for differences

There may be a variety of reasons for a significant difference between combined uncertainty estimates. These include the following:

— genuine differences in performance between laboratories;

— failure of a model to include all the significant effects on the measurement;

— overestimation or underestimation of a significant contribution to uncertainty.

Annex A
(informative)

Approaches to uncertainty estimation

A.1 The GUM approach

The *Guide to the expression of uncertainty in measurement* (GUM) provides a methodology for evaluating the measurement uncertainty associated with a result y from a model of the measurement process. The GUM methodology is based on the recommendations of the International Bureau of Weights and Measures (BIPM), sometimes referred to as *Recommendation INC-1* (1980). These recommendations first recognize that contributions to uncertainty may be evaluated either by the statistical analysis of a series of observations ("Type A evaluation") or by any other means ("Type B evaluation"), for example using data such as published reference material or measurement standard uncertainties or, where necessary, professional judgement. Separate contributions, however evaluated, are expressed in the form of standard deviations, and, where necessary, combined as such.

The GUM implementation of the BIPM recommendations begins with a measurement model of the form $y = f(x_1, x_2, ..., x_N)$, which relates the measurement result y to the input quantities x_i. The GUM then gives the uncertainty $u(y)$ for the case of independent input quantities as specified in Equation (A.1):

$$u(y) = \sqrt{\sum_{i=1}^{N} c_i^2 u^2(x_i)} \tag{A.1}$$

where

c_i is a sensitivity coefficient evaluated from $c_i = \partial y / \partial x_i$, the partial differential of y with respect to x_i;

$u(x_i)$ and $u(y)$ are *standard uncertainties* (that is, measurement uncertainties expressed in the form of standard deviations) in x_i and y respectively.

Where the input quantities are not independent, the relationship is more complex, as specified in Equation (A.2):

$$u(y) = \sqrt{\sum_{i=1}^{N} c_i^2 u^2(x_i) + \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} c_i c_j u(x_i, x_j)} \tag{A.2}$$

where

$u(x_i, x_j)$ is the covariance between x_i and x_j;

c_i and c_j are the sensitivity coefficients as described for Equation (A.1).

In practice, the covariance is often related to the correlation coefficient r_{ij} as specified in Equation (A.3):

$$u(x_i, x_j) = u(x_i) u(x_j) r_{ij} \tag{A.3}$$

where $-1 \leqslant r_{ij} \leqslant 1$.

In cases involving strong non-linearity in the measurement model, Equation (A.1) is expanded to include higher order terms; this issue is covered in more detail in the GUM.

After calculation of the combined standard uncertainty using Equations (A.1) to (A.3), an expanded uncertainty is calculated by multiplying $u(y)$ by a coverage factor k, which may be chosen on the basis of the estimated degrees of freedom for $u(y)$. This is dealt with in detail in Clause 12.

In general, it is implicit in the GUM approach that the input quantities are measured or assigned. Where effects arise that are not readily defined in terms of measurable quantities (such as operator effects), it is convenient either to form combined standard uncertainties $u(x_i)$ that allow for such effects or to introduce additional variables into the expression $f(x_1, x_2, ..., x_N)$.

Because of the focus on individual input quantities, this approach is sometimes called a "bottom-up" approach to uncertainty evaluation.

The physical interpretation of $u(y)$ is not entirely straightforward, since it may include terms which are estimated by judgement and $u(y)$ may accordingly be best regarded as characterizing a "degree-of-belief" function, which may or may not be observable in practice. However, a more straightforward physical interpretation is provided by noting that the calculation performed to arrive at $u(y)$ actually results in the standard deviation which would be obtained if all input variables were indeed to vary at random in the manner described by their assumed distributions. In principle, this would be observable and measurable under conditions in which all input quantities were allowed to vary at random.

A.2 Collaborative study approach

A.2.1 Basic model

Collaborative study design, organization and statistical treatment are described in detail in Parts 1 to 6 of ISO 5725. The simplest model underlying the statistical treatment of collaborative study data is given (using the same symbols as ISO 5725) in Equation (A.4):

$$y = m + B + e \tag{A.4}$$

where

   m is the expectation for y;

   B is the laboratory component of bias under repeatability conditions, assumed to be normally distributed with standard deviation σ_L;

   e is the random error under repeatability conditions, assumed to be normally distributed with standard deviation σ_w.

Additionally, B and e are assumed to be uncorrelated.

The application of Equation (A.1) to this simple model gives Equation (A.5) for a single result y:

$$u^2(y) = u^2(B) + u^2(e) \tag{A.5}$$

Noting that σ_L^2 and σ_w^2 are the variances associated with B and e respectively and that these are estimated by the between-laboratory variance s_L^2 and the repeatability variance s_r^2 obtained in an inter-laboratory study, so that $u(B) = s_L$ and $u(e) = s_r$, gives Equation (A.6) for the combined standard uncertainty $u(y)$ associated with the result:

$$u^2(y) = s_L^2 + s_r^2 \tag{A.6}$$

By comparison with ISO 5725-2:1994, Equation (A.6) is just the estimated reproducibility standard deviation s_R.

Since this approach concentrates on the performance of the complete method, it is sometimes referred to as a "top-down" approach.

Note that each laboratory calculates its estimate of m from an equation $y = f(x_1, x_2, ...)$ assumed to be the laboratory's best estimate of the measurand value y. Now, if $y = f(x_1, x_2, ...)$ is a complete measurement model used to describe the behaviour of the measurement system, it is expected that the variations characterized by s_L and s_r arise from variation in the quantities $x_1, ..., x_n$. If it is assumed that reproducibility conditions provide for random variation in all significant influence quantities, and taking into account the physical interpretation of $u(y)$ above, it follows that $u(y)$ in Equation (A.6) is an estimate of $u(y)$ as described by Equation (A.1) or (A.2).

The first principle on which this International Standard is based is accordingly that *the reproducibility standard deviation obtained in a collaborative study is a valid basis for measurement uncertainty evaluation.*

A.2.2 Incorporating trueness data

Trueness is generally measured as bias with respect to an established reference value. In some collaborative studies, the trueness of the method with respect to a particular measurement system (usually the SI) is examined by study of a certified reference material (CRM) or measurement standard with a certified value $\hat{\mu}$ expressed in that system's units (ISO 5725-4:1994). The resulting statistical model is specified by Equation (A.7):

$$y = \hat{\mu} + \delta + B + e \tag{A.7}$$

where

$\hat{\mu}$ is a reference value;

δ is the "method bias".

The collaborative study will lead to a measured bias $\hat{\delta}$ with associated standard deviation $s_{\hat{\delta}}$ calculated as specified in Equation (A.8):

$$s_{\hat{\delta}} = \sqrt{\frac{s_R^2 - (1 - 1/n)s_r^2}{p}} \tag{A.8}$$

where

p is the number of laboratories;

n is the number of replicates in each laboratory.

The uncertainty $u(\hat{\delta})$ associated with that bias is given by Equation (A.9):

$$u^2(\hat{\delta}) = s_{\hat{\delta}}^2 + u^2(\hat{\mu}) \tag{A.9}$$

where $u(\hat{\mu})$ is the uncertainty associated with the certified value $\hat{\mu}$ used for trueness estimation in the collaborative exercise.

Where the bias estimated during the trial is included in the calculation of results in laboratories, the uncertainty associated with the estimated bias should, if not negligible, be included in the uncertainty budget.

A.2.3 Other effects — The combined model

In practice, of course, s_R and $u(\hat{\delta})$ do not necessarily include variation in all the effects that influence a measurement result. Some important factors are missing by the nature of the collaborative study, and some

may be absent or under-estimated by chance or design. The second principle on which this International Standard is based is that *effects not observed within the context of the collaborative study shall be demonstrably negligible or explicitly allowed for.*

This is most simply accomplished by considering the effects of deviations x_i' from the nominal value x_i required to provide the estimate of y and assuming approximate linearity of effects. The combined model is then specified in Equation (A.10):

$$y = \hat{\mu} + \delta + B + \sum c_i x_i' + e \tag{A.10}$$

where the summed term is over all effects other than those represented by B, δ and e.

Examples of such effects might include sampling effects, test item preparation, and variation in composition or type of individual test items. Strictly, this is a linearized form of the most general model; where necessary, it is possible to incorporate higher order terms or correlation terms exactly as described by the GUM.

Noting that centring x_i' has no effect on the $u(x_i)$, so that $u(x_i') = u(x_i)$, it follows that the uncertainty associated with y estimated from Equation (A.10) is given by Equation (A.11):

$$u^2(y) = s_L^2 + s_r^2 + u^2(\hat{\delta}) + \sum c_i^2 u^2(x_i) \tag{A.11}$$

where the summation is limited to those effects not covered by other terms.

In the context of method-performance evaluation, it may be noted here that intermediate precision conditions can also be described by Equation (A.10), though the number of terms in the summation would be correspondingly larger because fewer variables would be expected to vary randomly under intermediate conditions than under reproducibility conditions. In general, however, Equation (A.10) applies to any precision conditions subject to suitable incorporation of effects within the summation. In an extreme case, of course, where the conditions are such that the terms s_r and s_L are zero and uncertainty in overall bias is not determined, Equation (A.11) becomes identical to Equation (A.1).

There are two corollaries.

— First, it is necessary to demonstrate that the quantitative data available from the collaborative study are directly relevant to the test results under consideration.

— Second, that even where the collaborative study data are directly relevant, additional studies and allowances may be necessary to establish a valid uncertainty estimate, making due allowance for additional effects [the x_i' in Equation (A.10)]. In allowing for additional effects, it is assumed that Equation (A.1) will apply.

Finally, this International Standard, in asserting that a measurement uncertainty estimate may be reliably obtained from a consideration of repeatability, reproducibility and trueness data obtained from the procedures in all parts of ISO 5725, makes the same assumptions as ISO 5725.

a) Where reproducibility data are used, it is assumed that all laboratories are performing similarly. In particular, their repeatability precision for a given test item is the same, and the laboratory component of bias [(represented by the term B in Equation (A.10)] is drawn from the same population as sampled in the collaborative study.

b) The test material(s) distributed in the study is/are homogeneous and stable.

A.3 Relationship between approaches

The foregoing discussion describes two apparently different approaches to the evaluation of uncertainty. The GUM approach, at one extreme, predicts the uncertainty in the form of a variance on the basis of variances associated with inputs to a mathematical model. The other uses the fact that, if those same influences vary representatively during the course of a reproducibility study, the observed variance is a direct estimate of the same uncertainty. In practice, the uncertainty values found by the different approaches are different for a variety of reasons, including

a) incomplete mathematical models (i.e. the presence of unknown effects);

b) incomplete or unrepresentative variation of all influences during reproducibility assessment.

Comparison of the two different estimates is therefore useful as an assessment of the completeness of measurement models. Note, however, that observed repeatability or some other precision estimate is very often taken as a separate contribution to uncertainty, even in the GUM approach. Similarly, individual effects are usually at least checked for significance or quantified prior to assessing reproducibility. Practical uncertainty estimates therefore often use some elements of both extremes.

Where an uncertainty estimate is provided with a result to aid interpretation, it is important that the deficiencies in each approach be remedied. The possibility of incomplete models is, in practice, usually addressed by the provision of conservative estimates, the explicit addition of allowances for model uncertainty. In this International Standard, the possibility of inadequate variation of input effects is addressed by the assessment of the additional effects. This amounts to a hybrid approach, combining elements of both "top-down" and "bottom-up" evaluations.

Annex B
(informative)

Experimental uncertainty evaluation

B.1 Practical procedure for estimating sensitivity coefficients

Where an input quantity x_i may be varied continuously throughout a relevant interval, it is convenient to study the effect of such changes directly. A simple procedure, assuming an approximately linear change of result with x_i, is as follows.

a) Select a suitable interval over which to vary variable x_i, which should centre on the best estimate (or on the value specified by the method).

b) Carry out the complete measurement procedure (or that part of it affected by x_i) at each of five or more levels of x_i, with replication if required.

c) Fit a linear model to the results, using x_i as abscissa and the measurement result as ordinate.

d) Use the slope of the line so found as the coefficient c_i in Equation (A.1) or Equation (14).

This approach may show different sensitivity coefficients for different test items. This may be an advantage in comprehensive studies of a particular item or class of test items. However, where the sensitivity coefficient is to be applied to a large range of different cases, it is important to verify that the different items behave sufficiently similarly.

B.2 Simple procedure for evaluating uncertainty due to a random effect

Where an input quantity x_j is discontinuous and/or not readily controllable, an associated uncertainty may be derived from analysis of experiments in which the variable varies at random. For example, the type of soil in environmental analysis may have unpredictable effects on analytical determinations. Where random errors are approximately independent of the level of the quantity of interest, it is possible to examine the dispersion of error arising from such variations, using a series of test items for which a definitive value is available or where a known change has been induced.

The general procedure is then as follows.

a) Carry out the complete measurement on a representative selection of test items, in replicate, under repeatability conditions, using equal numbers of replicates for each item.

b) For each observation, calculate the signed difference from the known value.

c) Analyse the results (classified by the quantity of interest) with ANOVA, using the resulting sums of squares to form estimates of the intra-group component of variance s_w^2 and the inter-group component of variance s_b^2. The standard uncertainty $u(x_j)$ arising from variation in x_j is equal to s_b.

NOTE   When different test items or classes of test item react differently to the quantity concerned (i.e. the quantity and test item class interact), the interaction will increase the value of s_b. A detailed treatment of this situation is beyond the scope of this International Standard.

Annex C
(informative)

Examples of uncertainty calculations

C.1 Measurement of carbon monoxide (CO) in automobile emissions

C.1.1 Introduction

Before being put on the market, passenger cars are required to be type-tested to check that the vehicle type complies with regulatory requirements concerning the emission by the motor and the exhaust system of carbon monoxide pollutant gas. The upper limit for approval is specified as 2,2 g/km. The test method is described in the European Directive 70/220 where the following specifications appear.

— The driving cycle (Euro 96) is given as a function of the speed (in km/h), the time (in s) and engaged gear. The car to be tested is put on a specified roller bench to perform the cycle.

— The measuring equipment is a specified CO analysis unit.

— The environment is controlled by using a specified pollution-monitoring cell.

— The personnel have undergone specified training.

Such a test of compliance can be performed in the test laboratory of a production unit of a car manufacturer or in an independent test laboratory.

C.1.2 Collaborative study data

Before adopting and routinely using such a test method, it is necessary to evaluate the factors or sources of influence on the results of the test method (and consequently on the uncertainty of the test results). This is done from experiments conducted in different laboratories. In order to control the test method, an inter-laboratory experiment is designed and conducted according to ISO 5725-2:1994. The purpose of this inter-laboratory experiment is to evaluate the precision of the test method when applied routinely in a given set of test laboratories. The evaluation of precision is made from the data collected with the inter-laboratory experiment, with statistical analysis conducted according to ISO 5725-2:1994. The study is conducted such that every participant undertakes all the processes necessary to carry out the measurement, and all relevant influence factors are accordingly taken into account.

It has been established that the repeatabilities of the laboratories are not significantly different and that the repeatability standard deviation of the test method can be estimated as 0,22 g/km. The reproducibility standard deviation of the test method can be estimated as 0,28 g/km.

C.1.3 Control of bias

The evaluation of trueness (control of bias against a reference) poses methodological and technical questions. There is no "reference car" in the sense of a reference material; trueness must accordingly be controlled by calibration of the test system. For example, the calibration of a CO analysis unit can be made with reference gas and the calibration of the roller bench can be made for quantities such as time, length, speed and acceleration. From a knowledge of emission rates at various speeds and from similar information, it is confirmed that the uncertainties associated with these calibrations do not lead to significant uncertainty contributions associated with the measurement result (that is, all calculated uncertainties are very much less than the reproducibility standard deviation). Bias is accordingly considered to be under due control.

C.1.4 Precision

Typical duplicated test runs by a laboratory have established that the repeatability is approximately 0,20 g/km. This is within the repeatability range found in the inter-laboratory study; the precision is accordingly considered to be under good control.

C.1.5 Relevance of test items

The scope of the method establishes it as suitable for all vehicles within the scope of "passenger cars". While most vehicles achieve compliance relatively easily, and the uncertainty tends to be smaller at lower emission levels, the uncertainty is important at levels close to the regulatory limit. It was therefore decided to take the uncertainty estimated near the regulatory limit as a reasonable, and somewhat conservative, estimate of uncertainty for lower levels of CO emission. Note that where a test shows a vehicle to have emitted substantially more than the limit, it might prove necessary to undertake additional uncertainty studies if comparisons are critical. In practice, however, such a vehicle would not in any case be offered for sale without modification.

C.1.6 Uncertainty estimate

Since the prior studies have established due control of bias and precision within the testing laboratory, and no factors arise from operations not conducted during the collaborative study, the reproducibility standard deviation is used for estimating the uncertainty standard deviation, leading to an expanded uncertainty of $U = 0,56$ g/km, quoted with a coverage factor $k = 2$ which gives a level of confidence of approximately 95 %.

NOTE   The interpretation of results with uncertainties in the field of compliance testing is considered in ISO 10576-1.

C.2 Determination of meat content

C.2.1 Introduction

Meat products are regulated to ensure that the meat content is accurately declared. Meat content is determined as a combination of nitrogen content (converted to total protein) and fat content. The present example shows the principle of combining different contributions to uncertainty, each of which itself arises chiefly from reproducibility estimates, as described in Clause 11.

C.2.2 Basic equations

Total meat content w_{meat} is defined in Equation (C.1):

$$w_{meat} = w_{pro} + w_{fat} \tag{C.1}$$

where

w_{pro} is the total meat protein, expressed as percentage by mass;

w_{fat} is the total fat content, expressed as percentage by mass.

Meat protein w_{pro} is calculated from Equation (C.2):

$$w_{pro} = 100 \, w_{mN} / f_N \tag{C.2}$$

where

f_N is a nitrogen factor specific to the material;

w_{mN} is the total meat nitrogen content.

In this instance, w_{mN} is identical to the total nitrogen content, w_{tN}, as determined by Kjeldahl analysis.

C.2.3 Experimental steps in meat-content determination

The experimental steps involved in the determination of the meat content are as follows.

a) Determine the fat content, w_{fat}.

b) Determine the nitrogen content, w_{mN}, using the Kjeldahl method (mean of duplicate measurements).

c) Calculate the total meat protein content, w_{pro}, using f_N [Equation (C.2)].

d) Calculate the total meat content, w_{meat} [Equation (C.1)].

C.2.4 Uncertainty components

The components of uncertainty to consider are those associated with each of the quantities listed in C.2.3. The most significant relate to w_{pro}, which constitutes some 90 % by mass of w_{meat}. The largest uncertainties associated with w_{pro} arise from the following:

a) uncertainty in the factor f_N owing to incomplete knowledge of the material;

b) variations in the reproducibility of the method, both from run to run and in detailed execution over the long term;

c) uncertainty associated with method bias;

d) uncertainty in fat content w_{fat}.

NOTE   Uncertainties a), b) and c) are associated with the sample, the laboratory and the method, respectively. It is often convenient to consider each of these three factors when identifying gross uncertainties, as well as any necessary consideration of the individual steps in the procedure.

C.2.5 Evaluating uncertainty components

C.2.5.1   Uncertainty associated with f_N

The uncertainty associated with f_N can be estimated from a published range of values. Reference [22] gives the results of an extensive study of nitrogen factors in beef, which show a clear variation between different sources and cuts of meat. Reference [22] also permits calculation of an observed standard deviation for f_N of 0,052 and a relative standard deviation of 0,014 for a large range of sample types.

NOTE   The nitrogen factors determined in Reference [22] used the Kjeldahl method and are accordingly directly applicable for the present purpose.

C.2.5.2   Uncertainty associated with w_{tN}

Information in two collaborative trials[23],[24] allows an estimate of the uncertainty arising from errors in the reproducibility or the execution of the method. Close examination of the trial conditions shows first that each was conducted over a broad range of sample types and with a good, representative range of competent laboratories and, second, that the reproducibility standard deviation s_R correlates well with the level of nitrogen. For both trials, the best-fit line is given by $s_R = 0,021\, w_{tN}$. The same study also shows that the repeatability standard deviation is approximately proportional to w_{tN}, with $s_r = 0,018\, w_{tN}$, and an inter-laboratory term $s_L = 0,011\, w_{tN}$.

The method specifies that each measurement is duplicated and the average taken. The repeatability term, which is an estimate of the repeatability of *single* results, must accordingly be adjusted to account for the effect of averaging two results within the laboratory (see the comment relating to s_r in Table 1). The uncertainty $u(w_{tN})$ associated with the nitrogen content is accordingly

$$u\left(w_{tN}\right) = w_{tN}\sqrt{s_L^2 + \frac{s_r^2}{2}} = w_{tN}\sqrt{0{,}011^2 + \frac{0{,}018^2}{2}} = 0{,}017 w_{tN} \tag{C.3}$$

Equation (C.3) forms the best estimate of the uncertainty in w_{tN} arising from reasonable variations in execution of the method.

The repeatability value is also used as a criterion for accepting the individual laboratory's precision; the method specifies that results should be rejected if the difference falls outside the relevant 95 % confidence interval (approximately equal to $1{,}96 s_r\sqrt{2}$). This check ensures that the intra-laboratory precision for the laboratory undertaking the test is in accordance with that found in the collaborative study.

NOTE   If this check fails more frequently than about 5 % of the time, it is likely that precision is not under sufficient control and action is required to amend the procedure.

Some consideration also needs to be given to uncertainty associated with w_{tN} arising from unknown bias within the method. In the absence of reliable reference materials, comparison with alternative methods operating on substantially different principles is an established means of estimating bias. A comparison of Kjeldahl and combustion methods for total nitrogen across a range of different sample types established a difference of $0{,}01\ w_{tN}$. This is well within the ISO Guide 33 criterion of $2\sigma_D$ [Equation (4)], confirming that uncertainties associated with bias are adequately accounted for within the reproducibility figures.

C.2.5.3   Uncertainty associated with w_{fat}

Additional collaborative trial data for fat analysis[25] provide a reproducibility standard deviation estimate of $0{,}02\ w_{fat}$. The analysis is again undertaken in duplicate and the results accepted only if the difference is within the appropriate repeatability limit, ensuring that the laboratory precision is under control. Prior verification work on a suitable reference material for fat determination establishes that uncertainties associated with bias are adequately accounted for by the reproducibility figures.

C.2.6 Combined uncertainty

Table C.1 shows the individual values and the uncertainties calculated using the above figures.

Table C.1 — Uncertainty budget for meat content

Quantity	Value of x_i % (mass fraction)	$u(x_i)$	$u(x_i)/x_i$
Fat content, w_{fat}	5,50	0,110	0,020
Nitrogen content, w_{mN}	3,29	0,056	0,017
Nitrogen factor, f_N	3,65	0,052	0,014
Meat protein, w_{pro}	90,1	$90{,}1 \times 0{,}022 = 1{,}98$	$\sqrt{0{,}017^2 + 0{,}014^2} = 0{,}022$
Total meat content, w_{meat}	95,6	$\sqrt{1{,}98^2 + 0{,}110^2} = 1{,}98$	0,021

A level of confidence of approximately 95 % is required. This is provided by multiplying the combined standard uncertainty by a coverage factor k of 2, giving (on rounding to two significant figures) an expanded uncertainty U on the meat content of $U = 4{,}0$ %; that is, $w_{meat} = 95{,}6 \pm 4{,}0$ %.

NOTE   "Meat content" can legitimately exceed 100 % in some products.

C.3 Uncertainty for AOAC method 990.12: Aerobic Plate Count

C.3.1 Introduction

The method is a microbiological method for monitoring microbial activity in foodstuffs[27]. The method uses bacterial culture plates of dry medium and water-soluble gel. Samples are added to culture plates at a rate of 1,0 ml per plate and spread over a growth area of approximately 20 cm². Plates are incubated and colonies counted. The measurand is the number of colony-forming units found. For nonzero counts, the conventional reporting units are $\log_{10}$(count), that is, the logarithm to the base 10 of the number of colony-forming units (CFU) found. Uncertainty estimates are desired for three food groups: shellfish, flour and vegetables.

The example here is based on published data from A2LA Guidance Document 108 (A2LA G108, 2007)[28], used by kind permission of the American Association for Laboratory Accreditation.

C.3.2 Collaborative study data

The method was validated by a collaborative study that used eight laboratories, six foods with different levels of contamination, two samples per food, and two replicates per sample. The data analysis was consistent with ISO 5725-2, and the validation study included all steps in the testing process, except for a step involving choice of an exact sub-sample size (measured samples were provided in the collaborative study). Table C.2 shows the reported estimates of repeatability and reproducibility relative standard deviation for the three foods relevant to the uncertainty evaluation requirement, given as percentages.

Table C.2 — Selected collaborative study data for aerobic plate count

Food	Reproducibility relative standard deviation %	Repeatability relative standard deviation %
Shrimp	11,1	9,8
Vegetables	9,2	6,3
Flour	5,8	5,3

Note that the repeatability and reproducibility data are all expressed as relative standard deviations, relative to the mean observed value for $\log_{10}$(count). This is convenient for this particular method, which tends to show dispersion approximately proportional to level and approximately consistent relative standard deviation.

C.3.3 Control of bias

To establish whether laboratory bias is within that expected, the laboratory carries out a comparison study with a reference laboratory. Results for vegetables and shrimp are always within 10 % (corresponding to $\Delta_l < 0{,}1\,\bar{x}$, $\bar{x}$ being the mean of the relevant observations). A comparison with a flour sample shows results 5 % apart (corresponding to $\Delta_l \leqslant 0{,}05\,\bar{x}$). These deviations are clearly consistent with the reproducibility standard deviations; bias is therefore judged to be acceptable.