GUIDELINE FOR CALCULATING MEASUREMENT UNCERTAINTY IN QUANTITATIVE FORENSIC INVESTIGATIONS

DOCUMENT TYPE:	REF. CODE:	ISSUE NO:	ISSUE DATE:
GUIDELINE	**QCC-GDL-001**	**001**	**15.04.2024**

CONTENT

This Guideline for calculating measurement uncertainty in quantitative forensic investigations was created through an ENFSI QCLG-QCC project (2019 - 2023). The following ENFSI members contributed to the content of the Guideline:

- Mr Ivo Alberink (Netherlands, NFI);
- Mrs Rosario Garcia Repetto (Spain, National Institute of Toxicology and Forensic Sciences. Seville Dept);
- Mr Leen van der Ham (Netherlands, NFI);
- Mrs Joyce KLU (UK, University of Dundee);
- Mr Tuomas Korpinsalo (Finland, Police Finland);
- Mrs Sonja Menges (Germany, BKA);
- Mr Anders Nordgaard (Sweden, LIU ;
- Mrs Sabine Schneiders (Germany, BKA);
- Mrs Luisa Soria (Spain, National Institute of Toxicology and Forensic Sciences. Seville Dept);
- Mr Jos Tóth (Netherlands, NFI);
- Mrs Daniela Vaz (Portugal, Scientific Police Laboratory of the Judiciary Police).

1 INTRODUCTION

All over the world, the last decades have seen great progress in the professional development of investigation methods used in forensic technical investigation. Society, national legislation but also international cooperation in the fight against crime are increasingly pushing for assurance of the quality of this forensic examination by introducing accredited quality management systems (QMS).

Consequence of this development is that applicants for forensic technical investigation expect the expert reports submitted to them to contain reliable results.

The users of the forensic expert reports have to interpret the contents correctly. This implies that the results and conclusions should also include the degree of reliability of the measurement data provided.

Enriching results of quantitative measurements with the calculated measurement uncertainty and the explanation thereof in the expert reports is were appropriate desirable and/or necessary.

Naturally, statistics and, as a derivative thereof, the performance of measurement uncertainty calculations have been part of the professional training of forensic experts. Nevertheless, forensic experts often experience the performance of measurement uncertainty calculations as a difficult necessity.

This guideline aims to provide information and practical help to anyone who is looking for applicable and understandable knowledge and worked out examples on how to apply measurement uncertainty in forensic technical methods.

2 AIM

The aim of this guideline is to share Best Practice for a practical approach covering quantitative measurements aspects. This in relationship with requirements of ISO 17025 [1] concerning areas of forensic science where uncertainty of measurements must be addressed.

The examples in this guideline are straightforward, even for the more complex examples. This allows a wide application of this guide. The uncertainty measurement calculations are based on recognized literature (section 9) including the experience of the authors of this Guideline (appendix AIII).

A step-by-step approach is present in the examples which allows understanding of the used measurement uncertainty calculations. The examples are offering the reader therefore a roadmap to calculate measurement uncertainty for his own methods.

3 SCOPE

The scope of this guideline focuses on measurement uncertainty calculations as applied in quantitative chemical and physical forensic investigation methods

Uncertainty of measurements in forensic qualitative methods is a separate area of interest and is not explained in this guideline.

The manner in which information, originating from i.e. sampling aspects, results of validation studies or proficiency tests (PT) can be used for calculations of measurement uncertainty, are present in the worked-out examples as included in this guideline.

The statistical models and formulas provided in chapter 7 are not to be interpreted as normative but serve as a framework to be used for calculating measurement uncertainty. There are several ways of doing the latter, and all these ways cannot be collected into one set of formulas. The examples provided in chapter 8 demonstrate this variety of ways to calculate measurement uncertainty, and are not direct examples of applying formulas provided in chapter 7. Moreover, the examples provided in sections 8.5-8.7 are more complex/advanced examples, added in order to illustrate the diversity of questions there may be about quantifying measurement uncertainty.

4 DEFINITIONS

The naming of definitions for specific focus areas that are used when validating (forensic) investigation methods and calculating measurement uncertainty helps to ensure that everyone has the same starting position to perform these activities.

Although not exhaustive, the most frequently used terms and parameters are listed in the following paragraphs, they are taken from several sources (see Appendix 1).

4.1 General Definitions

Measurement uncertainty
Several definitions for measurement uncertainty are available in literature.

E.g. EURACHEM [2] gives the following definition in part 2.5:
“non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.

NOTE: In general, for a given set of information, it is understood that the measurement uncertainty is associated with a stated quantity value attributed to the measurand. A modification of this value results in modification of the associated uncertainty.”

Others state measurement uncertainty as a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand'[1]. This parameter could be a standard deviation or another part of an interval indicating a certain confidence range.

Most important in this connection is that not only the single measurement has to be considered but also the overall result of the test, so that all components are embraced. Some may be obtained by interpreting the statistical spread of results of a series, others have to be worked out from complementary methods regarding sampling plans or experience.

Testing results should be the best approximation to the true value. Statistical random and systematic factor effects contribute to the uncertainty of measurement of the testing results.
The latter effects should be eliminated as far as possible by using correction factors for instance. [3]

[1] Uncertainty in Measurement, Introduction and Examples, Kallner A, eJIFCC vol 13 no1: http://www.ifcc.org/ ejifcc/vol13no1/1301200103.htm (probably based on GUM)

4.2 Definitions related to measurement uncertainty calculations

Uncertainty sources
General sources of uncertainty[2] include: equipment, unit under test, operator, method, calibration and environment.

Uncertainty components
General components of uncertainty[3] are: Repeatability, Reproducibility, Stability, Bias, Drift, Resolution and Certified Reference Material.

Repeatability
Repeatability is the measurement precision under a set of repeatable conditions, means the variation among repeated measurements made on the same object (identical samples of it) using the same instrument, the same operator, the same laboratory conditions etc. To perform a repeatability test, one must continually repeat the measurement process under the same conditions until you record your desired number of samples.

Reproducibility
Reproducibility[3] refers to the variation among measurements made on the same object (identical samples of it) using different instruments or different operators or different laboratory conditions etc. for each measurement degree of agreement between measurements or observations conducted on identical samples under different investigating situations like different operators, time and date, environmental conditions etc.

Within lab reproducibility
Within lab reproducibility[4] is the precision obtained within a single laboratory over a longer period of time (generally at least several months) and takes into account changes like different analysts, different apparatus if available, different batches of reagents/ standards, different times, different environmental conditions.

Between lab reproducibility
Between lab reproducibility[5] expresses the precision between the measurement results obtained at different laboratories.

Stability
Stability is the variation among measurements made on the same object (identical samples of it) under the same conditions (with respect to instruments, operators, laboratory conditions, etc.), but at different dates. It may be included in reproducibility assuming different dates is a kind of different laboratory conditions.

Bias
Bias is a quantitative term describing the difference between the average of measurements made on the same object and its true value[6]. Calculation of BIAS is possible with help of information received by calibration and standard measurements, blind (reference) samples and results of proficiency tests.
The bias of an analytical method is usually determined by study of relevant reference materials or by spiking studies. Bias may be expressed as analytical recovery (value observed divided by value expected).

2 https_www.isobudgets.com
3 Based on Https//www.definitions.net/definition/Reproducibility
4 Adapted from https://www.favv-afsca.be/labos/erk-alg/_documents/03-11-2008-procedureENLAB-P-508-Measurement-uncertainty-v.01_en.pdf and https://sisu.ut.ee/lcms_method_validation/41-precision-trueness-accuracy (august 2022)
5 https://sisu.ut.ee/lcms_method_validation/41-precision-trueness-accuracy (august 2022)
6 https:// www.itl.nist.gov/div898/handbook/

Bias should be shown to be negligible or corrected for, but in either case
the uncertainty associated with the determination of the bias remains an essential component of the overall uncertainty.[2]

Drift
Drift refers to a continuous or incremental change over time in indication, due to changes in metrological properties of a measuring instrument.[4]

NOTE Instrumental drift is related neither to a change in a quantity being measured nor to a change of any recognized influence quantity.

Resolution
Resolution is the smallest change in a quantity being measured that causes a perceptible change in the corresponding indication.[4]

NOTE Resolution can depend on, for example, noise (internal or external) or friction. It may also depend on the value of a quantity being measured.

Reference standard/material (RM)
Material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process. [5]

NOTE: Properties can be quantitative or qualitative, e.g. identity of substances or species.

Certified Reference Material (CRM)
Certified Reference material (CRM) characterized by a metrologically valid procedure for one or more specified properties, accompanied by an CRM certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability. [5]

Standard uncertainty[7]
Uncertainty of the result of a measurement expressed as a standard deviation

Combined measurement uncertainty
All uncertainties generated are usually expressed in the form of standard uncertainties. The root of the sum of the squared standard uncertainties results in the combined measurement uncertainty.

Expanded uncertainty[7]
Quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.

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

[7] Basic definitions of uncertainty – NIST: https://physics.nist.gov › glossary,30/08/2022

Coverage factor[7]
Numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty.

Type A evaluation (of uncertainty) [7]
method of evaluation of uncertainty by the statistical analysis of series of observations

Type B evaluation (of uncertainty) [7]
method of evaluation of uncertainty by means other than the statistical analysis of series of observations

Precision
Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions. [4]

NOTE 1 Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.

NOTE 2 The 'specified conditions' can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement (see ISO 5725-1:1994).

NOTE 3 Measurement precision is used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility.

NOTE 4 Sometimes "measurement precision" is erroneously used to mean measurement accuracy.

Intermediate Precision Condition of Measurement
Condition of measurement, out of a set of conditions that includes the same measurement procedure, same location, and replicate measurements on the same or similar objects over an extended period of time, but may include other conditions involving changes. [4]

NOTE 1 The changes can include new calibrations, calibrators, operators, and measuring systems.

NOTE 2 A specification for the conditions should contain the conditions changed and unchanged, to the extent practical.

NOTE 3 In chemistry, the term "inter-serial precision condition of measurement" is sometimes used to designate this concept.

Intermediate Precision
Measurement precision under a set of intermediate precision conditions of measurement. [4]

Relative Standard Deviation (RSD)
A special form of the standard deviation, obtained from dividing the sample standard deviation by the absolute value of the sample mean. [6]
It is commonly reported as a percentage and it gives an idea about how precise your data is.[8]

5 REQUIREMENTS OF QUALITY MANAGEMENT STANDARD ISO 17025-2017

The ISO IEC 17025 [1] is the general used quality management standard within forensic investigation institutions ("testing laboratories" according to 17025). To summarize, chapter 7.6. of the standard contains the following requirements concerning measurement uncertainty:
Testing laboratories using accredited quantitative methods shall determine the corresponding measurement uncertainty. This determination shall include all significant contributions, including contributions resulting from sampling, using appropriate methods. If no precise calculation is possible, an estimation based on the underlying theoretical principles or practical experience of the method performance shall be made.

[8] https://www.statisticshowto.com/relative-standard-deviation/ (august 2022)

The measurement uncertainty in the measurement data must be determined for each method where the reported results can influence the interpretation by the customer or where legal requirements must be met.

Some quantitative test methods show little or no variation in the mode of operation or the matrix and/or concentrations of the component being tested. For such methods it is not necessary to calculate the measurement uncertainty per individual study. However, the institute should be able to demonstrate that the critical aspects which may influence the measurement uncertainty are under control.

Standardized methods, for example methods published by a national standardisation agency, shall be considered as validated. The therein published measurement uncertainty can be used directly, if the institute can demonstrate that it meets the requirements of these accepted methods.

By verifying these ISO 17025 requirements with the intention and goal concerning the own calculations of measurement uncertainty in quantitative forensic investigations it can become determined that these requirements are reached.

6 THE PROCESS OF MEASUREMENT UNCERTAINTY CALCULATIONS

6.1 Basics

Primarily, the effort and the procedure for determination of the measurement uncertainty is subject to the requirements for the measured value. They may result from statutory provisions, risk assessments, the customer's requirements, etc.

The measurement uncertainty of a measurement procedure may contain numerous components. It can be assessed e.g. by means of a modelling approach or an integrative approach.

6.2 Modelling approach:

In the modelling approach (a “bottom-up approach”), a model of the measurement is created, as described in GUM [7]. In the course of such action, the measurement procedure can be divided into individual modules for which individual uncertainties can be determined. A cause-effect diagram (fishbone or Ishikawa diagram) is recommended for depicting the individual components.

For details on the procedure, see [2].

The modelling approach either delivers a measurement uncertainty for the overall procedure or uncertainty contributions from individual modules of the overall procedure which are combined for a total uncertainty pursuant to the law of propagation of uncertainty.

This example given in the guideline use the modelling approach:

- Quantifying Delta-9-Tetrahydrocannabinol (THC) in Blood example using the Measurement Uncertainty Calculator (MUCalc) (Chapter 8.5)

6.3 Integrative approach

In this indirect approach (a “top-down approach”), several sources of uncertainty are determined in an integrative manner. Usually, the results of the quality assurance sample analytics are used to determine the measuring accuracy. Uncertainty components not sufficiently considered during quality assurance must be determined in addition.

Trueness of measurement is determined using the results of analytics of certified matrix reference material if possible. The results from comparative measurements, performance tests or recovery experiments may also be used.

Common approaches according to Guidelines on measurement uncertainty, CXG 54-2004, revised in 2021 [8] are:

- Single lab validation: uncertainty of results obtained using the same procedure in a single laboratory under varying conditions
- Interlaboratory validation: uncertainty of results obtained using the same procedure in different laboratories
- Proficiency testing: uncertainty of results obtained using the same sample(s) in different laboratories

Examples for this approach are as well described in detail e.g. in NORDTEST Technical Report TR 569 "Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories" [9].

The examples given in this guideline, that use types of the integrative approach are:

- Calibration of thermometers (Chapter 8.1)
- Quantification of MDMA in powders, mixtures and tablets by high performance liquid chromatography (HPLC-DAD), (Chapter 8.2)
- Quantitative determination of cocaine in seizures by HPLC DAD method, (Chapter 8.3)
- Determination of ethanol in blood using headspace gas chromatography with flame ionization detector (HS-GC-FID), (Chapter 8.4)
- Predicting net weights of khat mardoufs (Chapter 8.6)
- Velocity estimation on a speeding car in video images (Chapter 8.7)

Whatever approach is used, it should be scientifically accepted. None of the following described methods may be said to be better than any other. [8]

6.4 Sources of uncertainty

Although it is assumed that a “true value” of a quantity being measured exists, this true value is unknown so is the measurement error. By evaluating the measurement uncertainty an interval can be given within quantitative values will lie with a stated coverage probability. [8]

Significant uncertainty contributions may arise from many possible sources, including examples such as sampling, matrix effects and interferences, environmental conditions, uncertainties of masses and volumetric equipment, reference values, approximations and assumptions incorporated in the measurement method and procedure, as well as random variation. Such uncertainty contributions having no significant influence on the total uncertainty can be disregarded.

NOTE: the focus should lie on the identification and evaluation of the main components of measurement uncertainty especially on systematic components as they cannot be reduced by repeating measurements! [8]

In order to assess the measurement uncertainty, all steps of the process performed in the lab must be considered individually, if applicable.

Thus, the following inherently relevant sources of uncertainty could be compiled for a certain examination procedure, among others:

(1) Weighing uncertainties
(2) Volume uncertainties of pipettes and volumetric flasks
(3) Fluctuations of density
(4) Temperature effects
(5) Metrological traceability
(6) Homogeneity of the sampling material
(7) Matrix of the sampling material
(8) (Complex) steps in preparation of the sample
(9) Stability of the measuring signal
(10) Calibration of measuring device

If the assessment finds that e.g. the uncertainty contributions 1) to 5) do not have significant influence on the overall uncertainty in comparison to the uncertainty contributions 6) to 10), they may be disregarded.

6.5 Process

In principle uncertainty estimation is simple and the following step-by-step approach (see [2]), summarises the tasks needed in order to obtain an estimate of the uncertainty associated with a measurement result, please see there for further details.

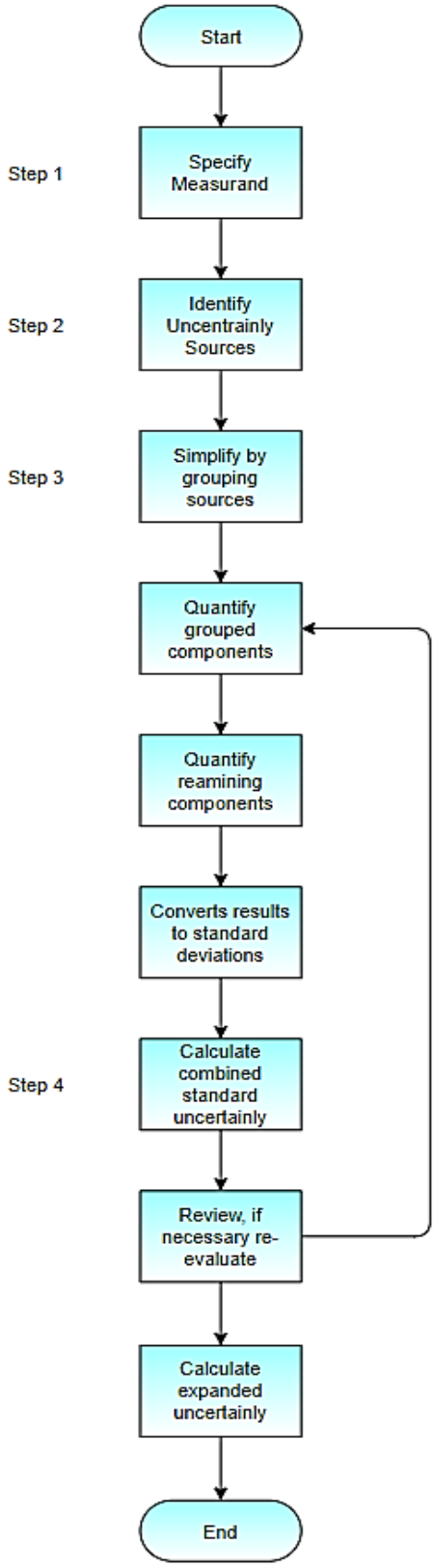

Figure 6.5-1: The uncertainty estimation process

6.6 Frequency of determination of the measuring uncertainty

On principle, the measurement uncertainties of the observed measured variables determined with a test procedure shall be assessed once; this can be performed e.g. within the scope of validation.

The assessment must be reviewed, e.g. if

- internal or external quality assurance measures indicate there are issues
- new findings are determined
- significant changes are made to the analysis procedure
- new or other analytical equipment is used

7 STATISTICAL MODELS FOR DETERMINATION OF MEASUREMENT UNCERTAINTY BASED ON INTRA-LAB RESULTS OR PROFICIENCY TESTS

Note that what is described in this document is a guideline, and that alternative approaches may be acceptable as well. It is preferable that alternative approaches can be found in scientific literature.

In Annex I, it is described in what various ways determination of measurement uncertainty based on intra-lab results or proficiency tests may take place, and what the logic behind this is. In the current section we describe this from a practical point of view, concentrating on which formulas may be used for what situations.

Overall, the situation is such that 1 or more measurements are performed in order to estimate some real (nominal) number. We will refer to the first as $X_1,...,X_n$ and to the second as m. The measurements are assumed to follow a so-called *normal* or Gaussian distribution, cf. [ref Gauss], with m as its *mean value* and σ as its *standard deviation*. The square of this is called the *variance*.

If the number of measurements is 1 and the standard deviation is known, based on the outcome x of the measurement, an approximate 95% confidence interval may be determined of the unknown real value m by evaluation of the inequality

$$|x-m| \leq 2\cdot\sigma.$$ [9]

In the case that there are more measurements, one may use the *mean* $\bar{X}=(X_1+...+X_n)/n$ in order to estimate m. Typically the mean has a lower standard deviation than σ, namely, it has a standard deviation of $\sigma/\sqrt{n}$. An approximate 95% confidence interval for m is determined by the inequality

$$|\bar{x}-m| \leq 2\cdot\sigma/\sqrt{n}$$

In general the standard deviation is unknown. This is dealt with by using an estimation S instead of σ, where

[9] For an exact interval the number (coverage factor) 1,96 is used instead of 2.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

If the assumption of normality holds, then it is a general result that the random variable $(\bar{X} - m)/(S/\sqrt{n})$ has a so-called *t distribution* with *n*-1 degrees of freedom. Coverage factors for *t* distributions can be found in statistical tables that are easily available.

An approximate 95% confidence interval for *m* is determined by the inequality

$$|\bar{x}\text{-}m| \leq t_{n-1\,;\,97.5\%}\cdot s/\sqrt{n},$$

with $t_{n-1\,;\,97.5\%}$ the 97.5% percentile of a *t* distribution with *n*-1 degrees of freedom.

In the standard casework that this guideline concentrates on, besides the above there is usually also a systematic *bias* component involved in measurements, and it is unknown. In the Annex it is described how to obtain 95% confidence intervals for such situations.

In the framework of *within-laboratory* estimation of measurement uncertainty, repeated measurements are performed within the laboratory on a material for which the measurand is known, for example certified reference material (CRM) provided by an institute of standards or unit conducting proficiency tests. We denote the value of this measurand as m_{CRM}, which is given by the supplier, together with the variance σ^2_{CRM}. Now if the measurements are again denoted as x_i, it is reasonable to look at the differences between the measurements and the measurand, $d_i = x_i - m_{\text{CRM}}$, and the so-called *mean square deviation* (MSD), i.e.

$$\text{MSD}_w = \frac{1}{p}\sum_{i=1}^{p} d_i^2 = \frac{1}{p}\sum_{i=1}^{p}(x_i - m_{\text{CRM}})^2$$

is defined. In the literature it is common to refer to the root mean square deviation (RMS), which is the square root of MSD, i.e. $RMS = \sqrt{MSD}$. It is commonly suggested to use the following standard deviation of $\bar{x}$-*m*:

$$u = \sqrt{\text{MSD}_w + \sigma^2_{\text{CRM}} + \frac{s^2_{R_w}}{n}},$$

with $s^2_{R_w}$ denoting the intermediate precision of the lab, which is for example determined as the variance within the lab of a series of measurements on a Shewhart card. A 95% confidence interval then is given via:

$$|\bar{x} - m| \leq 2 \cdot \sqrt{MSD_w + \sigma^2_{\text{CRM}} + \frac{s^2_{R_w}}{n}}$$

In the framework of determination of measurement uncertainty based on *proficiency tests* we have the following. Here a lab participates in *p* tests, generating results x_i for *i*=1,...,*p*, which are compared to the mean reported results $\bar{y}_i$ over the other laboratories that participate. Now the differences are calculated as

$$d_i = x_i - \bar{y}_i,$$

and the counterpart of MSD above is

$$MSD = \frac{1}{p}\sum_{i=1}^{p} d_i^2 = \frac{1}{p}\sum_{i=1}^{p}(x_i - \bar{y}_i)^2$$

and the standard uncertainty is taken as

$$u = \sqrt{MSD + \frac{s_{R_w}^2}{n}}$$

A 95% confidence interval then is given via:

$$|\bar{x} - m| \leq 2 \cdot \sqrt{MSD + \frac{s_{R_w}^2}{n}}$$

In similar approaches an extra term is introduced compensating for the fact that the terms $\bar{y}_i$ are considered as the ground truth of the nominal values m_i.

A different approach to the above separates the bias contribution from the standard deviation *u* that is used.

It expresses the bandwidth of the confidence intervals in terms of

$$\widehat{|B|}_w + 2 \cdot \sqrt{\sigma_{\text{CRM}}^2 + \frac{s_{R_w}^2}{n}}$$

in the case of within lab calculation of measurement uncertainty, and

$$\widehat{|B|} + 2 \cdot \sqrt{\frac{s_{R_w}^2}{n}}$$

in case of proficiency testing. Here $\widehat{|B|}$ is the mean absolute value, that is,

$$\widehat{|B|} = \frac{1}{p}\sum_{i=1}^{p} |x_i - \bar{y}_i|$$

The corresponding formula for the case of within-lab MU calculation is

$$\widehat{|B|}_w = \frac{1}{p}\sum_{i=1}^{p} |x_i - m_{\text{CRM}}|$$

7.1 Relative uncertainty

In the case where a model is used with *relative uncertainty* the formulas are a bit more involved. The model that is used is typically that

$$X_i / m = 1 + b + E_i$$

with E_i being some random error term that is from a normal distribution with mean 0 and standard deviation

CV (coefficient of variation). In the framework of within-laboratory estimation of measurement uncertainty, repeated measurements are again performed within the laboratory on a material for which the measurand is known, for example certified reference material (CRM) provided by an institute of standards or unit conducting proficiency tests. The value of this measurand is given as m_{CRM}, together with the relative standard deviation $\%\sigma$. Now if the measurements are again denoted as x_i, it is reasonable to look at the relative differences between the measurements and the measurand, $d_i = (x_i - m_{\mathrm{CRM}})/m_{\mathrm{CRM}} = x_i/m_{\mathrm{CRM}} - 1,$ and the

$$MSD_{w,r} = \frac{1}{p}\sum_{i=1}^{p} d_i^2 = \frac{1}{p}\sum_{i=1}^{p}(x_i/m_{\mathrm{CRM}} - 1)^2$$

with $RMS_{w,r} = \sqrt{MSD_{w,r}}$. With n measurements, it is suggested to use the following standard deviation of $\bar{x}/m_{\mathrm{CRM}} - 1$:

$$u = \sqrt{\mathrm{MSD}_{w,r} + CV_{\mathrm{CRM}}^2 + CV^2/n},$$

where CV is determined within the lab by means of series of relative measurements. A 95% confidence interval then is given via:

$$\left|\frac{\bar{x}}{m} - 1\right| \leq 2 \cdot \sqrt{\mathrm{MSD}_{w,r} + CV_{\mathrm{CRM}}^2 + CV^2/n}$$

In the framework of determination of measurement uncertainty based on proficiency tests we have the following. The lab participates in p tests, generating results X_i for i=1,...,p, which are compared to the mean reported results $\bar{Y}_i$ over the other laboratories that participate.

Now the relative differences are calculated as

$$d_i = \frac{x_i}{\bar{y}_i} - 1,$$

and the counterpart of MSD above is

$$MSD_r = \frac{1}{p}\sum_{i=1}^{p} d_i^2 = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{x_i}{\bar{y}_i} - 1\right)^2$$

and with n measurements the standard uncertainty is taken as

$$u = \sqrt{MSD_r + CV^2/n}\,.$$

A 95% confidence interval then is given via:

$$\left|\frac{\bar{x}}{m} - 1\right| \leq 2 \cdot \sqrt{MSD_r + \frac{CV^2}{n}}$$

In similar approaches an extra term is introduced compensating for the fact that the terms $\bar{y}_i$ are considered as the ground truth of the nominal values m_i. A different approach to the above

separates the bias contribution from the standard deviation *u* that is used. It expresses the bandwidth of the confidence intervals in terms of

$$|\widehat{B}|_{w,r} + 2 \cdot \sqrt{CV_{\text{CRM}}^2 + CV^2/n}$$

in the case of within lab calculation of measurement uncertainty, and

$$|\widehat{B}|_r + 2 \cdot \sqrt{CV^2/n}$$

in case of proficiency testing. Here $|\widehat{B}|_{w,r}$ and $|\widehat{B}|_r$ are the mean absolute relative values, that is,

$$|\widehat{B}|_{w,r} = \frac{1}{p}\sum_{i=1}^{p}\left|\frac{x_i}{m} - 1\right| \quad \text{and} \quad |\widehat{B}|_r = \frac{1}{p}\sum_{i=1}^{p}\left|\frac{x_i}{\bar{y}_i} - 1\right|$$

7.2 Example with interpretation

In the example, the task is to calculate the measurement uncertainty for measurements of conductivity (in units of Siemens per meter [S/m]) made on Swedish “10 kronor”-coins, with the purpose to detect fake coins.

7.2.1 Within-laboratory estimation

In this case, a specimen of a genuine coin is sent to a laboratory, with the information that its conductivity is 15.9 *S*/*m* with standard deviation 0.05 *S*/*m*. This means that $m_{\text{CRM}} = 15.9$ and $\sigma_{\text{CRM}}^2 = 0.05^2$=0.0025.

The laboratory makes $p = 20$ measurements on the coin rendering the following results:

15.80, 15.75, 15.72, 15.67, 15.88, 15.79, 16.03, 16.03, 15.96, 16.05,
15.93, 15.92, 16.04, 15.92, 16.04, 16.12, 16.01, 15.80, 15.71, 15.66

At laboratory *A*, Shewhart charts have been used for a long time capturing the intermediate precision of measurements giving an estimate $s_{Rw} = 0.264$.

7.2.1.1 *Absolute uncertainty*

From the data we calculate

$$MSD_w = \frac{1}{p}\sum_{i=1}^{p}(x_i - m_{CRM})^2 = \frac{1}{20}\cdot[(15.80 - 15.9)^2 + (15.75 - 15.9)^2 + \cdots + (15.66 - 15.9)^2] \approx 0.01980$$

Hence, a 95% confidence interval for the true value of the conductivity of a “10 kronor”-coin using one single measurement is

$$|x - m| \leq 2 \cdot \sqrt{0.01980 + 0.0025 + 0.264^2} \approx 0.61.$$

A measurement should thus be reported with a margin ± 0.61 *S/m*.

Using n measurements the corresponding confidence interval is

$$|\bar{x} - m| \leq 2 \cdot \sqrt{0.01980 + 0.0025 + \frac{0.264^2}{n}}.$$

If for instance the average of two measurements is 15.20 S/m, the margin to be reported is $\pm 2 \cdot \sqrt{0.01980 + 0.0025 + 0.264^2/2} \approx \pm 0.48$ *S/m*.

The alternative treating of bias is to calculate

$$\widehat{|B|}_w = \frac{1}{p} \cdot \sum_{i=1}^{p} |x_i - m_{\text{CRM}}| =$$
$$= \frac{1}{20} \cdot [|15.80 - 15.9| + |15.75 - 15.9| + \cdots |15.66 - 15.9|] \approx 0.1235.$$

and a 95% confidence interval for the true value of the conductivity of a “10 kronor”-coin using one single measurement is

$$|x - m| \leq \widehat{|B|} + 2 \cdot \sqrt{\sigma_{\text{CRM}}^2 + s_{\bar{R}_w}^2} = 0.1235 + 2 \cdot \sqrt{0.0025 + 0.264^2} \approx 0.66.$$

A measurement should thus be reported with a margin ± 0.66 *S/m*.

Using n measurements the corresponding confidence interval is

$$|\bar{x} - m| \leq 0.1235 + 2 \cdot \sqrt{0.0025 + \frac{0.264^2}{n}}.$$

So, with an average 15.20 *S/m* of two measurements, the margin to be reported is

$\pm 0.1235 + 2 \cdot \sqrt{0.0025 + 0.264^2/2} \approx \pm 0.51$ *S/m*.

7.2.1.2 *Relative uncertainty*

At laboratory *A*, the coefficient of variation reflecting intermediate precision is estimated as $CV = 0.0165\ (1.65\%)$. The coefficient of variation for the certified reference measurement can be calculated as $CV_{\mathrm{CRM}} = \sigma_{\mathrm{CRM}}/m_{\mathrm{CRM}} = 0.05/15.9 \approx 0.0031\ (0.31\%)$.

From the data we calculate

$$MSD_{w,r} = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{x_i}{m_{\mathrm{CRM}}} - 1\right)^2 = \\ = \frac{1}{20}\cdot\left[\left(\frac{15.80}{15.9} - 1\right)^2 + \left(\frac{15.75}{15.9} - 1\right)^2 + \cdots \left(\frac{15.66}{15.9} - 1\right)^2\right] \approx 7.8\cdot 10^{-5}.$$

The relative 95% expanded uncertainty for one single measurement then becomes

$$2\cdot\sqrt{\mathrm{MSD}_{w,r} + CV_{\mathrm{CRM}}^2 + CV^2} = 2\cdot\sqrt{7.8\cdot 10^{-5} + (0.05/15.9)^2 + 0.0165^2} \approx 0.038 = 3.8\%.$$

With an average 15.20 *S*/*m* of two measurements the relative 95% expanded uncertainty becomes

$$2\cdot\sqrt{\mathrm{MSD}_{w,r} + CV_{\mathrm{CRM}}^2 + \frac{CV^2}{2}} = 2\cdot\sqrt{7.8\cdot 10^{-5} + (0.05/15.9)^2 + \frac{0.0165^2}{2}} \approx 0.030 = 3.0\%.$$

For this result, the margin to be reported is $\pm 0.030\cdot 15.20 \approx \pm 0.46$ *S*/*m*.

The alternative treating of bias is to calculate

$$|\widehat{B}|_{w,r} = \frac{1}{p}\sum_{i=1}^{p}\left|\frac{x_i}{m} - 1\right| = \frac{1}{20}\left[\left|\frac{15.80}{15.9} - 1\right| + \left|\frac{15.75}{15.9} - 1\right| + \cdots + \left|\frac{15.66}{15.9} - 1\right|\right] \approx 0.0078.$$

and the relative 95% expanded uncertainty for one single measurement then becomes

$$|\widehat{B}|_{w,r} + 2\cdot\sqrt{CV_{\mathrm{CRM}}^2 + CV^2} = 0.0078 + 2\cdot\sqrt{(0.05/15.9)^2 + 0.0165^2} \approx 0.041 = 4.1\%.$$

With an average 15.20 *S*/*m* of two measurements the relative 95% expanded uncertainty becomes

$$|\widehat{B}|_{w,r} + 2\cdot\sqrt{CV_{\mathrm{CRM}}^2 + \frac{CV^2}{n}} = 0.0078 + 2\cdot\sqrt{(0.05/15.9)^2 + \frac{0.0165^2}{2}} \approx 0.032 = 3.2\%.$$

For this result, the margin to be reported is $\pm 0.032\cdot 15.20 \approx \pm 0.49$ *S*/*m*.

7.2.2 Using results from proficiency test

In this case, “10 kronor”-coins, all genuine, are sent to laboratories participating in 9 proficiency tests for measurement of conductivity.
The measurement results from the 9 tests are summarised in Table 7.2-1.

Test (i)	*Reported conductivity from laboratory A (x_i)*	*Average reported conductivity from the other labs ($\bar{y}_i$)*
1	16.3 S/m	16.2 S/m
2	15.8 S/m	15.7 S/m
3	16.0 S/m	16.1 S/m
4	15.2 S/m	15.5 S/m
5	15.7 S/m	15.9 S/m
6	16.1 S/m	16.0 S/m
7	15.9 S/m	15.9 S/m
8	16.0 S/m	16.1 S/m
9	15.9 S/m	15.9 S/m

Table 7.2-1: Results from proficiency test of measurements on Swedish “10 kronor”-coins.

Now, assume laboratory *A* is supposed to calculate their measurement uncertainty with help from the results from the other labs participating in the proficiency test.

7.2.2.1 *Absolute uncertainty*

The number of tests, $p = 9$ and MSD is calculated from the data in Table 7.2-1 as

$$MSD = \frac{1}{p} \times \sum_{i=1}^{p} d_i^2 = \frac{1}{9} \sum_{i=1}^{9} (x_i - \bar{y}_i)^2 =$$
$$= \frac{1}{9} \cdot [(16.3 - 16.2)^2 + (15.8 - 15.7)^2 + \cdots + (15.9 - 15.9)^2] \approx 0.020.$$

As was given in section 7.2.1, laboratory A’s estimate of the uncertainty due to repeatability within laboratory is $s_{Rw} = 0.264$. Hence, a 95% confidence interval for the true value of the conductivity of a “10 kronor”-coin using one single measurement is

$$|x - m| \leq 2 \cdot \sqrt{0.020 + 0.264^2} \approx 0.60.$$

A measurement should thus be reported with a margin ± 0.60 *S/m*.

Using n measurements the corresponding confidence interval is

$$|\bar{x} - m| \leq 2 \cdot \sqrt{0.020 + \frac{0.264^2}{n}}.$$

With an average 15.20 *S/m* of two measurements the margin to be reported is

$\pm 2 \cdot \sqrt{0.020 + 0.264^2/2} \approx \pm 0.47$ *S/m*.

The alternative treating of bias is to calculate

$$\widehat{|B|} = \frac{1}{p}\sum_{i=1}^{p}|x_i - \bar{y}_i| = \frac{1}{9}\cdot[|16.3 - 16.2| + |15.8 - 15.7| + \cdots |15.9 - 15.9|] \approx 0.1111.$$

and a 95% confidence interval for the true value of the conductivity of a “10 kronor”-coin using one single measurement is

$$|x - m| \leq \widehat{|B|} + 2 \cdot \sqrt{s_{R_w}^2} = 0.111 + 2 \cdot \sqrt{0.264^2} \approx 0.64$$

A measurement should thus be reported with a margin ± 0.64 *S/m*.

Using n measurements the corresponding confidence interval is

$$|x - m| \leq 0.111 + 2 \cdot \sqrt{\frac{0.264^2}{n}}$$

Thus, with an average 15.20 *S/m* of two measurements, the margin to be reported is $\pm 0.1111 + 2 \cdot \sqrt{0.264^2/2} \approx \pm 0.48$ *S/m*.

7.2.2.2 *Relative uncertainty*

As was given section 7.2.1.2, at laboratory *A*, the coefficient of variation reflecting intermediate precision is estimated as $CV = 0.0165\ (1.65\%)$.

From the data in Table 7.2-1 we calculate

$$MSD_r = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{x_i}{\bar{y}_i} - 1\right)^2 = \frac{1}{9}\cdot\left[\left(\frac{16.3}{16.2} - 1\right)^2 + \left(\frac{15.8}{15.7} - 1\right)^2 + \cdots + \left(\frac{15.9}{15.9} - 1\right)^2\right] \approx 8.1 \cdot 10^{-5}.$$

The relative 95% expanded uncertainty for one single measurement then becomes

$$2 \cdot \sqrt{MSD_r + CV^2} = 2 \cdot \sqrt{8.1 \cdot 10^{-5} + 0.0165^2} \approx 0.038 = 3.8\%.$$

With an average 15.20 S/m of two measurements the relative 95% expanded uncertainty becomes

$$2 \times \sqrt{MSD_r + \frac{CV^2}{2}} = 2 \cdot \sqrt{8.1 \cdot 10^{-5} + \frac{0.0165^2}{2}} \approx 0.029 = 2.9\%.$$

For this result, the margin to be reported is $\pm 0.029 \cdot 15.20 \approx \pm 0.44$ *S/m*.

The alternative treating of bias is to calculate

$$\widehat{|B|}_r = \frac{1}{p}\sum_{i=1}^{p}\left|\frac{x_i}{\bar{y}_i} - 1\right| = \frac{1}{9}\cdot\left[\left|\frac{16.3}{16.2} - 1\right| + \left|\frac{15.8}{15.7} - 1\right| + \cdots \left|\frac{15.9}{15.9} - 1\right|\right] \approx 0.0070.$$

The relative 95% expanded uncertainty for one single measurement then becomes

$$|\widehat{B}|_r + 2 \cdot \sqrt{CV^2} = 0.0070 + 2 \cdot \sqrt{0.0165^2} = 0.040 = 4.0\%.$$

With an average 15.20 S/m of two measurements the relative 95% expanded uncertainty becomes

$$|\widehat{B}|_r + 2 \cdot \sqrt{\frac{CV^2}{2}} = 0.0070 + 2 \cdot \sqrt{\frac{0.0165^2}{2}} = 0.030 = 3.0\%.$$

For this result, the margin to be reported is $\pm 0.030 \cdot 15.20 \approx \pm 0.46$ *S/m*.

8 EXAMPLES OF CALCULATIONS OF MEASUREMENT UNCERTAINTY

8.1 Calibration of thermometers

At the National Bureau of Investigation Forensic Laboratory in Finland, DNA reagents are stored in refrigerators. To ensure proper storage temperature, digital thermometers are used to monitor the temperature within the refrigerators. Every couple of years, the thermometers are calibrated by comparing them against two reference thermometers to confirm they are still accurate and also to determine their measurement uncertainty. This example demonstrates the calculations for the calibration and the determination of the measurement uncertainty for a single thermometer, using an integrative approach (see chapter 6).

The thermometer to be calibrated is a digital thermometer that displays the temperature to one decimal precision. The thermometer is calibrated using the ice point method where ice is mixed with cold water in order to produce ice water with 0 °C temperature. To this end, the ice water is measured 4 times with the thermometer to be calibrated as well as one of the reference thermometers. The second reference thermometer is only used to verify that the measurements from the first reference are correct and only one measurement is made with this thermometer. The reading from the second thermometer is not used for the calculations and is therefore excluded from the following.

The four measurements in degrees Celsius with both thermometers are as follows:

Thermometer	Measurement 1	Measurement 2	Measurement 3	Measurement 4
1321-014	0.4	0.1	0.2	-0.1
Reference 648	-0.1	0.1	0.0	-0.1

Table 8.1-1: Measurements results obtain from the thermometer in calibration and the reference thermometer.

The measurement model is given by the formula

$$t = m + \delta r + \delta c,$$

where t is the final temperature, m is the measured value, δr is the correction due to rounding and δc is the correction due to systematic error detected in calibration. Here, δr is assumed to be zero since the rounding error should produce no systematic bias. Regardless, the uncertainty of the rounding error must still be taken into account. From this measurement model, the total uncertainty of the measurement $u(t)$ can be derived as

$$u(t) = \sqrt{u(m)^2 + u(\delta r)^2 + u(\delta c)^2},$$

where the square root is taken over the added squared standard uncertainties of each individual component in the measurement model. Thus, to determine the overall measurement uncertainty, it is enough to know the uncertainties of each individual component.

As in practice only single measurements are made with the thermometers, the uncertainty related to the measurement $u(m)$ is determined based on experience accordingly to the recommendations from the Finnish center of metrology MIKES. The uncertainty $u(m)$ is assumed to be uniformly distributed in an interval of length L centered around m. Here it is

assumed that L is 1 degree Celsius, i.e. there is no more than ± 0.5 degrees error due to the inaccuracy of the measurement.

Based on these assumptions, the standard uncertainty of the measurement can be calculated as the standard deviation of the corresponding uniform distribution

$$u(m) = \frac{L}{2\sqrt{3}} = \frac{1}{2\sqrt{3}} \approx 0.3.$$

Uncertainty of the rounding correction $u(\delta r)$ is determined in a similar way. As the thermometer measures temperature at one decimal precision, the value read from the thermometer is ± 0.05 degrees from the actual measured value. Thus, the uncertainty can be considered to be uniformly distributed along an interval of length 0.1 centered around the actual value and it can be calculated as

$$u(\delta r) = \frac{0.1}{2\sqrt{3}} \approx 0.03.$$

Finally, there is the correction δc from the calibration itself. This term represents the deviation of the measurements of the thermometer to be calibrated from the reference thermometer and it has its own uncertainty $u(\delta c)$ due to the uncertainties in calibration of the reference thermometer itself as well as due to the finite number of repeat measurements during the actual calibration process.

Firstly, the means of the measurements given in the table earlier, in degrees Celsius, are

$$\bar{x}_{cal} = \frac{0.4 + 0.1 + 0.2 - 0.1}{4} = 0.15$$

for the thermometer to be calibrated and

$$\bar{x}_{ref} = \frac{-0.1 + 0.1 + 0.0 - 0.1}{4} = -0.025$$

for the reference. The corresponding (corrected) standard deviations are given by

$$s_{cal} = \sqrt{\frac{\sum_{i=1}^{4}(x_i - \bar{x}_{cal})^2}{4-1}} = \sqrt{\frac{(0.4 - 0.15)^2 + (0.1 - 0.15)^2 + (0.2 - 0.15)^2 + (-0.1 - 0.15)^2}{3}} \approx 0.208$$

for the thermometer to be calibrated and by

$$s_{ref} = \sqrt{\frac{\sum_{i=1}^{4}(x_i - \bar{x}_{ref})^2}{4-1}} = \sqrt{\frac{(-0.1 + 0.025)^2 + (0.1 + 0.025)^2 + (0.0 + 0.025)^2 + (-0.1 + 0.025)^2}{3}} \approx 0.096$$

for the reference. From these values, the standard uncertainties of the means are obtained by dividing the standard deviations by the square root of the number of the observations:

$$u(\bar{x}_{cal}) = \frac{s_{cal}}{\sqrt{4}} \approx 0.104,$$

$$u(\bar{x}_{ref}) = \frac{s_{ref}}{\sqrt{4}} \approx 0.048.$$

The correction term δc is a bias term, which reflects how far away from the "true" value, indicated by the reference thermometer, the values from the thermometer to be calibrated tend to be on average.

Its value is obtained by subtracting the mean of the measurements from the thermometer to be calibrated from the reference measurement mean, i.e. $\delta c = \bar{x}_{ref} - \bar{x}_{cal} = -0.025 - 0.15 = -0.175$. The uncertainty of this term then comprises the standard uncertainties of the means, the calibration uncertainty of the reference thermometer and the uncertainties induced by rounding for both thermometers. The calibration of uncertainty of the reference thermometer is given by its calibration report and, in this instance, is $u(\delta c_{ref}) = 0.01$ and centered at zero. As the effect of rounding is the same in both cases, the value for this uncertainty $u(\delta r)$ calculated earlier can be used for both instances. Thus, the uncertainty is given by

$$u(\delta c) = \sqrt{u(\bar{x}_{cal})^2 + u(\bar{x}_{ref})^2 + u(\delta c_{ref})^2 + 2u(\delta r)^2} \approx 0.12.$$

It should be noted, that while the uncertainties for the means were divided by the square root of the observations, similar treatment is not applied to the other terms. This is because the formula for the uncertainty of the mean assumes independent identically distributed errors which cannot be guaranteed for the correction terms. Furthermore, there could be and probably are other sources of error that are not accounted for in these calculations. Therefore, in order to err on the side of caution, the uncertainties are added as is, possibly slightly inflating the total uncertainty.

Now all the components for calculating the complete measurement uncertainty is available and it is given by

$$u(t) = \sqrt{u(m)^2 + u(\delta r)^2 + u(\delta c)^2} \approx 0.31.$$

Finally, the expanded measurement uncertainty is obtained from the previous by multiplying the complete measurement uncertainty with a coverage factor. At the NBI Forensic Laboratory, coverage factor of 2 is used to obtain approximately 95% uncertainty intervals. This gives $2u(t) \approx 0.63$ as the expanded uncertainty. Thus, the final measurement, according to the measurement model specified earlier, should be

$$t = m - 0.175, \pm 0.63.$$

8.2 Quantification of MDMA in powders, mixtures and tablets by high performance liquid chromatography (HPLC-DAD)

The MU estimation used in this example is an integrative approach (see Chapter 6) and is based on the Nordtest approach [9].

Uncertainty components that are considered:

- Precision component (within-laboratory reproducibility)
- Bias component (lab bias)

8.2.1 Step 1. Specify measurand

Quantification of MDMA in powders, mixtures and tablets by HPLC.

8.2.2 Step 2. Quantify precision component (R_W)

Control limits are set to +/- 6.3% (rel.).

They represent the intermediate precision on the 99% confidence level and result from 68 measurements of a control sample (see supplement data, table 8.2-1).

2,1% rel. RSD*3=6,3%rel

8.2.3 Step 3. Quantify bias component

a) The bias results (% rel.) from m=7 interlaboratory comparisons (21-24 participants) are:

-6.3, +4.5, -1.7, +3.9, -0.4, -2.9 and +1.8 (see supplement data, table 8.2-2).

They are calculated by

$$bias_i = \frac{V_i - V_{R_i}}{V_{R_i}} \cdot 100 \text{ (\% rel.)}$$

where V_{R_i} is the assigned (nominal) value and V_i is the laboratory result in the i-th interlaboratory comparison.

The root mean square of the bias is:

$$RMS_{bias} = \sqrt{\frac{\sum bias_i^2}{m}} = \sqrt{\frac{89{,}9}{7}} = 3.6 \text{ (\% rel.)}$$

b) The uncertainty of the nominal value is calculated by

$$u(C_{ref}) = \frac{S_R}{\sqrt{n}}$$

where S_R is the mean of the between laboratory RSD and n the mean number of participants. With S_R = 5.8% and n = 22 (values taken from the 7 interlaboratory comparisons) the uncertainty of the nominal value is

$$u(C_{ref}) = \frac{5.8}{\sqrt{22}} = 1.2 \text{ (\% rel.)}$$

8.2.4 Step 4. Convert components to standard uncertainty u(x)

Confidence intervals and similar distributions can be converted to standard uncertainty.

Lab precision:

u(Rw) = 6.3/3 = 2.1 (% rel.)

Since determinations are done in duplicate standard uncertainty reduces to

$$u(Rw) = \frac{2.1}{\sqrt{2}} = 1.5 \text{ (\% rel.)}$$

$$u(bias) = \sqrt{RMS_{bias}^2 + (u(C_{ref}))^2}$$

$$= \sqrt{3.6^2 + 1.2^2} = 3.8 \text{ (\% rel.)}$$

8.2.5 Step 5. Calculate combined measurement uncertainty uC

Standard uncertainties can be summed by taking the square root of the sum of the squares.

$$u_C = \sqrt{(u(R_W))^2 + (u(bias))^2} = \sqrt{1.5^2 + 3.8^2} = 4.1 \text{ (\% rel.)}$$

8.2.6 Step 6. Calculate expanded uncertainty

$$U = k * u_C$$

k = 2 (95% confidence level)

u_C = 4.1% (rel.)

U = 2 *4.1 = 8.2 (% rel.) (95% confidence level)

Supplement data:

No.	result	No.	result	No.	result	No.	result
1	65,2	**21**	63,8	**41**	66,2	**61**	68,6
2	65,7	**22**	63,9	**42**	66,9	**62**	66,5
3	65,6	**23**	63,1	**43**	67,5	**63**	66,1
4	67,1	**24**	64,1	**44**	67,4	**64**	65,6
5	65,8	**25**	63,9	**55**	69,2	**65**	68,8
6	66,5	**26**	63,7	**46**	66,8	**66**	66,4
7	65,3	**27**	63,3	**47**	65,8	**67**	65,4
8	65,3	**28**	63,6	**48**	66,7	**68**	65,7
9	64,8	**29**	64,1	**49**	65,7		
10	65,1	**30**	64	**50**	65,7		
11	66,4	**31**	65,6	**51**	67		
12	65	**32**	64,9	**52**	68,4		
13	65,3	**33**	67,6	**53**	66,1		
14	64,8	**34**	67,4	**54**	64,3		
15	64,6	**35**	65,2	**55**	67,2		
16	64,8	**36**	65	**56**	66,2		
17	64	**37**	64,9	**57**	64,7		
18	64,9	**38**	67,1	**58**	68,2		
19	64,2	**39**	65,4	**59**	67,2		
20	65,4	**40**	65,2	**60**	66,3		
Mean all (m)	65,7						
Std dev all (s)	1,4						
Rel. std dev	RSD= (s/m)*100 =1,4/65,7*100= 2,1						

Table 8.2-1: Data from mean value control charts.

Proficiency test: Quantification of MDMA in tablets and powders

Year	participants (n)	Assigned value (% base by weight) without outliers	RSD (%)	Lab result (%)	Bias (% rel.)	Bias`2
1	24	14,2	7,5	13,3	-6,3380	40,2
2	21	17,9	4,8	18,7	4,4693	20,0
3	23	29,3	2,6	28,8	-1,7065	2,9
4	23	23,3	7,9	24,2	3,8627	14,9
5	23	26,1	5,1	26,0	-0,3831	0,1
6	22	13,7	9,5	13,3	-2,9197	8,5
7	21	33,5	3,3	34,1	1,7910	3,2
Mean	**22,4**		**5,8**			
Sum Bias`2						**89,9**
RMS bias	**3,6**					

Table 8.2-2: Data from proficiency tests.

8.3 Quantitative determination of cocaine in seizures by HPLC DAD method

[2, 7, 9, 10-16]

8.3.1 General information

Method: Quantitative determination of cocaine in seizures by HPLC DAD method
Measurand: Cocaine seizures concentration
Units: % (g/100g)

8.3.2 Brief method description

HPLC-DAD is a major separation technique commonly used in forensic drug analysis. For ease of sample preparation, best reproducibility, and detectability, reversed phase chromatography is recommended for the analysis of cocaine. Dissolve an appropriate amount of standard or sample in ACN: aqueous buffer (80:20, v/v), targeting a concentration of the cocaine between 0.01-0.40 mg/ml (standard solutions) and use of peak area for HPLC-DAD quantitation. For linear regression, the ordinary least squares model is appropriate.

Seized samples were accurately weighted and directly diluted as above and then filtered through PVDF syringe filter before injection into the HPLC-DAD system.

8.3.3 Strategy of uncertainty calculation

The strategy for uncertainty calculation presented here is a top-down approach which uses information obtained from method validation with certified reference materials (CRM).

The top-down approach is an empirical approach which includes both imprecision and bias components of uncertainty. Uncertainties arising from random (within-laboratory precision) and systematic (bias) effects are treated alike.

Repeatability and intermediate reproducibility are terms included in within-laboratory precision. To assess repeatability, six CRM independent aliquots were analysed in the same day under the same conditions.

The estimation of the precision under intermediate conditions includes all causes of variation expected in the routine application of the analytical method: different days, different technicians, different glassware, different chemicals, different calibration curves. Validation experiments under intermediate conditions were performed in four non-consecutive days, performing six replicates each day. Bias is the difference between certified value and the average of the results obtained under intermediate conditions.

8.3.4 Uncertainty calculation

The steps needed for uncertainty calculation are:

8.3.4.1 *Step 1.- Data collection*

It is necessary to gather the following information regarding validation experiments:

- Certified reference materials: Certified value, uncertainty, and coverage factor (k)
- Number of replicates analysed in repeatability conditions (n)
- Number of replicates analysed in reproducibility conditions (intermediate precision) (N)
- Average of results in reproducibility conditions ($\bar{X}$)
- Standard deviation of repeatability (Sr)
- Standard deviation of reproducibility (SR)
- Bias (B)

In the example presented here, analytical method's data are:

Certified value (% w/w)	70,3
Uncertainty of certified value (% w/w)	0,60
Coverage factor (K)	2
Replicates in repeatability conditions (n)	6
Replicates in reproducibility conditions(N)	24
Average of results in reproducibility conditions ($\bar{X}$)	69,20
Standard deviation of repeatability (S_r) (% w/w)	0,83
Standard deviation of reproducibility (S_R) (% w/w)	1,58
Bias (B) (% w/w)	1,09

Table 8.3-1: Data from the analytical method.

8.3.4.2 *Step 2- Uncertainty estimation*

The combined standard uncertainty (u_c) is defined as an estimated standard deviation equal to the positive square root of the total variance obtained by combining all the uncertainty components. These components are the uncertainty associated to bias and the uncertainty related to method precision.

$$u_c = \sqrt{u_{bias}^2 + u_{precision}^2}$$

Bias uncertainty: It has been considered that bias uncertainty has two components: uncertainty of the reference value and uncertainty of the correction term

$$u_{bias} = \sqrt{u_{RV}^2 + u_{corr}^2}$$

- Uncertainty of reference value (uRV):

  Each CRM was provided with their corresponding certificate. The certificate reports the uncertainty associated to CRM certified value and the coverage factor (K=2 for 95% confidence). Equation applied is:

$$u_{RV} = \frac{U_{certificate}}{K}$$

  In the example:

$$u_{RV} = \frac{0{,}6}{2} = 0{,}3$$

- Uncertainty correction term:

  The uncertainty of this term is calculated considering that method bias reflects how far away method results are from the certified values. A rectangular distribution is assumed, and the following formula applied:

$$u_{corr} = \frac{Bias}{\sqrt{3}}$$

  In the example:

$$u_{corr} = \frac{1{,}09}{\sqrt{3}} = 0{,}63$$

Therefore, bias uncertainty is in the example:

$$u_{bias} = \sqrt{u_{RV}^2 + u_{corr}^2} = \sqrt{0{,}3^2 + 0{,}63^2} = 0{,}70$$

Precision uncertainty: The uncertainty related to method precision, has two components: repeatability and reproducibility uncertainties:

$$u_{precision} = \sqrt{u_{repeatability}^2 + u_{reproducibility}^2}$$

- Repeatability precision:

  The uncertainty is calculated using the repeatability standard deviation and the number of analyses performed in repeatability experiments using the formula:

$$u_r = \frac{S_r}{\sqrt{n}}$$

  In the example:

$$u_r = \frac{0{,}83}{\sqrt{6}} = 0{,}34$$

- Reproducibility precision:

  Uncertainty associated to reproducibility is a type A component and is calculated dividing the standard deviation for reproducibility by the square root of the total number of experiments performed in during validation:

$$u_R = \frac{S_R}{\sqrt{N}}$$

  In the example:

$$u_R = \frac{1{,}58}{\sqrt{24}} = 0{,}32$$

Therefore, precision uncertainty is in the example:

$$u_{precision} = \sqrt{u^2_{repeatability} + u^2_{reproducibility}} = \sqrt{0{,}34^2 + 0.32^2} = 0{,}47$$

Combined standard uncertainty: As previously stated, uncertainty contributions are summed up to a combined standard uncertainty:

$$u_c = \sqrt{u^2_{bias} + u^2_{precision}} = \sqrt{0{,}70^2 + 0{,}47^2} = 0{,}84$$

8.3.4.3 *Step 4.- Expanded uncertainty estimation*

Once combined uncertainty (u_c) is obtained, expanded uncertainty (U) is calculated as follows:

$$U = u_c \; x \; K$$

K is the coverage factor. In this case, to obtain the 95% of the confidence, K=2 is used.

In the example:

$$U = u_c \; x \; K = 0{,}84 \; x2 = 1{,}7 \text{ \%w/w}$$

Uncertainty is expressed in absolute form, if a relative form is needed it can be calculated as:

$$U(\%) = \frac{U}{\overline{X}} x \; 100 = \frac{1{,}7}{69{,}2} \; x \; 100 = 2{,}5\%$$

The contribution of uncertainty due to bias only has to be introduced if CI≥2. The following equation is used to determine u_{bias}:

$$u_{bias} = \frac{Bias}{\sqrt{3}} = \frac{V_c - V_M}{\sqrt{3}}$$

Effective degrees of freedom are calculated using Welch-Satterthwaite equation:

$$v_{eff} = \frac{u_c^4}{\sum_{i=1}^{n} \frac{u(x_i)^4}{v_i}}$$

Where $u(x_i)$ is each individual standard uncertainty and v_i is the degree of freedom of each standard uncertainty.

Expanded Uncertainty (U) is obtained by multiplying the combined standard uncertainty (u_c) by an appropriate coverage factor (k). The coverage factor is obtained from t distribution table with a level of confidence of 95%.

$U = u_C x\ k = u_c\ x\ t\alpha_{/2\upsilon}$

8.4 Determination of ethanol in blood using headspace gas chromatography with flame ionization detector (HS-GC-FID)

[2, 7, 9, 10-14, 17, 18]

8.4.1 General information

Method: Determination of ethyl alcohol in blood samples by gas chromatography with FID detector using headspace technique
Measurand: Blood alcohol concentration (BAC)
Units: g/L

8.4.2 Brief method description

This procedure outlines a simple method for the detection and quantitation of ethanol in blood samples by “Headspace” gas chromatographic procedure. Static headspace offers a reliable, simple, and accurate way to quantitate volatile compounds in a variety of liquid matrices. Practically, the sample and an internal standard are added to a vial and the vial is sealed. The sealed bottle is then placed into a heated sample carousel of the Headspace Analyzer. As temperature increases volatile compounds are released from the solution into the “headspace” above the liquid. The headspace is then sampled and analysed for the presence of the targeted analytes via gas chromatography.

8.4.3 Strategy of uncertainty calculation

The strategy for uncertainty calculation presented here is a top-down approach which uses information obtained from long-term participation in proficiency testing (PT) for method validation and uncertainty estimation. To apply this strategy is mandatory that:

- The test items in PT should be reasonably representative of the routine test items. For example, the type of material and range of values of the measurand should be appropriate
- The number of PT rounds is appropriate: a minimum of 6 different trials over an appropriate period is recommended to get a reliable estimate
- The assigned values, defined as the values attributed to a particular property of the proficiency test items, have an appropriate uncertainty. When the assigned value is calculated as a consensus value, the number of laboratories participating should be sufficient for reliable characterisation of the material
- Participation of the laboratory in the PT round was satisfactory (z-score≤2)

8.4.4 Uncertainty calculation

The steps needed for uncertainty calculation are:

8.4.4.1 *Step 1.- Data collection*

It is necessary to gather the following information:

- Number of PT items included in calculation (p)
- Assigned value of each PT item, which is used as a reference value (Yi)
- Assigned value's uncertainty (uAV) or assigned value's standard deviation (Syi)
- Number of participants laboratories (Ni): number of laboratories whose results were used to obtain assigned value (consensus value)
- Result obtained by the laboratory (X_i)

If sufficient data are available, different ranges of the property studied can be established. The minimum number of data needed to obtain statistically significant results is ten.
In the example presented here, the data collected (BAC= 0,5-1,0 g/L) through a period of 10 years, are:

Proficiency test item (i)	**N_i**	**Y_i**	**S_{Yi}**	**X_i**
1	48	0,63	0,06	0,68
2	58	0,95	0,05	0,98
3	47	0,76	0,05	0,78
4	54	0,58	0,02	0,58
5	44	0,55	0,06	0,52
6	43	0,56	0,05	0,57
7	55	0,54	0,03	0,57
8	46	0,57	0,05	0,58
9	56	0,56	0,03	0,56
10	46	0,75	0,06	0,75
11	55	0,72	0,04	0,68
12	47	0,86	0,07	0,87
13	49	0,8	0,09	0,77
14	51	0,5	0,04	0,49
15	49	0,65	0,07	0,63
16	53	0,8	0,05	0,79
17	47	0,72	0,05	0,75
18	41	0,58	0,05	0,59
19	44	0,68	0,05	0,68
20	46	0,91	0,09	0,91
21	51	0,51	0,04	0,50
22	62	0,64	0,07	0,64
23	55	0,78	0,06	0,79

Table 8.4-1: Data from Proficiency tests

8.4.4.2 *Step 2.- Bias and reproducibility calculation:*

Once all the data are collected, the absolute and relative differences between the result obtained by the laboratory (X_i) and the assigned value (Y_i) are calculated. The mean value of the results ($\bar{X}$), the sum of the absolute Di ($\sum|D_i|$), the sum of squares of the Di ($\sum D_i^2$) and the standard deviation of di (Sd) are then calculated.

Assigned value (Reference value)	Laboratory result	Differences (di)	Relative differences (Di)
Y_1	X_1	$d_1=Y_1-X_1$	$D_1= 100 \times d_1/ Y_1$
Y_2	X_2	$d_2=Y_2-X_2$	$D_2= 100 \times d_2/ Y_2$
…	…	..	…
Y_n	X_n	$d_n= Y_n-X_n$	$D_n= 100 \times d_n/ Y_n$

Table 8.4-2: Methodology used in the determination of bias and reproducibility

Therefore, in our example:

Proficiency test item (i)	Y_i	X_i	d_i	D_i (%)
1	0,63	0,68	-0,05	-7,94
2	0,95	0,98	-0,03	-3,16
3	0,76	0,78	-0,02	-2,63
4	0,58	0,58	0	0
5	0,55	0,52	0,03	5,45
6	0,56	0,57	-0,01	-1,79
7	0,54	0,57	-0,03	-5,56
8	0,57	0,58	-0,01	-1,75
9	0,56	0,56	0	0
10	0,75	0,75	0	0
11	0,72	0,68	0,04	5,56
12	0,86	0,87	-0,01	-1,16
13	0,80	0,77	0,03	3,75
14	0,5	0,49	0,01	2,00
15	0,65	0,63	0,02	3,08
16	0,80	0,79	0,01	1,25
17	0,72	0,75	-0,03	-4,17
18	0,58	0,59	-0,01	-1,72
19	0,68	0,68	0	0
20	0,91	0,91	0	0
21	0,51	0,50	0,01	1,96
22	0,64	0,64	0	0
23	0,78	0,79	-0,01	-1,28

$\bar{X}$	$\sum(\lvert D_i \rvert)$	$\sum(D_i - \bar{D})^2$	Sd

0,68	54,2	106,2	0,021

Table 8.4-3: Data used in the determination of bias and reproducibility

As stated before, when a laboratory participates satisfactorily in PT exercises for a long time, it is possible to use those results to perform method validation and obtain validation parameters such as bias and reproducibility.

Bias: Calculated as a quotient between the sum of absolute relative differences ($|D_i|$) and the number of PT rounds included (p).

$$B(\%) = \frac{\sum|D_i|}{p}$$

In our example:

$$B(\%) = \frac{54{,}2}{23} = 2{,}4\%$$

Reproducibility: Calculated as the standard deviation of relative differences (S_D):

$$S_D(\%) = \sqrt{\frac{\sum_{i=1}^{n_d}(\boldsymbol{D_i} - \bar{\boldsymbol{D}})^2}{p-1}}$$

In our example:

$$S_D(\%) = \sqrt{\frac{106{,}2}{23-1}} = 2{,}2\%$$

8.4.4.3 *Step 3.- Uncertainty estimation*

The combined standard uncertainty (uc) is defined as an estimated standard deviation equal to the positive square root of the total variance obtained by combining all the uncertainty components. These components are the uncertainty associated to bias and the uncertainty related to method precision.

$$u_c = \sqrt{u_{bias}^2 + u_{precision}^2}$$

Bias uncertainty: Using the laboratory's experience in PT, the uncertainty associated to bias can be estimated combining three components:

- Uncertainty related to assigned values of each PT item (Yi) This uncertainty is the uncertainty of the reference value (uRV).
- Uncertainty associated with the mean of the results obtained by the laboratory for the PT samples (uMV).
- Uncertainty associated with the difference (di) between laboratory's result and assigned value. This is the uncertainty associated to a correction term (u_D).

The formula applied is:

$$u_{bias} = \sqrt{u_{RV}^2 + u_{MV}^2 + u_D^2}$$

- Uncertainty of the reference value:

  In some cases, PT provider may report the uncertainty associated to the assigned value for each PT item as an expanded uncertainty, indicating the coverage factor used. In these cases, the standard uncertainty of each assigned value (u_{AVi}) is simply calculated by dividing expanded uncertainty by coverage factor.

  Once the standard uncertainty of each PT item is obtained, the uncertainty of the reference value is calculated as the quotient between the quadratic sum of the uncertainty of each PT item and the number of PT samples (p), as follows:

$$u_{RV,i} = \sqrt{\frac{\sum u_{AV,i}^2}{p}}$$

  However, it is possible that PT provider does not report uncertainties but reports the standard deviation of the assigned value for each PT item. In these cases, the standard uncertainty for each PT item is calculated dividing the standard deviation reported by the square root of the number of participating laboratories (N_i).

$$u_{AV,i} = \frac{S_{Y,i}}{\sqrt{Ni}}$$

  The standard deviation of the reference value is calculated as has been explained before.

  In our example, PT provider reported the standard deviation of the assigned value for each PT item. Therefore, the standard uncertainty for each PT item is calculated as follows:

Proficiency test item (i)	N	S_{Yi}	u_{VA}
1	48	0,06	0,009
2	58	0,05	0,007
3	47	0,05	0,007
4	54	0,02	0,003
5	44	0,06	0,009
6	43	0,05	0,008
7	55	0,03	0,004
8	46	0,05	0,007
9	56	0,03	0,004
10	46	0,06	0,009
11	55	0,04	0,005
12	47	0,07	0,010
13	49	0,09	0,013
14	51	0,04	0,006
15	49	0,07	0,010
16	53	0,05	0,007
17	47	0,05	0,007
18	41	0,05	0,008
19	44	0,05	0,008
20	46	0,09	0,013
21	51	0,04	0,006
22	62	0,07	0,009
23	55	0,06	0,008

Table 8.4-4: Data used in the determination of the uncertainty of the reference value.

The uncertainty of reference value obtained is:

$$u_{RV} = \sqrt{\frac{\sum u_{AV}^2}{p}} = \sqrt{\frac{0{,}015}{23}} = 0{,}008$$

- Uncertainty of the mean value

  The uncertainty of the mean value is calculated using the standard deviation of the differences (Sd) and the number of PT items analysed (p):

$$u_{MV} = \frac{S_d}{\sqrt{p}}$$

  In our example:

$$u_{MV} = \frac{S_d}{\sqrt{p}} = \frac{0{,}021}{\sqrt{23}} = 0{,}045$$

- Uncertainty of the correction term:

The differences (d_i) between a laboratory's results and the assigned values reflects how far away the laboratory results are from the "true values". Therefore, they can be used to calculate the uncertainty due to the correction term.

The differences are estimates made in the form of maximum range(±d), being the shape of its distribution unknown. Therefore, a rectangular distribution is assumed, and the following formula is applied:

$$u_d = \frac{d}{\sqrt{3}}$$

Method bias has been previously calculated as a relative form (B%), therefore, the uncertainty associated with the differences can be expressed as:

$$u_d = \frac{B(\%) \times \bar{X}}{100\ x\sqrt{3}}$$

Being $\bar{X}$ the mean value of our results (X_i).

In our example:

B (%)	$\bar{X}$	**u_D**
2,4	0,68	0,0094

Therefore, in our example, the uncertainty associated to the bias is:

$$u_{bias} = \sqrt{u_{RV}^2 + u_{MV}^2 + u_d^2} = \sqrt{0{,}008^2 + 0{,}045^2 + 0{,}009^2} = 0{,}013$$

Precision uncertainty (reproducibility within laboratory): As previously stated, in the present strategy, reproducibility within laboratory has been calculated from PT results. Uncertainty associated to the reproducibility (u_{RW}) is calculated using the formula:

$$u_{precision} = u_{RW} = \frac{S_D(\%)\ x\ \bar{X}}{100\ x\ \sqrt{n}}$$

Where $S_D(\%)$ is the method reproducibility, $\bar{X}$ is the mean value of the laboratory's result and n is the number of analyses performed in each PT sample.

In our example, as each sample was analysed in duplicate, n is 2.

S_D **(%)**	$\bar{X}$	**u_{RW}**
2,2	0,68	0,011

Therefore, the combined standard uncertainty is:

$$u_c = \sqrt{u_{bias}^2 + u_{precision}^2} = \sqrt{0{,}013^2 + 0{,}011^2} = 0{,}02$$

8.4.4.4 *Step 4.- Expanded uncertainty estimation*

Once combined uncertainty (u_c) is obtained, expanded uncertainty (U) is calculated as follows:

$$U = u_c \; x \; K$$

Where K is the coverage factor. In this case, to obtain the 95% of the confidence, K=2 is used.

Therefore,

$$U_c = u_c \; x \; K = 0{,}02 \; x \; 2 = 0{,}04 \; \mathrm{g/L}$$

Uncertainty is expressed in absolute form, if a relative form is needed it can be calculated as:

$$U_c(\%) = \frac{U}{\bar{X}} x \; 100 = \frac{0{,}04}{0{,}68} \; x \; 100 = 5{,}4\%$$

8.5 Quantifying Delta-9-Tetrahydrocannabinol (THC) in Blood example and MUCalc

8.5.1 Introduction

A *bottom-up* approach was used to estimate the measurement uncertainty of Delta-9-Tetrahydrocannabinol (THC) in Blood (see Chapter 6). The methodology described here underpins the development of a Measurement Uncertainty Calculator MUCalc (https://uod.ac.uk/lrcfsmucalc) by the Leverhulme Research Centre for Forensic Science (LRCFS). MUCalc is an open access white-box software calculator developed using the Shiny package in R. It is user friendly and displays in detail, all the methods and formulas used in the calculation for easy understanding and verification of results.

By white-box, it provides a more detailed methodical analysis with a transparent step-by-step calculation of how each uncertainty component is estimated in a more user-friendly easy to follow approach. These formulas and calculations are displayed on screen and in a summary report, generated by software, making it easy for users to understand and cross examine every result generated by MUCalc. It also provides reference links to published articles to assist users make informed parameter choices.

The current version of the software is suitable for toxicological samples (e.g. blood, urine) and seized drug samples (e.g. drug tablets, leaves, powder). Uncertainty components are calculated in accordance with the standards of International Organization for Standardization ISO/IEC 17025 [1] for

1. Homogeneity
2. Calibration Curve
3. Method Precision
4. Calibration Standard
5. Sample Preparation

Each of these uncertainty sources/components is explained in detail in annex II. In addition to calculating Uncertainty of Homogeneity, a Homogeneity Test is carried out to test whether there is a statistically significant difference between group means of samples using a one-way analysis of variance (ANOVA).

MUCalc offers the choice to fit a linear regression or a quadratic regression to a calibration curve data with the option to specify weights if weighted least square regression is desired.

Each uncertainty component is calculated separately and then combined to derive the Combined Uncertainty. If data is uploaded for all components, the Combined Uncertainty is calculated using all components. An uncertainty component can be excluded from the Combined Uncertainty by simply not uploading any data for that component. The Combined Uncertainty is multiplied by a Coverage Factor to derive an Expanded Uncertainty.

MUCalc has the option to calculate a Coverage Factor when a confidence level is specified or allows one to specify a Coverage Factor directly. Where a confidence level is specified, MUCalc calculates an Effective Degrees of Freedom using the Welch-Satterthwaite equation. The derived Effective Degrees of Freedom along with the specified confidence level is used to read a value termed Coverage Factor, from a T-Distribution table.

MUCalc summarises all results in a single tab and can also generate a pdf report giving in detail, data supplied and calculations performed by the software. The development of MUCalc is ongoing, detailed information on the current version is available at

https://doi.org/10.5281/zenodo.3944694 and a live version available at https://uod.ac.uk/lrcfsmucalc. See specifications of MUcalc (Annex II). The calculations described in this THC example below can be done using MUCalc, and are published in Klu et al.(2021) [19].

8.5.2 Strategy of uncertainty calculation

8.5.2.1 *Step 1 - Specifying the Measurand*

The measurand is the concentration of the THC analyte (µg/L) in a blood sample expressed using the relationship:

$$x_{\mathrm{THC}} = \frac{x_{cs}}{V} \times \ f_{\mathrm{precision}} \ (\mu\mathrm{g/L}) \qquad (1)$$

where x_cs is the amount of THC in the case sample, V is the volume of the case sample and $f_{precision}$ is the correction factor for method precision.

8.5.2.2 *Step 2 - Identify the Sources of Uncertainty*

With reference to Equation (1), the sources of uncertainties associated with quantifying THC in blood are identified using the cause and effect diagram displayed in Figure 8.5-1. The main uncertainty sources are from method precision, sample volume, calibration curve and the preparation of calibration standards.

In the next sections, each of these uncertainty sources are quantified in detail and combined to obtain an overall measure of uncertainty.

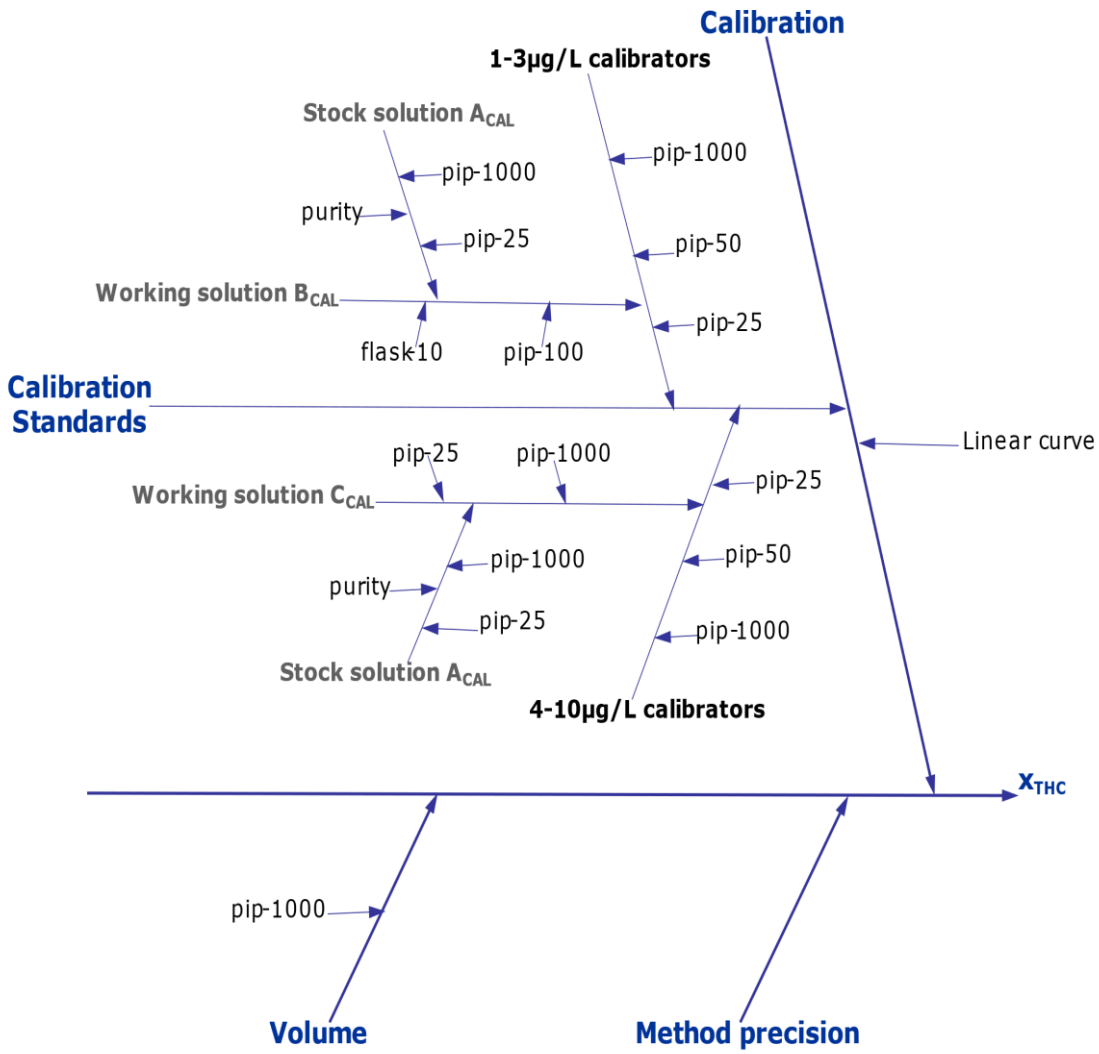

Figure 8.5-1: Cause and effect diagram for identifying the sources of uncertainty in quantifying THC in blood. A 25 µL pipette is denoted with pip-25 and a 10 mL volumetric flask is denoted with flask-10.

8.5.2.3 *Step 3 - Quantifying Uncertainty Sources*

For simplicity, the uncertainty associated with the preparation of calibration standards was calculated separately from that of the calibration curve.

Uncertainty of the Calibration Standards

The uncertainty associated with the calibration standards combines the uncertainties stated on the Certificates of Analysis of the Certified Reference Materials (CRMs) and the inaccuracies of all measuring equipment (e.g. pipettes and volumetric flasks) used to dilute the CRMs and spike blank blood samples when preparing a calibration curve. The structure of how the THC CRM was diluted to make other solutions in preparing the calibration curve is displayed in Figure 8.5-2.

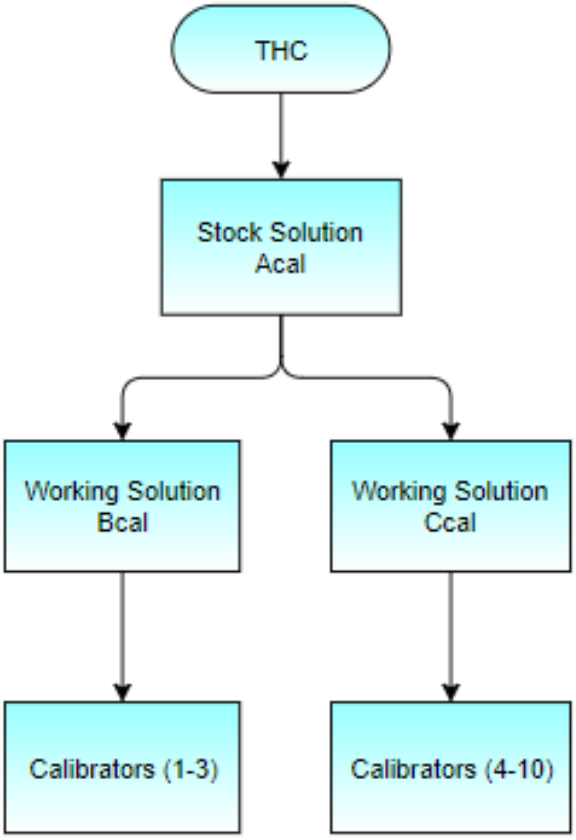

Figure 8.5-2: The structure of THC dilution process for the preparation of the calibration curve.

The volume, tolerance, and coverage factor (k) of pipettes and volumetric flasks used as given in the manufacturer's reference material, are given in Table 8.5-1, along with the number of times each pipette and volumetric flask was used in the preparation process for each standard solution.

The standard uncertainty (u) of THC as well as that of pipettes and volumetric flasks is given by $u = \mathrm{Tolerance}/k$, and the relative standard uncertainty (RSU) is given by $u = u/\mathrm{Volume}$. These are summarised in Table 8.5-2, together with the uncertainty associated with the preparation of calibration standards.

THC CRM

	Purity (mg/mL)	Tolerance (mg/mL)	Coverage factor(*k*)
THC	1	0.033	2

Solutions

	Pipette/ Flask	Volume µL	Tolerance µL	Coverage factor(*k*)	Times used
Stock Solution A_{CAL}					
	pip-25	25	0.30	2	1
	pip-1000	1000	5	2	2
Working Solution B_{CAL}					
	pip-50	50	0.30	2	1
	pip-1000	1000	5	2	7
Working Solution C_{CAL}					
	pip-100	100	0.3	2	1
	flask-10	10000	25	√33	1
Calibration Standards					
1-3 µg/L (Cal_{1-3})	pip-25	25	0.30	2	9
	pip-50	50	0.30	2	1
	pip-1000	1000	5	2	5
4-10 µg/L (Cal_{4-10})	pip-25	25	0.30	2	7
	pip-50	50	0.30	2	3
	pip-1000	1000	5	2	5

Table 8.5-1: Data on THC CRM purity, pipette and flask used for solutions preparation.

RSU of THC CRM, volumetric flasks and pipettes

$u(\mathbf{Purity}) = \frac{0.033}{2} = \mathbf{0.0165},$	$u_r(\mathbf{Purity}) = \frac{0.0165}{1} = \mathbf{0.0165}$
u(pip-25)$= \frac{0.30}{2} = \mathbf{0.15},$	$u_r(\mathbf{pip-25}) = \frac{0.15}{25} = \mathbf{0.006}$
u(pip-50)$= \frac{0.30}{2} = \mathbf{0.15},$	u_r(pip-50) $= \frac{0.15}{50} = \mathbf{0.003}$
(pip-100)$= \frac{0.3}{2} = \mathbf{0.15},$	u_r(pip-100)$= \frac{0.15}{100} = \mathbf{0.0015}$
u(pip-1000) $= \frac{5}{2} = \mathbf{2.5},$	u_r(pip-1000) $= \mathbf{2.5/1000} = \mathbf{0.0025}$
u(flask-10) $= \frac{25}{\sqrt{3}} = \mathbf{14.43376},$	u_r(flask-10)$= \frac{14.43376}{10000} = \mathbf{0.00144}$

RSU of working standard solution

$$u_r(\mathbf{A_{CAL}}) = \sqrt{u_r(\mathbf{Purity})^2 + u_r(\mathbf{pip-25})^2 + 2 \times u_r(\mathbf{pip-1000})^2}$$

$$= \sqrt{0.0165^2 + 0.006^2 + 2 \times 0.0025^2}$$

$$= \mathbf{0.01791}$$

$$u_r(\mathbf{B_{CAL}}) = \sqrt{u_r(\mathbf{A_{CAL}})^2 + u_r(\mathbf{pip-50})^2 + 7\ ur(\mathbf{pip-1000})^2}$$

$$= \sqrt{0.01791^2 + 0.003^2 + 7 \times 0.0025^2}$$

$$= \mathbf{0.019327}$$

$$u_r(\mathbf{C_{CAL}}) = \sqrt{u_r(\mathbf{A_{CAL}})^2 + u_r(\mathbf{pip-100})^2 + u_r(\mathbf{flask-10})^2}$$

$$= \sqrt{0.01791^2 + 0.0015^2 + 0.00144^2}$$

$$= \mathbf{0.018}$$

RSU of calibration standards 1-3 µg/L and 4-10 µg/L

$$u_r(\mathbf{Cal_{1-3}}) = \sqrt{u_r(\mathbf{B_{CAL}})^2 + 9 \times u_r(\mathbf{pip-25})^2 + u_r(\mathbf{pip-50})^2 + 5 \times u_r(\mathbf{pip-1000})^2}$$

$$= \sqrt{0.019327^2 + 9 \times 0.006^2 + 0.003^2 + 5 \times 0.0025^2}$$

$$= 0.02716$$

$$u_r(\mathbf{Cal_{4-10}})$$
$$= \sqrt{u_r(\mathbf{C_{CAL}})^2 + 7 \times u_r(\mathbf{pip-25})^2 + 3 \times u_r(\mathbf{pip-50})^2 + 5 \times u_r(\mathbf{pip-1000})^2}$$

$$= \sqrt{0.018^2 + 7 \times 0.006^2 + 3 \times 0.003^2 + 5 \times 0.0025^2}$$

$$= \mathbf{0.0252}$$

Table 8.5-2: Calculations of RSU for THC CRM, volumetric flasks, pipettes, stock and working solutions used for calibration standards 1-3 µg/L and 4- 10 µg/L.

The RSU associated with the preparation of calibration standards was obtained by combining the RSU of $u_r(\mathrm{Cal}_{1-3})$ and $u_r\ (\mathrm{Cal}_{4-10})$ as:

$$u_r(\mathrm{CalStd}) = \sqrt{u_r(\mathrm{Cal}_{1-3})^2 + u_r(\mathrm{Cal}_{4-10})^2}$$

$$= \sqrt{0.02716^2 + 0.0252^2}$$

$$= 0.0371$$

Uncertainty of the Calibration Curve

The uncertainty associated with the fitted calibration curve is estimated using the error propagation formula:

$$u(\mathrm{CCur}) = \frac{S_{y/x}}{b_1}\sqrt{\frac{1}{r_{cs}} + \frac{1}{n} + \frac{(x_{cs} - \bar{x})^2}{S_{xx}}}, \quad (2)$$

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}; \quad (3)$$

Where:

$\boldsymbol{S_{y/x}}$) Is the residual or standard error of regressing *y* on *x*
$\boldsymbol{b\_1}$ is the slope of the regression line
$\boldsymbol{r\_cs}$ is the number of replicates made on the case sample to determine $\boldsymbol{x\_cs}$
$\boldsymbol{n}$ is the number of measurements used to generate the calibration curve
$\boldsymbol{x\_cs}$ is the mean amount of THC in the case sample
$(\boldsymbol{x})^{-}$ is the mean value of the different calibration standards
$\boldsymbol{x\_i}$ is the target calibrator concentration at the $\boldsymbol{i}$th level
$\boldsymbol{S\_xx}$ is the sum of squares deviation of $\boldsymbol{x}$ given by $\sum\_\boldsymbol{i}\blacksquare(\boldsymbol{x\_i} - \boldsymbol{x}^{-})\text{^}\boldsymbol{2}$

The relative standard uncertainty is given by

$$u_r\ (\mathrm{CCur}) = \frac{u(\mathrm{CCur})}{x_{cs}} \quad (4)$$

Consider the calibration curve data of peak area ratios for 10 concentration levels {1, 1.5, 2, 2.5, 3, 4, 5, 6, 8 and 10} µg/L given in Table 8.5-3 along with the coefficients of the linear regression and the sum of squared deviations.

Concentration (x)	Peak Area Ratios (y)	$(x - \bar{x})^2$	$\hat{y} = b_0 + b_1 x$	$(y - \hat{y})^2$
1	0.50936	10.89000	0.46247	0.00220
1.5	0.73972	7.84000	0.72863	0.00012
2	1.00815	5.29000	0.99479	0.00018
2.5	1.24273	3.24000	1.26095	0.00033
3	1.53580	1.69000	1.52711	0.00008
4	2.09479	0.09000	2.05943	0.00125
5	2.50074	0.49000	2.59175	0.00828
6	3.06545	2.89000	3.12407	0.00344
8	4.15375	13.69000	4.18871	0.00122
10	5.34078	32.49000	5.25336	0.00764
$\bar{x}$		$S_{xx} = \sum(x - \bar{x})^2$		$\sum(y - \hat{y})^2$
4.3		78.6		0.02474
		Intercept b_0	-0.06985	
		Slope b1	0.53232	
		R2	0.9989	
		n	10	

Table 8.5-3: 10 concentration levels versus Peak area ratios, linear regression coefficients and the sum of squares of regression for the calibration curve data.

The standard error of regression can be computed using Equation (3) and values from Table 8.5-3 as

$$S_{y/x} = \sqrt{\frac{0.02474}{10 - 2}}$$
$$= 0.05561$$

From the calibration curve data in Table 8.5-3, at each calibration level, one replicate is analysed for generating the calibration curve according to laboratory protocol. To obtain a more reliable estimate of the standard error, standard errors from previous calibration curve data can be pooled. Pooling the errors gives a better estimate for the standard error of regression by taking into account different laboratory conditions over different days. The standard errors of a further 10 calibration curve data sets is summarised in Table 8.5-4, and the pooled standard error of regression, $S_{p_{(y/x)}}$, is calculated using Equation (5) as:

$$S_{p_{(y/x)}} = \sqrt{\frac{\sum(n-1)S_{y/x}^2}{\sum(\mathrm{n}-1)}}$$
$$= \sqrt{\frac{0.38558}{98}} \quad (5)$$
$$= 0.06273$$

Substituting the pooled standard error, $S_{p_{(y/x)}}$, as an as an estimate for $S_{y/x}$, the uncertainty of the calibration curve from Equation (2) becomes

$$u(\mathrm{CCur}) = \frac{S_{p_{(y/x)}}}{b_1}\sqrt{\frac{1}{r_{cs}} + \frac{1}{n} + \frac{(x_{cs} - \bar{x})^2}{S_{xx}}},$$

$$= \frac{0.06273}{0.53232}\sqrt{\frac{1}{2} + \frac{1}{10} + \frac{(2 - 4.3)^2}{78.6}},$$

$$= 0.09626$$

For a given case sample, two replicates are taken $r_{cs} = 2$ and the average reported. For an average concentration reading of $x_{cs} = 2$ µg/L, the relative standard uncertainty of the calibration curve using Equation (4) is given by

$$u(\mathrm{CCur}) = \frac{0.09626}{2}$$

$$= 0.04813$$

$\boldsymbol{n}$	$\boldsymbol{n-1}$	$S_{y/x}$	$\boldsymbol{(n-1)S_{y/x}^2}$
10	9	0.05561	0.02784
0	9	0.05127	0.02366
10	9	0.03796	0.01297
10	9	0.07499	0.05061
10	9	0.04149	0.01549
10	9	0.04626	0.01926
10	9	0.05563	0.02786
10	9	0.04353	0.01705
10	9	0.11674	0.12265
10	9	0.04294	0.01660
9	8	0.08031	0.05160
	$\sum(\boldsymbol{n-1})$		$\boldsymbol{\sum(n-1)S_{y/x}^2}$
	98		0.38558

Table 8.5-4: The standard error and sum of squares deviation of 11 different calibration curves.

Uncertainty of the Method Precision

The quality control (QC) data for evaluating the uncertainty of the method precision is summarised in Table 8.5-5. Blank blood samples were spiked with THC at three concentration

levels: 2 µg/L (low), 5 µg/L (medium), and 10 µg/L (high). For each concentration level, three replicates were analysed over eleven separate days using a freshly prepared calibration line each day.

The uncertainty associated with the method precision *u*(Precision) is estimated for each concentration level 2µg/L (Low), 5 µg/L (Medium) and 10 µg/L (High) using a pooled standard deviation (S_p) approach given by

$$u(\text{Precision}) = \frac{S_p}{\sqrt{r_{cs}}} \tag{9}$$

where S_p, similar to Equation (5) is given by

$$S_P = \sqrt{\frac{\sum_i (v_i \times S_i^2)}{\sum_i v_i}},$$

v_i is the degrees of freedom of the ith sample, S_i is standard deviation of the ith sample and

Concentration	**Peak Area Ratios**										
(g/L)	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 7**	**Run 8**	**Run 9**	**Run 10**	**Run 11**
Low											
	2.198	1.825	2.144	2.108	2.065	1.810	1.993	1.829	1.786	2.044	1.851
2	1.988	1.920	2.166	2.052	2.002	1.806	1.942	1.768	1.880	1.810	1.822
	2.161	1.851	2.182	1.972	2.152	1.795	1.931	1.826	1.785	1.896	1.701
Mean	2.11567	1.86533	2.16400	2.04400	2.07300	1.80367	1.95533	1.80767	1.81700	1.91667	1.79133
Std. Dev	0.11210	0.04910	0.01908	0.06835	0.07532	0.00777	0.03308	0.03439	0.05456	0.11836	0.07956
Medium											
	4.885	5.067	4.893	4.986	4.884	4.377	4.969	4.475	4.801	4.731	4.405
5	4.869	5.266	5.037	4.906	4.913	4.672	4.641	4.549	4.535	4.718	4.472
	4.806	5.086	5.141	4.867	4.863	4.684	4.737	4.388	4.611	4.709	4.402
Mean	4.85333	5.13967	5.02367	4.91967	4.88667	4.57767	4.78233	4.47067	4.64900	4.71933	4.42633
Std. Dev	0.04177	0.10982	0.12454	0.06067	0.02511	0.17389	0.16863	0.08059	0.13701	0.01106	0.03958
High											
	9.952	9.945	9.851	10.306	10.054	9.219	9.493	9.732	9.327		8.609
10	9.910	10.235	9.940	10.299	9.616	9.249	9.091	9.322	8.988	10.972	
	10.002	9.941	9.740	10.840	10.473	9.275	9.225	9.224	9.255	11.199	8.936
Mean	9.95467	10.04033	9.84367	10.48167	10.04767	9.24767	9.26967	9.42600	9.19000	11.08550	8.77250
Std. Dev	0.0466	0.16860	0.10020	0.31035	0.42854	0.02802	0.20469	0.26950	0.17860	0.16051	0.23122

r_cs is the number of case sample replicates.

The relative standard uncertainty of the method precision, $u\_r\,(\text{Precision})$, is calculated by dividing the standard uncertainty by its nominal value (NV) or by the mean concentration of replicates on NV ($\bar{x}\_NV$).

Table 8.5-5: Quality control data for concentration levels 2 µg/L (Low), 5 µg/L (Medium) and 10 µg/L (High) over 11 different days with three replicates each concentration level.

$$u\_r\ (\text{Precision}) = \frac{u(\text{Precision})}{\text{NV}},$$

Concen-tration	Nominal Value (NV) µg/L	Standard. Deviation S	Degrees of Freedom ν	$q = S^2 \times \nu$	Pooled S $S_p = \sqrt{\sum q / \sum \nu}$	Case Sample Replicate r_{cs}	Standard Uncertainty (SU) $u = S_p / r_{cs}$	Relative SU $u_r = u/\text{NV}$
Low	2	0.11210	2	0.02513	0.06832	2	0.04831	0.02415
		0.04910	2	0.00482				
		0.01908	2	0.00073				
		0.06835	2	0.00934				
		0.07532	2	0.01135				
		0.00777	2	0.00012				
		0.03308	2	0.00219				
		0.03439	2	0.00236				
		0.05456	2	0.00595				
		0.11836	2	0.02802				
		0.07956	2	0.01266				
			$\sum \nu = 22$	$\sum q = 0.10268$				
Medium	5	0.04177	2	0.00349	0.10412	2	0.07362	0.01472
		0.10982	2	0.02412				
		0.12454	2	0.03102				
		0.06067	2	0.00736				
		0.02511	2	0.00126				
		0.17389	2	0.06047				
		0.16863	2	0.05687				
		0.08059	2	0.01299				
		0.13701	2	0.03754				
		0.01106	2	0.00024				
		0.03958	2	0.00313				
			$\sum \nu = 22$	$\sum q = 0.23851$				
High	10	0.04606	2	0.00424	0.22525	2	0.15927	0.01593
		0.16860	2	0.05685				
		0.10020	2	0.02008				
		0.31035	2	0.19263				
		0.42854	2	0.36728				
		0.02802	2	0.00157				
		0.20469	2	0.08379				
		0.26950	2	0.14526				
		0.17860	2	0.06380				
		0.16051	1	0.02576				
		0.23122	1	0.05346				
			$\sum \nu = 20$	$\sum q = 1.01474$				

Table 8.5-6: Uncertainty of the method precision calculation for concentration levels 2 µg/L (low), 5 µg/L (medium), and 10 µg/L (high).

The calculations for the uncertainty of method precision are detailed in Table 8.5-6 for each concentration level. From Table 8.5-6, the RSU of method precision for a given x_cs is the value

with the closet nominal value (NV) to x_cs. Hence, for $x\_cs = 2$, the closet nominal value is $\mathrm{NV} = 2$ and the uncertainty

$$u_r(\text{Precision}) = 0.02415.$$

Uncertainty of the Sample Volume

The RSU of the volume $u(V)$ of case blood sample is equivalent to the uncertainty of the pipette used which is the pipette pip-1000. From Tables 8.5-1 & 8.5-2, the pipette pip-1000 has volume 1000 L with a tolerance of 5 L and a reference certificate coverage factor of 2.

$$u(V) = 5/2 = 2.5\ \mu\text{L}$$

$$u_r(V) = \frac{u(V)}{V} = \frac{2.5}{1000} = 0.0025$$

8.5.2.4 *Step 4 - Combined and Expanded Uncertainty*

Calculating the Combined Uncertainty

The concentration of THC in case sample from Equation (1) is $x_{THC} = 2/1 = 2$ µg/L. The combined uncertainty, u_c, is obtained by combing all the individual uncertainty components as follows:

$$\frac{u_c}{x_{THC}} = \sqrt{u_r(\text{Precision})^2 + u_r(\text{CalStd})^2 + u_r(\text{CCur})^2 + u_r(V)^2}$$

Hence,

$$u_c = x_{THC} \times \sqrt{u_r(\text{Precision})^2 + u_r(\text{CalStd})^2 + u_r(\text{CCur})^2 + u_r(V)^2}$$

$$= 2 \times \sqrt{0{:}02415^2 + 0{:}0371^2 + 0{:}04813^2 + 0{:}0025^2}$$

$$= 0.131\ \mu\text{g/L}$$

The Effective Degrees of Freedom and Coverage Factor

To obtain a suitable coverage factor k, the effective degrees of freedom ν_{eff} is calculated using the Welch-Satterthwaite equation generally defined as:

$$V_{\text{eff}} = \frac{u_c^4}{\sum_l u_{(l)}^4/\nu_l} \qquad (6)$$

where u_c is the combined uncertainty, $u_{(l)}$ is the individual standard uncertainty component l combined to obtain u_c, and ν_l is the degrees of freedom for each uncertainty component l. With the use of relative standard uncertainties for the combined uncertainty, Equation (6) becomes

$$V_{\text{eff}} = \frac{\left[\frac{u_c}{x_{\text{THC}}}\right]^4}{\sum_l u_{r(l)}^4/\nu_l}$$

$$= \frac{\left[\frac{u_c}{x_{\mathrm{THC}}}\right]^4}{\frac{u_r(\mathrm{Precision})^4}{v_{(\mathrm{Precision})}} + \frac{u_r(\mathrm{CalStd})^4}{v_{(\mathrm{CalStd})}} + \frac{u_r(\mathrm{CCur})^4}{v_{(\mathrm{CCur})}} + \frac{u_r(\mathrm{V})^4}{v_{(\mathrm{V})}}}$$

$$= \frac{\left[\frac{0.131}{2}\right]^4}{\frac{0.02415^4}{22} + \frac{0.0371^4}{\infty} + \frac{0.04813^4}{8} + \frac{0.0025^4}{\infty}}$$

$$= 26.8$$

The degrees of freedom for the preparation of calibration standards and sample volume are unknown and therefore $v_{(\mathrm{CalStd})} = v_{(V)} = \infty$. From the t-distribution table with a 99.7% confidence interval, a coverage factor of $k_{V_{\mathrm{eff}},99.7\%} = 3$ is chosen for calculating the expanded uncertainty.

Calculating the Expanded Uncertainty

Finally, the expanded uncertainty is obtained by multiplying the coverage factor by the combined uncertainty:

$$U = k \times u_c$$

$$= 3 \times 0.131$$

$$= 0.393 \mu g/L$$

The concentration of THC in the case sample is given by 2± 0.393 µg/L.

8.6 Predicting net weights of khat mardoufs

Khat is a flowering plant (*Catha Edulis*) growing naturally in East Africa and on the Arabic peninsula. It is consumed as a drug (chewing), since it contains cathinone (an alkaloid), and is classified as a drug of abuse by WHO, however not considered to be a seriously dangerous drug. In Sweden, khat is classified as an illicit drug, hence it is illegal to import and have in possession.

Khat is prepared in bundles of suitable size for chewing. In the distribution of the drug, several bundles are packed into a so-called *mardoufs*. When smuggled, sacs with about 100 mardoufs are shipped.

To avoid drying of the material, a mardouf is wrapped in banana leaves. When customs (or police authorities) should deem upon the amount of khat that is seized, there is a tedious work of removing all banana leaves before the material can be weighed. Therefore, a model for estimating the weights of the wrappings from the gross weight of the seizure to approximate the net weight has been developed (integrative approach, see Chapter 6).

8.6.1 Linear regression model

In a study at NFC of a total of 260 seized mardoufs their gross weights and wrapping weights of the covers were measured. A scatter plot is shown in Figure 8.6-1a.

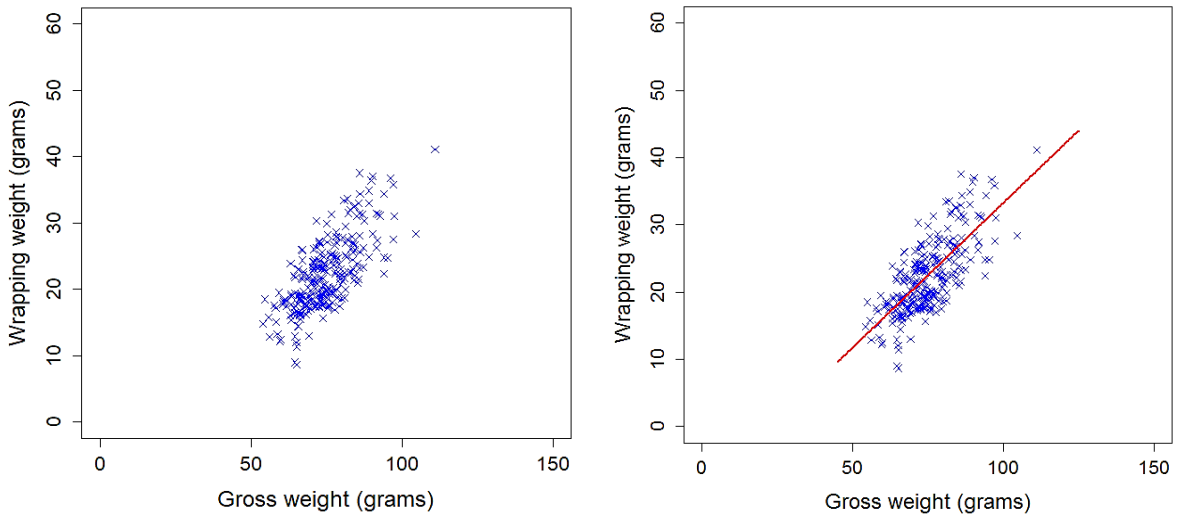

Figure 8.6-1: Wrapping weight plotted against gross weight for 260 seized mardoufs of khat (a); and a least-squares fitted regression line to the points (b).

The relationship between wrapping weight and gross weight seems to be quite linear, although there is quite an amount of variation not explained by a linear relationship. A robust model would therefore be a linear regression model. Let *ww* denote the wrapping weight of a and *gw* denote the gross weight of a randomly selected mardouf. The linear regression model is then expressed as

$$ww = a + b \cdot gw + e$$

where *b* is the slope of the line (of the theoretical linear relationship), *a* is the intercept (where the theoretical line crosses the vertical axis, and *e* stands for the deviation from the line (explaining why not all points lie strictly on a line – usually referred to as the error term).). For the subsequent analysis, we assume that for different pairs of ww and gw the corresponding deviations (the *e*'s) are independent and identically distributed with zero mean and constant variance, σ_e^2 (hence also independent of gw).

8.6.2 Prediction of net weight and its uncertainty

Now, once parameter estimates have been obtained, we can predict the *net weight*, $nw = gw - ww$ of a newly seized mardouf with gross weight gw_0 as

$$\widehat{nw}_0 = gw_0 - \widehat{ww}_0 = gw_0 - \left(\hat{a} + \hat{b} \cdot gw_0\right) = \left(1 - \hat{b}\right) \cdot gw_0 - \hat{a}$$

Moreover, $\hat{a}$ and $\hat{b}$ are unbiased estimates of *a* and *b* respectively and hence

$$E(\widehat{nw}_0) = (1 - b) \cdot E(gw_0) - a$$

Thus, a potential bias in this prediction is due to whether there is a bias in the measured gross weight of the new mardouf.

The uncertainty of $\widehat{nw}_0$ will however depend both on the uncertainty of the parameter estimates and the uncertainty of the measured gross weight of the new mardouf. For the deduction of the uncertainty, we use the simplification that the uncertainty of the parameter estimates stems from the uncertainty of the measured wrapping weights $ww_1, ww_2, \ldots, ww_n$ *only* (*n* = 260 in the example). This is how inference for linear regression is regularly pursued – it is a model combining the conditional mean of the wrapping weight given a gross weight and the "random" deviation from that mean.

We could consider contribution from the uncertainty of the measured gross weights $gw_1, gw_2, \ldots, gw_n$, but consequently with much more involved deduction of the uncertainty of $\widehat{nw}_0$. The scatterplot of wrapping weights against gross weights in Figure 8.6-1b reveals that the deviations from an assumed underlying straight line are far from being explained by measurement error only – the banana leaves used cannot be tailored to the mardouf they should cover, while the uncertainties of the measured gross weights would simply be due to measurement error.

Hence, the difference in deduced uncertainty for the expression $\hat{a} + \hat{b} \cdot gw$ between using and not using uncertainty in the measured gross weights can be considered negligible compared to the level of uncertainty coming primarily from the parameter estimates $\hat{a}$ and $\hat{b}$. Notwithstanding, we will include the measurement uncertainty of the gross weight of the new mardouf even though its contribution is expected to be small.

An estimate of the variance of $\widehat{nw_0}$ can be deduced to (using variance calculations from least-squares fitting of a regression model):

$$\widehat{Var(nw_0)} =$$

$$= \frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(gw_i - \overline{gw})^2}\left(\hat{\sigma}_{gw_0}^2 + \left(gw_0 + \widehat{bias_0}\right)^2 + (\overline{gw})^2 - 2 \cdot \overline{gw} \cdot \left(gw_0 + \widehat{bias_0}\right)\right)$$

$$+\left(1 - \hat{b}\right)^2 \cdot \hat{\sigma}_{gw_0}^2 + \frac{\hat{\sigma}_e^2}{n}$$

where $\hat{\sigma}_{gw_0}^2$ and $\widehat{bias_0}$ are taken from an analysis of the uncertainty of the measurement gw_o, and $\hat{\sigma}_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(ww_i - \hat{a} - \hat{b} \cdot gw_i\right)^2$ (usually referred to as the mean square sum of residuals or errors, *SSE*). The expression for the standard uncertainty then becomes

$$u = \sqrt{\frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(gw_i - \overline{gw})^2}\left(\hat{\sigma}_{gw_0}^2 + \left(gw_0 + \widehat{bias_0}\right)^2 + (\overline{gw})^2 - 2 \cdot \overline{gw} \cdot \left(gw_0 + \widehat{bias_0}\right)\right) + \left(1 - \hat{b}\right)^2 \cdot \hat{\sigma}_{gw_0}^2 + \frac{\hat{\sigma}_e^2}{n}}$$

Now, assume we have measured the gross weight of a new mardouf to be 65.0 grams. The estimated bias of this measurement is rounded off to zero (so small it can be neglected). The standard uncertainty is calculated as $u_{gw_0} = \sqrt{\hat{\sigma}_{gw_0}^2} \approx 0.02$ grams.

From the 260 pairs of measured gross weights and wrapping weights plotted in Figure 8.6-1a we obtain the following:

$\sum_{i=1}^{n}(gw_i - \overline{gw})^2 = 22404$; $\hat{\sigma}_e^2 \approx 14.09$; $\overline{gw} = 74.8$ grams ; $\hat{a} \approx -9.70$ $\hat{b} \approx 0.427$

and the standard uncertainty is

$$u \approx \sqrt{\frac{14.09}{22404}\left(0.02^2 + (65.0 + 0)^2 + 74.8^2 - 2 \cdot 74.8 \cdot (65.0 + 0)\right) + (1 - 0.427)^2 \cdot 0.02^2 + \frac{14.09}{260}} \approx$$

$$\approx 0.339$$

8.6.3 Expanded uncertainty

For obtaining the expanded uncertainty, we must remember that $\widehat{nw_0}$ is used as a prediction of the *actual* wrapping weight of the new mardouf, while the uncertainty deduced above is for the estimation of its *expected* or mean value.
The prediction error is

$$\widehat{nw_0} - nw_0 = \widehat{nw_0} - (gw_0 - ww_0) = \widehat{nw_0} - \left((1 - b) \cdot gw_0 - a - e_0\right)$$

where e_0 is the unknown deviation from the linear relationship between nw_0 and gw_0. This term would give a large contribution to the expanded uncertainty compared to $Var(\widehat{nw_0})$.

The expected (mean) value of the prediction error can easily be shown to be zero, and an estimate of the variance of the prediction error can be deduced to

$$\widehat{Var(nw_0 - nw_0)} =$$

$$= \frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(gw_i - \overline{gw})^2}\left(\hat{\sigma}_{gw_0}^2 + \left(gw_0 + \widehat{bias}_0\right)^2 + (\overline{gw})^2 - 2 \cdot \overline{gw}\left(gw_0 + \widehat{bias}_0\right)\right) + \hat{\sigma}_e^2 \cdot \left(1 + \frac{1}{n}\right)$$

The standard uncertainty of the prediction error is thus

$$u_{pe} = \sqrt{\frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(gw_i - \overline{gw})^2}\left(\hat{\sigma}_{gw_0}^2 + \left(gw_0 + \widehat{bias}_0\right)^2 + (\overline{gw})^2 - 2 \cdot \overline{gw}\left(gw_0 + \widehat{bias}_0\right)\right) + \hat{\sigma}_e^2 \cdot \left(1 + \frac{1}{n}\right)}$$

The prediction error is a linear combination of the product of two random variables ($b - \hat{b}$ and gw_o) and two other random variables ($(a - \hat{a})$ and e_0 respectively). It is therefore not normally distributed, and it is not possible to apply any t-distribution to deduce the expanded uncertainty. We may use Chebyshev's inequality to obtain an interval for the prediction error with coverage at least 95%.

Chebyshev's inequality applied to the prediction error is

$$P\left(|\widehat{nw}_0 - nw_0| > k \cdot \sqrt{Var(|\widehat{nw}_0 - nw_0|)}\right) < \frac{1}{k^2}$$

Setting $1/k^2$ to 0.05 gives $k = \sqrt{1/0.05} \approx 4.47$

Hence, an expanded uncertainty for the predicted net weight – considering that the uncertainty u_{pe} is based on an estimate of the true variance of the prediction error – can be set to

$$U = 4.5 \cdot \sqrt{\frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(gw_i - \overline{gw})^2}\left(\hat{\sigma}_{gw_0}^2 + \left(gw_0 + \widehat{bias}_0\right)^2 + (\overline{gw})^2 - 2 \cdot \overline{gw}\left(gw_0 + \widehat{bias}_0\right)\right) + \hat{\sigma}_e^2 \cdot \left(1 + \frac{1}{n}\right)}$$

With the numbers used to calculate the uncertainty of $\widehat{nw}_0$ in the previous section the expanded uncertainty becomes

$$U \approx 4.5 \cdot \sqrt{\frac{14.09}{22404}\left(0.02^2 + (65.0 + 0)^2 + 74.8^2 - 2 \cdot 74.8 \cdot (65.0 + 0)\right) + 14.09 \cdot \left(1 + \frac{1}{260}\right)} \approx$$

$$\approx 17.0$$

The net weight of a mardouf weighing 65.0 grams is thus $(1 - 0.427) \cdot 65.0 - (-9.70) \approx 46.9$ grams with a 95% error margin of 17.0 grams.

We note that the impact of the uncertainty (0.02) of the measured weight of 65.0 grams is negligible.

8.7 Velocity estimation on a speeding car in video images

In the previous examples it often was described how to deal with measurement uncertainty in cases where many reference measurements are available, such as through control charts. In casework where this is not the case it may be relevant what the (combined) measurement uncertainty is as well though and publications are found in literature as well. One example is that of velocity estimation on a speeding car in video images, see [20, 21]. Here limited controlled experiments are performed, based on which measurement uncertainty is determined (integrative approach, see Chapter 6).

Closed Circuit TV (CCTV) systems often record vehicle motion prior to incidents. From the footage an estimate of the average velocity of the vehicle between two frames can be calculated. Estimation may be based on estimating the travelled distance of the car between two images and the time elapsed on the camera system, estimation of the velocity is by division of the two. In [20] and [21] it is described how it is possible to quantify the corresponding measurement uncertainty.

The measurement uncertainty on the reported velocity may be derived using validation recordings of a car driving by at known velocity. If for example a statistical linear regression model is applicable, confidence or (Bayesian) probability intervals can be determined for the unknown velocity of the speeding car. An example of this is as follows.

8.7.1 Case

A car driving by is recorded by a camera system (CCTV-system). In figure 8.7-1, two consecutive images are shown of the car driving by. Just beyond the view of the camera the car hits a motorbike and the driver of the motorbike dies. The question is: what was the velocity of the car driving by in the video?

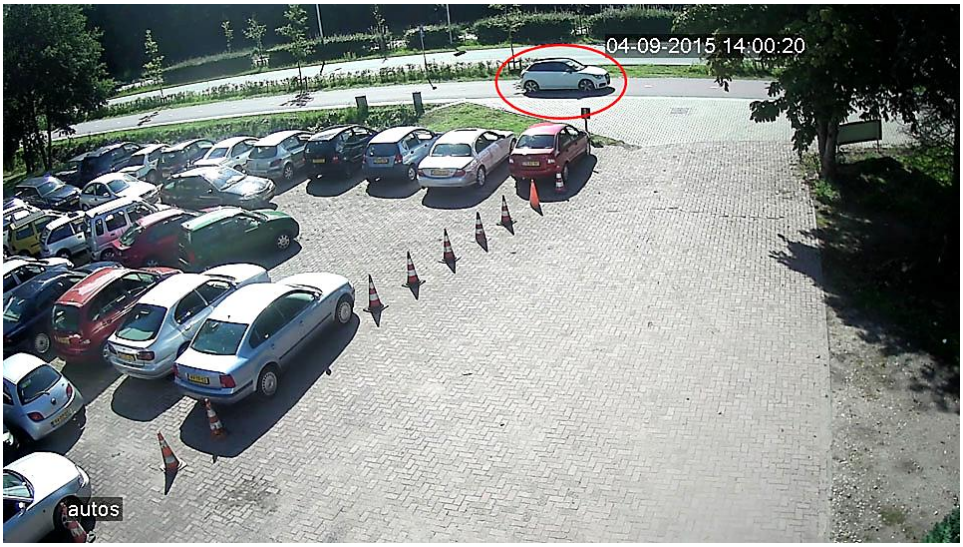

Figure 8.7-1: Two consecutive images of the car driving by in the case.

8.7.2 Method

The average velocity of the car is calculated from the video images using a 3D-model from the scene and the car. To calculate the measurement uncertainties an investigation at the scene took place. Test drives with a similar car were carried out at different velocities. The ground truth of the velocity was recorded by a data logger and the test drives were recorded (reference recordings) by the same CCTV-system that recorded the incident. The velocity of the car in the reference recordings is measured using the same method (3D-model) as was used for the incident recording. The differences between the measured velocity (from the images) and real velocity (from the data logger) is used to calculate the systematical and random error. These two errors are used to calculate the measurement uncertainties and calculate a confidence interval for the velocity of the car in the images of the incident. The statistical model to obtain the confidence interval is described in [20]. In figure 8.7-2, two consecutive images are shown of a car driving by in a test drive.

Figure 8.7-2: Two consecutive images of a car driving by in a test drive.

8.7.3 Results

Results of the test drives are given in Table 8.7-1.

Reference drive	Measured (km/h)	Real (km/h)
1	60.9	61.6
2	70.3	71
3	79.5	81.5
4	88.8	90.5
5	98.2	99.8
6	32.4	33
7	40.4	41.3
8	49.7	50.5
9	59.2	60.9
10	68.8	70.3
11	79.5	81.5
12	89.3	91.5
13	98.3	100.1
14	30.1	30.8
15	39.2	40.3
16	59.5	60.9
17	68.6	70.6
18	78.0	79.8
19	89.0	90.3
20	97.9	99.6
21	30.0	30.5

Table 8.7-1: Results for the measured and real velocities in the test drives.

In Figure 8.7-3 the results are depicted.

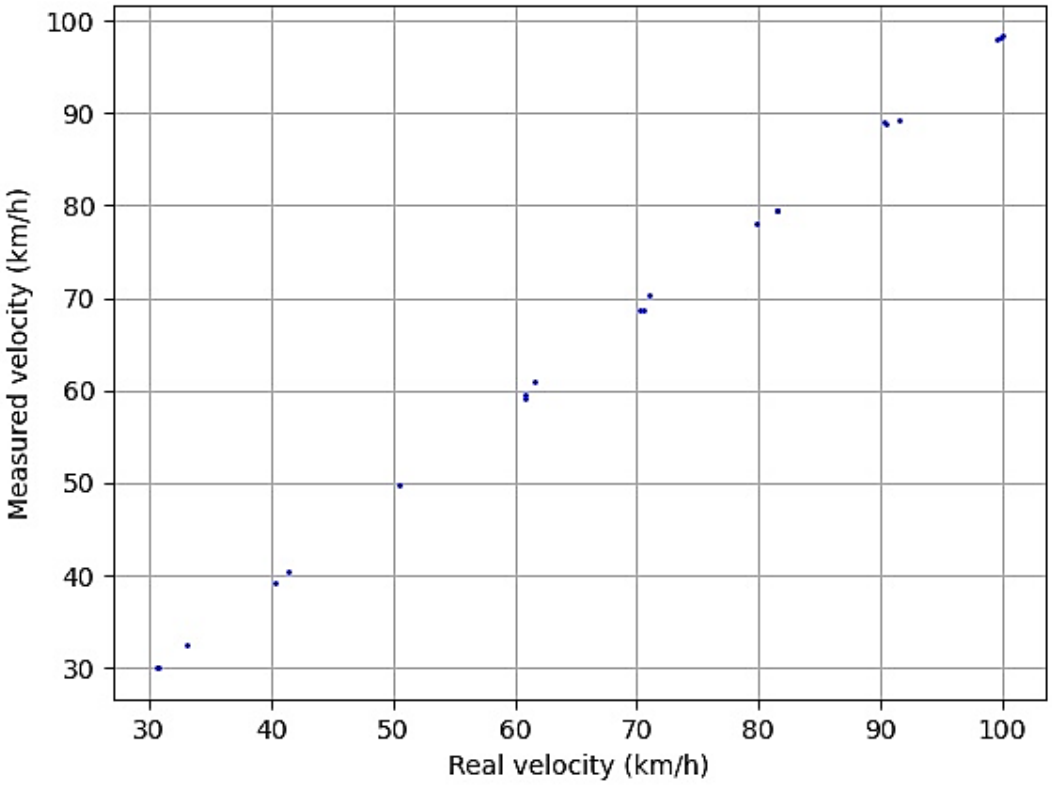

Figure 8.7-3: Illustration of the results of 21 test drives. On the x-axis, the real velocity of the car is given, on the y-axis the calculated / measured velocity.

Overall the results are as follows:

- Measured average velocity of the questioned car in the images: v_{image} = 62.8 km/h;
- Number of test drives: n = 21;
- Average difference from measured velocity – real velocity over all test drives: $\bar{\delta}$ = - 1.36 km/h;
- Standard deviation from average difference: s = 0.54 km/h.

The average difference of $\bar{\delta}$ = -1.36 km/h means that the measured velocity in the images is on average systematically underestimated by 1.36 km/h. The measurement of the average velocity of the car is compensated for this systematical error and therefore becomes:

$$v_{incindent} = v_{image} - \bar{\delta} = 62.8 \text{ km/h} + 1.36 \text{ km/h} = 64.2 \text{ km/h}.$$

The boundaries of the confidence interval can be derived by calculating the contribution of the random error v_{random}:

$$v_{random} = s\, \xi_{n-1,confidence} \sqrt{1 + \frac{1}{n}}$$

where $\xi_{n-1,\,confidence}$ is the quantile corresponding to the Student's t distribution with n-1 degrees of freedom and given level of confidence. In the reference [20] quantiles are given for (n-1) = 1,…,∞ and, confidence level 90% , 95%, 97.5% and 99%, for 2-sided intervals.

With the number of test drives being n = 21, so (n-1) = 20, gives a quantile of $\xi_{20,95\%}$ = 2.09 for a 95% confidence interval. The random error becomes:

$$v_{random} = 0.54 \times 2.09 \times \sqrt{(1+(1/21))} = 1.16 \text{ km/h}.$$

The upper limit of the 95% confidence interval can now be calculated by

$$v_{incindent} + v_{random} = 64.2 \text{ km/h} + 1.16 \text{ km/h} = 65.3 \text{ km/h}$$

(rounded to one decimal after full digit calculation) and the lower limit by

$$v_{incindent} - v_{random} = 64.2 \text{ km/h} – 1.16 \text{ km/h} = 63.0 \text{ km/h}.$$

Therefore, the 95% confidence interval for the average velocity of the car is:

$$v_{incident} = [63.0, 65.3] \text{ km/h}.$$

In [21], a Bayesian approach is described in which based on a linear regression model, probability distributions are generated for the model parameters and subsequently the velocity of a car in a case.

9 REFERENCES

1. ISO/IEC 17025: 2017, General requirements for the competence of testing and calibration laboratories
2. EURACHEM/CITAC Guide CG4, Quantifying Uncertainty in Analytical Measurement, edition 3, 2012.
3. QCC-UM-001: 2006, Guidance for Uncertainty of Measurement in Quantitative Analyses or testing (UM)
4. JCGM 200:2012, International vocabulary of metrology – Basic and general concepts and associated terms, 2008 version with minor corrections (VIM), 3rd edition
5. ISO Guide 30:2015, Reference materials – Selected terms and definitions
6. ISO Guide 3534-1:2006, Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability.
7. JCGM 100: 2008, Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM).
8. CXG 54:2004, Guidelines on measurement uncertainty, revised in 2021.
9. Nordtest technical report 537, Handbook for calculation of measurement uncertainty in environmental laboratories, edition 4, 2017.
10. ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results-Part 1: General principles and definitions.
11. ISO 5725-2:2019, 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.
12. ISO 21748:2017, Guide to the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.
13. Eurolab Technical Report no. 1/2006, Guide to the evaluation of measurement uncertainty for quantitative test results.
14. Eurolab technical report no. 1/2007, Measurement uncertainty revisited: alternative approaches to uncertainty evaluation.
15. Magnusson, B. and Ellison, S.2008. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. Anal Bioanal Chem 390: 201-213.
16. O'Donnell, G.E. and Hibbert, D.B. 2005. Treatment of bias in estimating measurement uncertainty. Analyst 130: 721-729.
17. Laso J and Peris A. 2014. Use of Proficiency testing results in method validation. 8th workshop on proficiency testing in analytical chemistry, microbiology and laboratory medicine. Eurachem. Berlin https://www.eurachem.org/images/stories/workshops/2014_10_PT/pdf/book_of_abstracts.pdf
18. Instituto Nacional de Seguridad e Higiene en el Trabajo (2015) Determinación de la incertidumbre de medida de agentes químicos. Utilización de los resultados de participación en programas de ensayo de aptitud en la estimación de la incertidumbre de medida de un procedimiento analítico (CR-09/2015) https://www.insst.es/documents/94886/359418/CR-09-2015.pdf/2fcc6b11-baf9-4e60-9ac5-d432587b3ac8?version=1.0&t=1528890557590
19. Klu, J.K., et al. 2021. Measurement uncertainty in quantifying delta-9-tetrahydrocannabinol (THC) in blood using SPE and LC/MS/MS. Forensic Science International 322: 110744.
20. Hoogeboom, B. and Alberink, I. 2010. Measurement uncertainty when estimating the velocity of an allegedly speeding vehicle from images. J. For. Sci. 55(5): 1347 – 1351.
21. Mieremet, A., Alberink, I., Hoogeboom, B. and Vrijdag, D. 2018. Probability intervals of speed estimations from video images: The Markov Chain Monte Carlo approach. For. Sci. Int. 288: 29–35.
22. Gullberg, R. G.2014. Uncertainty in Blood and Breath Alcohol Measurements. In Garriot's Medicolegal Aspects of Alcohol, eds. Y. Caplan and B. Goldberger, 6th ed. Lawyers & Judges Pub C0, Inc.

23. Franco de Oliveira, S. and M. Yonamine, 2016. Measurement uncertainty for the determination of amphetamines in urine by liquid-phase microextraction and gas chromatography-mass spectrometry. Forensic Science International. 265: 81-88.
24. Gullberg, R.G. 2012. Estimating the Measurement Uncertainty in Forensic Blood Alcohol Analysis. Journal of Analytical Toxicology 36(3): 153-161.
25. EURACHEM/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling, a guide to methods and approaches, M. Ramsey, S. Ellison, and P. Rostron, Editors. 2019.
26. Horwitz, W. 1990. Nomencature for sampling in analytical chemistry (Recommendations 1990). Pure & Appl. Chem. 62(6): 1193-1208.
27. Ellison, S.L.R. 2015. Homogeneity studies and ISO Guide 35:2006. Accreditation and Quality Assurance. 20(6): 519-528.
28. Kumar, A. and D.K. Misra. 2020. A Review on the Statistical Methods and Implementation to Homogeneity Assessment of Certified Reference Materials in Relation to Uncertainty. MAPAN. 35(3): 457-470.
29. Linsinger, T.P.J. 2020. Evaluation of CRM homogeneity in cases of insufficient method repeatability: Comparison of Bayesian analysis with substitutes for ANOVA based estimates. Analytica Chimica Acta: X. 5: 100049.
30. Ulrich, J.C. and J.E. de Souza Sarkis. 2013. Estimation of uncertainty to the homogeneity study in the certification of a candidate certified reference material in fish matrix. International Congress of Metrology.
31. Van-der-Veen, A. and J. Pauwels. 2000. Uncertainty calculations in the certification of reference materials. 1. Principles of analysis of variance. Accreditation and Quality Assurance. 5(12): 464-469.
32. Burke, S. 2001. Regression and calibration. LC-GC Europe Online Supplement statistics and data analysis: 13-18.
33. Hibbert, D. 2006. The uncertainty of a result from a linear calibration. Analyst. 131(12): 1273-1278.
34. Kadis, R. 2017. Evaluation of the measurement uncertainty: Some common mistakes with a focus on the uncertainty from linear calibration. Journal of Chromatography A. 1499: 226-229
35. Shaun Burke, *Regression and Calibration*.
36. Gu, H., Liu, G., Wang, J., Aubry, A. F., & Arnold, M. E. (2014). Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay performance. *Analytical Chemistry*, 86(18), 8959-8966.

10 AMENDMENTS AGAINST PREVIOUS VERSION

Not applicable. This is the first version.

ANNEX I. STATISTICAL BACKGROUND TO DETERMINATION OF MEASUREMENT UNCERTAINTY BASED ON IN-HOUSE VALIDATION OR PROFICIENCY TESTING

The term “measurement uncertainty” reveals that what is measured is not to be treated as exact. Expressing it differently, it means that if we make another measurement, we do not expect to get exactly the same value as with the first measurement. This in turn means that there is some unpredictable variation in the outcomes of measurements, and variation as a concept is what constitutes the base for statistical science.

Although there is unpredictable variation present, we must try as far as possible to quantify how big it is and to what extent it affects the interpretation of measured values. If we for instance should make a quantification of the purity of some drug powder, e.g. amphetamine, our measurement may give a value like 57.3%. This can be from a single quantification analysis, or it can be the average of a few repeated analyses. No matter which, it is important to know how accurate this measurement is with respect to what the true purity of the powder is. The unpredictable variation may be large meaning that the true purity may with high certainty be in a quite wide interval, say between 45% and 70%, which means that a reported value of 57.3% is not that accurate. But it can also be in a narrower interval, say between 55% and 59% in which case 57.3% would be considered accurate, especially with respect to future decisions in the judicial process that will make use of the reported result from this quantification.

Therefore, outcomes of measurements should always be accompanied by appreciated uncertainties of the reported values to guide the end-user on the accuracy of them.

AI.1 PRECISION AND ACCURACY

The two terms precision and accuracy are sometimes confused (and may also be confused with the term resolution). We revisit the definition of precision and accuracy in terms of measurement uncertainty:

Precision is how much a measurement may vary with respect to all sources of variation. To quantify the precision, measures of dispersion from statistical models are used, of which the most common is variance (or standard deviation).

Accuracy is how close to the target value a measurement is. This has not only to do with the variation of the measurement, but also whether there are systematic deviations or not. A measurement may have high precision but may still have low accuracy. To quantify the accuracy both measures of location and dispersion from statistical models must be used.

AI.2 RANDOM VARIABLES

A quantity - like a measurement - the value of which is not fixed and cannot be exactly predicted on forehand is referred to as a random variable. There can however be knowledge about how it varies – which values it can attain and with which probabilities these values are attained. This is referred to as the probability distribution of the random variable. One important characteristic of such a probability distribution is its *mean* or *average*, which is the “centre” of the distribution. In this text we will use the term *average* since this is the used term in several other guidelines and quality documents on measurement uncertainty. However, it is important to separate the average value of a random variable from the average of a fixed number of measurements (a

sample average or *arithmetic mean*). The latter is a quantity based on collected data, while the former is a theoretical and usually unknown quantity. This is the reason to why in the statistical literature the term *mean* or *expected value* is preferred.

Another important characteristic of a probability distribution is its *variance*, which measures the dispersion of the possible values (the range of the probability distribution).

In statistical models these two measures can be straightforwardly calculated for sums of random variables – here applied to combination of components of measurement uncertainty. When the random variables do not affect each other in their values, referred to as *independent* in statistical theory, their averages and variances can be summed to obtain the average and variance of their sum, which is an important property used in combining components of measurement uncertainty.

Related to the variance is the standard deviation, which is simply the square root of the variance. This measure is easier to interpret since it is on the same scale as the random variable itself, but in contrast to the variance, the standard deviation of a sum of random variables is not the sum of the standard deviations for the individual random variables. To obtain the standard deviation of a sum of several independent random variables, we must first calculate the sum of the variances and then take the square root this sum to obtain the standard deviation.

In mathematical notation we can write it like the following:

X is a random variable, its average is denoted μ, its variance is denoted σ^2 and its standard deviation is (naturally) denoted σ.

If $Y = a \cdot X + b$, where a and b are any constants, then $\mu_Y = a \cdot \mu + b$, $\sigma_Y^2 = a^2 \cdot \sigma^2$, and $\sigma_Y = |a| \cdot \sigma$, where $|a|$ is the absolute value of a (e.g. $|2| = 2, |-2| = 2$).

If $X_1, X_2, \dots, X_n$ are *n* independent random variables with averages $\mu_1, \mu_2, \dots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$; and $Sum = X_1 + X_2 + \cdots + X_n$, then the average of Sum is

$$\mu_{Sum} = \mu_1 + \mu_2 + \cdots + \mu_n \ ,$$

the variance of Sum is

$$\sigma_{Sum}^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 \ ,$$

and the standard deviation of Sum is

$$\sigma_{Sum} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2} \ .$$

If $X_1, X_2, \dots, X_n$ have equal averages ($= \mu$) and equal variances ($= \sigma^2$) their arithmetic mean $\bar{X}_{(n)} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ has average μ, variance $\frac{\sigma^2}{n}$, and standard deviation $\frac{\sigma}{\sqrt{n}}$.

AI.3 MEASUREMENTS AND MEASURANDS – STATISTICAL MODEL

To understand how the different components of calculating measurement uncertainty occur, we must formalise the measurement situation and introduce terms representing these components. From now on we will use the term “estimate” instead of “calculate”, since estimate means (numerical) approximation of an unknown quantity, while calculate is a more general term.

A measurement is from a data point of view the obtaining of a numerical value. We choose to denote this value x.
This measurement is treated as an approximation of the true value of the *measurand* (the quantity to be measured). We denote this value m.

The relationship between x and m constitutes the ground from which the components of an estimated measurement uncertainty are identified and estimated. This relationship should account for

- bias – the systematic difference between the measurement and the measurand, which is the difference between the average of (an infinite) number of measurements made on the same object and the true value of the measurand (of this object)

- repeatability – the measurement precision under a set of repeatable conditions (same apparatus, sample, operator, room temperature etc.)

- reproducibility – the degree of agreement between measurements made on the same object (identical samples of it) under different investigating situations (e.g. using different apparatus, operators, time and date, environmental conditions etc.)

- stability – how stable the measurement process is over time, i.e. how the accuracy of the measurements changes

Stability may be included in reproducibility assuming different dates is a kind of investigation situation. High degrees of repeatability and reproducibility respectively actually mean low variation among the measurements, but in a statistical model it is impractical do use such degrees as components since they then would have to be inverted to show their contribution to the uncertainty. Therefore, repeatability and reproducibility are represented by the variance they contribute with.

In the statistical literature repeatability is usually referred to as unexplained variation (or sometimes random error), while reproducibility is referred to as explained variation since it is possible to trace it back to its source (different instruments, different operators etc.). Stability is related to time and is of main interest for studying the performance of a particular instrument (or a general measurement setup) over time to estimate drift. When the measurement uncertainty of a particular result should be estimated it is therefore more practical to include it in the degree of reproducibility.

AI.3.1 The General Model

A statistical model for the relationship between the measurement X and the measurand m can be written

$$X = m + B + D + E \quad (1)$$

where B (naturally) stands for the deviation between X and m due to the bias, D stands for the deviation between X and m due to the degree of reproducibility, i.e. due to that a certain instrument, operator etc. have been used, and E stands for the deviation between x and m due to the degree of repeatability, i.e. due to the unexplained variation between repeated measurements on the same object under the same conditions.
In a traditional setup the components m and B in expression (1) are both fixed, but their values are not known. The term B would then be the actual deviation between the measurement x and the value of the measurand m due to a systematic error. However, when measurement uncertainty in general should be appreciated it has to be considered that the bias may vary from case to case. Hence, the term B should rather be considered as a random contribution to the measurement that is separate from the contributions due to reproducibility or repeatability. This also means that seen over all possible measurements that can be taken there is an average bias, which we can denote b. This can be written as

$$B = b + \delta$$

where b is the average bias (of measurements of the current measurand) and δ is the random deviation with average 0 and variance σ_B^2.

The components D and E are both in the analysis of measurement uncertainty considered to be random with zero as average value. However, their variances must be considered when estimating measurement uncertainty.

From (1) we can see that the measurement error is

$$Q = X - m = B + D + E \quad (2)$$

AI.3.2 A simplified model

The separation of variation due to the degrees of reproducibility and repeatability respectively that is used in expressions (1) and (2) allows for separate estimation of the contributions to uncertainty from these components. However, in practice the variation due to the degree of repeatability is usually small compared to the variation due to the degree of reproducibility. Moreover, to obtain robust estimates of these two components a quite comprehensive experimental design is needed, which is not feasible to set up for each kind of measurement situation. Therefore, it is common to let $D + E$ in expressions (1) and (2) be represented by one component only, that accounts for the variation due to both the degree of reproducibility and the degree of repeatability. In literature on measurement uncertainty this is often referred to as variation due to the degree of *reproducibility within laboratory*. Here this component will be denoted R_w (where the subscript "w" stands for "within laboratory") and the simplified model is expressed as

$$X = m + B + R_w \quad (3)$$

and the measurement error can be written

$$Q = X - m = B + R_w \quad (4)$$

If the bias term B would be considered as constant then the variance of the measurement X (and of the measurement error Q) would be the variance of R_w, which can be denoted $\sigma_{R_w}^2$. However, treating the bias as a random component the correct expressions for the variance of the measurement and the measurement error are

$$\sigma_X^2 = \sigma_B^2 + \sigma_{R_w}^2 \quad (5)$$

and

$$\sigma_Q^2 = \sigma_B^2 + \sigma_{R_w}^2 \quad (6)$$

respectively (provided we can assume that the components B and R_w are independent).
A graphical illustration of expressions (3)-(5) is shown in Figure AI.1.

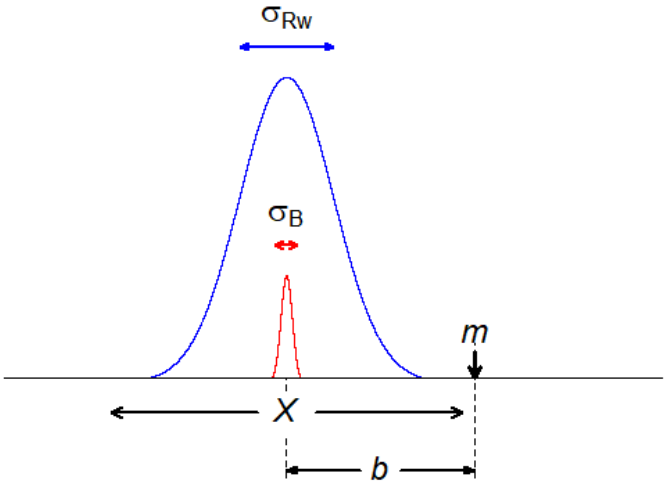

Figure AI- 1: The measurement X varies primarily due to the degree of reproducibility illustrated with the blue curve (probability distribution) and with the standard deviation σ_{R_w}, but may also have a bias with average b and standard deviation σ_B (illustrated with the red curve). The measurand m deviates from the average of X (when bias is present). Note that m can also be to the left of the average of X.

AI.4 UNCERTAINTY AND EXPANDED UNCERTAINTY

In literature on measurement uncertainty the *standard uncertainty* is usually defined as the square root of the so-called *combined uncertainty*, where the latter is a sum of the estimated variances from the different variables contributing to the variation of the measurement X plus a squared contribution from the bias. With the simplified model in expression (3) above the standard uncertainty, u, could be written

$$u = \sqrt{u_B^2 + s_{R_w}^2} \quad (7)$$

where $s_{R_w}^2$ would be a (standard) estimate of $\sigma_{R_w}^2$ obtained within the laboratory and u_B^2 represents the bias contribution.

AI.4.1 Expanded uncertainty

The *expanded uncertainty* is defined as the standard uncertainty multiplied with a so-called *coverage factor* to account for a pre-specified range of possible values that the measurement X can take. To understand this better, we give a short account for how the standard deviation of a random variable relates to the range of possible values for that variable.

For any random variable X with average μ and standard deviation σ the following holds:

The interval $\mu \pm 1.5 \cdot \sigma$ comprises at least 55 % of all possible values of X
The interval $\mu \pm 2 \cdot \sigma$ comprises at least 75% of all possible values of X
The interval $\mu \pm 3 \cdot \sigma$ comprises at least 88% of all possible values of X
The interval $\mu \pm 4 \cdot \sigma$ comprises at least 93% of all possible values of X

A mathematical result called ***Chebyshev's inequality*** states that $\mu \pm k \cdot \sigma$ comprises at least $100 \cdot (1 - 1/k^2)$% of all possible values of X for $k > 1$. k is referred to as the coverage factor.

However, if X is normally distributed the intervals become narrower:

$\mu \pm \sigma$ comprises about 68% of all possible values
$\mu \pm 2 \cdot \sigma$ comprises about 95% of all possible values (8)
$\mu \pm 3 \cdot \sigma$ comprises about 99.87 % of all possible values

This can be used to motivate that for a single measurement X that is assumed to be normally distributed with standard deviation σ, the interval $X \pm 2 \cdot \sigma$ covers the average μ of X with 95% *confidence*, a so-called 95% *confidence interval*.

Nevertheless, the theoretical standard deviation of X is rarely known (if ever), and so it must be estimated from available data. Using the simplified model in expression (3) above and assuming there is no bias, the standard deviation of X is estimated by s_{R_w} in expression (7). If that estimate is based on sufficiently many measurements (say at least 50) typically from a control chart the coverage factor 2 would give approximately 95% confidence, i.e. $X \pm 2 \cdot s_{R_w}$ is an approximate 95% confidence interval for the measurand m. However, it is very important to remember that if the number of measurements used for this estimation is much fewer than 50, the confidence will be much lower. See also section 6.

In this confidence interval it is assumed that X is a single measurement. However, when it is the average of n measurements (where n may often be equal to 2 or 3), we must divide s_{R_w} with $\sqrt{n}$ (cf. section 2).

When bias is present the situation is more complicated. Using expression (5) above we could estimate the standard deviation of X with $\sqrt{s_B^2 + s_{R_w}^2}$ (or $\sqrt{\frac{s_B^2}{n} + \frac{s_{R_w}^2}{n}}$ if an average of n measurements is used) , where s_B would be an estimate of the standard deviation of the bias B, also based on sufficiently many measurements. But it is very difficult to obtain such an estimate. Moreover, the interval $X \pm 2 \cdot \sqrt{s_B^2 + s_{R_w}^2}$ would be a 95% confidence interval for $m + b$, and we still would not know the value of b.

We present here two alternative ways of including the contribution from bias into the expanded uncertainty. The first is to estimate b and σ_B together and include this estimate into the expression for the standard uncertainty, and then compute the expanded uncertainty. This means that the coverage factor would correspond to a higher confidence than what is given by the intervals (8) above. The second is to estimate b and add its absolute value to the expanded

uncertainty deduced with the assumption of no bias. This means that the total expanded uncertainty will cover a larger range of values than is set by the coverage factor.

AI.4.2 Joint estimation of the bias contribution

To estimate the bias contribution, we need a set of measurements where the measurand is known. This can be done in several ways.

AI.4.2.1 Within-laboratory estimation

One way is to make repeated measurements within the laboratory on a material for which the measurand is known. This could be a certified reference material (CRM) provided by an institute of standards or unit conducting proficiency tests. Denote the value of this measurand m_{CRM}. Now, assume we have made *p* measurements on this material and denote these measurements $x_1, x_2, \dots, x_p$. For the differences between the measurements and the measurand, i.e. $d_1 = x_1 - m_{\mathrm{CRM}},\ d_2 = x_2 - m_{\mathrm{CRM}}, \dots,\ d_p = x_p - m_{\mathrm{CRM}}$ it is common to compute the so-called *mean square deviation* (MSD), i.e.

$$MSD_w = \frac{1}{p} \cdot \sum_{i=1}^{p} d_i^2 = \frac{1}{p} \cdot \sum_{i=1}^{p} (x_i - m_{\mathrm{CRM}})^2 \tag{9}$$

(where the subscript "w" refers to within-laboratory estimation, see subsection AI.4.2.2.). In the literature, it is also common to refer to the *root mean square deviation* (RMS), which is simply the square root of MSD, i.e. $RMS = \sqrt{MSD}$.

Applying the simplified model (3), i.e. $X = m_{\mathrm{CRM}} + B + R_w$, with $B = b + \delta$ where δ has average zero and variance σ_B^2, it can be shown – assuming B and R_w are independent random variables that MSD_w is on the average equal to $\sigma_X^2 + b^2 = \sigma_B^2 + \sigma_{R_w}^2 + b^2$ (taking into account the random fashion of the measurements)[10].

Hence, using the square root of MSD_w as standard uncertainty would on the average include the expected bias. Nevertheless, since the number of measurements, p, usually cannot be that high, it is not possible to use the standard coverage factors (from (7)) to calculate the expanded uncertainty.

However, when a certified reference material is delivered to the laboratory, it usually comes with the material provider's assessment of uncertainty, i.e. m_{CRM} is obtained using a high-accuracy measurement method, but some dispersion cannot be avoided. It can be assumed that this measurement method is free from bias, but there is a small standard deviation, σ_{CRM} that should not be ignored. This standard deviation is delivered by the material provider either on absolute or relative form (see further section AI.5), and is used in the estimation of measurement uncertainty as a known quantity, i.e. should not be estimated by the laboratory.

Some literature (see e.g. [9]) suggests using as standard uncertainty[11]

[10] With $E(\cdot)$ denoting the expected value (average), $E\left[\frac{1}{p}\sum_{i=1}^{p}(X_i - m_{\mathrm{CRM}})^2\right] =$
$E\left[\frac{1}{p}\sum_{i=1}^{p}(X_i - (m_{\mathrm{CRM}} + b) - b)^2\right] = E\left[\frac{1}{p}\sum_{i=1}^{p}\left(\left(X_i - (m_{\mathrm{CRM}} + b)\right)^2 + b^2 - 2\left(X_i - (m_{\mathrm{CRM}} + b)\right)b\right)\right] = \frac{1}{p}\left[\sum_{i=1}^{p} E\left[\left(X_i - (m_{\mathrm{CRM}} + b)\right)^2\right] + pb^2 - 2b \cdot 0\right] = Var(X_i) + b^2 = \sigma_X^2 + b^2 = \sigma_B^2 + \sigma_{R_w}^2 + b^2$

[11] The notation used here is different from the one used in other literature, but the components have their counterparts.

$$u_w = \sqrt{MSD_w + \sigma_{\text{CRM}}^2 + s_{R_w}^2} \tag{10}$$

and for 95% confidence compute the expanded uncertainty as

$$U_w = 2 \cdot \sqrt{MSD_w + \sigma_{\text{CRM}}^2 + s_{R_w}^2} \tag{11}$$

While this would give a confidence greater than 95% (provided $s_{R_w}^2$ is based on sufficiently many observations), it should be noted that the contribution from the term R_w in the simplified model is double counted, why the expanded (and standard) uncertainty may be too large compared to the quality standard used at the laboratory. When an average of n measurements is used the corresponding expressions for the standard uncertainty and expanded uncertainty are $u_w = \sqrt{MSD_w + \sigma_{\text{CRM}}^2 + \frac{s_{R_w}^2}{n}}$ and $U_w = 2 \cdot \sqrt{MSD_w + \sigma_{\text{CRM}}^2 + \frac{s_{R_w}^2}{n}}$ respectively.

Example 1
Suppose a method of measuring refractive index of glass at a laboratory has over a longer time shown a standard deviation of 0.00025 for the variation due to reproducibility within laboratory, i.e. $s_{R_w} = 0.00025$. A reference glass with a certified refractive index of 1.52000 (m_{CRM}) with reported standard deviation 0.0000002 is measured p =10 times with the following result:

1.51996 1.52009 1.52006 1.52049 1.52008

1.52008 1.51967 1.52008 1.51981 1.52001

$MSD_W = \frac{1}{10} \cdot [(1.51996 - 1.52000)^2 + (1.52009 - 1.52000)^2 + \cdots + (1.52001 - 1.52000)^2] \approx 4.18 \cdot 10^{-8}$

The expanded uncertainty with coverage 95% then becomes

$U_w = 2 \cdot \sqrt{4.18 \cdot 10^{-8} + 0.0000002^2 + 0.00025^2} \approx 0.0065$

AI.4.2.2 Using proficiency tests

Another way is to use results from proficiency tests in which the laboratory together with other laboratories has participated. In these test samples from a certified reference material are sent out to the participating laboratories, however, not necessary with detailed information about its certified value. Assume the laboratory has taken part in p such tests for the measurand of interest.

In test i ($i = 1,2, \ldots, p$) assume there are n_i other laboratories participating and that we therefore have n_i reported measurement results ($y_{i1}, y_{i2}, \ldots, y_{in_i}$) from these. For the current laboratory the reported result is denoted x_i. Now, with $d_i = x_i - \bar{y}_i$, where $\bar{y}_i$ is the sample average of the n_i reported results from the other laboratories[12], the counterpart of MSD_0 in expression (9) would be

$$MSD = \frac{1}{p} \cdot \sum_{i=1}^{p} d_i^2 = \frac{1}{p} \cdot \sum_{i=1}^{p} (x_i - \bar{y}_i)^2 \tag{12}$$

[12] $\bar{y}_i = \frac{1}{n_i} \cdot \sum_{j=1}^{n_i} y_{ij}$

With some algebra it can be shown that MSD is on the average equal to $\sigma_X^2 + b^2 + \Delta$, where Δ denotes an average of the average squared distances between the measurand value m_i and $\bar{y}_i$ over the p proficiency tests[13].

Following the arguments in subsection AI.4.2.1 the standard uncertainty could be taken as

$$u = \sqrt{MSD + s_{R_w}^2} \tag{13}$$

and the expanded uncertainty as

$$U = k \cdot \sqrt{MSD + s_{R_w}^2} \tag{14}$$

with k being the coverage factor used, keeping in mind that this again is an overestimation of the uncertainty. Note that here we have no counterpart to σ_{CRM}. It is common practice to neither provide m_{CRM} nor σ_{CRM} to prevent laboratories from tuning their results to conform as well as possible to the expected. In some literature (e.g. [9]) methods for appreciating this contribution to the dispersion is provided, though. When an average of n measurements is used the corresponding expressions for the standard uncertainty and expanded uncertainty are $u = \sqrt{MSD + \frac{s_{R_w}^2}{n}}$ and $U = 2 \cdot \sqrt{MSD + \frac{s_{R_w}^2}{n}}$ respectively.

[13] With the model $X_i = m_i + B + R_w$ and with $E(\cdot)$ denoting the expected value (average),

$E\left[\frac{1}{p}\sum_{i=1}^{p}(X_i - \bar{Y}_i)^2\right] = E\left[\frac{1}{p}\sum_{i=1}^{p}(X_i - m_i + m_i - \bar{Y}_i)^2\right]$

$= E\left[\frac{1}{p}\sum_{i=1}^{p}[(X_i - m_i)^2 + (m_i - \bar{Y}_i)^2 + 2(X_i - m_i)(m_i - \bar{Y}_i)]\right] = E\left(\frac{1}{p}\sum_{i=1}^{p}(X_i - m_i)^2\right)$

$+E\left(\frac{1}{p}\sum_{i=1}^{p}(m_i - \bar{Y}_i)^2\right) + \frac{2}{p}\sum_{i=1}^{p}E(X_i - m_i)E(m_i - \bar{Y}_i) \approx \sigma_x^2 + b^2 + E\left(\frac{1}{p}\sum_{i=1}^{p}(m_i - \bar{Y}_i)^2\right)$, assuming $E(m_i - \bar{Y}_i) \approx 0$, since the variance of X_i does not depend on i and X_i and $\bar{Y}_i$ are independent and using the deduction in footnote 9.

Example 2
Return to Example 1 with refractive indices. Suppose the laboratory has taken part in 10 proficiency tests and that their reported results are the values given in Example 1, i.e.

1.51996 1.52009 1.52006 1.52049 1.52008 1.52008 1.51967 1.52008 1.51981 1.52001

Now, the reported results from the other laboratories are summarised as

Proficiency test (i)	*No. participating labs (n_i)*	*Average of reported results ($\bar{y}_i$)*
1	10	1.52004
2	8	1.51992
3	11	1.52001
4	7	1.52003
5	9	1.51997
6	10	1.52010
7	9	1.52002
8	8	1.51986
9	10	1.51999
10	8	1.52000

$MSD = \frac{1}{10} \cdot [(1.51996 - 1.52004)^2 + (1.52009 - 1.51992)^2 + \cdots + (1.52001 - 1.52000)^2] \approx 4.65 \cdot 10^{-8}$
The expanded uncertainty with coverage 95% then becomes $U = 2 \cdot \sqrt{4.65 \cdot 10^{-8} + 0.00025^2} \approx 0.00066$

AI.4.3 Separating the bias contribution from the expanded uncertainty

MSD and MSD_w as defined in the previous sections account for both the average bias and the dispersion in bias as parameters in the model.

We use the same notation for the number of measurements/proficiency tests and the (reported) measurements as in the previous subsections.

Whether we use measurements within the laboratory (as in AI.4.2.1) or reported results from proficiency tests (as in AI.4.2.2), we obtain a set of differences between the measurement from the current laboratory and a reference value for the measurand. Instead of computing the mean square of these differences we can compute their mean absolute value, MAD (*mean absolute deviation*):

$$MAD_w = \frac{1}{p} \cdot \sum_{i=1}^{p} |d_{0i}| = \frac{1}{p} \cdot \sum_{i=1}^{p} |x_{0i} - m_{\mathrm{CRM}}| \quad (15a)$$

$$MAD = \frac{1}{p} \cdot \sum_{i=1}^{p} |d_i| = \frac{1}{p} \cdot \sum_{i=1}^{p} |x_i - \bar{y}_i| \quad (15b)$$

where the subscript "w" stands for that the MAD is computed from within-laboratory measurements only (cf. subsection AI. 4.2.1).

MAD_w and MAD both serve as predictions of the absolute bias, $|B|$ and can be denoted $\widehat{|B|}$. Now, the expanded uncertainty based on variance components only is for the simplified model (3) $U_V = k \cdot \sqrt{s_{R_w}^2}$ with k being the coverage factor used. For within-laboratory estimation we would add the variance component due to the dispersion reported by the material provider, i.e.

σ_{CRM}, giving $U_{V,w} = k \cdot \sqrt{\sigma_{\mathrm{CRM}}^2 + s_{R_w}^2}$. The *total uncertainty* (referred to as *total allowable error* in [22]) can then be computed as

$$U_{T,w} = |\widehat{B}| + U_V = |\widehat{B}| + k \cdot \sqrt{\sigma_{\mathrm{CRM}}^2 + s_{R_w}^2} \tag{16a}$$

$$U_T = |\widehat{B}| + U_V = |\widehat{B}| + k \cdot \sqrt{s_{R_w}^2} \tag{16b}$$

Example 3
Return to Example 1 and 2.

Using the laboratory's internal measurements and the certified value m_{CRM} with dispersion σ_{CRM}, we compute

$MAD_w = \frac{1}{10}[|1.51996 - 1.52000| + |1.52009 - 1.52000| + \cdots + |1.52001 - 1.52000|] \approx 0.000145$

and the total 95% uncertainty is

$U_{T,w} = 0.000145 + 2 \cdot \sqrt{0.0000002^2 + 0.00025^2} \approx 0.00065$

Using results from proficiency tests, we compute

$MAD = \frac{1}{10}[|1.51996 - 1.52004| + |1.52009 - 1.51992| + \cdots + |1.52001 - 1.52000|] \approx 0.000165$

and the total 95% uncertainty is then

$U_T = 0.000165 + 2 \cdot \sqrt{0.00025^2} \approx 0.00067$

AI.4.4 Expressions for general models

Throughout this section we have used the simplified model (3) when deducing expressions. However, with more complicated models where the contributing components are more than B and R_w the expressions are of course expanded. Generally, the expanded uncertainty incorporating the bias contribution can be written

$$U = k \cdot \sqrt{MSD + u_{c,V}^2} \tag{17}$$

where $u_{c,V}$ stands for the combined uncertainty from all components for which stable estimates of their variances are available (for within-laboratory estimation the component σ_{CRM}^2 is included) and where MSD is obtained either from within-laboratory measurements or from results from proficiency tests.

For instance, with the model $X = m + B + S + D + E$, where S refers to the variation due to drift ((in)stability), D refers to the variation due to the degree of reproducibility and E refers to the variation due to the degree of repeatability, and where the four random components (B, S, D and E) are assumed independent, the expression for the expanded uncertainty would be $U = k \cdot \sqrt{MSD + (\sigma_{CRM}^2) + u_S^2 + u_D^2 + u_E^2}$.

Similarly, to obtain the total uncertainty the general expression can be written

$$U_T = |\widehat{B}| + k \cdot \sqrt{u_{c,V}^2} \tag{18}$$

AI.5 RELATIVE VARIATION

AI.5.1 Theoretical aspects

It is the rule rather than the exception that the variation of chemical and physical measurements depends on the numerical magnitude of the measurand, since the measurement equipment cannot be assumed to provide a constant precision over the (sometimes very wide) range of potential values of the measurands. Therefore, it is common to use and estimate relative standard deviations rather than their absolute counterparts. For a random variable (*cf.* section AI.2) with average μ and standard deviation σ, its relative standard deviation is defined as

$$\%\sigma = 100 \cdot \frac{\sigma}{\mu} \tag{19}$$

The ratio σ/μ is known as the *coefficient of variation*, CV. Hence, if the coefficient of variation is assumed known for a random variable with average μ, then the absolute standard deviation of this random variable can be calculated as $\sigma = \mu \cdot CV = \mu \cdot \%\sigma/100$. This implies that with a known CV the absolute standard deviation can be estimated from a sample average, $\bar{x}$, as $\hat{\sigma} = \bar{x} \cdot CV$.

However, for the sum of two random variables, e.g. in the simplified model (3) the sum $B + R_w$, the coefficient of variation is not – like for variances – the square root of sum of the squared coefficients of variation for the components of the sum, i.e. $CV_{B+R_w} \neq \sqrt{CV_B^2 + CV_{R_w}^2}$.

An alternative way to express the relative variation is to relate all standard deviations to the target value of the variable of interest. To exemplify, let X be a random variable the average value of which is supposed to be m (but there may be systematic deviation present), and X can be written as the sum of two independent variables, X_A and X_B, i.e. $X = X_A + X_B$. Then, letting $W = X/m$, the following relation is obtained:

$$W = \frac{X_A}{m} + \frac{X_B}{m} \tag{20}$$

The variance of W is then

$$\sigma_W^2 = \frac{\sigma_{X_A}^2}{m^2} + \frac{\sigma_{X_B}^2}{m^2} \text{ but also } \sigma_W^2 = \frac{\sigma_X^2}{m^2} \tag{21}$$

Here, we can interpret σ_X/m as the relative standard deviation (relative to the target value), which can be expressed in percent if multiplicated by 100, and we note that its square is equal to the sum of the squared counterparts for X_A and X_B.

AI.5.2 Application to measurement uncertainty

The expressions presented in section AI.4 for estimating the expanded uncertainty account for absolute uncertainty but can be modified to account for relative uncertainty.

AI.5.2.1 Using within-laboratory measurements

When within-laboratory measurements are used to estimate the bias contribution, the value of the measurand is assumed known (m_{CRM}). The counterpart of MSD_w (expression (9)) for estimating *relative* bias contribution is then the *mean square relative deviation*

$$MSD_{w,r} = \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{d_i}{m_{CRM}}\right)^2 = \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{x_i - m_{CRM}}{m_{CRM}}\right)^2 = \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{x_i}{m_{CRM}} - 1\right)^2 \quad (22)$$

(subscript "r" refers to relative). However, note that $MSD_{w,r}$ can be rewritten as

$$MSD_{w,r} = \frac{1}{m_{\mathrm{CRM}}^2} \cdot \frac{1}{p} \cdot \sum_{i=1}^{p} (x_i - m_{\mathrm{CRM}})^2 = \frac{1}{m_{\mathrm{CRM}}^2} \cdot MSD_w$$

Moreover, the standard deviation of the variation due to the degree of reproducibility within laboratory (R_w) relative to m_{CRM} is simply $\sigma_{R_w}/m_{\mathrm{CRM}}$, so the expression for *expanded relative uncertainty* with joint estimation of the bias contribution is

$$U_{w,r} = \frac{k}{m_{\mathrm{CRM}}} \cdot \sqrt{MSD_w + \sigma_{\mathrm{CRM}}^2 + s_{R_w}^2} = \frac{1}{m_{\mathrm{CRM}}} \cdot U_w \quad (23)$$

The counterpart of MAD_w (expression (15a)) is the *mean absolute relative deviation*

$$MAD_{w.r} = \frac{1}{p} \cdot \sum_{i=1}^{p} \left|\frac{d_i}{m_{\mathrm{CRM}}}\right| = \frac{1}{p} \cdot \sum_{i=1}^{p} \left|\frac{x_i - m_{\mathrm{CRM}}}{m_{\mathrm{CRM}}}\right| = \frac{1}{p} \cdot \sum_{i=1}^{p} \left|\frac{x_i}{m_{\mathrm{CRM}}} - 1\right| \quad (24)$$

But analogously to the case for $MSD_{w,r}$, $MAD_{w,r}$ can be rewritten as

$$MAD_{w,r} = \frac{1}{|m_{\mathrm{CRM}}|} \cdot \frac{1}{p} \cdot \sum_{i=1}^{p} |x_i - m_{\mathrm{CRM}}| = \frac{1}{|m_{\mathrm{CRM}}|} \cdot MAD_w = \frac{1}{m_{\mathrm{CRM}}} \cdot MAD_w$$

since m_{CRM} is assumed to be > 0.

Hence, the relative bias can be predicted as $\hat{B}_r = MAD_w/m_{\mathrm{CRM}}$ and the expression for *total relative uncertainty* is

$$U_{T,w,r} = \frac{MAD_w}{m_{\mathrm{CRM}}} + \frac{k}{m_{\mathrm{CRM}}} \cdot \sqrt{\sigma_{\mathrm{CRM}}^2 + s_{R_w}^2} = \frac{1}{m_{\mathrm{CRM}}} \cdot \left(MAD_w + k \cdot \sqrt{\sigma_{\mathrm{CRM}}^2 + s_{R_w}^2}\right) = \frac{1}{m_{\mathrm{CRM}}} \cdot U_{T,w} \quad (25)$$

When an average of n measurements is used, $s_{R_w}^2$ is replaced by $s_{R_w}^2/n$.

Example 4
Return to Example 1 and 3.

Using internal measurements at the laboratory and the certified value m_{CRM}, the expanded relative uncertainty and the total relative uncertainty become

$$U_{r,w} = \frac{1}{1.52000} \cdot U_w = \langle \text{See Example 1} \rangle = \frac{1}{1.52000} \cdot 0.00065 \approx 0.00042 \ (= 4.2\%)$$

$$U_{T,w,r} = \frac{1}{1.52000} \cdot U_{T,w} = \langle \text{See Example 3} \rangle = \frac{1}{1.52000} \cdot 0.00065 \approx 0.00042 \ (= 4.2\%)$$

However, if the laboratory is always reporting relative uncertainties, it may be the case that $s_{R_w}^2$ is not explicitly known, but the laboratory has a stable estimate of its relative standard deviation for reproducibility, i.e. $\%\sigma_{R_w}$ and the corresponding coefficient of variation is then $CV_{R_w} = \%\sigma_{R_w}/100$. Moreover, the uncertainty of m_{CRM} may also be provided as a relative standard deviation $\%\sigma_{CRM}$ which the corresponding coefficient of variation $CV_{CRM} = \%\sigma_{CRM}/100$. Then instead of dividing everywhere with m_{CRM} as is done in expressions (23) and (25) above, we simply obtain expressions for the expanded and total relative uncertainties as

$$U_{w,r} = k \cdot \sqrt{MSD_{w,r} + CV_{\mathrm{CRM}}^2 + CV_{R_w}^2} \tag{26}$$

and

$$U_{T,w,r} = MAD_{w,r} + k \cdot \sqrt{CV_{\mathrm{CRM}}^2 + CV_{R_w}^2} \tag{27}$$

Note that a prerequisite for expressions (26) and (27) to be valid is that the coefficients of variation used are based on longitudinal monitoring of the variation, so that we can assume that $CV_{\mathrm{CRM}} \cdot m_{\mathrm{CRM}} \approx \sigma_{\mathrm{CRM}}$ and $CV_{R_w} \cdot m_{\mathrm{CRM}} \approx \sigma_{R_w}$ (cf. the theoretical points about coefficients of variation in Section 5.1).

Example 5
Again, return to Example 1 and 3. By dividing by the certified value $m_{CRM} = 1.52000$, we obtain $MSD_{w,r} = 4.18 \cdot 10^{-8}/1.52000$ and $MAD_{w,r} = 0.000145/1.52000$. Now, assume that the relative standard deviation $\%\sigma_{R_w}$ is appreciated to be 0.016% and that the relative standard deviation $\%\sigma_{CRM}$ is given to be 0.000014%. Then, expanded relative uncertainty with coverage 95% is

$$U_{r,w} = 2 \cdot \sqrt{4.18 \cdot 10^{-8}/1.52000 + \left(\frac{0.000014}{100}\right)^2 + \left(\frac{0.016}{100}\right)^2} \approx 0.00046 \ (= 0.046\%)$$

and the total relative uncertainty is

$$U_{T,w,r} = 0.000145/1.52000 + 2 \cdot \sqrt{\left(\frac{0.000014}{100}\right)^2 + \left(\frac{0.016}{100}\right)^2} \approx 0.00042 \ (= 0.042\%)$$

When an average of n measurements is used, $CV_{R_w}^2$ is replaced by $CV_{R_w}^2/n$.

AI.5.2.2 Using proficiency tests

The situation is more complicated when results from proficiency tests are used. The counterpart of MSD (expression (12)) would be

$$MSD_r = \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{d_i}{\bar{y}_i}\right)^2 = \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{x_i - \bar{y}_i}{\bar{y}_i}\right)^2 = \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{x_i}{\bar{y}_i} - 1\right)^2 \tag{28}$$

and the counterpart of MAD (expression 15(b)) would be

$$MAD_r = \frac{1}{p} \cdot \sum_{i=1}^{p} \left|\frac{d_i}{\bar{y}_i}\right| = \frac{1}{p} \cdot \sum_{i=1}^{p} \left|\frac{x_i - \bar{y}_i}{\bar{y}_i}\right| = \frac{1}{p} \cdot \sum_{i=1}^{p} \left|\frac{x_i}{\bar{y}_i} - 1\right| \tag{29}$$

Since the divisors of the terms in the sum in expressions (28) and (29) vary with the proficiency tests, a straightforward deduction of expressions for the expanded uncertainty like the ones of expressions (23) and (25) is not possible. A workaround solution is provided in section 5.2.3, but here we instead assume that relative standard deviation $\%\sigma_{\mathrm{R_w}}$ (and coefficient of variation $\mathrm{CV}_{\mathrm{R_w}} = \%\sigma_{\mathrm{R_w}}/100$) are used for the contribution from reproducibility within laboratory. The expression for the expanded relative uncertainty with joint estimation of the bias contribution then becomes

$$U_r = k \cdot \sqrt{MSD_r + CV_{R_w}^2} \tag{30}$$

and the expression for the total relative uncertainty becomes

$$U_{T,r} = MAD_r + k \cdot CV_{R_w} \tag{31}$$

Note that analogously to the expressions in section 4.2.2 we have not included a component reflecting variation in the certified values. We again refer to literature (e.g. [1]) for methods appreciating this contribution to the dispersion.

Example 6
Return to Example 2 (and 1). Using the data from the proficiency tests we can calculate

$$MSD_r = \frac{1}{10} \cdot \left[\left(\frac{1.51996 - 1.52004}{1.52004}\right)^2 + \left(\frac{1.52009 - 1.51992}{1.51992}\right)^2 + \cdots + \left(\frac{1.52001 - 1.52000}{1.52000}\right)^2\right] \approx 3.06 \cdot 10^{-8}$$

and

$$MAD_r = \frac{1}{10} \cdot \left[\left|\frac{1.51996 - 1.52004}{1.52004}\right| + \left|\frac{1.52009 - 1.51992}{1.51992}\right| + \cdots + \left|\frac{1.52001 - 1.52000}{1.52000}\right|\right] \approx 0.000109$$

Assume (as in Example 5) that the relative standard deviation $\%\sigma_{R_w}$ is appreciated to be 0.016% and that the relative standard deviation $\%\sigma_{CRM}$ is given to be 0.000014%. Then, expanded relative uncertainty with coverage 95% is

$$U_r = 2 \cdot \sqrt{3.06 \cdot 10^{-8} + \left(\frac{0.016}{100}\right)^2} \approx 0.00047\ (= 0.047\%)$$

and the total relative uncertainty is

$$U_{T,r} = 0.000109 + 2 \cdot \frac{0.016}{100} \approx 0.00043\ (= 0.043\%)$$

When an average of n measurements is used, $CV_{R_w}^2$ is replaced by $CV_{R_w}^2/n$.

AI.5.2.3 Alternatives to using coefficients of variation

Expression (28) cannot be rewritten the same way as with expression (22) since the divisor ($\bar{y}_i$) varies with i. Moreover, the $\bar{y}_i$'s cannot be used to form an expression for the relative standard deviation based on s_{R_w}.

With the assumption that $\bar{y}_i$ is very close to the target value m_i, a workaround for this problem is to substitute the overall mean of the reported results from all proficiency tests for $\bar{y}_i$. The overall (unweighted) mean is

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{p} n_i \cdot \bar{y}_i}{\sum_{i=1}^{n} n_i} \tag{32}$$

However, if the n_i's are similar in value, $\bar{\bar{y}} \approx (1/p) \cdot \sum_{i=1}^{p} \bar{y}_i$ (the weighted mean[14]). With this substitution MSD_r (expression (28)) is approximated as

$$MSD_r \approx \frac{1}{p} \cdot \sum_{i=1}^{p} \left(\frac{x_i - \bar{y}_i}{\bar{\bar{y}}}\right)^2 = \left(\frac{1}{\bar{\bar{y}}}\right)^2 \cdot \frac{1}{p} \cdot \sum_{i=1}^{p} (x_i - \bar{y}_i)^2 = \left(\frac{1}{\bar{\bar{y}}}\right)^2 \cdot MSD \tag{33}$$

It should be noted that if we now estimate the relative standard deviation of the variation due to the degree of reproducibility by $s_{R_w}/\bar{\bar{y}}$, we introduce a random component ($\bar{\bar{y}}$) that jeopardizes further the use of the intervals of expression (8) for finding the coverage factor. We have earlier stated that s_{R_w} should be a stable estimate (based on 50 measurements or more). The division by $\bar{\bar{y}}$ will make $s_{R_w}/\bar{\bar{y}}$ less stable compared to s_{R_w}/m_{CRM}. However, since the estimated standard uncertainty will be an overestimation (cf. subsection 4.2), this is probably a negligible issue when joint estimation of the bias contribution is used.

An expression for the approximate relative expanded uncertainty using results from proficiency tests are used can thus be

$$U_r \approx \frac{k}{\bar{\bar{y}}} \cdot \sqrt{MSD + s_{R_w}^2} = \frac{1}{\bar{\bar{y}}} \cdot U \tag{34}$$

Substituting the overall mean for $\bar{y}_i$, MAD_r (expression (29)) is approximated as

$$MAD_r \approx \frac{1}{|\bar{\bar{y}}|} \cdot \frac{1}{p} \cdot \sum_{i=1}^{p} |x_i - \bar{y}_i| = \frac{1}{\bar{\bar{y}}} \cdot MAD \tag{35}$$

since we can assume that $\bar{\bar{y}} > 0$. Analogously to the deduction of expression (34), the total relative uncertainty could be estimated as

$$U_{T,r} \approx \frac{1}{\bar{\bar{y}}} \cdot \left(MAD + k \cdot \sqrt{s_{R_w}^2}\right) = \frac{1}{\bar{\bar{y}}} \cdot U_T \tag{36}$$

[14] It may look strange that $(1/p) \cdot \sum_{i=1}^{p} \bar{y}_i$ is referred to as a *weighted* mean when no weights are visible, but the point is that if the n_i's are different then $\bar{y}_i$'s based on fewer results (than the average of the n_i's) are upweighted in $(1/p) \cdot \sum_{i=1}^{p} \bar{y}_i$ while $\bar{y}_i$'s based on more results are downweighted.

However, since $\bar{\bar{y}}$ introduce further uncertainty into this expression (compared to expression (28)), the coverage factor may not give what is stated in the intervals (8). A remedy here could be to replace $\bar{\bar{y}}$ by the minimum of the $\bar{y}_i$s, which would increase $U_{T,r}$ to a level more reliable. When an average of n measurements is used, $s_{R_w}^2$ is replaced by $s_{R_w}^2/n$.

Example 7
Return to Example 2 and 3. From the results from the proficiency tests we can see that the number of participating laboratories between the tests vary. In the number of laboratories, the variation is not that big, but since the numbers are quite small (10 or lower), the relative variation between them must be considered substantial. Hence, we should use the overall mean in the calculations.

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{10} n_i \cdot \bar{y}_i}{\sum_{i=1}^{10} n_i} = \frac{10 \cdot 1.52004 + 8 \cdot 1.51992 + \cdots + 8 \cdot 1.52000}{10 + 8 + \cdots + 8} \approx 1.519997$$

With this mean we calculate

$$U_r = \frac{1}{1.519997} \cdot U = \langle \textit{See Example 2} \rangle = \frac{1}{1.519997} \cdot 0.00066 \approx 0.00043 = 0.043\%$$

and

$$U_{T,r} = \frac{1}{1.519997} \cdot U_T = \langle \textit{See Example 3} \rangle = \frac{1}{1.519997} \cdot 0.00067 \approx 0.00044 = 0.044\%$$

A more conservative estimate of $U_{T,r}$ is obtained by replacing $\bar{\bar{y}}$ by $min\{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{10}\} = 1.51986$, which gives

$$U_{T,r} = \frac{1}{1.51986} \cdot 0.00067 \approx 0.00044 = 0.044\%$$

(but obviously this makes no significant difference here – they start differing in the 7th decimal).

AI.5.3 Expressions for general models

Like in subsections AI.4.2 and AI.4.3, the expressions in the previous section are developed assuming the simplified model (3), i.e. $X = m + B + R_w$.

For more complicated models comprising several variance components in the combined uncertainty, the general expressions corresponding to expressions (23), (25), (30) and (31) would be

$$U_{w,r} = \frac{k}{m_{\text{CRM}}} \cdot \sqrt{MSD_w + u_{c,V}^2} \tag{37a}$$

$$U_{T,w,r} = \frac{1}{m_{\text{CRM}}} \cdot \left(MAD_w + k \cdot \sqrt{u_{c,V}^2} \right) \tag{37b}$$

$$U_r = k \cdot \sqrt{MSD_r + u_{c,V}^2} \tag{37c}$$

$$U_{T,r} = MAD_r + k \cdot \sqrt{u_{c,V}^2} \tag{37d}$$

where $u_{c,V}$ stands for the combined uncertainty from all components for which stable estimates of their variances are available (cf. subsection AI.4.4).

AI.6 A NOTE ON INTERVALS AND THE CHOICE OF COVERAGE FACTOR

In section AI.4.1 we stated that the interval $X \pm 2 \cdot \sigma$, where σ is the true standard deviation of X, is a 95% confidence interval for the average, μ, of X. The interpretation of such an interval is that it will cover μ with a confidence of 95%. Theoretically, if we would repeat calculating such an interval procedure for each measurement we take, 95% of the intervals will cover μ.

In practice, the standard deviation σ is not known. For a single study comprising a limited sample of n repeated measurements $(x_1, x_2, \ldots, x_n)$, we would use their sample average $\bar{x}_{(n)} = (1/n)\sum_1^n x_i$ as the reported result. The sample average seen as randomly varying from sample to sample has the nice property that its standard deviation is $\sigma/\sqrt{n}$ (hence, the more measurements it is built on, the more stable it will be). The corresponding 95% confidence interval would then be $\bar{x}_{(n)} \pm 2 \cdot \sigma/\sqrt{n}$, but σ is of course still generally unknown. Since we have several measurements, a natural consideration would be to replace σ by the sample standard deviation $s = \sqrt{\frac{1}{n-1} \cdot \sum_1^n \left(x_i - \bar{x}_{(n)}\right)^2}$ in the expression for the confidence interval. However, since this would introduce more variation (both $\bar{x}_{(n)}$ and s would vary from sample to sample), the standard coverage factors (2 above) do not apply.

The solution to this problem is to use the so-called (***Student's***) *t-distribution*[15] with which the coverage factor to be used will depend on n. It shall be said, though, that for relatively small values of n the coverage factors are substantially higher than the ones used when σ is known. However, the *t*-distribution only applies to measurements that are normally distributed. Moreover, when the standard deviation (or rather the variance σ^2) is decomposed into several contributing components (that is typical for appreciating source of measurement error), the *t*-distribution does not apply[16]. Several approximations have been suggested over the years, where the approximation lies in finding a *t*-distribution that is close the underlying distribution that would apply but is not deductible[17]. However, for this to work the variance components must have estimates that from a random point of view vary according to so-called *central* χ^2*-distributions* ("chi-square").

For a standard uncertainty of the kind that we have taken up in the previous sections, e.g. one that can be written $u = \sqrt{MSD + u_{c,V}^2}$, that last requirement is not fulfilled since MSD does not possess such a distribution. Moreover, if one or several components of $u_{c,V}^2$ are estimated using coefficients of variation (see subsection AI.5.1) these distributions do not apply either.

We strongly advise that substitutions for unknown variance components should be stable estimates based on larger sets of previous measurements. They do then serve as negligibly varying substitutes and the coverage factors valid for the theoretical confidence interval can be applied with good approximation.

When working with models using relative standard deviations it is not advised to work with t-distributions with so-called 'effective degrees of freedom'. Besides the fact that the model of sums of chi-square type distributions is not met if MSD is included in the sum under the square root, the fact that relative measurements are random in both numerator and denominator makes the model of Satterthwaite, which is already complicated, basically unapplicable.

[15] This distribution was deduced by the Irish statistician W. Gosset in the early 1900s (but he published his results using the pseudonym "Student") *Biometrika* 6: 1:25

[16] In the statistical literature this is referred to as *Behren-Fisher's problem*.

[17] Using Satterthwaite's approach to finding the so-called *degrees of freedom* for the sum of the estimated variance components. *Biometrics* 2(6):110-114

ANNEX II. STATISTICAL SPECIFICATION OF MUCALC VERSION 3.1

AII.1 INTRODUCTION

The International Vocabulary of Metrology [4] defines uncertainty as “a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used”. The quantification of the uncertainty of a measurand gives a range of values that contain the true value.

Any analytical measurement has uncertainty associated with it. A forensic toxicologist reports the concentration of an analyte, e.g., THC, in $\mu g/L$ together with its uncertainty accumulated through the process of producing the measurement [23, 24]. The concentration measurement together with its uncertainty is considered when comparing it to a legal limit in court to determine the severity of an offence.

The core statistical method for measure uncertainty (MU) described in this document has been published in [19]. The sources of uncertainty that affect MU of concentration of an analyte are: (1) homogeneity, (2) calibration curve, (3) method precision, (4) calibration standards, and (5) sample preparation. The calculation of the MU for these sources are described in dedicated sections. The aggregation of the MU and its applicate to calculate a confidence interval of the true concentration is given in the next section.

l	Sources of uncertainty
1	Homogeneity
2	Calibration curve
3	Method precision
4	Calibration standards
5	Sample preparation

Table AII-1: Indices for the sources of uncertainty

AII.2 HOMOGENEITY

Homogeneity/heterogeneity, as defined in EURACHEM/CITAC guide [25] and IUPAC recommendations 1990 [26], is:

"The degree to which a property or constituent is uniformly distributed throughout a quantity of material. Note:

- A material may be homogeneous with respect to one analyte or property but heterogeneous with respect to another.
- The degree of heterogeneity (the opposite of homogeneity) is the determining factor of sampling error."

The property/constituent measured here is the concentration of analyte.

In this section in addition to the specification of the RSU of Homogeneity, $u_r(1)$, a test for homogeneity is described because is one of the features of MUCalc.

AII.2.1 Homogeneity test

A one-way analysis of variance (ANOVA) is used to test the null hypotheses, H_0, of equality of means among sample groups against the alternative hypothesis, H_1, that at least two of the group means differ, on the assumption that samples are normally distributed, have equal variance and are independent [27-30].

The data consists of measurements $N = n_1 + n_2 + \cdots + n_j$ measurements $\{X_{i,j}: i = 1,2, \dots, n_j; j = 1,2, \dots, k\}$ of k groups with means $m_1, m_2, \dots m_k$. The hypotheses of interest are

$$H_0: m_1 = m_2 = \cdots = m_k$$
$$H_1: m_i \neq m_j \ \ for\ some\ i, j$$

The test is based on a statistic F_s and a critical value F_c, both described below. H_0 is rejected if $F_s \geq F_c$. If $F_s < F_c$, we fail to reject the null hypothesis, which means that there is no evidence that the means differ. There should be at least two groups to compare with at least two replicates in each group.

The statistic F_s is defined as

$$F_s = \frac{MSS_B}{MSS_W},$$

where,

$$MSS_B = \frac{\sum_{j=1}^{k} n_j \left(\underline{X}_j - \underline{X}_T\right)^2}{k-1} \quad \text{and} \quad MSS_W = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(X_{ij} - \underline{X}_j\right)^2}{N-k}. \tag{1}$$

$\underline{X}_j$ is the mean of measurement in group j

$$\underline{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}$$

$\underline{X}_T$ is the grand mean of all measurements

$$\underline{X}_T = \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{ij} \tag{2}$$

Under H_0, F_s follows and F distribution with degrees of freedom $\nu_W = k - 1$ and $\nu_B = N - k$. The critical value F_c is such that $Pr\ (X < F_c) = 1 - \alpha$, where X follows an F distribution with degrees of freedom ν_B and ν_W. F_c is obtained from the inverse of the CDF of an F-Distribution, available in most statistical packages.

AII.2.2 Homogeneity uncertainty

Homogeneity uncertainty quantifies the uncertainty associated with the between-group homogeneity where differences among sample groups are of interest [27-31]. Homogeneity uncertainty is measured with the RSU for $l = 1$,

$$u_r(1) = \frac{u(1)}{\underline{X}_T}.$$

$\underline{X}_T$ is the grand mean, Eqn. (2),

$$u(1) = \max\{u_a(1), u_b(1)\},$$

Where

$$u_a(1) = \sqrt{\frac{MSS_B - MSS_W}{n_0}},$$

And

$$u_b(1) = \sqrt{\frac{MSS_W}{n_0}} \times \sqrt{\frac{2}{k(n_0 - 1)}}.$$

MSS_W and MSS_B are defined in Eqn. (1), while n_0 is

$$n_0 = \frac{1}{k-1}\left[N - \frac{\sum_{j=1}^{k} n_j^2}{N}\right].$$

AII.3 CALIBRATION CURVE

A calibration curve in this context is a graph that describes the relationship between instrument response/peak area ratio, Y, and concentration, X. Uncertainty of calibration curve arises when a regression is used to generate a calibration model, the calibration curve is used in reverse form where concentration X is predicted from the instrument response Y. This prediction has an associated uncertainty termed uncertainty of calibration curve and it is expressed as a confidence interval that can be calculated using the RSU $u_r(2)$.

MUCalc provides three types of calibration curves: linear regression, weighted linear regression and quadratic linear regression.

AII.3.1 Linear regression

Linear regression is the most commonly used statistical method in calibration [32]. It is used when the relationship between peak area ratio and concentration is linear and satisfies the assumptions of a regression model: homoscedasticity and normality.

A linear regression is of the form

$$y = b_0 + b_1 x + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

where x, y are the explanatory and response variables. The data for estimating the parameters are a set of concentrations of the standard, $\{x_i : i = 1,2, \dots, n\}$ and their associated peak area ratios $\{y_i : i = 1,2, \dots, n\}$. The specification of $u_r(2)$ requires the following quantities:

i. the average concentration $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;
ii. the set of predicted peak height ratio $\{\hat{y}_i : i = 1,2, \dots, n\}$;
iii. the estimated slope $\hat{b}_1$;
iv. the standard error of regression

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}};$$

v. the sum of squares deviation

$$S_{xx} = \sum_{i}^{n}(x_i - \bar{x})^2.$$

The RSU of calibration curve, $u_r(\boldsymbol{x}_s)$, for a given case sample mean concentration, $\boldsymbol{x}_s$, is

$$u_r(\boldsymbol{x}_s) = \frac{u(\boldsymbol{x}_s)}{\boldsymbol{x}_s},$$

where $u(\boldsymbol{x}_s)$ is the standard uncertainty of the calibration curve [33],

$$u(\boldsymbol{x}_s) = \frac{S_{y/x}}{b_1}\sqrt{\frac{1}{r_s} + \frac{1}{n} + \frac{(\boldsymbol{x_s} - \bar{x})^2}{S_{xx}}},$$

and r_s is the number of replicates made on the test sample to determine $\boldsymbol{x}_s$.

AII.3.2 Weighted linear regression

Weighted linear regression (WLR) is used if the standard deviation of peak area ratio correlates with the magnitude of the concentration. The specification of RSU of calibration curve $u_r(2)$ requires the following quantities either in addition to or that differ from a linear regression:

i. a set of weights $\{W_i : i = 1,2, \dots, n\}$;
ii. a set of standardised weights $\{w_i : i = 1,2, \dots, n\}$ where

$$w_i = W_i \times \frac{n}{\sum_{i=1}^{n} W_i};$$

iii. the standard error

$$S_{w_{y/x}} = \sqrt{\frac{\sum_{i=1}^{n} w_i(y_i - \hat{y}_i)^2}{n-2}};$$

iv. the sum of squares deviation

$$S_{xx_w} = \sum_{i}^{n} w_i \, (x_i - \bar{x})^2.$$

The RSU $u_r(2)$ for a mean concentration $\boldsymbol{x_s}$ is

$$u_r(2) = \frac{u(2)}{\boldsymbol{x}_s}$$

Where [35,36]

$$u(2) = \frac{S_{w_{y/x}}}{b_1} \sqrt{\frac{1}{w_s r_s} + \frac{1}{n} + \frac{(\boldsymbol{x_s} - \bar{x}_w)^2}{S_{xx_w}}}.$$

$\bar{x}_w$ is the weighted mean value of concentrations given by

$$\bar{x}_w = \frac{1}{n} \sum_{(i=1)6n} w_i x_i,$$

and w_s is the standardised weight of $\boldsymbol{x_s}$.

AII.3.3 Quadratic regression

Quadratic regression is used when the peak height ration follows a quadratic polynomial as a function of concentration. Using the approach described by [33], a quadratic regression can be written as

$$y - \bar{y} = b_1(x - \bar{x}) + b_2(x^2 - \overline{x^2}) + \epsilon, \quad \epsilon \sim N(0, \sigma^2),$$

to make the regression curve to start from the origin ($b_0 = 0$).

Given an instrument response of case sample peak area ratio, y_s, the level of concentration $\boldsymbol{x_s}$ is estimated by solving for x as while treating $\bar{x}$ and $\overline{x^2}$ as constants,

$$\hat{\boldsymbol{x}}_{\boldsymbol{s}} = \frac{-b_1 \sqrt{b_1^2 - 4b_2(\bar{y} - y_s - b_1\bar{x} - b_2\overline{x^2})}}{2b_2}. \qquad (3)$$

The standard uncertainty of calibration curve, $u(2)^2$ is the same as $Var(\hat{\boldsymbol{x}}_{\boldsymbol{s}})$ and is obtained by applying Taylor's theorem

$$u(2)^2 = \left(\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{db_1}\right)^2 Var(b_1) + \left(\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{db_2}\right)^2 Var(b_2) + \left(\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{d\bar{y}}\right)^2 Var(\bar{y}) + \left(\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{dy_s}\right)^2 Var(y_s) + 2\left(\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{db_1}\right)\left(\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{db_2}\right) Cov(b_1, b_2)$$

The partial derivatives are obtained by differentiating Eqn. (3) with respective to b_1, b_2, $\bar{y}$ and y_s:

$$\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{db_1} = \frac{-1 + {}^{1}/_{2}\, D^{-1/2}(2b_1 + 4b_2\bar{x})}{b_2}$$

$$\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{db_2} = \frac{b_1 - D^{1/2}}{2b_2^2} + \frac{{}^{1}/_{2}\, D^{-1/2}(4y_s - 4\bar{y} + 4b_1\bar{x} + 8b_2\overline{x^2})}{2b_2}.$$

$$\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{d\bar{y}} = -D^{-1/2}.$$

$$\frac{d\hat{\boldsymbol{x}}_{\boldsymbol{s}}}{dy_s} = D^{-1/2}.$$

D is the discriminant of x,

$$D = b_1^2 - 4b_2(\bar{y} - y_s - b_1\bar{x} - b_2\overline{x^2}).$$

$Var(b_1)$, $Var(b_2)$ and $Cov(b_1, b_2)$ can be obtained from the covariance matrix described in section AII.9.

The variance of $\bar{y}$ and y_s are

$$Var(\bar{y}) = \frac{S_{y/x}^2}{n} \text{ and } Var(y_s) = \frac{S_{y/x}^2}{r_s}$$

where

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-3}}.$$

AII.3.4 Pooled standard error of regression

If quality control data is available that considers different laboratory conditions over different days, the standard error of regression of these calibration curves can be pooled together to obtain a better estimate. The pooled estimate can be calculated when the same type of regression model is used in all the calibration curves.

The pooled standard error of regression for m calibration curves each with n_j data points and standard error of regression $S_{y/x_{(j)}}$, $j = 1,2,\ldots,m$, is

$$S_p = \sqrt{\frac{\sum_{j=1}^{m}(n_j - 1)S_{y/x_{(j)}}^2}{\sum_{j=1}^{m}(n_j - 1)}}.$$

S_p replaces $S_{y/x}$ in the calculation of the standard uncertainty of calibration, $u(2)$.

AII.4 METHOD PRECISION

Method precision measures the variability in the results of a repeated experiment under similar conditions. Given the repeated analysis of a given sample, the variability associated with a method can be quantified.

The uncertainty of method precision, $u(3)$, for a case sample with mean concentration $\boldsymbol{x_s}$ is obtained from a set of method precision uncertainties $\{u(3, NV): NV = 1,2, \dots, n_{NV}\}$ calculated for a set of concentration nominal values. Specifically, the standard uncertainty for method precision is set to $u(3) = u(3, NV^*)$ where NV^* is the closest value to $\boldsymbol{x_s}$:

$$NV^* = \arg\min_{NV}\{|NV - \boldsymbol{x_s}|: NV = 1,2, \dots, n_{NV}\}.$$

The data for calculating $u(3, NV^*)$ consist of concentrations measurements obtained from n_{runs} and n_i measurement in each run. Specifically,

$$\{x_{NV^*,i,j}: i = 1,2, \dots, n_{runs}; j = 1,2, \dots, r_s\}$$

The standard deviation for each run, $S_{NV,i}$. The spooled standard deviation is then calculated

$$S_{p(NV^*)} = \sqrt{\frac{\sum_{i=1}^{n_{runs}} S_{NV^*,i}^2 \times (r_s - 1)}{\sum_{i=1}^{n_{runs}} (r_s - 1)}}.$$

The standard uncertainty of method precision is then calculated as [34],

$$u(3) = u(3, NV^*) = \frac{S_{p,NV^*}}{\sqrt{r_s}}.$$

The RSU of method precision is then calculated as

$$u_r(3) = \frac{u(3)}{\bar{x}_{NV^*}},$$

where $\bar{x}_{NV^*}$ is the mean concentration of all samples for NV^* across all runs,

$$\bar{x}_{NV^*} = \frac{1}{n_{runs} r_s} \sum_{i=1}^{n_{runs}} \sum_{j=1}^{r_s} x_{NV^*,i,j}$$

AII.5 CALIBRATION STANDARD

The uncertainty associated with calibration standard combines the uncertainty from the reference materials and the solution preparation. The uncertainty from the reference materials is stated in the certificates of analysis of certified reference materials (CRMs) while the uncertainties in solution preparation comes from inaccuracies of the measuring equipment, e.g. pipettes and volumetric flasks, used to dilute CRMs and spike blank samples when preparing solutions for a calibration curve.

The quantification of the uncertainties in the preparation process requires information on the steps involved in the solution preparation and details of the equipment used. These steps may

be different from laboratory to laboratory. Table AII-2 shows an example of the structure of calibration standard preparation.

Solution	Notation	Parent Solution	Equipment	Number of Equipment	Volume Tolerance Coverage	No. of Times Used
Stock Solution A	SSA	RSS	$Eq_{SSA,i}$	$N_{Eq,SSA}$	$Eq_{SSA,i,vol}$ $Eq_{SSA,i,tol}$ $Eq_{SSA,i,cov}$	$N_{Eq,SSA,i}$
Working Solution B	WSB	SSA	$Eq_{WSB,i}$	$N_{Eq,WSB}$	$Eq_{WSB,i,vol}$ $Eq_{WSB,i,tol}$ $Eq_{WSB,i,cov}$	$N_{Eq,WSB,i}$
Working Solution C	WSC	SSA	$Eq_{WSC,i}$	$N_{Eq,WSC}$	$Eq_{WSC,i,vol}$ $Eq_{WSC,i,tol}$ $Eq_{WSC,i,cov}$	$N_{Eq,WSB,i}$
Calibrator Range 1	CR_1	WSC	$Eq_{CR1,i}$	$N_{Eq,CR1}$	$Eq_{CR1,i,vol}$ $Eq_{CR1,i,tol}$ $Eq_{CR1,i,cov}$	$N_{Eq,CR1,i}$
Calibrator Range 2	CR_2	WSB	$Eq_{CR2,i}$	$N_{Eq,CR2}$	$Eq_{CR2,i,vol}$ $Eq_{CR2,i,tol}$ $Eq_{CR2,i,cov}$	$N_{Eq,CR2,i}$

Table AII-2: Solutions and their required information. In addition, the reference standard solution (RSS) purity (RSS_purity), tolerance (RSS_tol) and coverage (RSS_cov) are also required.

For example, the preparation for stock solution A, denoted SSA, requires a set of equipment consisting of $N_{Eq,SSA}$ pipettes and flasks, denoted $\{Eq_{SSA,i}: i = 1,2,\dots,N_{Eq,SSA}\}$. There are other quantities needed for each equipment and these are volume ($Eq_{SSA,i,vol}$), coverage ($Eq_{SSA,i,cov}$), and the number of times that this equipment is used ($N_{Eq,SSA,i}$).

A tree diagram can be drawn from the input file to represent the dependence of solution preparation where an arrow from solution 1 to solution 2 means that solution 2 is made from solution 1. For example, in
Figure AII-1, *RSS* is the parent solution of *SSA* which means that SSA is made from RSS.

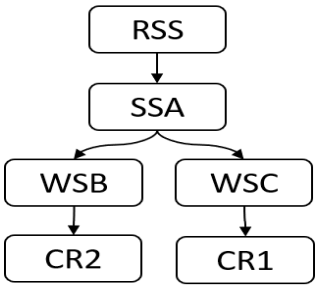

Figure AII-1: A diagrammatic representation of the preparation of solutions from other solutions based on the input file.

The diagram in
Figure *AII-1* is used to describe the calculation of the RSU of calibration standards, $u_r(4)$:

$$u_r(4) = \sqrt{\{u_r(CR_1)\}^2 + \{u_r(CR_2)\}^2}$$

The calculation of $u_r(CR_i)$, $i \in \{1,2\}$, is achieved by calculating first the RSU of the founder node *RSS* and then the RSU of the solutions in following the arrows from *RSS*.

The RSU for *RSS* is calculated with

$$u_r(RSS) = \frac{u(RSS)}{RSS_{purity}} = \frac{\frac{RSS_{tol}}{RSS_{cov}}}{RSS_{purity}}.$$

The RSU for the rest of the solutions is obtained iteratively based on the RSU of the parent solution and on the RSU of all the equipment used:

$$u_r(Solution) = \sqrt{\{u_r(ParentSolution)\}^2 + \sum_{i=1}^{N_{Eq,Solution,i}} \left[\left\{u_r\left(Eq_{Solution,i}\right)\right\}^2 + N_{Eq,Solution,i}\right]}$$

The RSU for equipment $Eq_{Solution,i}$ is calculated with the formula,

$$u_r\left(Eq_{Solution,i}\right) = \frac{\frac{Eq_{Solution,i,tol}}{Eq_{Solution,i,cov}}}{Eq_{Solution,i,vol}}.$$

Using the formula for the RSU of a solution applied to *SSA*,

$$u_r(SSA) = \sqrt{\{u_r(RSS)\}^2 + \sum_{i=1}^{N_{Eq,SSA}} \left[\left\{u_r\left(Eq_{SSA,i}\right)\right\}^2 \times N_{Eq,SSA,i}\right]}.$$

Then

$$u_r(WSB) = \sqrt{\{u_r(SSA)\}^2 + \sum_{i=1}^{N_{Eq,WSB}} \left[\left\{u_r\left(Eq_{WSB,i}\right)\right\}^2 \times N_{Eq,WSB,i}\right]},$$

and,

$$u_r(WSC) = \sqrt{\{u_r(SSA)\}^2 + \sum_{i=1}^{N_{Eq,WSC}} \left[\left\{u_r\left(Eq_{WSC,i}\right)\right\}^2 \times N_{Eq,WSC,i}\right]}.$$

Finally,

$$u_r(CR1) = \sqrt{\{u_r(WSC)\}^2 + \sum_{i=1}^{N_{Eq,WSC}} \left[\left\{u_r\left(Eq_{CR_1,i}\right)\right\}^2 \times N_{Eq,CR_1,i}\right]},$$

and,

$$u_r(CR2) = \sqrt{\{u_r(WSB)\}^2 + \sum_{i=1}^{N_{Eq,WSC,i}} \left[\{u_r(Eq_{CR_2,i})\}^2 \times N_{Eq,CR_2,i}\right]}.$$

AII.6 SAMPLE PREPARATION

The RSU of sample preparation, $u_r(5)$, combines uncertainty sources from the use of equipment in preparing a sample, such as weighing balance, pipette, and volumetric flask.
The RSU of sample preparation is

$$u_r(5) = \sqrt{\sum_{i=1}^{N_{Eq,sample,i}} \left[\{u_r(Eq_{sample,i})\}^2 \times N_{Eq,sample,i}\right]}.$$

where

$$u_r(Eq_{sample,i}) = \frac{\frac{Eq_{sample,i,tol}}{Eq_{sample,i,cov}}}{Eq_{sample,i,cap}}.$$

and $N_{Eq,sample,i}$ is the number of times that equipment $Eq_{sample,i}$ is used in the preparation of a given sample.

AII.6.1 Combined Uncertainty

The combined uncertainty u_c is obtained by combining all the individual uncertainty components for which data is uploaded for.

If data is uploaded for all the uncertainty components; Homogeneity, Calibration Curve, Method Precision, Calibration Standard and Sample Preparation, relative standard uncertainty is computed for each uncertainty component using the methods described above in section AII.1-AII.5, and are combined to obtain the overall uncertainty of the analytical process

For l uncertainty sources/components with individual standard uncertainty u(l), the combined uncertainty, u_c, for a given case sample mean concentration, $\boldsymbol{x_s}$, is given by

$$u_c = \boldsymbol{x_s}\sqrt{\sum_l u_r(l)^2}\,, \qquad (1)$$

AII.7 COVERAGE FACTOR AND EFFECTIVE DEGREES OF FREEDOM

A coverage factor is a number chosen to determine the level of confidence to be associated with data points within a desired standard deviation. Alternatively, A coverage factor, k, for specified level of confidence, CL, and effective degrees of freedom, v_{eff}, is a number, $k_{v_{\text{eff}},CL\%}$, usually greater than one from which an expanded uncertainty, U_{exp}, is obtained when multiplied by a combined standard uncertainty, u_c.

To determine a suitable coverage factor, a specified level of confidence, CL, is required along with knowledge about the effective degrees of freedom, v_{eff}, of all uncertainty components.

An effective degree of freedom is computed using the Welch-Satterthwaite equation [32] given by

$$v_{\text{eff}} = \frac{\left(\frac{u_c}{x_s}\right)^4}{\sum \frac{u_r(l)^4}{v_{(l)}}}$$

The derived effective degrees of freedom along with the specified $CL\%$ is used to read a value termed coverage factor, $k_{v_{\text{eff}},CL\%}$, from the T-Distribution. Alternatively, MUCalc allows one to specify a number directly for the coverage factor.

AII.8 EXPANDED UNCERTAINTY

The expanded uncertainty, U_{exp}, is the final step of measurement uncertainty computation. This is done to derive a confidence interval believed to contain the true unknown value.

The expanded uncertainty is computed by multiplying the combined uncertainty, u_c, with the coverage factor $k_{v_{\text{eff}},CL\%}$

$$U_{exp} = u_c \text{ x } k_{v_{\text{eff}},CL\%}$$

The percentage expanded uncertainty, $\%U_{exp}$, is given by

$$\%U_{exp} = \frac{U_{exp}}{\boldsymbol{x_s}} \text{ x } 100$$

The confidence interval of the true concentration is calculated as

$$\left(\boldsymbol{x_s} - u_c k_{v_{\text{eff}},CL\%},\ \boldsymbol{x_s} + u_c k_{v_{\text{eff}},CL\%}\right).$$

AII.9 COVARIANCE MATRIX

Given a regression model of the form

$$y = b_0 + b_1 x_1 + b_2 z$$

The design matrix is of the form

$$X=\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & x_4 & \cdots & x_n \\ z_1 & z_2 & z_3 & z_4 & \cdots & z_n \end{bmatrix}$$

The covariance matrix is given by

$$\sigma^2 \times (X^T X)^{-1} = \begin{bmatrix} Var(b_0) & Cov(b_0, b_1) & Cov(b_0, b_2) \\ Cov(b_0, b_1) & Var(b_1) & Cov(b_1, b_2) \\ Cov(b_0, b_2) & Cov(b_1, b_2) & Var(b_2) \end{bmatrix}$$

where the standard error of regression $S_{y/x}$ is used as an estimate for the standard deviation σ.

AII.10 NOTATION

H_0, H_1	Null and alternative hypotheses in a statistical hypothesis testing setting
l	index for sources of uncertainty: homogeneity ($l = 1$), calibration curve ($l = 2$), method precision ($l = 3$), calibration standards ($l = 4$), and sample preparation ($l = 5$).
u_c	combined standard uncertainty
$u(l)$	standard uncertainty for uncertainty source l
$u_r(l)$	relative standard uncertainty for uncertainty source l
$v(l)$	degrees of freedom associated to source of uncertainty l
v_{eff}	effective degrees of freedom
$\boldsymbol{x_s}$	average case sample concentration
y_s	average case sample peak area ratio
MSS_B	mean square error between sample groups
MSS_W	mean square error within sample groups
F_c	critical value obtained from the inverse of the CDF of an F-Distribution
RSU :	relative standard uncertainty

ANNEX III. FIGURE CONTENT

ANNEX IV. TABLE CONTENT