Reference number of working document: **ISO/TC 69/SC 6 N 456**

Date: 2002-01-15

Reference number of document: **ISO/DTS 21748**

Committee identification: ISO/TC 69/SC 6

Secretariat: JISC/JSA

Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation

Warning

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Document type: Technical Specification
Document subtype: if applicable
Document stage: (30) Committee
Document language: E

Basic template BASICEN2 1999-02-12

Contents

Foreword v

Introduction vi

1 Scope 1

2 Normative references 1

3 Terms and definitions 2

4 Symbols (and abbreviated terms) 4

5 Principles 5
5.1 Individual results and measurement process performance 5
5.2 Applicability of reproducibility data 5
5.3 Basic model 5
5.4 Repeatability data 6

6 Process for evaluating uncertainty using repeatability, reproducibility and trueness estimates 6

7 Establishing the relevance of method performance data to measurement results from a particular measurement process. 6
7.1 Demonstrating control of bias 6
7.1.1 General requirements 6
7.1.2 Methods of demonstrating control of bias 7
7.1.3 Detection of significant bias 8
7.2 Verification of precision 8
7.3 Continued verification of performance 8

8 Establishing relevance to the test item 9
8.1 Sampling. 9
8.2 Sample preparation and pre-treatment. 9
8.3 Changes in test item type 9
8.4 Variation of uncertainty with level of response 10
8.4.1 Adjusting s_R. 10
8.4.2 Changes in other contributions to uncertainty 10

9 Additional factors 10

10 Combining contributions 10

11 Uncertainty budgets based on collaborative study data 11

12 Evaluation of uncertainty for a combined result 13

13 Expression of uncertainty information 13
13.1 General expression 13
13.2 Choice of coverage factor 13
13.2.1 Level of confidence desired 13
13.2.2 Degrees of freedom associated with the estimate 13

14 Comparison of method performance figures and uncertainty data 14

Annex A (Informative) Rationale 15

A.1 Approaches to uncertainty estimation 15
A.1.1 The GUM approach 15
A.1.2 Collaborative study approach 16
A.1.2.1 Basic Model 16
A.1.2.2 Including Trueness data 16
A.1.2.3 Other effects: The combined model 17

A.2 Relationship between approaches 18

Annex B (Informative) Experimental Uncertainty evaluation 19
B.1 Practical experiment for estimating sensitivity coefficients. 19
B.2 Simple experiment for evaluating uncertainty due to a random effect. 19
Annex C (Informative) Examples 20
C.1 Estimation of the uncertainty in a measurement of carbon monoxide (CO) pollutant emissions in the car industry 20
C.1.1 Introduction 20
C.1.2 Collaborative study data 20
C.1.3 Control of bias 20
C.1.4 Precision 21
C.1.5 Relevance of test items 21
C.1.6 Uncertainty estimate 21
C.2 Determination of meat content 21
C.2.1 Introduction 21
C.2.2 Method 21
C.2.3 Uncertainty Components 22
C.2.4 Evaluating Uncertainty Components 22
C.2.5 Combined uncertainty 23
C.2.6 References for example C2 24
Bibliography 25

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 3.

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.

This document is being issued in the Technical Specification series of publications (according to the ISO/IEC Directives, Part 1, 3.1.1.1) as a “prospective standard for provisional application” in the field of measurement and testing because there is an urgent need for guidance on how standards in this field should be used to meet an identified need.

This document is not to be regarded as an “International Standard”. It is proposed for provisional application so that information and experience of its use in practice may be gathered. Comments on the content of this document should be sent to the ISO Central Secretariat.

A review of this Technical Specification will be carried out not later than 3 years after its publication with the options of: extension for another 3 years; conversion into an International Standard; or withdrawal.

Attention is drawn to the possibility that some of the elements of TS 21748 may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 21748:2002 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

Introduction

Knowledge of the uncertainty of 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 real risk of either over- or under-interpretation 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 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, is a widely adopted standard approach, but applies poorly in the absence of a comprehensive model of the measurement process. A very wide range of standard test methods are, however, subjected to collaborative study according to Part 2 of ISO 5725:1994. The present document provides an appropriate and economic methodology for estimating uncertainty for the results of these methods which complies fully with the relevant BIPM principles whilst taking advantage of method performance data obtained by collaborative study.

The dispersion of results obtained in a collaborative exercise may also usefully be compared with measurement uncertainty estimates obtained via 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.

Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation

1 Scope

The document gives guidance for

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

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

It is recognised that ISO 5725-3:1994 provides additional models for studies of intermediate precision. 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 document.

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

The present document assumes that recognised, 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 document are primarily for guidance. It is recognised that while the recommendations presented do form a valid approach to the evaluation of uncertainty for many purposes, other suitable approaches may also be adopted.

In general, references to measurement results, methods and processes in this document should be understood to apply also to testing results, methods and processes.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Specification. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this technical specification are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 3534-1, *Statistics – Vocabulary and symbols – Parts 1 Probability and general statistical terms*

ISO 3534-2, *Statistics – Vocabulary and symbols – Parts 2 Statistical quality control*

ISO 3534-3, *Statistics – Vocabulary and symbols – Parts 3 Design of experiments*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results – Parts 1 General principles and definitions*

ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results – Parts 2 Basic method for the determination of repeatability and reproducibility of a standard measurement method*

ISO 5725-3, *Accuracy (trueness and precision) of measurement methods and results – Parts 3 Intermediate measures of the precision of a standard measurement method*

ISO 5725-4, *Accuracy (trueness and precision) of measurement methods and results – Parts 4 Basic methods for the determination of the trueness of a standard measurement method*

ISO 5725-5, *Accuracy (trueness and precision) of measurement methods and results – Parts 5 Alternative methods for the determination of the precision of a standard measurement method*

ISO 5725-6, *Accuracy (trueness and precision) of measurement methods and results – Parts 6 Use in practice of accuracy values*

Guide to the expression of uncertainty in measurement (GUM), 1995, BIPM/IEC/IFCC/ISO/IUPAC/IUPAP/OIML

International vocabulary of basic and general terms in metrology (VIM), 1993, BIPM/IEC/IFCC/ISO/IUPAC/IUPAP/OIML

ISO/IEC Guide 33:1989, *Uses of certified reference materials*

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

ISO/IEC Guide 43-2:1997, *Proficiency testing by interlaboratory comparisons - Parts 2 Selection and use of proficiency testing schemes by laboratory accreditation bodies*

ISO/IEC 17025:1999, *General requirements for the competence of calibration and testing laboratories*

3 Terms and definitions

For the purposes of this Technical Specification, the following terms and definitions apply. In addition, reference is made to “intermediate precision conditions” which are discussed in detail in ISO 5725:1994 Part 3.

3.1
Uncertainty (of Measurement)
‘parameter, associated with the result of a measurement, that characterises 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 characterised by experimental standard deviations. The other components, which also can be characterised 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.’

[GUM 1995]

3.2
Standard Uncertainty
‘**$u(x_i)$ -** uncertainty of the result of a measurement expressed as a standard deviation.’

[GUM]

3.3
Combined Standard Uncertainty
'**$u_c(y)$ -** 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 co-variances of these other quantities weighted according to how the measurement result varies with these quantities.'

[GUM]

3.4
Expanded Uncertainty
'**U -** quantity defining an interval about a result of a measurement that may be 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 characterised 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 An expanded uncertainty U is calculated from a combined standard uncertainty u_c and a coverage factor k using:

$$U = ku_c$$

[GUM]

3.5
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. '

[GUM]

3.6
Uncertainty Budget
A list of sources of uncertainty and their associated standard uncertainties, compiled with a view to evaluating a combined standard uncertainty for a measurement result.

NOTE The list may additionally include information such as sensitivity coefficients, degrees of freedom for each standard uncertainty, and an identification of the means of evaluating each standard uncertainty in terms of Type A or type B.

3.7
Bias
'The difference between the expectation of the test results and an accepted reference value.

NOTE 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 accepted reference value is reflected by a larger bias value.'

[ISO 3534-1]

3.8
Trueness
'The closeness of agreement between the average value obtained from a large set of test results and an accepted reference value.

NOTE The measure of trueness is normally expressed in terms of bias. The reference to trueness as "accuracy of the mean" is not generally recommended. '

[ISO 3534-1]

3.9
Precision
The closeness of agreement between independent test results obtained under stipulated conditions.

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

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

NOTE 3 “Independent test results” means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular examples of extreme stipulated conditions.

[ISO 3534-1]

3.10
Repeatability
‘Precision under repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.’

[ISO 3534-1]

3.11
Repeatability Standard Deviation
‘The standard deviation of test results obtained under repeatability conditions.

NOTE This is a measure of dispersion of the distribution of test results under repeatability conditions. Similarly “repeatability variance” and “repeatability coefficient of variation” could be defined and used as measures of the dispersion of test results under repeatability conditions. ’

[ISO 3534-1]

3.12
Reproducibility
‘Precision under reproducibility conditions, i.e. conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

NOTE A valid statement of reproducibility requires specification of the conditions changed. Reproducibility may be expressed quantitatively in terms of the dispersion of the results.’

[ISO 3534-1]

3.13
Reproducibility Standard Deviation
‘The standard deviation of test results obtained under reproducibility conditions.

NOTE This is a measure of dispersion of the distribution of test results under reproducibility conditions. Similarly “reproducibility variance” and “reproducibility coefficient of variation” could be defined and used as measures of the dispersion of test results under reproducibility conditions. ’

[ISO 3534-1]

4 Symbols (and abbreviated terms)

Symbols and abbreviated terms are described in their first occurrence in the text.

5 Principles

5.1 Individual results and measurement process performance

5.1.1 Measurement uncertainty is a property of individual results. Repeatability, reproducibility, and bias, by contrast, relate to the performance of a measurement or testing *process*. For studies under ISO 5725:1994, 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 document, the measurement method is assumed to be implemented in the form of a single detailed procedure (as defined in the International Vocabulary of Basic and General Terms in Metrology). It is implicit in this guide 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 (section 7)

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

5.2 Applicability of reproducibility data

The first principle on which the present guide is based is that the reproducibility standard deviation obtained in a collaborative study is a valid basis for measurement uncertainty evaluation. The second principle is that effects not observed within the context of the collaborative study must be demonstrably negligible or explicitly allowed for. The latter principle is catered for by an extension of the basic model used for collaborative study.

5.3 Basic model

5.3.1 The statistical model on which this guidance is based is:

$$y = \mu + \delta + B + \left(\sum c_i x'_i\right) + e \qquad (1)$$

where y is an observed result, μ the (unknown) expectation of ideal results, δ a term representing bias intrinsic to the measurement method in use, B the laboratory component of bias and e an error term. B and e are assumed normally distributed with expectation zero and variance σ_L^2 and σ_r^2 respectively. These terms form the model used by ISO 5725 for the analysis of collaborative study data. y is assumed to be calculated from y=f(x_1, x_2,… x_n) . x'_i is a deviation from the nominal value of x_i, and c_i the linear coefficient $\partial y / \partial x_i$. Since the method bias δ, laboratory bias B and error e are overall measures of error, the summation $\sum c_i x'_i$ is over those effects subject to deviations *other* than those leading to δ, B, or e and provides a method for incorporating effects of operations which are not carried out in the course of a collaborative study. x'_i are assumed normally distributed with expectation zero and variance $u(x_i)^2$. The rationale for this model is given in detail in Annex A for information.

5.3.2 Given the model of equation (1), the uncertainty $u(y)$ associated with an observation can be estimated as

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

where s_L^2 is the estimated variance of B, s_r^2 the estimated variance of e, $u(\hat{\mu})^2$ the uncertainty associated with δ due to the uncertainty of estimating δ by measuring a reference measurement standard or reference material with certified value μ. $u(x_i)$ is the uncertainty associated with x'_i. Given that the reproducibility standard deviation s_R is given by $s_R^2 = s_L^2 + s_r^2$, equation (2) may be abbreviated to

$$u(y)^2 = s_R^2 + u(\hat{\mu})^2 + \sum c_i^2 u(x_i)^2 \qquad (2a)$$

5.4 Repeatability data

It will be seen that repeatability data is used in the present guide primarily as a check on 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 is also employed in the calculation of the reproducibility component of uncertainty (see below). This guide does not, however, describe the application of repeatability data in the absence of reproducibility data.

6 Process for evaluating uncertainty using repeatability, reproducibility and trueness estimates

6.1 The principles on which this document is based 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 at a).

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

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

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

These different steps are described in more detail in the following sections.

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

6.2 Where the precision differs in practice from that expected from the studies at a), the associated contributions to uncertainty should be adjusted. Section 8.4 describes adjustments to reproducibility estimates for the common case where the precision is approximately proportional to level of response.

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

The results of collaborative study yield a set of performance figures (s_R, s_r, and, in some circumstances, a 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 precision and bias, and by continued performance checks (quality control and assurance). These are treated separately below.

7.1 Demonstrating control of bias

7.1.1 General requirements

7.1.1.1 The laboratory should demonstrate that bias in its implementation of the method is under control, that is, that its 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 this is not the case, the resulting uncertainty contributions should be amended in accordance with the provisions of section 8.

7.1.1.2 In general, a bias check constitutes a comparison between laboratory results and some reference value(s), and constitutes an estimate of B. Equation A6 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 $0.2s_R$) and the following guidance accordingly assumes negligible uncertainties associated with the bias check. Where this is the case, and no evidence of bias is found, equation (2a) 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 (2a), for example by including additional terms in the uncertainty budget.

7.1.2 Methods of demonstrating control of bias

Bias control may be demonstrated, for example, by any of the following methods:

7.1.2.1 Study of a certified reference material or measurement standard

The laboratory should perform replicate measurements on the reference standard under repeatability conditions, to form an estimate Δ_l (= laboratory mean – certified value) of bias on this material (note that this is not, in general, the same measurement standard as that used in assessing trueness for the method. Further, Δ_l is not generally equal to B). Following ISO/IEC Guide 33 (with appropriate changes of symbols), the measurement process is considered to be performing adequately if

$$|\Delta_l| < 2.\sigma_D. \qquad (3)$$

σ_D is estimated by s_D:

$$s_D^2 = s_L^2 + \frac{s_w^2}{n_l} \qquad (4)$$

with n_l the number of replicates, s_w the within-laboratory standard deviation derived from these replicates or other repeatability studies, and s_L the between-laboratory standard deviation. (Note that this procedure assumes that the uncertainty associated with the reference value is small compared to σ_D). This is taken to be confirmation that the laboratory 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 more stringent criterion than equation 3.

NOTE 2 Where the method is known to be non-negligibly biased (from collaborative trueness studies) 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.

7.1.2.2 Comparison with a definitive test method of known uncertainty

The laboratory should test a suitable number of test items using both the reference method and the test method. A significance test is performed on the result. Where the measured bias is not statistically significant, equation (2a) is used without change.

7.1.2.3 Comparison with other laboratories using the same method (for example, proficiency testing as defined in ISO/IEC Guide 43)

If a testing laboratory l participates in additional collaborative exercises 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 independently assigned value and assigned uncertainty. The procedure of section 7.1.2.1 then applies exactly.

b) The intercomparison generates q (one or more) 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 $\overline{\Delta}_y = \sum_{i=1,n_l} (\hat{y}_i - y_i)/q$ and standard deviation s(Δ_y) with respect to the consensus means. This value should be subjected to a t-test to determine whether the bias is statistically significant, and also inspected for practical significance compared to the original collaborative study reproducibility s_R. For the t-test, $t = \left|\overline{\Delta}_y\right| / \left(s(\Delta)/\sqrt{q}\right)$, t_{crit} is the value of Student's t for 2 tails, 95% confidence and q-1 degrees of freedom. If the bias so determined is not significant, the bias is considered as under control, and equation (2a) is used without change.

NOTE 1 This procedure assumes that the consensus value is based on a large number of results compared to q, leading to a negligible uncertainty in the assigned value. Where this is not the case, it is prudent to increase the final uncertainty, for example by adding an additional term to the uncertainty budget.

NOTE 2 In some proficiency schemes, all returned results are converted to z-scores by subtraction of the mean value and division by the standard deviation for proficiency testing (ISO/IEC Guide 43). Where this is the case, and the standard deviation for proficiency testing is equal to or less than s_R for the method, a mean z-score between $\pm 2/\sqrt{q}$ over q rounds provides sufficient evidence of bias control.

7.1.3 Detection of significant bias

As noted in the Scope, the present Guide is applicable only where the bias is demonstrably under control. Where a significant bias is detected, 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.

7.2 Verification of precision

7.2.1 The test laboratory 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 a repeatability standard deviation (by pooling results if necessary) of s_i. with ν_i degrees of freedom. s_i should be compared, using an F-test if necessary, with the repeatability standard deviation s_r obtained in the collaborative study.

7.2.2 If s_i is found to be significantly greater than s_r, the laboratory concerned should either identify and correct the causes, or use s_i in place of s_r in all uncertainty estimates calculated using this guide. Note particularly that this will involve an increase in the estimated value of s_R, as $s_R = \sqrt{s_L^2 + s_r^2}$ is replaced by $s_R' = \sqrt{s_L^2 + s_i^2}$ (s_R' denoting the adjusted estimate). Similarly, where s_i is significantly smaller than s_r, the laboratory may also use s_i in place of s_r, giving a smaller estimate of uncertainty.

NOTE 1 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 section 8.4) for the dependence.

NOTE 2 Where a specific value of precision is required, ISO/IEC 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 target precision value.

7.3 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

- 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.

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

8 Establishing relevance to the test item

In collaborative study or estimation of intermediate measures of precision under ISO 5725, 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 homogenised 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.

8.1 Sampling.

8.1.1 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 2859 parts 0-4 (for attributes), ISO 3951 (Sampling procedures for variables) or other standards for specific purposes.

8.1.2 Inhomogeneity is typically investigated experimentally via homogeneity studies which can yield a variance estimate, usually from an ANOVA analysis of replicate results on several test items, in which the between-item component of variance s_{inh}^2 represents the effect of inhomogeneity. Where test materials are found to be significantly inhomogeneous (after any prescribed homogenisation), 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 is to 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.

8.2 Sample preparation and pre-treatment.

In most studies, samples are homogenised, and may additionally be stabilised, 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 positive). 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 distribution, and a standard uncertainty estimated by division by the appropriate factor.

8.3 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 must either be predicted on the basis of established effects arising from bulk properties (which then lead to uncertainties estimated using equation (A1)) or investigated by systematic or random change in test item type or composition (see Annex B).

8.4 Variation of uncertainty with level of response

8.4.1 Adjusting s_R.

It is common to find that some or most contributions to uncertainty for a given measurement are dependent on the level of response. ISO 5725:1994 provides for three simple case where the reproducibility standard deviation s_R (obtained from the published collaborative study data) is approximately described by

$$s_R = b\mathrm{m} \tag{5}$$

$$s_R = a + b\mathrm{m} \tag{6}$$

$$s_R = c\mathrm{m}^d \tag{7}$$

where a, b, c and d are empirical coefficients derived from study of five or more different test items with different mean responses m (a and c must be positive). Where this is the case, the uncertainty should be based on a reproducibility estimate calculated using the appropriate model.

Where the provisions of section 7.2 apply, 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

$$s'_R = (a + bm)(\sqrt{s_L^2 + s_i^2} / \sqrt{s_L^2 + s_w^2}) \tag{8}$$

where s_R' has the same meaning as in section 7.2.

8.4.2 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.

9 Additional factors

Section 8 covers the main factors which 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 attainable in routine practice was not adequately covered within the selection of conditions during the collaborative study.

The effects of factors held constant or varied insufficiently during collaborative studies should be estimated separately, either by 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 (2a)).

10 Combining contributions

Equation 2a gives the following general expression for the estimation of combined standard uncertainty $u_c(y)$ in a result y.

$$u_c^2(y) = \hat{s}_R^2 + u^2(\hat{\delta}) + \sum_{i=k,n} c_i^2 u^2(x_i) \tag{9}$$

where $u(\hat{\delta})$ is calculated according to

$$u(\hat{\delta}) = s_{\hat{\delta}} = \sqrt{\frac{s_R^2 - (1-1/n)s_r^2}{p}} \qquad (10)$$

(equation A8 explains further). u(B) does not appear in the equation because the uncertainty in B is s_L, already included in $\hat{s}_R^2$. The subscript i covers effects identified under sections 8 and 9 (assuming these have indices running contiguopusly from k 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 <0.2s_R lead to changes of under 0.02s_R in the overall uncertainty estimate.

11 Uncertainty budgets based on collaborative study data

The present guide essentially assumes only one model for the results of a measurement or test; equation (2a). The evidence required to support continued reliance on the model may come from a variety of sources, but while the uncertainties associated with the tests involved remain negligible, equation (2a) is used. However, there are some different situations in which the form of equation (2a) changes slightly; particularly where the reproducibility or repeatability terms depend on the response. The following tables summarise the uncertainty budgets for conditions where i) the uncertainty is essentially independent of the response and ii) where the uncertainty depends on the response.

Table 1 - Uncertainty contributions independent of response

Effect	Associated standard uncertainty in y	Comment
μ	$u(\hat{\mu})$	Only included if the collaborative study bias is corrected for and the uncertainty is non-negligible.
δ	$u(\delta)$	Only included if the collaborative study bias is corrected for and the uncertainty is non-negligible.
B	s_L	See below
e	s_r	If an average of n_r complete replicates of the method[a] is used in practice on a test item, the uncertainty associated with e becomes $s_r / \sqrt{n_r}$
x_I	$c_i u(x_i)$	See section 9 and Annex B

a 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 [a]

Effect	Associated standard uncertainty in y	Comment
μ	$\frac{\partial y}{\partial \mu} u(\hat{\mu})$	Only included if the collaborative study bias is corrected for and the uncertainty is non-negligible. (The differential is included to cover cases where the correction is not a simple addition or subtraction)
δ	$\frac{\partial y}{\partial \mu} u(\delta)$	As above
B, e	$s_R = b\mathrm{m}$ or $s_R = a + b\mathrm{m}$ or $s_R = c\mathrm{m}^d$	where a and b are the coefficients of the appropriate established relationship between s_R and the mean response m, as in equations 6 to 7. This combined estimate should be used instead of the separate estimates associated with B and e below when the separate dependencies of s_L and s_r on m have not been established.
B	$s_L = a_L + \mathrm{m} b_L$	a_L and b_L are coefficients of a presumed linear relationship between s_L and the mean response m, analogous to equation 6. This form is only available when the dependence of s_L on m has been established. Where it has not, use the combined estimate associated with B and e above.
e	$s_r = a_r + \mathrm{m} b_r$	where a_r and b_r are coefficients of a presumed linear relationship between s_r and the mean response m, analogous to equation 6. If an average of n_r complete replicates of the method[b] is used in practice on a test item, the uncertainty associated with e becomes $s_r / \sqrt{n_r}$ This form is only available when the dependence of s_r on m has been established. Where it has not, use the combined estimate associated with B and e above.
x_I	$c_i u(x_i)$	See section 9 and Annex B

[a] The following assumes a simple linear dependence of the form in equation 6

[b] The method may itself mandate replication; n_r relates to repetition of the whole method including any such replication.

12 Evaluation of uncertainty for a combined result

12.1 A “Combined result” is formed from the results of a number of different tests (each characterised by collaborative study). For example, a calculation for “meat content” would typically combine protein content calculated from a nitrogen determination with a fat content and a moisture content, each determined by different standard methods.

12.2 Uncertainties $u_c(y_i)$ for each contributing result y_i may be obtained using the principles previously described in this guide, or directly using equations (A1) or (A2) as appropriate. Where, as is often the case, the input values y_i are independent, the combined uncertainty $u_c(Y)$ for the result $Y = \mathrm{f}(y_1, y_2, ...)$ is given by

$$u(Y) = \sqrt{\sum_{i=1,n} c_i^2 u^2(y_i)} \qquad (11)$$

Where the results y_i are not independent, due allowance should be made for correlation by reference to the GUM (which uses equation A2).

13 Expression of uncertainty information

13.1 General expression

Uncertainties may be expressed as combined standard uncertainties $u(y)$ or as combined expanded uncertainties, $U(y) = ku(y)$ 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.

13.2 Choice of coverage factor

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

13.2.1 Level of confidence desired.

For most practical purposes, combined expanded uncertainties should be quoted to correspond approximately with 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 on choosing k.

13.2.2 Degrees of freedom associated with the estimate

13.2.2.1 For most practical purposes, when approximately 95% confidence is required and the number of degrees of freedom in the dominant contributions to uncertainty are large (>10) choosing k=2 provides a sufficiently reliable indication of the likely range of values. However, there are circumstances in which this might lead to significant under-estimation, notably those where one or more significant terms in equation (9) are estimated with under about 7 degrees of freedom.

13.2.2.2 Where one such term $u_i(y)$, with ν_i degrees of freedom, is dominant (an indicative level is $u_i(y) \geq 0.7\, u(y)$) it is normally sufficient to take the effective number of degrees of freedom ν_{eff} associated with $u(y)$ as ν_i.

13.2.2.3 Where several significant terms are of approximately equal size and all have limited degrees of freedom (i.e. <<10), apply the Welch-Satterthwaite equation (equation 12) to obtain the effective number of degrees of freedom ν_{eff}.

$$\frac{u_c^4(y)}{\nu_{eff}} = \sum_{i=1,m} \frac{u_i^4(y)}{\nu_i} \qquad (12)$$

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

NOTE In many fields of measurement, 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.

14 Comparison of method performance figures and uncertainty data

14.1 Evaluation of measurement uncertainty following this guide will provide a standard uncertainty which, while based primarily on reproducibility or intermediate precision estimates, makes due allowance for factors which do not vary during the study on which these precision estimates are based. In principle, the resulting standard uncertainty $u_c(y)$ 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 following test procedure is recommended. Note, however, that the procedure is based on two important assumptions. The first is that, however estimated, a standard uncertainty $u_c(y)$ with ν_{eff} effective degrees of freedom follows the usual distribution for a standard deviation s with n-1 degrees of freedom (that is, s^2 is distributed as χ^2 with n-1 degrees of freedom). This assumption permits use of an ordinary F-test. However, because combined uncertainties may include terms from a variety of distributions, and also terms with difference variances, the test should be treated as indicative and the level of confidence implied should be viewed with due caution. The second assumption is that 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.

14.2 To compare two estimates $u_c(y)_1$ and $u_c(y)_2$, with effective degrees of freedom ν_1 and ν_2, using a level of confidence α (e.g. for 95% confidence, $\alpha = 0.05$):

a) Calculate $F = (u_c(y)_1 / u_c(y)_2)^2$

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 $F > F_{crit}$, $u_c(y)_1$ should be considered significantly greater than $u_c(y)_2$. If $(1/F) > F_{crit}$, $u_c(y)_1$ should be considered significantly less than $u_c(y)_2$.

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

— Genuine differences in performance between laboratories,

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

— Over- or under-estimation of a significant contribution to uncertainty.

Annex A
(Informative)

Rationale

A.1 Approaches to uncertainty estimation

A.1.1 The GUM approach

The Guide to the expression of uncertainty in measurement (GUM), published by ISO, provides a methodology for evaluating the measurement uncertainty $u(y)$ in a result y from a comprehensive mathematical model of the measurement process. The 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 recognise that contributions to uncertainty may be evaluated either by 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 these recommendations begins with a measurement model of the form $y = f(x_1, x_2 \ldots x_n)$, which relates the measurement result y to the input variables x_i. The GUM then gives the uncertainty $u(y)$, as:

$$u(y) = \sqrt{\sum_{i=1,n} c_i^2 u(x_i)^2} \qquad \text{(A1)}$$

for the case of independent variables, 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.

Where the variables are not independent, the relationship is more complex:

$$u(y) = \sqrt{\sum_{i=1,n} c_i^2 u(x_i)^2 + \sum_{\substack{i,k=1,n \\ i \neq k}} c_i c_k u(x_i, x_k)} \qquad \text{(A2)}$$

where $u(x_i,x_k)$ is the covariance between x_i and x_k, and c_i and c_k are the sensitivity coefficients as described for equation (A1). In practice, the covariance is often related to the correlation coefficient r_{ik} using

$$u(x_i,x_k) = u(x_i).u(x_k).r_{ik} \qquad \text{(A3)}$$

where $-1 \le r_{ik} \le 1$.

In cases involving strong non-linearity in the measurement model, equations (A1) and (A2) are expanded to include higher order terms; this issue is covered in more detail in the GUM.

Following calculation of the combined standard uncertainty using equations A1-A3, an expanded uncertainty is calculated by multiplying $u(y)$ by a coverage factor k, which may be chosen on the basis of the estimated number of degrees of freedom for $u(y)$. This is dealt with in detail in section 13.

In general, it is implicit in the GUM approach that the inputs are measured or assigned values. 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 $u(x_i)$ which allow for such effects or to introduce additional variables into $f(x_1, x_2 \ldots x_n)$.

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 characterising 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 calculates the variance 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.1.2 Collaborative study approach

A.1.2.1 Basic Model

Collaborative study design, organisation and statistical treatment is described in detail in ISO 5725 parts 1-6. The simplest model underlying the statistical treatment of collaborative study data is

$$y = m + B + e \qquad \text{(A4)}$$

where m is the expectation for y, B the laboratory component of bias under repeatability conditions, assumed to be normally distributed with mean 0 and standard deviation σ_L, and e the random error under repeatability conditions, assumed normally distributed with mean 0 and standard deviation σ_w. B and e are additionally assumed to be uncorrelated. Applying equation (A1) to this simple model, and noting that σ_w is estimated by the repeatability standard deviation s_r obtained in an interlaboratory study, gives, for a single result y,

$$u(B) = s_L \; ; u(e) = s_r \qquad \text{(A5)}$$

and the combined standard uncertainty $u(y)$ in the result is given by

$$u(y)^2 = u(B)^2 + u(e)^2 = s_L^2 + s_r^2 \qquad \text{(A6)}$$

which, by comparison with ISO 5725:1994, 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 equation used to describe the behaviour of the measurement system and hence calculate m, it is expected that the variations characterised 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 A6 is an estimate of $u(y)$ as described by equations A1 or A2.

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

A.1.2.2 Including 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 μ expressed in that system's units (ISO 5725 part 4, 1994 refers). The resulting model is accordingly

$$y = \mathrm{m} + \delta + B + e \qquad \text{(A7)}$$

where 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

$$s_{\hat{\delta}} = \sqrt{\frac{s_R^2 - (1 - 1/n)s_r^2}{p}} \qquad \text{(A8)}$$

where p is the number of laboratories and n the number of replicates in each laboratory. The uncertainty $u(\hat{\delta})$ associated with that bias is given by

$$u(\hat{\delta})^2 = s_{\hat{\delta}}^2 + u(\hat{\mu})^2 \qquad \text{(A9)}$$

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.1.2.3 Other effects: The combined model

In practice, of course, s_R and $u(\hat{\delta})$ do not necessarily include variation in all the effects operating on a given 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 guide is based is that *effects not observed within the context of the collaborative study must be demonstrably negligible or explicitly allowed for.*

This is most simply accomplished by considering the effects of deviations x'_i from the nominal value of x_i required to provide the estimate of y and assuming approximate linearity of effects. The combined model is then

$$y = \mathrm{m} + \delta + B + \left(\sum c_i x'_i\right) + e \qquad \text{(A10)}$$

where the summed term is over all effects other than those represented by m, B, δ and e. Examples of such effects might include sampling effects, test item preparation, or variation in composition or type of individual test items. Strictly, this is a linearised 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 in y estimated from equation (A10) is given by

$$u(y)^2 = s_L^2 + s_r^2 + u(\hat{\mu})^2 + \sum c_i^2 u(x_i)^2 \qquad \text{(A11)}$$

where, again, 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 (A10), 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 (A10) 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_R are zero and uncertainty in overall bias is not determined, equation (A11) becomes identical to equation (1).

The corollaries are i) that it is necessary to demonstrate that the quantitative data available from the collaborative study are directly relevant to the test results under consideration and ii) 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 A10). In allowing for additional effects, it is assumed that equation (A1) will apply.

Finally, 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 ISO 5725:1994, this document makes the same assumptions as ISO 5725:1994.

a) Where reproducibility data is used, it is assumed that laboratories are all performing similarly. In particular, their repeatability precision for a given test item is the same, and that the laboratory bias (represented by the term δ in equation A10) is drawn from the same population as sampled in the collaborative study.

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

The following sections provide a methodology for verifying that additional effects are negligible and, where they are not, incorporating the resulting uncertainties into an uncertainty estimate for the result.

A.2 Relationship between approaches

The foregoing discussion describes two apparently different approaches to the evaluation of uncertainty. In extreme cases, one, the GUM approach, predicts the uncertainty in the form of a variance on the basis of variances associated with inputs to a mathematical model. The second 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 accordingly 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 accordingly 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 are 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 the present Guide, 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 "top-down" and "bottom-up" evaluation.

Annex B
(Informative)

Experimental Uncertainty evaluation

B.1 Practical experiment for estimating sensitivity coefficients.

Where a variable x_i may be varied continuously through a relevant range, 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 range over which to vary variable x_i which should centre on the typical value (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 through the results, using x_i as abscissa and the measurement result as ordinate.

d) The gradient of the line so found forms the coefficient c_i in equations (A1) or (9).

It may be convenient to express the gradient in relative terms $c_i' = c_i / \bar{y}_i$ where $\bar{y}$ is the mean, or central value of y used in the effect study (this is interpretable as fractional or percentage change in the result per unit change in x_i). The relevant contribution to uncertainty in a result y is then conveniently found by multiplying the term $u(x_i)$ by y where c_i' is the coefficient expressed in relative terms.

NOTE 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 experiment for evaluating uncertainty due to a random effect.

Where an input variable 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 variable 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:

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) Calculate the estimated error in each observation.

c) Analyse the results (classified by the variable of interest) using ANOVA, using the resulting sums of squares to form estimates of the within-group component of variance s_w^2 and the between-group component of variance s_b^2. The standard uncertainty $u_y(x_j)$ 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 variable concerned (i. e. the variable and test item class interact) this simple approach can only be applied to each individual class of test item.

Annex C
(Informative)

Examples

C.1 Estimation of the uncertainty in a measurement of carbon monoxide (CO) pollutant emissions in the car industry

C.1.1 Introduction

Before being put on the market, passenger cars have 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 approval limit is an upper specification limit of 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 expressing speed (in km/h) in terms of time (in s) and engaged gear. The cycle is conducted on a specific roller bench on which is put the tested car.

— The measuring equipment is a specific CO analysis unit.

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

— The personnel is specifically trained.

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 using routinely such a test method, it is necessary to evaluate the factors or sources of influence on the accuracy 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 interlaboratory experiment is designed and conducted according to ISO 5725. The purpose of this interlaboratory 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 interlaboratory experiment, with statistical analysis conducted according to ISO 5725. The study is conducted such that all the processes necessary to undertake the measurement are undertaken by every participant; all relevant are accordingly taken into account.

It is shown that the repeatability standard deviation of the test method can be estimated (the repeatabilities of the laboratories are not shown as significantly different) as 0.22 g/km, and 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 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, acceleration. From a knowledge of emission rates at various speeds and from similar information, it is be confirmed that the uncertainties associated with these calibrations do not lead to significant uncertainties in the final 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

The laboratory has shown that in typical duplicated test runs, its repeatability is approximately 0.20 g/km. This is within the repeatability found in the interlaboratory 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'. Since the uncertainty is important at levels close to the regulatory limit, most vehicles achieve compliance relatively easily, and the uncertainty tends to be smaller at lower levels, it is 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 showed that a vehicle emitted very substantially more than the limit, it might prove necessary to undertake additional uncertainty studies if comparisons were 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 value of the reproducibility standard deviation is used for estimating the uncertainty standard deviation, leading to an expanded uncertainty of

U= 0.56 g/km (k=2)

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 accordingly shows the principle of combining different contributions to uncertainty, each of which itself arises chiefly from reproducibility estimates, as described at section 12.

C.2.2 Method

Total meat content M_{tot} is given by

$$M_{tot} = P_{meat} + F_{tot} \qquad \text{[C1]}$$

where

P_{meat} = total meat protein (%w/w)

F_{tot} = total fat content (%w/w).

Meat protein P_{meat} (in %) is calculated from

$$P_{meat} = 100\, N_{meat} / f_N \qquad \text{[C2]}$$

where f_N is a nitrogen factor specific to the material, and N_{meat}, the total meat nitrogen content. In this instance N_{meat} is identical to the total nitrogen content N_{tot} determined by Kjeldahl analysis.

The table shows the experimental steps involved.

Table C.1 — Experimental steps in Meat Content determination

Step	Description	Quantities (eq C1, C2)
1	Determine fat content	F_{tot}
2	Determine nitrogen content by Kjeldahl	N_{meat}
3	Calculate 'defatted meat' content (eq C2)	P_{meat}, f_N
4	Calculate total meat content (eq. C1)	M_{tot}

C.2.3 Uncertainty Components

The components of uncertainty to consider are those associated with each of the parameters in the table above. The most significant relate to P_{meat}, which constitutes some 90% of M_{tot}. The largest uncertainties in P_{meat} arise from

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

b) the reproducibility of the method, which is subject to variations both from run to run and in detailed execution in the long term and

c) the uncertainty associated with method bias.

NOTE These uncertainties 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.

In addition, the uncertainty in fat content F_{tot} needs to be considered.

C.2.4 Evaluating Uncertainty Components

a) Uncertainty in f_N

The uncertainty in f_N can be estimated from a published range of values. Reference C1 gives the results of an extensive study of nitrogen factors in beef, which show a clear variation between different sources and cuts of meat; the extremes are 3.57 and 3.73 (for defatted meat). The mean value is 3.65 (also the centre point of the range). The observed range is 0.16; if taken as a rectangular distribution, this gives a standard deviation of 0.16/2√3=0.046 (0.013 as relative standard deviation). Reference C1 also permits calculation of the observed standard deviation over a large range of samples, giving a value of 0.052 (0.014 as relative standard deviation).

b) Uncertainty in N_{tot}

Information in two collaborative trials[C2,C3] allows an estimate of the uncertainty arising from imperfect reproducibility or 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 level of nitrogen. For both trials, the best fit line is given by

$$s_R = 0.021\ N_{tot}$$

Thus the figure of 0.021, which is actually a relative standard deviation, is a good estimate of the uncertainty in N_{tot} arising from reasonable variations in execution of the method.

The same study also shows that the repeatability standard deviation is approximately proportional to N_{tot}, with s_r = 0.018 N_{tot}, and a between-laboratory term s_L=0.011. The repeatability value can in principle be used as a criterion for accepting the individual laboratory's precision. However, the method specifies that each measurement is duplicated, and that results should be rejected if the difference falls outside the relevant 95% confidence interval (approximately $1.96 s_r \sqrt{2}$). This check ensures that the within-laboratory precision for the laboratory undertaking the test is in accordance with that found in the collaborative study.

Some consideration also needs to be given to uncertainty in N_{tot} 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 samples of different types found a difference of $0.01 N_{tot}$. This is well within the ISO Guide 33 criterion of $2\sigma_D$ (equation 3) confirming that uncertainties associated with bias are adequately accounted for within the reproducibility figures.

c) Uncertainty in F_{tot}

Additional collaborative trial data for fat analysis[C4] provides a reproducibility standard deviation estimate of $0.02 F_{tot}$. 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.2.5 Combined uncertainty

The table below shows the individual values and the uncertainties calculated using the above figures.

Table C.2 — Uncertainty budget for Meat Content

Parameter	Value V	u	u/V
Fat content F_{tot} (%)	5.5	0.11	0.02
Nitrogen content N_{meat}	3.29	0.072	0.021
Nitrogen factor f_N	3.65	0.051	0.014
Meat protein P_{meat} (%)	90.1	90.1 x 0.025 =**2.3**	$\sqrt{0.021^2 + 0.014^2}$ =0.025
Meat content M_{tot} (%)	95.6	$\sqrt{2.3^2 + 0.11^2}$ =**2.3**	

A level of confidence of approximately 95% is required. This is provided by multiplying the total uncertainty by a coverage factor k of 2, giving an estimated uncertainty U on the meat content of U = 4.6%; that is, M_{tot} = 95.6±4.6%

C.2.6 References for example C2

C1. Analytical Methods Committee: *Analyst* **118** 1217 (1993)

C2. B. Shure, P. A. Corrao, A. Glover, A. J. Malinowski: *J. AOAC* **65** 1339 (1982)

C3. M. King-Brink, J. G. Sebranek: *J. AOAC Int*. **76**, 787 (1993)

C4. D. Breese Jones: US Department of Agriculture Circular no. 183, August 1931

Bibliography

[1] ISO 2859-0:1995 Sampling procedures for inspection by attributes -- Part 0: Introduction to the ISO 2859 attribute sampling system, ISO, Geneva

[2] ISO 2859-1:1999 Sampling procedures for inspection by attributes -- Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection, ISO, Geneva

[3] ISO 2859-2:1985 Sampling procedures for inspection by attributes -- Part 2: Sampling plans indexed by limiting quality (LQ) for isolated lot inspection, ISO, Geneva

[4] ISO 2859-3:1991 Sampling procedures for inspection by attributes -- Part 3: Skip-lot sampling procedures, ISO, Geneva

[5] ISO 2859-4:1999 Sampling procedures for inspection by attributes -- Part 4: Procedures for assessment of stated quality levels. ISO, Geneva

[6] ISO 3951:1989 Sampling procedures and charts for inspection by variables for percent nonconforming. ISO, Geneva

[7] Kaarls R., BIPM Proc.-Verb. Com. Int. Poids et Mesures 1981 **49** A1-A12