

CODEx ALIMENTARIUS COMMISSION



Food and Agriculture
Organization of the
United Nations



World Health
Organization

Viale delle Terme di Caracalla, 00153 Rome, Italy - Tel: (+39) 06 57051 - E-mail: codex@fao.org - www.codexalimentarius.org

Agenda Item 5

CRD04

ORIGINAL LANGUAGE ONLY

JOINT FAO/WHO FOOD STANDARDS PROGRAMME CODEx COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING

42nd Session
Budapest, Hungary

13 – 16 June 2023

with report adoption on 20 June 2023 (virtual)

REVISED DRAFT INFORMATION DOCUMENT PROCEDURES FOR THE ESTIMATION OF MEASUREMENT UNCERTAINTY

(Analysis of comments prepared by Germany)

1 Introduction

A measurement result should ~~always~~ be accompanied by information regarding its uncertainty. Such information provides an indication of the quality of the measurement result and allows meaningful comparison to other measurement results or reference values. Without a statement of measurement uncertainty, a measurement result is essentially incomplete and cannot be properly interpreted.

This document provides ~~guidance information~~ regarding those sources of uncertainty which originate in the laboratory itself, i.e. in connection with the procedures and conditions starting with the laboratory sample and ending with the measurement result. In particular: the question of sampling uncertainty and the extent to which laboratory samples are representative of the content in the container will not be addressed in this information document. Such questions are addressed in CXG 50-2004 [13].

Measurement uncertainty is defined as a parameter "...that characterizes the dispersion of the values that could reasonably be attributed to the measurand", see 2.2.3 in GUM [1]. This document aims to clarify what is meant by in this definition and to provide the information which is necessary to understand how different approaches for the evaluation of measurement uncertainty relate to one another. This should allow the reader to make informed decisions regarding the best procedure to adopt in any given case.

Accordingly, the present document provides background information and clarifies basic notions which are central to a correct evaluation and interpretation of measurement uncertainty. First, the top-down and bottom-up approaches are described and compared. Then, the basic model for the top-down approach is presented. This constitutes a convenient framework within which to elucidate some of the basic conceptual aspects of measurement uncertainty. In the course of the discussion, the term *measurand* will be explained and the relationship between the top-down and bottom-up approaches will be further clarified on the basis of a more general classification of uncertainty sources. The question of the statistical uncertainty in estimating dispersion parameters – such as standard deviation values – will be addressed; and the effect of the number of observations on this statistical uncertainty will be examined. Specific designs for the evaluation of the different components of the top-down approach will then be provided, including designs for the evaluation of subsampling and matrix effects. Finally, examples will illustrate how measurement uncertainty influences sampling plans.

2 Top-down versus bottom-up approaches

The term "bottom-up approach" is used to denote any approach in which the measurement uncertainty is calculated on the basis of an equation expressing the relationship between input variables and the measurement result. In the phrasing from Section 4.1.1 of the *Guide to the expression of uncertainty in measurement* (GUM) [1]: In most cases, a measurand Y is not measured directly, but is determined from N other quantities X_1, X_2, \dots, X_N through a functional relationship (model) f :

$$Y = f(X_1, X_2, \dots, X_N)$$

It must be emphasized that, in this approach, the measurement result Y is *calculated* from the input variables X_1, X_2, \dots, X_N . Analyte concentration is an example of a measurement result; optical density, peak area and signal height are examples of input variables.

An alternative approach – described e.g. in EURACHEM/CITAC Guide CG4 [2] and in ISO 21748 [4] – consists in making use of available *method validation* data. In the words of Section 7.6.1 in the EURACHEM Guide [2]: “A collaborative study carried out to validate a published method [...] is a valuable source of data to support an uncertainty estimate.” In this approach, there is no “functional relationship” between input variables and the measurement result. Rather, results are obtained under different measurement conditions, and total observed variation is partitioned into individual components. This approach is often referred to as the *top-down* approach.

In order to obtain measures of precision which can subsequently be used to “support an uncertainty estimate” following the top-down approach, two main types of experiments can be conducted: single-lab (in-house) and multi-lab (collaborative) studies. It must be emphasized that precision measures obtained in these two types of studies are not always comparable. Nonetheless, if relevant uncertainty sources have not been taken into account, it is often expedient to complement the information from a multi-lab study by subsequent single-lab experiments.

The main distinction between the two approaches is that whereas the bottom-up approach starts from a physico-chemical consideration of the actual measurement mechanism, the top-down approach starts from a data set in which the variation between different measurement results is directly observable. In this sense, it can be said that the bottom-up approach is *theoretical* whereas the top-down approach is *empirical*.

A related distinction is that, in the bottom-up approach, the starting point is the relationship between the measurement result and input variables, whereas, in the top-down approach, the starting point is the relationship between total variation and individual components of variation.

Finally, another distinction between both approaches is that while the number of components in the top-down approach is usually low¹, the number of input variables in the bottom-up approach can be quite high. For this reason, in the bottom-up approach, it will often be impractical to conduct an experiment in which estimates for the uncertainties associated with all the input variables can be reliably obtained. Indeed, the bottom-up approach explicitly allows the inclusion of *prior information* regarding the size of the errors which can be expected to arise in connection with each source (Type B evaluation).

In the case of the bottom-up approach, there are two options for the calculation of the combined (i.e. total) measurement uncertainty: [performing a linear approximation and the Monte Carlo Method \(MCM\)](#).

Bottom-up approach: linear approximation

~~The first option consists in performing a linear approximation. This~~ The first option is often referred to as the law of propagation of uncertainty. In the case that there are no correlations between the different input variables, the combined measurement uncertainty – expressed as a standard deviation – is obtained as follows:

$$u_c = \sqrt{\sum_{i=1}^N c_i \cdot u_i^2}$$

where u_c denotes the combined uncertainty, u_i denotes the uncertainty associated with input variable i and c_i denotes the corresponding sensitivity coefficient, usually obtained via partial differentiation, i.e. $c_i = \left(\frac{\partial f}{\partial x_i}\right)^2$ ($c_i = \left(\frac{\partial f}{\partial x_i}\right)^2$), see 5.1.2 and 5.1.3 in GUM [1].

Bottom-up approach: Monte Carlo method

The second option consists in applying a Monte Carlo method (MCM). This can be briefly described as a computer simulation of the measurement process or (in statistical terms) as “repeated sampling from the probability density function PDFs of the X_i and the evaluation of the model in each case,” see 5.9.1 in [3]. This option is also referred to as the propagation of distributions. In practice, the implementation of this option requires software, since the number of simulation runs (i.e. the number of times each input variable is sampled) is typically on the order of 10^6 . If the model function f is highly nonlinear, the used of MCM is recommended. For instance, in the case of standard addition, the model is

$$Y = \frac{a}{b}$$

In this model, b denotes the slope parameter, calculated as

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

¹ The number of components follows directly from the experimental design of the method validation study.

where the x_i denote the added standard concentrations (with mean value \bar{x}) and the y_i denote the corresponding response values (with mean value \bar{y}); and a denotes the intercept, calculated as

$$a = \bar{y} - b \cdot \bar{x}.$$

The uncertainty values of the individual x_i variables are taken from the certificates of the reference standard substances of materials, while the uncertainty values for the y_i variables are obtained from the regression analysis² (residual standard deviation).

For such a model, the results obtained via linear approximation and via MCM can differ considerably. The MCM calculation will also show whether the distribution of the measurand is asymmetric. For instance, in the case of standard addition, the distribution for the measurand $Y = \frac{a}{b}$ is typically right-skewed:

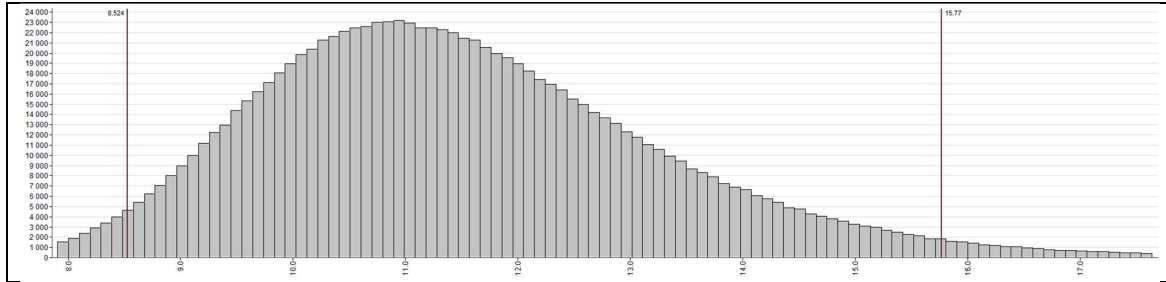


Figure 1: Right-skewed distribution for the standard addition measurand $Y = \frac{a}{b}$ obtained via 10^6 MCM simulation runs.

Top-down approach

In the case of the top-down approach, the total measurement uncertainty is obtained by summing different variance components, such as between-laboratory variance and repeatability variance. The number of replicate measurements should be taken into consideration. For instance, in the simplest case, the total standard uncertainty is obtained as

$$u = \sqrt{s_L^2 + \frac{s_r^2}{n_r}}$$

where s_L denotes the between-laboratory standard deviation, s_r denotes the repeatability standard deviation and n_r denotes the number of replicates whose mean value is taken as the final measurement result. For further information, the reader is referred to ISO 21748 [4].

3 Basic model for the top-down approach

In this section, the basic model for the top-down approach is discussed. The model is premised on the assumption that data from an interlaboratory validation study (also known as a collaborative study) are available. Such a study is conducted in order to characterize the performance of an analytical method. In particular, the characterization of the *precision*³ of an analytical method can be used “to support an uncertainty estimate”. The reader is referred to the ISO 5725 series – in particular to Part 2 [5] – for background information.

The basic model is as follows:

Measurement value Y

$$= \text{true value} + \text{method bias (average across labs and matrices)} + \text{matrix-specific bias} \\ + \text{laboratory bias} + \text{repeatability error}$$

For further details, the reader is referred to [6] and [7]. In the case of an in-house method, the laboratory bias term is replaced by a term representing intermediate effects (whether via measurements performed on different days, or via a factorial design), see ISO TS 23471 [20] and Section 9.1 below.

In the following, the individual terms of the basic model are discussed.

² Strictly speaking, a linear regression approach that takes into account errors in observations in both the x- and y-axes (e.g. Deming regression) should be applied.

³ Precision is defined (paraphrasing 2.15 in [8]) as the degree of agreement between independent measurement results obtained under specified conditions. For instance, reproducibility precision characterizes the agreement between results from different laboratories, while repeatability precision characterizes the agreement between results obtained under near-identical conditions in the same laboratory. Precision can be used to derive a measurement uncertainty estimate – but it must not be confused with measurement uncertainty.

True value

In general, the true value is not known. It can be estimated by averaging e.g. across methods, samples and laboratories (this will often result in a mere approximation). However, it is crucial to note that in the GUM [1], measurement uncertainty is defined *without any reference to a true value*; rather, it is defined as a parameter "... that characterizes the dispersion of the values which could reasonably be attributed to the measurand", see 2.2.3 in GUM [1]. This definition has since been adopted in all other relevant standards and guidance documents (EURACHEM [2], VIM [8]). This does not mean that the true value no longer plays a role in the evaluation of measurement uncertainty. However, it is not the (unavailable) difference between true value and measurement result, but *the uncertainty of bias correction* which must be taken into account in the evaluation of measurement uncertainty. In other words, the focus shifts from the (unavailable) true value itself to the uncertainty in the estimation of the bias. Note that if a certified reference value is available along with a reference uncertainty value, the latter can be included in the uncertainty of bias correction.

Method bias (average across labs and matrices)

The method bias ~~across both labs and matrices~~ can be estimated by averaging across laboratories ~~and matrices~~. A prerequisite is that, within each participating laboratory, measurements are performed under appropriate conditions, with appropriate equipment and trained personnel. Any available information from (certified) reference materials should be taken into consideration. As explained in the discussion of the true value, the corresponding contribution to the calculation of measurement uncertainty will consist in the uncertainty in the estimate of this bias.

Matrix-specific bias (matrix mismatch)

In many cases, a method's bias depends on the matrix being examined. In other words: bias varies from one matrix to another. Such effects occur when the extraction of analyte is affected by the matrix, so that a part of the analyte is not recovered; or when a part of the matrix is extracted along with the analyte and interacts with the measurement's physico-chemical mechanism, resulting in a bias. The corresponding component of total variability is referred to as the matrix-mismatch component. It is important to note that all the uncertainty sources listed in Section 7 contribute to this term of the basic model. See Uhlig (2023) [24] for further information.

Laboratory bias

In many cases, a method's bias depends on the laboratory which is performing the measurement. In other words, the bias varies from laboratory to laboratory. It is often impractical or insufficient to estimate laboratory bias and perform a correction of laboratory bias⁴. Accordingly, in order to take laboratory bias into consideration in estimating measurement uncertainty, it is often expedient to resort to the expected range of laboratory bias, characterized via the~~The corresponding component of total variability is called the~~ laboratory standard deviation (one of the components of the reproducibility precision). A prerequisite is that, within each laboratory that participated in the validation study, measurements were performed under appropriate conditions, with appropriate equipment and trained personnel.

Repeatability error

This term represents variation across replicate measurements (i.e. independent measurements performed under near-identical test conditions).

Note regarding the case that the precision depends on the concentration level!:

When there is a known relationship between precision (e.g. in-house reproducibility) and concentration, it is possible to apply an approach based on a clear distinction between, on the one hand, random variation between test results at a given concentration level, and, on the other, the range of values which can "reasonably be attributed to the measurand," i.e. the measurement uncertainty. This approach is described in Uhlig (2023) [25] and gives rise quite naturally to asymmetrical measurement uncertainty intervals in cases of relatively high-poor precision (say, greater than 10 %) and heteroscedasticity (e.g. constant *relative* in-house reproducibility). This approach is also described in Annex E of ISO TS 23471 [20].

Note regarding bias and (certified) reference materials

Bias can be taken into account in two different ways in measurement uncertainty.

Case 1: Uncertainty of bias correction

⁴ This is principally due to the absence of appropriate reference material, intermediate precision effects on bias estimates and the fact that laboratory bias may vary from one matrix to the next.

If (certified) reference material is available, then an estimate of bias can be obtained, and a bias correction can be performed. If a bias correction is performed, the uncertainty of the bias correction should be included in the measurement uncertainty. In the simplest case, the uncertainty of the bias correction has two components: the uncertainty of the mean value \bar{y} obtained in the experiment, and the uncertainty of the reference value x . (In this simplest case, the estimate of bias is $\bar{y} - x$). The uncertainty of the reference value x may be a value available from a certificate (e.g. in the case of certified reference material); however, it may also be derived from precision data from a validation study.

Case 2: Predication range of bias

If several estimates of bias are available, then the expected range of bias can be included in measurement uncertainty. This is typically done in the top-down approach, when the measurement uncertainty estimate is derived from the precision data of a collaborative study. Indeed, for each sample, the difference between each laboratory mean and the overall mean can be considered an estimate of laboratory bias, and the between-laboratory standard deviation s_L characterizes the expected range of laboratory bias. Similarly, if data obtained on the basis of several matrices are available, the variation of (method and/or laboratory) bias across matrices can be characterized. In this approach, the availability of (certified) reference material is not necessary; nor is the performance of bias correction required. (Though the interpretation of the variation of bias will depend on whether correction was previously performed and on whether the variation was calculated in relation to a reference value.)

4 Specifying the measurand

The concept “measurand” clearly plays a central role in the definition of measurement uncertainty and will shed further light on the connection between validation data and measurement uncertainty.

Leaving aside the technicalities of the definition of a measurand⁵, it is sufficient to note that the specification of a measurand has three separate components:

- specification of a property, e.g. *mean arsenic concentration*. Note that the concept “analyte” corresponds to this part of the specification of the measurand
- specification of a phenomenon, body or substance which the property is associated with, e.g. *a given batch of apple juice*. Note that the concept “matrix”, used in the previous section, corresponds to this part of the specification of the measurand
- and specification of a reference framework regarding the manner in which the property is characterized, e.g. [ng/ml]

Loosely phrased, specifying a measurand thus involves stating (1) *what* is to be measured, (2) *what* is it to be measured *in*, and (3) *how* should the measurement result be expressed in order to ensure comparability to other measurement results or relevant values?

In particular, the specification of the measurand should include information as to whether analyte concentration is to be measured in a laboratory sample or in a “larger sample” or a batch of products in a container. Only in the latter case is *sampling* uncertainty relevant (see Section 7 for an overview of the different sources of uncertainty). Similarly, if measurement results from several laboratory samples are used to assess the conformity of bulk material from a container, it is the measurement uncertainty of the mean value across the results corresponding to the individual laboratory samples which is relevant.

More generally, while measurement uncertainty is always determined on the basis of the laboratory sample, it is nevertheless important to include all available information about the laboratory sample in the evaluation of measurement uncertainty, e.g.

- Where does the material come from (e.g. container)?
- Have other samples from the same origin been tested?
- What is the intended use of the measurement result (e.g. conformity assessment for the individual laboratory sample or for the container)?

⁵ In the VIM [8], measurand is defined (definition 2.3) as “quantity intended to be measured”. Quantity, in turn, is defined (definition 1.1) as “property of a phenomenon, body, substance, where the property has a magnitude that can be expressed as a number and a reference”. An example given directly under this definition is “amount-of-substance concentration of ethanol in wine sample *i*”. The term “reference” in this definition is explained in NOTE 2 as: “A reference can be a measurement unit, a measurement procedure, a reference material, or a combination of such.”

For example, determining the contribution to uncertainty which arises from the material's heterogeneity (e.g. fundamental variability, see Section 9.4) may require a considerable amount of work, depending on the analyte, concentration and grain/particle size. If the origin of the material is known, it may be possible to use previously obtained results regarding the heterogeneity contribution to uncertainty instead of obtaining a new estimate from scratch.

The specification of the measurand should also make it possible to determine whether bias/recovery correction is required, and what form this correction should take. For example, if the measurand is specified in terms of the amount of analyte recovered, then recovery correction may not be appropriate. On the other hand, if the measurand is specified in terms of the total amount of analyte present in a test sample, then recovery correction may be necessary.

Finally, it may be impractical or impossible to provide an exhaustive specification of the measurand. For this reason, it may be necessary to include an extra component of measurement uncertainty, called "definitional uncertainty" (see definition 2.27 in VIM [8]), in order to account for any ambiguity ("finite amount of detail") in the specification of the measurand. However, in most cases, the definitional uncertainty can be considered negligible.

5 Relation between measurand and validation data

If the results of a validation study are to be used to determine measurement uncertainty, it must be ensured that the study refers to the same measurand.

Example 1: Measurement uncertainty is being evaluated in a given laboratory for a measurand specified in terms of analyte concentration in test samples. The analytical method used has been validated for the same analyte, but on the basis of extracts rather than test samples. In other words, the measurand for the validation study is analyte concentration in extracts. It follows that the measurand for which measurement uncertainty must be evaluated is different from the measurand from the validation study. Accordingly, the measurement uncertainty cannot be evaluated on the basis of the characterization of the dispersion of measurement results from the validation study.

Example 2: Measurement uncertainty is being evaluated in a given laboratory for a measurand which is specified in terms of a range of matrices. The analytical method used has been validated for the same analyte, but for only one of the matrices. It follows that the measurand for which measurement uncertainty must be evaluated is different from the measurand from the validation study. Accordingly, the measurement uncertainty cannot be evaluated on the basis of the characterization of the dispersion of measurement results from the validation study (the matrix [bias-mismatch](#) term is missing, [see Section 9.2](#)).

The conditions under which validation data can be used to support a measurement uncertainty estimate can be stated as follows:

If...

the measurement result is obtained using a validated method

and the *measurand* is included in the scope of the validation

and precision ([in particular, in-house reproducibility standard deviation](#)) within the laboratory which is evaluating measurement uncertainty is comparable to the method's precision as characterized in the validation study

then...



the precision estimates from the validation study can be used in the calculation of measurement uncertainty.

Note: These precision estimates can also be used to derive an estimate of the uncertainty of method bias correction.

In order to check and provide evidence of competence in the application of the method and to ensure adequate precision in the laboratory which is evaluating measurement uncertainty, it may be necessary to perform a verification study.

The reader is referred to Section 7 in EURACHEM [2] for further guidance regarding using validation data in the evaluation of measurement uncertainty.

6 Empirical versus rational methods

In the definition of the measurand, the specification of the property must include sufficient information to allow an appropriate reference (see 1.1 in the VIM [8]) to be selected. In particular, it is important to distinguish between

- Empirical method (type I methods in the Codex system)
- Rational method (type II-IV methods in the Codex system)

In Section 5.4 of EURACHEM [2], the following explanation is provided: “*In analytical measurement, it is particularly important to distinguish between measurements intended to produce results which are independent of the method used, and those which are not so intended. The latter are often referred to as empirical methods or operationally defined methods.*”

In Section 5.5 of the same document, it is explained that non-empirical methods are sometimes called rational methods. This distinction is closely related to that between *operationally defined* and *non-operationally defined* measurands found in Section 9.2.3 of ISO Guide 35 [9]. The reader is also referred to Section 3.1 in the EURACHEM Guide to Metrological Traceability in Chemical Measurement [21].

As far as the evaluation of measurement uncertainty is concerned, this distinction has the following important implication: for *empirical* methods (*operationally defined* measurands), there is no method bias term in the basic model for the top-down approach described in Section 3. (Please note that the bottom-up approach does not allow the distinction *method* versus *other* bias components).

7 Uncertainty sources in the top-down and bottom-up approaches

In the *top-down* approach, total variation observed in a data set is partitioned into different components. In the *bottom-up* approach, the total uncertainty is obtained from uncertainty values associated with individual input variables. The following question arises: what is the *relationship* between the components from a top-down model and the uncertainty sources included in a bottom-up model?

In order to answer this question, an overview of different types of uncertainty sources – *independently of the approach* – is now provided. The intention is to distinguish broad categories of uncertainty sources. Apart from shedding further light on the relationship between the top-down and bottom-up approaches, this overview may prove useful for determining which sources may be relevant in any given case, and whether all relevant sources have been included in the evaluation of measurement uncertainty.

Sources of uncertainty are conveniently classified under six main headings:

- Sampling (The question of sampling uncertainty is not addressed in the present document. The reader is referred to CXG 50-2004 [13])
- Storage/transportation
- Subsampling
- Measurement conditions
- Measurement procedure
- Computational effects

Source of uncertainty	Role in measurement uncertainty
<i>Sampling</i>	<p>If the measurand is defined in terms of e.g. analyte concentration in a container or in a batch of products, then sampling is required, and its contribution to measurement uncertainty must be assessed, see Section 7.6 in ISO 17025 [10].</p> <p>If the measurand is defined in terms of a single test material (laboratory sample), then there is no contribution to uncertainty due to sampling. There may be a contribution from subsampling, however (i.e. obtaining test portions from the laboratory sample).</p> <p><i>Fundamental variability</i> is one of the “subcomponents” of sampling uncertainty, see the discussion in Section 9.4.</p>
<i>Storage/transportation</i>	<p>If different storage or shipping conditions have an effect on measurement results, then the corresponding contribution to the total uncertainty must be taken into account.</p>
<i>Subsampling</i>	<p>This term denotes taking test portions from the laboratory sample. If the latter is not homogeneous (finely ground in case of solid matter, mixed or agitated in case of liquids and semi-solids), then it cannot be ensured that the subsampling uncertainty is negligible. Accordingly, appropriate homogenisation is required before subsampling in order to reduce this uncertainty source.</p> <p><i>Fundamental variability</i> is one of the “subcomponents” of subsampling uncertainty, see the discussion in Section 9.4.</p>
<i>Measurement conditions</i>	<p>It must be emphasized that the term measurement as used here includes any sample preparation and clean-up procedures.</p> <p>If different measurement conditions (e.g. different time of year, different technician, different reagents, different equipment) contribute to measurement uncertainty, this source must be taken into consideration.</p>
<i>Measurement procedure</i>	<p>This term denotes the intrinsic or irreducible uncertainty component associated with the physical/chemical/biochemical mechanisms involved in the measurement procedure (including sample preparation and clean-up procedures), e.g. extraction efficiency. The input variables in the bottom-up approach can be considered to belong under this heading.</p>
<i>Computational effects</i>	<p>Inaccurate calibration model and calculation methods, peak integration procedures and rounding will also contribute to measurement uncertainty.</p>

Note regarding quantitative microbiological methods and estimating measurement uncertainty in accordance with ISO 19036

In ISO 19036, a top-down approach is described. This approach distinguishes three components of measurement uncertainty: the technical uncertainty, the matrix uncertainty and the distributional uncertainty.

The technical uncertainty represents sources of uncertainty represented above under the categories *Measurement conditions* and *Measurement procedure*. The technical uncertainty thus arises from the implementation of the method *per se*. The technical uncertainty does not include any variation arising from the heterogeneity of the laboratory sample. Such variation is represented by the matrix uncertainty. The matrix uncertainty as defined in ISO 19036 thus corresponds to the subsampling source of uncertainty listed above. Conceptually, it has no relation whatsoever to the matrix-specific bias term (matrix mismatch) described in Section 3. Finally, the distributional uncertainty arises from the fact that the number of discrete cells (colony forming units) may vary from test portion to test portion, even if the laboratory sample is perfectly homogenous. The distributional uncertainty is thus related to the fundamental variability discussed in Section 9.4 below. It should be mentioned that the following uncertainty components are not covered by ISO 19036:

- method or lab bias (there are no “true values” for microbiological measurands)
- sampling

Note regarding subsampling:

In the top-down approach, any estimate of measurement uncertainty must take into consideration at least the following two components: laboratory bias and repeatability. For non-destructive methods, any sub-sampling variation contributes to the repeatability component. In order to reflect sub-sampling variation found in routine samples, “real” samples must be used in the validation study. If this is not practicable (e.g. because the samples differ too much from lab to lab), and homogenous test material is used, then the sub-sampling component of repeatability must be estimated in a separate experiment. The sub-sampling component must not be confused with the matrix bias (matrix mismatch) component, which may vary considerably from lab to lab, thus inflating the between-laboratory component.

8 Requirements regarding data size

If a standard deviation is calculated on the basis of a series of measurement results, how well does it characterize the actual dispersion of the values? Indeed, if several measurement series are performed and a separate standard deviation value is calculated for each, these standard deviation values will differ. In other words, a given standard deviation, obtained on the basis of empirical data, only represents an *estimate* of the “true” standard deviation.⁶

The confidence interval for a standard deviation can be obtained by means of the following Excel formula: $\text{SQRT}((N-1)/\text{CHISQ.INV}(p, N-1))$, where p is the probability value (e.g. 0.025 or 0.975) and N is the number of laboratories or the number of tests inside the single laboratory. This Excel formula corresponds to the following mathematical formulas for the lower and upper limits (LCL and UCL) of a 95 % confidence interval given a standard deviation estimate s : $\text{LCL} = \frac{\sqrt{N-1}}{\sqrt{\chi^2_{(N-1, 0.975)}}} \cdot s$ and $\text{UCL} = \frac{\sqrt{N-1}}{\sqrt{\chi^2_{(N-1, 0.025)}}} \cdot s$, where $\chi^2_{(v, p)}$ denotes the p -quantile of a chi-squared distribution with v degrees of freedom.

It is recommended that standard deviations be computed on the basis of a minimum of $N = 12$ values (corresponding to $v = 11$ degrees of freedom for the estimation of the standard deviation), in which case $\chi^2_{(N-1, 0.975)} = \chi^2_{(11, 0.975)} = 21.92$ and $\chi^2_{(N-1, 0.025)} = \chi^2_{(11, 0.025)} = 3.82$, and the confidence interval for the standard deviation is $[0.71 \cdot s, 1.70 \cdot s]$.

As far as the simultaneous estimation of e.g. between-laboratory (or between-matrix) standard deviation and repeatability standard deviation is concerned, this recommendation means that measurement results from at least 12 laboratories (or matrices) should be available, each with at least two replicates per laboratory (or matrix).

It is required that data from at least 8 laboratories must be available (see Section 6.3.4 in ISO 5725-1 [18] where 8-15 laboratories is proposed as a “common” figure).

In the case that different uncertainty sources are *simultaneously* taken into consideration, say in the bottom-up approach, the requirement regarding data size can be applied via the Welch-Satterthwaite formula, see [1] Annex G, G.4.1. More specifically: take the case that 2 different uncertainty sources are included in the calculation of the combined uncertainty, u_1 and u_2 . Say that each was obtained by applying the formula for the sample standard deviation on the basis of n_1 and n_2 measurement results, respectively. The number of degrees of freedom for the combined uncertainty can then be computed as

⁶ Table 3 in CXG 59 [11] provides expected ranges for standard deviation estimates calculated from empirical data for different values of N (number of observations). Please note that expected ranges must not be confused with the confidence intervals.

$$\text{Degrees of freedom for combined uncertainty} = \frac{(u_1^2/n_1 + u_2^2/n_2)^2}{\frac{(u_1^2/n_1)^2}{n_1 - 1} + \frac{(u_2^2/n_2)^2}{n_2 - 1}}$$

The recommendation is to ensure a minimum of 11 degrees of freedom for the combined uncertainty.

In the case that prior information is used for an individual u_i value (Type B variable) and that no information regarding data size is available, it is suggested to use $n_i = 7$; the uncertainty which corresponds to this data size is intended to reflect the fact that, in the case of Type B variables, distributional assumptions are often based on “educated guesses.”

Example of the application of the Welch-Satterthwaite formula

Take the case that measurement uncertainty must be evaluated on the basis of the following functional relationship, where the measurement result Y is expressed as a function of 4 input variables:

$$Y = f(X_1, X_2, X_3, X_4) = X_1 + X_2 + X_3 + X_4$$

Table 1: Data size and uncertainty values for the input variables

Input variable	Type	n	u^2
X_1	A	3	4
X_2	B	30	15
X_3	B	30	15
X_4	B	Not available Take $n_4 = 7$	5

The Welch-

The Satterthwaite formula can now be applied.

$$\begin{aligned} & \text{Degrees of freedom for combined uncertainty} \\ &= \frac{(u_1^2/n_1 + u_2^2/n_2 + u_3^2/n_3 + u_4^2/n_4)^2}{\frac{(u_1^2/n_1)^2}{n_1 - 1} + \frac{(u_2^2/n_2)^2}{n_2 - 1} + \frac{(u_3^2/n_3)^2}{n_3 - 1} + \frac{(u_4^2/n_4)^2}{n_4 - 1}} \\ &= 9.4 \end{aligned}$$

We would suggest users of the Welch-Satterthwaite formula refer to the GUM [1] Annex G for a discussion of the formula and the range within which the calculated degrees of freedom should lie, since incorrect implementations commonly fall outside of this range.

9 Simple procedures for evaluating uncertainty components

If validation data are incomplete (i.e. some of the relevant sources of uncertainty have not been characterized), further experiments must be conducted before the top-down approach can be applied.

For instance, in a collaborative study, each participating laboratory should ideally receive samples representing different matrices and different analyte concentrations. However, due to restrictions in material availability, collaborative studies are often conducted on the basis of a single sample per participant. In such a case, almost no conclusions can be drawn regarding the impact of matrix effects. Accordingly, the characterization of the matrix-specific bias term from the basic model must often be performed in a separate experiment.

In the following, simple procedures are described for characterizing different components of variation – such as the matrix-specific bias.

More sophisticated procedures for simultaneously estimating several components of variation are provided in [12]. The reader is also referred to ISO TS 23471 [20], in which study designs are described for the evaluation of data obtained from several concentration levels in one laboratory; and to ISO 5725-3 [19], in which, mainly, alternative study designs are described for the evaluation of data from one concentration level in several laboratories.

9.1 PROCEDURE FOR CHARACTERIZING IN-HOUSE VARIATION

If the analytical method is an in-house method, then an in-house (single-lab) validation study is conducted. If validation data are incomplete or unavailable, in-house components of variation can be characterized on the basis of a further experiment (or QC data, as long as such data are available and have an appropriate structure).

Total in-house variation is called intermediate precision and should reflect all relevant uncertainty sources except matrix mismatch⁷ – in particular, variation arising from different measurement conditions (i.e. operator, reagent batch, etc.) within the laboratory, along with repeatability.

The structure of the experimental or QC data must allow the distinction between in-house repeatability conditions and intermediate conditions (different day, different technician, different reagent batch, etc.). The uncertainty can then be calculated as follows:

$$u = \sqrt{s_I^2 - s_{r,inhouse}^2 + \frac{s_{r,inhouse}^2}{k}}$$

where s_I denotes the intermediate standard deviation, $s_{r,inhouse}$ denotes the repeatability estimate and k denotes the number of replicates whose mean value is taken as the final measurement result.

As explained in Section 8.0, it is recommended that, at a minimum, $N = 12$ different in-house measurement conditions (e.g. different days) be represented in the data set.

In the following example, we take the case that QC data are available for 20 different days. (If appropriate QC data are not available and a further experiment is required, $N = 12$ days are sufficient).

Table 2: In-house QC data for the calculation of intermediate (in-house) and repeatability standard deviation values

	Result 1	Result 2
Day 1	10.72	12.29

⁷ By definition, intermediate precision does not include the variation of bias across matrices (matrix mismatch) bias, see 2.22 in VIM [8]. If matrix bias/matrix mismatch is included, then the term in-house reproducibility is used.

Day 2	4.56	0.90
Day 3	8.79	9.75
Day 4	10.08	6.51
Day 5	12.29	11.32
Day 6	7.95	6.79
Day 7	13.06	14.54
Day 8	11.23	12.09
Day 9	7.31	9.51
Day 10	5.85	5.08
Day 11	7.48	9.12
Day 12	12.59	10.65
Day 13	7.55	6.59
Day 14	12.05	11.15
Day 15	4.86	6.48
Day 16	6.99	7.10
Day 17	7.40	6.75
Day 18	8.85	11.15
Day 19	11.93	10.17
Day 20	8.50	8.29

The between-day and repeatability standard deviation values are calculated as follows.

First we introduce the following notation: the days are indexed $i = 1, \dots, m$ (in this example, $m = 20$); the replicates within each day are indexed $j = 1, n$ (in this example, $n = 2$); and the individual measurement results are denoted x_{ij} .

First, compute the overall mean value \bar{x} , and the day-specific mean values \bar{x}_i . Then compute the between-day sum of squares⁸:

$$SSB = n \cdot \sum_{i=1}^m (\bar{x}_i - \bar{x})^2$$

and the within-day sum of squares:

$$SSW = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

The in-house repeatability standard deviation $s_{r,inhouse}$ is then obtained as

$$s_{r,inhouse} = \sqrt{\frac{SSW}{m \cdot (n - 1)}}$$

and the between-day standard deviation s_D is obtained as

$$s_D = \sqrt{\frac{1}{n} \left(\frac{SSB}{m-1} - s_{r,inh}^2 \right)}.$$

(If the value under the square root sign is negative, then $s_D = 0$.)

Finally, the intermediate (in-house) standard deviation is calculated as:

$$s_I = \sqrt{s_D^2 + s_{r,inhouse}^2}.$$

For the data from Table 2, the calculation results are as follows:

Table 3: Calculation of SSB and SSW on the basis of in-house QC data

⁸ The following are standard formulas for the one-way analysis of variance with random effects.

Overall mean value \bar{x}	Day-specific mean values \bar{x}_i	Differences $\bar{x}_i - \bar{x}$	SSB	Differences $x_{ij} - \bar{x}_i$	Differences $x_{ij} - \bar{x}_i$	SSW
8.91	11.51	2.60	283.05	-0.79	0.79	29.95
	2.73	-6.18		1.83	-1.83	
	9.27	0.36		-0.48	0.48	
	8.29	-0.61		1.79	-1.79	
	11.80	2.90		0.49	-0.49	
	7.37	-1.54		0.58	-0.58	
	13.80	4.90		-0.74	0.74	
	11.66	2.75		-0.43	0.43	
	8.41	-0.50		-1.10	1.10	
	5.46	-3.44		0.39	-0.39	
	8.30	-0.61		-0.82	0.82	
	11.62	2.72		0.97	-0.97	
	7.07	-1.83		0.48	-0.48	
	11.60	2.69		0.45	-0.45	
	5.67	-3.24		-0.81	0.81	
	7.05	-1.86		-0.06	0.06	
	7.08	-1.83		0.32	-0.32	
	10.00	1.09		-1.15	1.15	
	11.05	2.14		0.88	-0.88	
	8.40	-0.51		0.10	-0.10	

The following precision estimates are obtained:

Table 4: Precision estimates obtained from in-house QC data

$S_{r,inhouse}$	S_D	S_I
<u>1.22</u>	<u>2.59</u>	<u>2.86</u>

$S_{r,inhouse}$	S_D	S_I
<u>1.22</u>	<u>2.59</u>	<u>2.86</u>

9.2 PROCEDURES FOR CHARACTERIZING VARIATION OF BIAS ACROSS MATRICES (MATRIX MISMATCH)

In this section, a procedure for estimating the variation of bias across matrices is described. Such an estimate is necessary when

- on the one hand, a number of different matrices/sample types are included in the scope of the method
- and on the other hand, only few matrices/sample types were included in the validation study

It is assumed that, for any given matrix, heterogeneity between laboratory samples is negligible, and that the measurand is specified in terms of a number of matrices, from which N matrices are selected⁹. Selection should be based on the method's intended use/scope. As explained in Section 80, it is recommended that, at a minimum, $N = 12$ matrices be included.

A simple approach for characterizing variation of bias across matrices consists in spiking the N matrices and obtaining duplicate measurement results in a single laboratory for each matrix. In this manner, variation of bias between the matrices (matrix mismatch specific bias) can be distinguished from variation within each matrix (repeatability error). In this procedure, the matrix is modelled as a random effect, and the result is a standard deviation characterizing variation across all the matrices included in the specification of the measurand.

⁹ For instance, a number of different apple types, or a number of different cattle breeds.

Example**Table 5: Data from an experiment for the calculation of the matrix mismatch effects (variation of bias across matrices)-bias**

	MV1	MV2
Matrix 1	114.51	112.24
Matrix 2	120.25	111.59
Matrix 3	88.46	86.62
Matrix 4	118.93	102.35
Matrix 5	74.06	80.91
Matrix 6	117.50	102.69
Matrix 7	120.96	109.35
Matrix 8	96.05	92.92
Matrix 9	98.43	87.09
Matrix 10	107.99	117.42
Matrix 11	117.34	126.87
Matrix 12	76.56	109.79

Applying the same calculation procedure as in Section 9.1, the following precision estimates are obtained:

Table 6: Precision estimates for the calculation of matrix biasmismatch

S_r	S_{matrix}
9.53	12.24

For further information on matrix mismatchbias, see Uhlig (2023) [24] .

9.3 PROCEDURES FOR CHARACTERIZING BETWEEN-LABORATORY VARIATION

Procedure 1: Conduct an interlaboratory validation study with a minimum of $N = 12$ laboratories and with duplicate measurement results within each laboratory. It is necessary to ensure that heterogeneity between laboratory samples is negligible. In this manner, variation between the laboratories (lab bias) can be distinguished from variation within the laboratories (repeatability error). In this procedure, the laboratory is modelled as a random effect, and the result is a standard deviation characterizing variation across laboratories.

Example**Table 7: Data from an experiment for the calculation of the lab bias**

	MV1	MV2
Lab 1	0.981	1.238
Lab 2	0.182	0.601
Lab 3	1.107	0.994
Lab 4	1.471	1.532
Lab 5	1.169	0.674
Lab 6	0.491	1.271
Lab 7	1.717	0.970
Lab 8	0.931	1.171
Lab 9	1.017	1.248
Lab 10	0.909	0.723
Lab 11	0.812	1.312
Lab 12	1.375	1.719

Applying the same calculation procedure as in Section 9.1, the following precision estimates are obtained:

Table 8: Precision estimates for the calculation of lab bias

S_r	S_{lab}
0.30	0.23

Procedure 2: If PT data are available, and a sufficient number of participants (ideally, at least 12) have used the same method (ideally, with replicate measurements within each laboratory) – then these data can be used to characterize variation across laboratories. In order to ensure neutral data evaluation and avoid conflicts of interest, the data should come from PT schemes run by competent authorities.

9.4 PROCEDURES FOR CHARACTERIZING CONTRIBUTIONS TO UNCERTAINTY FROM SAMPLE PREPARATION STEPS AND FROM SUBSAMPLING WITHIN THE LABORATORY

Procedures for characterizing contributions from sample preparation steps and from the heterogeneity of the laboratory sample

In a collaborative study conducted in accordance with ISO 5725-2, the repeatability component may or may not reflect contributions from

- sample preparation (all steps performed starting from the laboratory sample in order to obtain the test portion)
- subsampling variation due to the heterogeneity of the laboratory sample

In particular, if the collaborative study was conducted on the basis of reference material which

- is homogenous
- is sent to the laboratories in the form of test portions requiring no further sample preparation steps

then the repeatability estimate will reflect neither of the source of uncertainty.

For this reason, if either of these sources of uncertainty is relevant (i.e. affects measurements in routine testing), “real” samples must be used in the validation study. If this is not practicable (e.g. because of stability issues), and homogenous test material is used instead, then the sample preparation and/or subsampling components of repeatability must be estimated in a separate experiment.

Note: The subsampling component must not be confused with the variation of bias across matrices (matrix mismatch) component, which may vary considerably from lab to lab, thus inflating the between-laboratory component (rather than the repeatability).

In the absence of fundamental variability, a simple procedure for estimating the sample preparation and/or subsampling components is as follows: conduct an in-house experiment with 12 “real” samples (routine samples). For each sample, 2 test results are obtained under repeatability conditions. Any uncertainty from heterogeneity and/or sampling steps will manifest itself as the “within-sample” estimate, following the evaluation scheme described in 9.1, 9.2 and 9.3 above.

Procedures for characterizing fundamental variability

Fundamental variability is a subcomponent of the repeatability error term from the basic model in Section 3 and denotes the irreducible variation between samples which remains even under the highest achievable degree of homogeneity. Fundamental variability reflects heterogeneity at the level of the sample’s constituent particles; it has an influence on the uncertainty of measurement results when the target analyte is located on sparsely distributed carrier particles¹⁰. Fundamental variability appears twice: first, during sampling, and second, during subsampling in the laboratory, i.e. extraction of a test portion after homogenization of the laboratory sample. In practice, nonnegligible fundamental variability can be reduced by modifying the testing procedure in two respects: first, by finer grinding or comminuting or mixing of the test material, and second, by increasing the test portion size.

It should be noted that, while a correct partitioning of observed variability between sampling, subsampling and other uncertainty components is achievable in theory, doing so is difficult in practice *when the fundamental variability is significant*. Take the case that the number of carrier particles in the laboratory sample collected from the container or batch of products varies randomly between 0 and 10. The fundamental variability between subsamples (test portions) will thus depend on which laboratory sample they were collected from. In such a situation, a correct characterization of fundamental variability would be quite involved. It would be much more efficient to ensure variation regarding carrier particle numbers between laboratory samples were negligible – in other words, to ensure that every single laboratory sample were representative of the container or batch of products, thus eliminating the sampling fundamental variability from the equation. Often, this may be achieved by increasing laboratory sample size; but a more general point is that a correct evaluation of fundamental variability requires an appropriate inclusion of the sampling step, i.e. a consideration of the different steps from sampling to analysis as one single process¹¹.

¹⁰ Fundamental variability is related to Pierre Gy’s fundamental error, see [26].

¹¹ Consider the following example: a 5 t container contains one single carrier particle, translating to 1 µg/kg analyte concentration. A 5 kg laboratory sample is collected from the container. Thus, with 99.9 % probability, the laboratory sample

The question thus arises: how can we decide whether fundamental variability is significant? Fundamental variability cannot be characterized by means of classical homogeneity studies such as the standard designs described in ISO 13528 [22] and Guide 35 [9]. Indeed, in these designs, it is not possible to distinguish fundamental variability from sample heterogeneity *per se*, so that the former may be mistaken for the latter.

The following procedure, originally proposed in Uhlig (2022) [23], allows a characterization of fundamental variability.

Step 1

Check whether one of the following criteria are met:

Criterion 1: The in-house repeatability standard deviation is larger than 3 times the expected value.

Criterion 2: The in-house repeatability standard deviation is larger than the Horwitz SD value.

Criterion 3: Conspicuous “upper” outliers are present in QC data. For instance, in the QC data provided in Table 2 (Section 9.1), the Day 7 value of 14.54 could be considered such an “upper” outlier. The presence of such outliers constitutes a further indication that the unexpectedly large observed variability may be due to fundamental variability.

If at least one of these criteria is met, proceed to Step 2.

Step 2

Conduct the following experiment:

1. Obtain 20 test results under repeatability conditions. Calculate the corresponding variance s_1^2 .
2. Increase test portion size by a factor k (e.g. triple test portion size, $k = 3$). If it is not possible or practical to increase test portion size, grinding and homogenizing a volume corresponding to a k -fold increase in test portion size prior to taking a test portion with the original size is another option.
3. Obtain 20 test results under repeatability conditions on the basis of the finely ground test material / increased test portion size. Calculate the corresponding variance s_2^2 .
4. If the ratio $\frac{s_1^2}{s_2^2}$ is greater than 2.17, then calculate the SD characterizing fundamental variability as follows:

$$s_F = \sqrt{\frac{k}{(k-1)} \cdot (s_1^2 - s_2^2)}$$

—If the ratio $\frac{s_1^2}{s_2^2}$ is less than 2.17, this is an indication that fundamental variability is not significant and does not need to be included in the estimate of measurement uncertainty.

5. _____

Example

Table 9: Test results from an experiment for the calculation of fundamental variability

	Experiment 1: Original test portion size	Experiment 2: Test portion size is tripled
Sample 1	14.0	15.1
Sample 2	11.9	13.8
Sample 3	10.5	11.8
Sample 4	14.9	14.0
Sample 5	13.1	11.4
Sample 6	9.5	15.7

will contain no carrier particle, and there will be no fundamental variability. However, with 0.1 % probability, the laboratory sample will contain the single carrier particle. In such a case, if a 500 g test portion is taken from the laboratory sample, then the analyte concentration in the test portion will be either 0 mg/kg (nine times out of ten) or 10 mg/kg (one time out of ten). This corresponds to a (Poisson) standard deviation of 1 mg/kg – which clearly constitutes a disproportionate estimate in relation to the situation in the container. This example shows how restricting the calculation of fundamental variability to the subsampling step can lead to gross misestimation.

Sample 7	15.6	12.4
Sample 8	18.3	11.5
Sample 9	12.5	12.1
Sample 10	16.4	13.7
Sample 11	18.0	15.8
Sample 12	14.0	12.5
Sample 13	13.0	12.8
Sample 14	20.8	15.1
Sample 15	10.2	11.8
Sample 16	21.5	10.6
Sample 17	13.9	11.1
Sample 18	17.8	12.9
Sample 19	7.7	11.4
Sample 20	12.2	16.3

Note that, in Experiment 1, several conspicuously large values are obtained – an indication that fundamental variability is non-negligible.

The following variances and corresponding ratio are obtained:

Table 10: Variances and their ratio

s_1^2	s_2^2	s_1^2/s_2^2
13.54	3.05	4.44

As can be seen, the ratio s_1^2/s_2^2 is greater than the value 2.17. Accordingly, the fundamental variability is calculated as

$$s_F = \sqrt{\frac{3}{2} \cdot (s_1^2 - s_2^2)} = 3.97.$$

9.5 PROCEDURES FOR DETERMINING WHETHER THE ESTIMATE OF MEASUREMENT UNCERTAINTY DERIVED FROM PRECISION DATA FROM A COLLABORATIVE STUDY IS COMPLETE

In general, the workload associated with a *reliable* quantitative assessment of a measurement uncertainty estimate will prove prohibitive. For this reason, a qualitative assessment is proposed, consisting in addressing the two following questions.

Are uncertainty contributions from subsampling and sample preparation included in the repeatability estimate and/or reproducibility estimate?

Whether or not subsampling is included depends on the sample material provided to the laboratories. If the material is grinded/homogenized, subsampling uncertainty is not included.

Whether or not sample preparation is completely included depends which sample preparation steps have already been performed by the organizer of the collaborative study.

If these sources of uncertainty were not duly reflected in the collaborative study, additional experiments as described in Section 9.4 are required.

Are matrix mismatch effects reflected in the reproducibility estimate?

In general, this is not the case. The influence of matrix mismatch can only determined via comparison with reference values or via spiking experiments, see the procedure proposed in Section 9.2. Another example can be found in Annex F of ISO TS 23471 [20], where equipment and operator effects along with matrix mismatch effects arising from different lots and storage conditions are analyzed by means of a factorial design.

10 Influence of measurement uncertainty on sampling plans: examples

In the *General guidelines on sampling* [13], it is stated that “Codex Methods of Sampling are designed to ensure that fair and valid sampling procedures are used when food is being tested for compliance with a particular Codex commodity standard”. Sample size and acceptance number / acceptability constant for inspection by attributes / variables are determined on the basis of procedures and sampling plans described

in ISO standards and/or Codex guidelines. While measurement uncertainty may be considered irrelevant for inspection by attributes, its impact on inspection by variables must be accounted for.

In the introduction to ISO 3951-1:2013, it is stated that “[i]t is assumed in the body of this part of ISO 3951 that measurement error is negligible [...]”. Nonetheless, procedures for increasing the sample size are provided in Annex B of ISO 3951-1 [14] and Annex P of ISO 3951-2 [15] for the case that measurement uncertainty is non-negligible. It is important to note that these procedures are only applicable if “the measurement method is unbiased, i.e. the expected value of the measurement error is zero” (see Annex P.1 in ISO 3951-2:2013 [15]). In such a case, total variability is expressed as

$$\sigma_{total} = \sqrt{\sigma^2 + \sigma_m^2}$$

where σ denotes the process standard deviation and σ_m denotes the measurement standard deviation.

If σ_m is non-negligible (i.e. greater than one tenth of the sampling standard deviation s or process standard deviation σ), the sample size n must be increased to either $n^* = n \cdot (1 + \gamma^2)$ where $\gamma = \sigma_m/\sigma$ (the process standard deviation σ is known) or $n^* = n \cdot (1 + \tilde{\gamma}^2)$ where $\tilde{\gamma}$ is an estimated upper bound of $\gamma = \sigma_m/\sigma$ (the process standard deviation σ is unknown). The acceptability constant k remains unchanged. For further details, see Annex P in ISO 3951-2:2013 [15].

Example

A lot of 500 items of pre-packaged mineral water is assessed for sodium content. If the measurement uncertainty is not taken into consideration, for an agreed AQL of 2.5 % (maximum concentration 200 mg/L), general inspection level II (default level) a sample of 30 items should be collected for assessment, (ISO 3951-2 [15], Annex A, Table A1 and Annex B, Table B1). The production is well under control and the control charts give a process standard deviation σ of 2 mg/L. The measurement uncertainty standard deviation σ_m is 1 mg/L and is thus non-negligible. With $\gamma = \sigma_m/\sigma = 0.5$ and $1 + \gamma^2 = 1.25$ the sample size must be increased to 38.

If there is a bias, the above procedure must be modified. One possibility would be to proceed as follows¹². The standard deviation of \bar{x} , the mean across the n measurement results, is expressed as

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2 + \sigma_0^2}{n} + \sigma_b^2}$$

where σ denotes the process standard deviation, σ_0 denotes the repeatability component of measurement uncertainty (calculated on the basis of the n items sampled from the lot), and σ_b represents available information (e.g. the between-lab standard deviation from a method validation study) used to estimate the bias term.

The modified procedure is as follows:

1. Increase the sample size under the assumption that there is no measurement error
2. Calculate $d = \frac{1}{n} - \frac{\sigma_b^2}{\sigma^2}$
3. If $d \leq 0$, inflated variability due to a bias cannot be compensated for via an increase in sample size.
4. If $d \leq \frac{1}{2n}$, bias compensation via an increase in sample size may not be appropriate due to the large number of samples required. It is then suggested to reduce bias or to use another measurement method.
5. If $d > \frac{1}{2n}$, calculate the new sample size as $n^* = \frac{1 + \frac{\sigma_0^2}{\sigma^2}}{d} = \frac{\sigma^2 + \sigma_0^2}{\frac{\sigma^2}{n} - \sigma_b^2}$

Example (continued from previous example)

It is now assumed that there is a method bias and that a σ_b estimate of 0.2 mg/L is available. Accordingly, on the basis of the previously calculated value of $n = 38$, d is calculated as $d = 0.016$. Since $d > \frac{1}{2n} = 0.013$, the new sample size is calculated as $n^* = 77$ (with $\sigma_0 = \sigma_m = 1$ mg/L).

Procedures for bulk sampling

Procedures for bulk sampling are provided in ISO 10725:2000 [17]. As in the case of sampling from packages, these procedures are only valid under the assumption that there is no method bias. Modified procedures for the case that there is a method bias are currently being developed. For now, the discussion is limited to the

¹² This modified procedure is taken from [the](#) current stage of development of Annex B of ISO/WD ISO 3951-6 [16].

case that there is no bias.

A *dominant* measurement uncertainty has an effect on the number of test samples per composite sample n_T as well as the number of measurements per test sample n_M . The measurement uncertainty is dominant when both the standard deviation of the sampling increment σ_I and the standard deviation between test samples σ_P are far less (one tenth or less) than the measurement standard deviation σ_M (i.e. the measurement uncertainty), which must be known and stable, see Annex B in ISO 10725 [17]. The number of sample increments per composite sample n_I remains unchanged, no matter whether the measurement uncertainty is dominant or not. The mass of the increments should be sufficiently large to offset the fundamental variability.

Example

A lot of wheat bulk material is to be assessed for cadmium content (maximum concentration e.g. 0.1 mg/kg). In this example, it is assumed that cadmium concentrations in the lot are homogeneous, resulting in very low standard deviations σ_I and σ_P , estimated as 0.0015 mg/kg and 0.002 mg/kg, respectively. Since the concentrations are very low, a relatively high measurement uncertainty $\sigma_M = 0.025$ mg/kg is obtained. The discrimination interval D (difference between agreed risk-based acceptance and rejection levels) is 0.02 mg/kg. The measurement standard deviation $\sigma_M = 0.025$ mg/kg is thus dominant (d_I is calculated as 0.075). The number of increments per composite sample is $n_I = 6$, the number of test samples per composite sample is $n_T = 2$ and the number of measurements per test sample is $n_M = 2$ (yielding a product $n_T \cdot n_M = 4$, which can be interpreted as a measure of the analytical workload). The combined overall standard deviation σ_0 is calculated as $\sqrt{\frac{n_T \cdot n_M}{n_I} \sigma_I^2 + n_M \sigma_P^2 + \sigma_M^2} \approx 0.03$ mg/kg and divided by the discrimination interval D in order to obtain the relative standard deviation $d_0 = \sigma_0/D \approx 1.26$. By means of Table B1 in Annex B of ISO 10725 [17], this relative standard deviation d_0 is used to determine the adjusted number of test samples per composite sample $n_T = 2$ (i.e. n_T remains the same) as well as the adjusted number of measurements per test sample $n_M = 3$, yielding a product $n_T \cdot n_M = 6$.

References

- [1] Evaluation of measurement data — Guide to the expression of uncertainty in measurement, JCGM 100:2008.
- [2] S L R Ellison and A Williams (eds.) EURACHEM/CITAC Guide CG4: Quantifying Uncertainty in Analytical Measurement, Third Edition, QUAM:2012.P1.
- [3] Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method, JCGM 101:2008.
- [4] ISO 21748:2017, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.
- [5] ISO 5725-2:2019/1994, Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.
- [6] B Jülicher, Petra Gowik and Steffen Uhlig (1998) Assessment of detection methods in trace analysis by means of a statistically based in-house validation concept, *The Analyst*.
- [7] B Jülicher, Petra Gowik and Steffen Uhlig (1998) A top-down in-house validation based approach for the investigation of the measurement uncertainty using fractional factorial experiments, *The Analyst*.
- [8] International vocabulary of metrology — Basic and general concepts and associated terms (VIM), JCGM 200:2012.
- [9] ISO Guide 35, Fourth edition (2017), Reference materials — Guidance for characterization and assessment of homogeneity and stability.
- [10] ISO/IEC 17025:2017, General requirements for the competence of testing and calibration laboratories.
- [11] CXG 59-2006, Guidelines on estimation of uncertainty of results.
- [12] S Uhlig and P Gowik (2018) Efficient estimation of interlaboratory and in-house reproducibility standard deviation in factorial validation studies, *Journal of Consumer Protection and Food Safety*.
- [13] CXG 50-2004, General guidelines on sampling.
- [14] ISO 3951-1:2022/46, Sampling procedures for inspection by variables — Part 1: Specification of single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL.

- [15] ISO 3951-2:2013, Sampling procedures for inspection by variables — Part 2: General specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection of independent quality characteristics.
- [16] ISO/~~DISWD~~ 3951-6:2019, Sampling procedures for inspection by variables — Part 6: Specification for single sampling plans indexed by limiting quality (LQ).
- [17] ISO 10725:2000, Acceptance sampling plans and procedures for the inspection of bulk materials.
- [18] ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions.
- [19] ISO 5725-3:1994, Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method. (A new revision is currently being prepared for publication.)
- [20] ISO TS 23471:2022, Experimental designs for the evaluation of uncertainty – Use of factorial designs for determining uncertainty functions.
- [21] S L R Ellison and A Williams (eds.) EURACHEM/CITAC Guide: Metrological Traceability in Chemical Measurement (Second Edition 2019).
- [22] ISO 13528:202245, Statistical methods for use in proficiency testing by interlaboratory comparison.
- [23] S Uhlig, B Colson and P Gowik (2022) A procedure for estimating fundamental variability, available as a preprint, doi: 10.20944/preprints202211.0460.v1
- [24] S Uhlig, B Colson and P Gowik (2023) Matrix mismatch and its estimation in method validation studies, submitted for publication.
- [25] S Uhlig, B Colson and P Gowik (2022) Measurement Uncertainty Interval In Case of a Known Relationship between Precision and Mean, available as a preprint, doi: 10.20944/preprints202208.0179.v1
- [25] ~~Gy P M (1979) Sampling of particulate materials – theory and practice. Elsevier, Amstedam.~~

General Comments

comments	Australia	Response
a. We believe the information document serves the purpose of supporting the revision of CXG 54 and its implementation by providing further explanation of the theory, some calculated examples but also references more detailed discussions in published standards or journal articles.		
b. We have suggested some modifications to the text to align more with the GUM and hopefully make the document clearer.		
c. There are some minor editorial changes required, as outlined under specific comments.		
Considering the complexity for laboratories to apply the principles expressed in the document, Brazil supports its adoption as an informative document.	Brazil	Okay
Regarding the text, we identified the need of updating some references as listed below:		
<ul style="list-style-type: none"> • Footnotes 2, 3 and 5 - pages 3, 4 and 9: the reference number for VIM document should be replaced from 7 to 8. 		
<ul style="list-style-type: none"> • Footnote 4 - page 8: the reference number should be replaced from 10 to 11. 		
<ul style="list-style-type: none"> • List of references: 		
<ul style="list-style-type: none"> - The last version of the document “ISO 5725-2:1994, Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method [5]” is 2019. 		
<ul style="list-style-type: none"> - The last version of the documents “ISO 3951-1:2016, Sampling procedures for inspection by variables — Part 1: Specification of single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL [14]” and “ISO TS 23471, Experimental designs for the evaluation of uncertainty – Use of factorial designs for determining uncertainty functions [20] and “ISO 13528:2015, Statistical methods for use in proficiency testing by interlaboratory comparison [22]” is 2022. 		
Chile considers that the document is adequate in its content, however, it is not clear if the document could be understood by all who consult it, because terminology is used that requires referring to the standards or attached bibliography, for a better comprehension.	Chile	We agree, but the decision lies presumably with the Secretariat.

It would be important that, after the approval of the document, the CCMAS could organize an on-line workshop for its dissemination and better understanding among the different member countries.		
Egypt suggests endorsement of this document from CCMAS 42	Egypt	
The European Union and its Member States (EUMS) congratulate Germany for doing an excellent job that provides the technical background necessary for the estimation of measurement uncertainty and the examples for illustrating different use cases; it will certainly support the guidance provided by CXG 54.	European Union	Okay, changes have been made to reflect these comments.
The content of the information document explains in a comprehensive manner relevant approaches to estimate measurement uncertainty (top-down and bottom-up), the models and assumptions governing those approaches and provides practical examples how to evaluate uncertainty components. Even if a note regarding sub-sampling has been added to the current version of the information document, it could still profit from a stronger focus on test methods validated by collaborative study as methods endorsed by CCMAS and included in CXS 234-1999 have to be validated by multi-lab studies.		
Method performance data resulting from collaborative studies do in a number of cases not include certain uncertainty sources. It would be an advantage if the document described how to identify additional influential factors on the measurement result, e.g., preparation steps related to transforming a laboratory sample into the test portion, sub-sampling from the laboratory sample, matrix variation, etc., that were not adequately covered in the collaborative study (reconciliation of potential uncertainty sources with the available collaborative study results) and quantify the uncertainty component arising from them. Lastly, it should briefly describe when and how to combine those additional standard uncertainties with the performance parameters (sr and sR) of the collaboratively trialed method.		
As some endorsed methods do not have the concerned commodity in their scope (e.g. AOAC 968.31, which has been validated for canned tomatoes, lima beans, and potatoes, but which also endorsed for Ca in canned strawberry), guidance and an example should be provided how to assess uncertainty due to the matrix mismatch.		
A paragraph on the role of certified reference materials for estimating measurement uncertainty, particularly the uncertainty of bias correction, would strengthen the information document as well.		
Agree with no comments	Iraq	
Germany has done a good job documenting some of the more detailed technical information underpinning the CXG 54 Guidelines that makes this information more accessible than (say) providing references to papers.	New Zealand	

In particular the document includes a range of useful examples that show how some of the calculations are done and some information is provided on the role of measurement uncertainty in acceptance sampling, that supports CXG 50.		
We think the reference in 9.1 to Section) should refer to Section 8		Indeed.
Estamos de acuerdo en general con el documento.	Paraguay	
The Philippines agrees on the proposed draft information document prepared by Germany on the GUIDELINES ON MEASUREMENT UNCERTAINTY (CXG 54-2004) and agrees to publish on the Codex website.	Philippines	
Rationale: The Information document was updated to take into account the comments received during CCMAS41. This document provides the information on how different approaches for the evaluation of measurement uncertainty relate to one another and have information regarding the best procedure to adopt in any given case. Furthermore, this new document provides information and clarifies basic understandings which are important for the correct evaluation and interpretation of measurement uncertainty. We also agree on the new note added at the end of Section 3 regarding the case that precision depends on concentration.		
"The examples presented on pages 15 to 16 of CL 2023/14/OCS-MAS offer valuable insights into the determination of sample sizes in the presence of measurement uncertainty."	Singapore	
The symbols described in the paragraphs throughout the document are raised (superscripts). This isn't typical and looks odd. It would be better if the symbols were in line with the text. This starts from section 2, where "Y" is raised relative to the text around it. "...in most cases a measurand Y is..."	ICUMSA	This has been corrected.

Specific Comments

	1. Introduction	Response
Japan	<p><u>This document provides information on example procedures to estimate measurement uncertainty. Since procedures are not limited to what are written in this document, Members may use other procedures for the estimation of measurement uncertainty than those described in this document.</u> A measurement result should always be accompanied by information regarding its uncertainty. Such information provides an indication of the quality of the measurement result and allows meaningful comparison to other measurement results or reference values. Without a statement of measurement uncertainty, a measurement result is essentially incomplete and cannot be properly interpreted.</p>	Not accepted. The procedures are not binding since the document is an information document. But the components of uncertainty described here and in CXG 54 should be considered "binding" in estimating uncertainty.
	Rationale	
	To clarify the nature of this document (information document), not a Codex guideline or standard.	To be decided by secretariat.
Mauritius	<p>A measurement result should always <u>ideally</u> be accompanied by information regarding its uncertainty. Such information provides an indication of the quality of the measurement result and allows meaningful comparison to other measurement results or reference values. Without a statement of measurement uncertainty, a measurement result is essentially incomplete and cannot be properly interpreted.</p>	Partially accepted. "Always" is deleted. Please note that the point here is not what is provided in a report. Rather the point is the availability of the MU estimate.
	<p>A measurement result should always be accompanied by information regarding its uncertainty. Such information provides an indication of the quality of the measurement result and allows meaningful comparison to other measurement results or reference values. Without a statement of measurement uncertainty, a measurement result is essentially incomplete and cannot be properly interpreted.</p>	
	<p>As per ISO 17025, clause 7.6, we need to identify contributions to MU. We need to evaluate MU or an estimate based on an understanding of the theoretical principles or practical experience of the performance of the method</p>	
	<p>As per clause 7.8.3.1, test reports shall however, where necessary for interpretation of test results, include, where applicable, the measurement uncertainty presented in the same unit as the measurand or in a term relative to the measurand when:</p>	

	it is relevant to the validity or application of test results	
	a customer's instruction so requires	
	or MU affects conformity to a specification limit	
	In view of above we are proposing alternate wording	
Japan	This document provides guidance <u>information</u> regarding those sources of uncertainty which originate in the laboratory itself, i.e. in connection with the procedures and conditions starting with the laboratory sample and ending with the measurement result. In particular: the question of sampling uncertainty and the extent to which laboratory samples are representative of the content in the container will not be addressed. Such questions are addressed in CXG 50-2004 [13].	Accepted
	Again, this is not a guidance but to provide information.	
Mauritius	This document provides guidance regarding those sources of uncertainty which originate in the laboratory itself, i.e. in connection with the procedures and conditions starting with the laboratory sample and ending with the measurement result. In particular: the question of sampling uncertainty and the extent to which laboratory samples are representative of the content in the container will not be addressed <u>addressed in this information document</u> . Such questions are addressed in CXG 50-2004 [13].	Accepted
	Measurement uncertainty is defined as a parameter "...that characterizes the dispersion of the values that could reasonably be attributed to the measurand", see 2.2.3 in GUM [1]. This document aims to clarify what is meant in <u>by</u> this definition and to provide the information which is necessary to understand how different approaches for the evaluation of measurement uncertainty relate to one another. This should allow the reader to make informed decisions regarding the best procedure to adopt in any given case.	Accepted
	2. Top-down versus bottom-up approaches	
	<u>2.1 Monte Carlo method (MCM).</u>	
Chile	It is suggested to clearly distinguish in the document what refers to the Monte Carlo Method (MCM); as this would make it easier to understand	Accepted

	The uncertainty values of the individual variables are taken from the certificates of the reference standard substances of materials, while the uncertainty values for the variables are obtained from the regression analysis (residual standard deviation).	
	Strictly speaking, the ordinary minimum squares method used in Chemical Metrology assumes that the X axis has no uncertainty, the random error is only attributable to the instrument (Y axis).	
	Of course, this is not correct.	
	Formally, if one wanted to incorporate the uncertainty of the calibrants, a linear regression with error in both axes should be used. That would be rigorous, but the proposed approximation can be accepted without major modifications.	Accepted
	The CCQM gas metrology group uses this type of regression with uncertainty in both axes	Noted
	3. Basic model for the top-down approach	
	<i>True value</i>	
Australia	The 'estimation by averaging' should be done with caution, thus suggest the following amendment. 'It can be estimated by averaging e.g. across methods, samples and laboratories, but must be done with caution, as in practice, unless participants for an interlaboratory study are chosen based on both expertise and similar methodologies, an average bias of zero is difficult to achieve.'	Accepted. Text has been mostly adopted. "This results in an approximation."
	2. Section 3, sub-section on individual term 'True value', first paragraph, third sentence. Suggested following amendment, 'However, it is crucial to note that in the GUM [1], measurement uncertainty is defined without any reference to a true value; rather, it is defined as a parameter "... that characterizes the dispersion of the values which could reasonably be attributed to the measurand", see 2.2.3 in GUM [1].'	Not accepted. The proposed text is already in the current version.
	3. Section 3, sub-section on individual term 'True value', first paragraph, last sentence. Either the last sentence is removed, as a discussion of reference material usage is better left when defining or discussing method biases. Alternatively, if retained suggest the following amendment. 'If a certified reference material and associated values is available, it will suggest a range where the true value might be, and give along with a reference uncertainty value, for inclusion the latter can be included in the uncertainty of bias correction.	Not accepted. It is not clear why a reference to "true value" should be added, since according to GUM, any such references are carefully avoided.
	<i>Method bias (average across labs and matrices)</i>	

Mauritius	When talking about method bias or laboratory bias from page 5 and 14, would it be a prerequisite that laboratories participating in the validation process through interlaboratories are (i) ISO accredited ones adopting the same/latest methods (ii) accredited for the test (iii) working in the recommended conditions both environmental and equipment use (iv) including any deviation that are permissible by the relevant ISO methods to ensure having all the deviations that could be met. These are important as we see from PT results, that even ISO accredited laboratories sometimes do not fall within the acceptable z scores. Thus it is important to define the laboratories to be involved for comprehensiveness.	Partially accepted. However, we accept (iii). It is a prerequisite that participating laboratories work under adequate conditions with adequate equipment and trained personnel. Accreditation is not a requirement.
	<i>Method bias (average across labs and matrices)</i>	
Australia	'Method bias (average across labs and matrices)' first sentence. Suggest amendment 'The method bias across both labs and matrices can be estimated by averaging across laboratories and matrices, but only becomes meaningful if an estimate of bias is available from use of a certified reference material.'	Partially accepted. "Information from CRM should be taken into consideration if available."
	<i>Laboratory bias</i>	
Australia	'Laboratory bias' last sentence. Suggest that if a definition of 'Laboratory standard deviation' is required it should be done separately, as bias and variance are not interchangeable. Thus removed the third second sentence from this paragraph. 'The corresponding component of total variability is called the laboratory standard deviation.'	Rejected. But the text has been modified to explain why a bias is described via a standard deviation.
	<i>Repeatability error</i>	
Australia	'Repeatability error', 'Note regarding the case that the precision depends on the concentration level:' Second last sentence in paragraph. We suggest the '...relatively high precision...' as high or low could be misinterpreted and suggest amendment to '...relatively high/low precision...'	Accepted
	5. Relation between measurand and validation data	
Australia	The conditions under which validation data can be used to support a measurement uncertainty estimate can be stated as follows:	Partially accepted. Typically, bias is not estimated in validation studies. For this reason, the focus here lies on precision. Nonetheless, in theory it is true that the uncertainty of bias correction can be estimated via the precision estimates. A note has been added to mention this possibility.

	We suggest that the conditions under which validation data can be used to support a measurement uncertainty estimate, should include 'precision and bias' and not just 'precision'. Thus, suggest following amendments ' and the measurand is included in the scope of the validation and plus precision and bias within the laboratory which is evaluating measurement uncertainty is comparable to the method's precision and bias as characterized in the validation study then... the precision and bias estimates from the validation study can be used in the calculation of measurement uncertainty. In order to check and provide evidence of competence in the application of the method and to ensure adequate precision and bias in the laboratory which is evaluating measurement uncertainty, it may be necessary to perform a verification study.'	
	7. Uncertainty sources in the top-down and bottom-up approaches	
Mauritius	Sources of uncertainty are conveniently classified under six main headings: Sources are described in ISO 19036 as follows: technical, matrix and distributional uncertainties- taking test portion from lab or test sample preparation of initial suspension serial dilution inoculation incubation, (technical operations and calculations- equipment, culture media and reagents) counting of colonies and/or detection of growth confirmation Are these all covered in the present document?	Yes they are. They fall under the categories measurement conditions and measurement procedure, and belong to the technical uncertainty. A note has been added.

	8. Requirements regarding data size	
Australia	In the case that different uncertainty sources are <i>simultaneously</i> taken into consideration, say in the bottom-up approach, the requirement regarding data size can be applied via the Satterthwaite formula. More specifically: take the case that 2 different uncertainty sources are included in the calculation of the combined uncertainty, and . Say that each was obtained by applying the formula for the sample standard deviation on the basis of and measurement results, respectively. The number of degrees of freedom for the combined uncertainty can then be computed as	
	We suggest alignment with the GUM, with a reference for further information. Thus, amend to 'In the case that different uncertainty sources are simultaneously taken into consideration, say in the bottom-up approach, the requirement regarding data size can be applied via the Welch-Satterthwaite formula see [1] Annex G, G.4.1.' We also suggest amending all the other uses of the 'Satterthwaite formula' to 'Welch-Satterthwaite formula'.	Accepted.
	In the case that prior information is used for an individual value (Type B variable) and that no information regarding data size is available, it is suggested to use ; the uncertainty which corresponds to this data size is intended to reflect the fact that, in the case of Type B variables, distributional assumptions are often based on "educated guesses."	
	We would consider the $n_i = 7$ could be conservatively low in some cases. Thus suggest the amendment '... suggest to use $n_i > 7$; the...'. This would also require the amendment in Table 1, by replacing '=' with '>', thus 'Not available, Take $n_4 > 7$.	Rejected. We do not consider $n_i - 1 = 6$ degrees of freedom as low, since one degree of freedom more (i.e. 7 degrees of freedom) corresponds to the case of 8 participants in a collaborative study (for $s_r \ll s_R$).
	Add a sentence to the end of this section (Section 8). 'We would suggest users of the Welch-Satterthwaite formula refer to the GUM [1] Annex G for a discussion on the formula and the range in which the calculated degrees of freedom should fall, since incorrect implementations commonly give answers outside of this range.'	Accepted
	9.1 Procedure for characterizing in-house variation	
	Procedure for characterizing in-house variation	

Canada	It is suggested that a statement be added to clarify that the equations in 9.1 are general for Section 9 and not specific to in-house variation since the formulas in Section 9.1 for characterizing in-house variation can also be applied to Sections 9.2 (across matrices) and 9.3 (between labs). For example, variability between the matrices (Smatrix) and labs (Slab) are calculated in the same way as between days (SD) in Table 4.	Accepted
	Procedure for characterizing in-house variation	
Australia	Suggest a reference is provided for the computation equations after table 2.	Accepted. Reference is made to "one-way analysis of variance with random effects". (footnote)
Canada	As explained in Section 08, it is recommended that, at a minimum, different in-house measurement conditions (e.g. different days) be represented in the data set.	Accepted.
	9.2 Procedures for characterizing variation across matrices (matrix mismatch)	
	Procedures for characterizing variation across matrices (matrix mismatch)	
Australia	Procedure 2, sentence 1. We believe more than one measurement result within each laboratory is required for separation of sr and slab. Thus suggest the following amendment 'If PT data are available, and a sufficient number of participants (ideally, at least 12) have used the same method and replicate measurement results within each laboratory – then these data can be used to characterize variation across laboratories.'	Partially accepted. The comment relates to 9.3 rather than 9.2. We do not require replicate measurements, since the repeatability component can also be obtained from other sources. We nevertheless have added that, ideally, replicate measurements are performed in each lab.
	9.4 Procedures for characterizing fundamental variability	
	Procedures for characterizing fundamental variability	
Canada	The fundamental variability Section (9.4) seems to be quite general, and thus some more details could be given. For example, might it be helpful to explain the Horwitz SD value, particularly for users new to this type of work?	Partially accepted. We will add a fundamental variability curve as a function of content, for different particle and test portion sizes.
Chile	Fundamental variability is a subcomponent of the repeatability error term from the basic model in Section 3 and denotes the irreducible variation between samples which remains even under the highest achievable degree of homogeneity. Fundamental variability reflects heterogeneity at the level of the sample's constituent particles ; it has an influence on the uncertainty of measurement results when the target analyte is located on sparsely distributed carrier particles. Fundamental variability appears twice: first, during sampling, and second, during subsampling in the laboratory, i.e. extraction of a test	

	<p>portion after homogenization of the laboratory sample. In practice, nonnegligible fundamental variability can be reduced by modifying the testing procedure in two respects: first, by finer grinding or comminuting or mixing of the test material, and second, by increasing the test portion size.</p>	
	Chile has doubts if this is equivalent to the fundamental error of Pierre Gy's Theory?	Accepted. We will add a reference.
Canada	$S_F = \sqrt{\frac{k}{(k-1)} \cdot (s_1^2 - s_2^2)}$	
	If s12/s22 is less than 2.17, does that mean fundamental variability is not significant and SF doesn't need to be estimated? If this understanding is correct, please provide a statement clarifying this, so that the reader understands what to do if the condition is <2.17.	Accepted. We will add a clarification.
	10. Influence of measurement uncertainty on sampling plans: examples	
Australia	This modified procedure is taken from current stage of development of Annex B of ISO/WD ISO 3951-6 [15].	Accepted
	Specific Comments – editorial	
	14. Section 2, paragraph 9, second sentence. For clarity suggest explaining the acronym used in the quotation using square brackets. Thus amend to 'This can briefly described as "repeated sampling from the PDFs[probability density functions] for the and the evaluation of the model in each case," see 5.9.1 in [3].'	Partially accepted, we simply wrote out probability density function in full.
	15. Section 3, last paragraph, second sentence. The reference is not consistent with the Reference section. Suggest amendment to "This approach is described in Uhlig (20232) [25] and..."	Accepted

	16. Section 9.1, paragraph 4, first sentence. The text reference appears incorrect as “Section 0” but correctly linked to Section 8. Thus, amend to “As explained in Section 08, it is	Accepted
Canada	Procedures for bulk sampling are provided in ISO 10725:2000 [17] . As in the case of sampling from packages, these procedures are only valid under the assumption that there is no method bias. Modified procedures for the case that there is a method bias are currently being developed. For now, the discussion is limited to the case that there is no bias.	Accepted
	Please clarify if the procedures described at the end of Section 10 are specific to bulk sampling. It is suggested to have a subsection or subheading before text that is pertinent to bulk sampling. For instance, the second example at the end of Section 10 does not appear to use the general guidelines summarized earlier in the same section. Please clarify which criteria and variables in the example apply specifically to this scenario (i.e., bulk sample of wheat), or provide more details if they are more broadly applicable in other scenarios.	Accepted.
	References	
Australia	Section ‘Reference’. While we all seek to implement an information document with the most recent research, the reference [23], [24], & [25] articles have not undergone peer review and unless specifically required should not be included	This is perhaps a question for the secretariat.
Chile	ISO 5725-2:1994: <u>2019</u> , Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.	Accepted
	outdated versions of ISO standards are mentioned in the references, which would be important to correct.	Accepted
	ISO 3951-1:2016 <u>2022</u> , Sampling procedures for inspection by variables — Part 1: Specification of single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL.	Accepted
	outdated versions of ISO standards are mentioned in the references, which would be important to correct.	Accepted

	ISO/WD 3951-6:2019, Sampling procedures for inspection by variables — Part 6: Specification for single sampling plans indexed by limiting quality (LQ).	
	The committee in Chile searched in ISO and only found "ISO/DIS 3951-6 Sampling procedures for inspection by variables — Part 6: Specification for single sampling plans for isolated lot inspection indexed by limiting quality (LQ)" also indicates that it is standard under development, that is, it is not yet official: https://www.iso.org/standard/78827.html	Accepted
	ISO TS 23471:2022, Experimental designs for the evaluation of uncertainty – Use of factorial designs for determining uncertainty functions.	Accepted
	ISO 13528:20152022, Statistical methods for use in proficiency testing by interlaboratory comparison.	Accepted
	Current version 2022	