P19-02 GUIDANCE ON THE IMPLEMENTATION OF THE CALA MEASUREMENT UNCERTAINTY POLICY

Revision 2.0

February 12, 2020

CALA

Trust, measured accurately

TABLE OF CONTENTS

1.0 BACKGROUND

The guidance in this document is intended to provide information, not as a prescriptive, step-by-step, procedure. Laboratories may use other methods to determine measurement uncertainty as long as the requirements of P19-01 CALA Measurement Uncertainty Policy and ISO/IEC 17025 are met. It is not possible to provide guidance or examples for all the various testing activities in the laboratory, however the general principles described herein applies to all types of testing activity.

2.0 DEFINITIONS

Accuracy (of measurement): (VIM 2.13): closeness of the agreement between the result of a measurement and a true value of the measurand

Note: Accuracy is a qualitative concept. The term precision should not be used for accuracy. An accepted reference value may be used in place of a true value in this definition.

Bias: (ISO 3534-1): the difference between the expectation of the test results from a particular laboratory and an accepted reference value.

Note: Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

Combined standard uncertainty: (GUM 2.3.4): standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities

Coverage factor: (GUM 2.3.6): numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty

Note: A coverage factor, *k*, is typically in the range of 2 to 3.

Error (of measurement): (VIM 2.16): result of a measurement minus a reference quantity value

Note: Since a true value cannot be determined, in practice a conventional true value is used. When it is necessary to distinguish *error* from *relative error*, the former is sometimes called *absolute error of measurement*. This should not be confused with *absolute value of error*, which is the modulus of the error.

Expanded uncertainty: (GUM 2.3.5): 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.

Note: 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 characterised by the measurement result and its combined standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions may be justified.

Confidence level: (GUM C.2.29): The value of the probability associated with a confidence interval or a statistical coverage interval.

Note: The value is often expressed as a percentage.

Measurand: (VIM 2.3): quantity intended to be measured.

EXAMPLE: Vapour pressure of a given sample of water at 20°C.

NOTE: The specification of a measurand may require statements about quantities such as time, temperature, and pressure.

Measurement: (VIM 2.1): process of experimentally obtaining one or more quantity values that can be reasonably attributed to a quantity.

Precision: (ISO3534-1): the closeness of agreement between independent test results obtained under stipulated conditions.

Note: Precision depends only on the distribution of random errors and does not relate to the true value or the specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

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

Repeatability: (VIM 2.21): closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement

Note: The conditions are called repeatability conditions. Repeatability conditions include: the same measurement procedure; the same observer; the same measuring instrument used under the same conditions; the same location; and, repetition over a short period of time.

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

Replicability: See Repeatability

Reproducibility: (VIM 2.25): closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement

Note: A valid statement of reproducibility requires specification of the conditions changed. The changed conditions may include but are not limited to: principle of measurement; method of measurement; operator; measuring instrument; reference standard; location; conditions of use; and, time.

Reproducibility may be expressed quantitatively in terms of the dispersion characteristics of the results.

Standard uncertainty: (GUM 2.3.1): uncertainty of the result of a measurement expressed as a standard deviation

Trueness: (ISO 3534-1): the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value

Note: The measure of trueness is usually expressed in terms of bias. Trueness has been referred to as accuracy of the mean. This usage is not recommended.

Type A evaluation of uncertainty: (GUM 2.3.2): method of evaluation of uncertainty by the statistical analysis of observations

Type B evaluation of uncertainty: (GUM 2.3.3): method of evaluation of uncertainty by means other than the statistical analysis of a series of observations

Uncertainty (of measurement): (GUM 2.2.3): parameter, associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand

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

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

It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion. This definition is that of the "Guide to the expression of uncertainty in measurement" in which its rationale is detailed (see in particular 2.2.4 and Annex D to the GUM).

3.0 COMMON APPROACHES TO ESTIMATION OF MEASUREMENT UNCERTAINTY.

Currently there are a number of approaches that can be taken in estimating the measurement uncertainty associated with testing. These include the mathematic modelling, single-laboratory validation, inter-laboratory validation and the use of proficiency testing data. The most common approach used by testing laboratories is the single laboratory validation approach described in further detail below.

3.1 Modelling Approach

The modelling approach to evaluation of uncertainty is described in chapter 8 of the GUM and will not be covered here. The procedure is based on a model designed to account for the interrelation of all the influence quantities that could significantly affect the measurand. Corrections are assumed to be included in the model to account for all recognised, significant systematic effects.

3.2 Single Laboratory Validation Approach

Laboratory validation and quality control data provide a valuable source of information that can be used to estimate measurement uncertainty. The single laboratory approach is based on using laboratory data to determine the uncertainty component for reproducibility within the laboratory (u_{Rw}) and uncertainty component for method and laboratory bias (u_{bias}).

Method validation replicate data is a source of data from repeat analyses run to establish precision estimates at different analyte concentration levels. The results from those run at low concentrations for the calculation of detection and quantitation limits can also be used to assess uncertainty at low analyte concentration ranges. The validation data can also serve as a source of information on the uncertainty contributed by other sources (such as analyst, instrument, temperature, time etc.) depending on how the validation work was planned and executed to include such variables. This is especially the case if ruggedness studies were incorporated as an integral part of the validation program to assess the effect of varying parameters likely to be significant sources of uncertainty. Where method validation is used to determine uncertainty, the laboratory should update the uncertainty estimate as more data becomes available. A thorough discussion of the use of method validation data in the estimation of uncertainty can be found in *VAM Project 3.2.1 Development and Harmonization of Measurement Uncertainty Principles; Part (d): Protocol for uncertainty evaluation from validation data,* by V.J. Barwick and S.L.R. Ellison, January 2000, Version 5.1.

3.3 Inter-laboratory Validation Approach

Where the major sources of variability are determined by interlaboratory validation studies such as described in ISO 5725-2, the precision and trueness data may be used to estimate the measurement uncertainty. This approach is detailed in ISO/TS 21748 *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation.*

The general approach used in this document requires the following:

- Estimates of the repeatability, reproducibility and trueness of the method in use, obtained by collaborative study as described in ISO 5725-2, be available from published information about the test method in use. These provide estimates of the intra-laboratory and inter-laboratory components of variance, together with an estimate of uncertainty associated with the trueness of the method.
- The laboratory confirms that its implementation of the test method is consistent with the established performance of the test method by checking its own bias and precision. This confirms that the published data are applicable to the results obtained by the laboratory.
- Any influences on the measurement results that were not adequately covered by the collaborative study be identified and the variance associated with the results that could arise from these effects be quantified.

3.4 The Use of Proficiency Testing Data

Proficiency testing data are a source of reproducibility SD (SD_R) that includes both intra- and inter-laboratory sources of uncertainty. It is larger than the intra-laboratory uncertainty (SD_r), known as repeatability, of a laboratory whose methods are in statistical control. In the absence of any other source of repeated data, reproducibility from proficiency testing and other round robin studies can be used as an estimate of measurement uncertainty. If PT samples are used in estimating uncertainty:

- The items should be representative of the routine samples in the laboratory
- The assigned values should have an appropriate uncertainty
- The number of PT rounds should be sufficient to obtain a reliable estimate (e.g., 6 rounds)
- The number of participating laboratories should be sufficient for reliable characterization of the sample when consensus values are used.

It is, however, very likely to be an overestimate of the intra-laboratory uncertainty. In some cases, PT results may also be used for bias detection and correction.

4.0 EVALUATION OF UNCERTAINTY COMPONENTS – TYPE A AND TYPE B EVALUATION

There are two main methods of evaluation the various components contributing to the overall uncertainty. The first method, termed Type A, estimates uncertainties through the statistical analysis of experimental laboratory data such as that from routine QA/QC work (e.g. duplicates, reference material usage, method validation studies, and proficiency testing (PT) and other inter-laboratory programs). This is the approach used by most analytical laboratories. One main advantage of the type A approach is that almost all the data required is readily available in the laboratory.

The second method, Type B, can also be characterized by standard deviations evaluated from assumed probability distributions based on experience or other information.

Type B evaluation of the standard uncertainty is the method of evaluation of uncertainty by means other than the statistical analysis of a series of observations. The standard uncertainty is evaluated by scientific judgement based on available information of possible variability. The information used may be obtained from

- previous measurement data;
- experience with or general knowledge of the behaviour and properties of relevant materials and instruments;
- manufacturer's specifications;
- data provided in calibration and other certificates;
- uncertainties assigned to reference data taken from handbooks.

Further details on Type B evaluations can be found in reference 4 at the end of this document. The proper use of the available information for a Type B evaluation of standard uncertainty of measurement calls for insight based on experience and general knowledge.

Either approach may be used and in some cases determination of the measurement uncertainty may involve a combination of both.

5.0 STEPS IN ESTIMATION MEASUREMENT UNCERTAINTY.

Step 1. Specify the measurand

Define the relationship between what is being measured and the input quantities on which it depends.

Step 2. Identify uncertainty sources.

List the possible sources of uncertainty. Using the method SOP and the final result-calculation equation, identify and list all potential sources of uncertainty

Consideration should be given to the different factors contributing to measurement uncertainty. These may include but not limited to;

- Sampling or subsampling – Where sampling (or sub-sampling) is treated as part of the test, the uncertainty arising from such sampling should be considered by the laboratory. The effects of variation between subsamples and inhomogeneity particularly in solid samples may be significant contributors to uncertainty.
- Transportation, storage and handling of samples.
- Preparation of samples
- Environmental and measurement conditions
- Personnel carrying out the test
- Variations in test procedure
- The measuring instruments
- Calibration standards or reference materials
- Software/method of generating result.
- Uncertainty arising from correction for systematic effects

Step 3. Quantify uncertainty components.

Estimate the standard uncertainty (u_x) for each source identified. Uncertainty from a number of separate sources can often be determined from QC data, validation studies, proficiency testing data, manufacturer's specifications, calibration certificates, reference standards, and reference data from handbooks. Identify and compile recent laboratory repeat analysis and other data selecting laboratory Quality Control and validation data that includes as many sources of variability as possible. The data used should take into consideration the range of analyte concentrations for the matrix analyzed.

Match each repeat data set with those sources of uncertainty that are likely to have varied during the collection of the repeat data and identify double counted sources of uncertainty.

Estimate the magnitude of any source of uncertainty that is not varied during the collection of any of the repeat data sets. It may be sometimes necessary to conduct additional tests.

Step 4. Calculate combined uncertainty

Tabulate each source of uncertainty and its associated SD, and/or relative SD (RSD) derived from the repeat data set(s) matched to it, or from the estimate made. Eliminate double counted sources. Using only those SDs that are 1/3 or more the size of the largest individual SD, calculate the combined standard uncertainty using standard propagation of error rules (the square root of the sums of squares of SDs known as the "root sum of squares" - RSS);

$$u_c = \sqrt{SD_1{}^2 + SD_2{}^2 + SD_3{}^2}$$

Step 5. Calculate expanded uncertainty

The Expanded Uncertainty(U) is derived by multiplying the Combined Standard Uncertainty(u_c) by a coverage factor "k". The value of k for 95% coverage is selected on the basis of the number of values "n" that are used for the calculation for the SDs. Use k = 2 or the appropriate Student's t factor for n-1 degrees of freedom and a 95% confidence level.

$$U = k \times u_c$$

Report the result with the expanded uncertainty and with a description of how the uncertainty was calculated.

6.0 DETERMINATION OF MEASUREMENT UNCERTAINTY

The determination of the measurement uncertainty is primarily based on the evaluation of two main components, the uncertainty due within laboratory reproducibility (***u_{Rw}***) and uncertainty due to method and laboratory bias (***u_{bias}***). They can be used to address most of the uncertainty sources identified in step 2 above.

6.1 Determination of Within Laboratory reproducibility(***u_{Rw}***)

6.1.1 Laboratory Control Samples (including CRM)

The uncertainty component for the within laboratory reproducibility can be estimated from laboratory control samples measured under reproducibility conditions (i.e. run on different days, different analysts). When a stable control sample covering the whole analytical process (including sample prep) and with similar matrix to test samples is analyzed, the within laboratory reproducibility can be estimated from the results as follows:

$$u_{Rw} = s_{Rw}$$

where s_{Rw} is the standard deviation of the control sample results.

Where the test method covers a wide concentration range, laboratory control samples should be analyzed at different concentration levels. A minimum of eight measurements should be used, however a greater number of measurements (>20) improves the reliability of the estimate.

6.1.2 Sample Duplicates

For some analyses, a stable control sample is not readily available. In such cases, sample duplicate insertion can be a valuable source of uncertainty data, known as replicability **(*s_{dup}*)**, that reflects the variability due to differences between analytical portions (non-homogeneity) and other factors that can vary between replicates (weighing, volumetric manipulations, and short term instrument drift are examples). The uncertainty resulting from variation between batches (s_{var}) is not accounted for by s_{dup}

Note: If the duplicates are measured in the same analytical run, as is usually the case, any uncertainty associated with the instrument set up and calibration is not included. More than 20 duplicate pairs should be run of samples of a similar concentration.

$$s_{dup} = \sqrt{\Sigma (R^2/2N)}$$

where R is the difference between duplicate pairs and N is the number of duplicate pairs. This should be calculated for low, medium and high concentration ranges to reflect the concentration dependence of the SD.

Alternatively, the RSD can be calculated (at low, medium and high concentration ranges as well) as:

$$RSD_{dupl} = \sqrt{\left\{\sum\left[(a_i - b_i)/\overline{X}_i\right]^2 / 2N\right\}}$$

where $(a_i - b_i)/\overline{X}_i$ is the relative difference between duplicates for sample "i" and N is the number of samples for which duplicates have been run. This value makes allowances for the concentration dependence of the SD for concentrations between those at which the calculation was made.

Where there is significant variability between batches or a stable quality control sample is not readily available, the within laboratory reproducibility is given by:

$$u_{Rw} = \sqrt{s_{dup}^2 + s_{var}^2}$$

where s_{dup} is the standard deviation obtained from laboratory duplicate results

and s_{var} is the uncertainty due to variation between batches.

6.2 Determination of uncertainty for method and Laboratory bias (u_{bias})

For the estimation of the uncertainty due to method and laboratory bias (u_{bias}), two components need to be estimated:

- the bias (ie. The difference for the nominal or certified value) and;
- the uncertainty of the nominal or certified reference value

6.2.1 Use of Reference Materials

Routine analysis of reference materials provides data that can be used to estimate the uncertainty associated with method and laboratory bias (u_{bias}). A minimum of eight measurements should be used, however a greater number of measurements improves the reliability of the estimate.

If several reference materials are used with varying concentrations, different values for bias will be obtained and u_{bias} is given by:

$$U_{bias} = \sqrt{{b_{rms}}^2 + {\bar{u}_{cref}}^2}$$

Where brms is the root mean square of the bias $= \sqrt{\sum(bias)^2 / n}$

and u_{cref} is the mean uncertainty of the reference values.

If only one reference material is used, u_{bias} is given by:

$$u_{bias} = \sqrt{(bias)^2 + \left(\frac{s_{bias}}{\sqrt{n}}\right)^2 + u_{cref}^2}$$

Where s_{bias} is the standard deviation of the reference values

6.2.2 Use of Proficiency Testing Results

Results from the analysis of proficiency testing samples can be used in a similar way to the results from analysis of reference to estimate bias. Here it is assumed that the assigned value for the PT study is a good estimate of the true value. Please note that the estimation of bias from the PT study generally has a higher uncertainty than if a CRM was used as the certified value of a CRM is better defined than an assigned value for a PT study.

The laboratory should have analyzed at least eight different samples over more than one round of proficiency testing.

Calculate the difference D between the measurement results for the different PT samples which may be positive or negative. Determine the root mean square of the differences, *Drms*

$$D_{rms} = \sqrt{\frac{\sum D^2}{n_{pt}}}$$

Where D is the difference between the measurement result and the assigned value of the PT sample and n_{pt} is the number of PT samples analysed.

The uncertainty of the assigned value should be obtained from the study organizer.

The standard uncertainty component for method and laboratory bias is given by .

$$u_{bias} = \sqrt{D_{rms}{}^2 + \overline{u}_{cref}{}^2}$$

Where $\mathbf{D_{rms}}$ is the root mean square of the differences and

$\mathbf{u_{cref}}$ is the mean uncertainty of the assigned values of the PT samples.

6.3 Calculation of the combined standard uncertainty

If there are no other uncertainty components, the combined standard uncertainty is given by

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

6.4 Calculation of the expanded uncertainty

The Expanded Uncertainty(U) is derived by multiplying the Combined Standard Uncertainty(u_c) by a coverage factor "k". The value of k for 95% coverage is selected on the basis of the number of values "n" that are used for the calculation for the SDs. Use k = 2 or the appropriate Student's t factor for n-1 degrees of freedom and a 95% confidence level.

$$U = k \times u_c$$

7.0 HIERARCHY OF DATA SELECTION FOR ESTIMATION OF UNCERTAINTY

Reference Samples, Spike Recovery, Method Validation Replicate and Sample Duplicates are typical sources of Laboratory Repeat Data Sets and can be sources of repeated measurements from which SD and RSD can be calculated. There are pros and cons to use of these sources.

The following hierarchy is presented to provide laboratories with guidance on which types of data they might use to estimate uncertainty within the laboratory. This list is given in order of priority from (I) Most Suitable, to (III) Least Suitable:

(I) Laboratory Control Samples (LCS) and Matrix Spikes: In cases where matrix specific LCS and/or matrix spike data are available, include uncertainty estimated from the standard deviation of the LCS or matrix spikes of more than 20 points collected from their insertion into routine analytical runs.

(II) Sample Replicate Data (Precision data): In cases where sample replicates are analyzed and there is sufficient data above the limit of quantitation, include sample replicate data to estimate uncertainty that incorporates sub-sample uncertainty as a source. Depending on the conditions for replicate measurements, two different standard deviations may be obtained:

- the within laboratory repeatability standard deviation obtained under repeatability conditions (i.e. same analyst, same equipment, short time period)
- the within laboratory reproducibility standard deviation obtained under within laboratory reproducibility conditions (i.e. different analyst, different equipment, long time period)

(III) Proficiency Testing Sample Data: In cases where the previous options are not available and where Proficiency Testing samples are analyzed with sufficient data above the limit of quantitation, pooled Proficiency Testing sample data can be used to estimate uncertainty.

Examples of the determination of the measurement uncertainty for analytical chemistry, microbiology and toxicology are described in the following appendices. However, the principles described can be applied to other types of testing.

APPENDIX 1: MEASUREMENT UNCERTAINTY FOR ANALYTICAL CHEMISTRY

A1.1 Aim

This appendix explains and expands on one approach that can be used to evaluate measurement uncertainty for analytical chemistry methods.

A1.2 Estimation of the Uncertainty for any Sources not Accounted for by Repeated Data

In the unusual cases where it is necessary to estimate uncertainties for any sources not accommodated by repeated data, the estimation of the uncertainty from these sources is based on information from manufacturer specifications that accompany instruments and equipment (such as volumetric ware), tabulated data from handbooks, experience from other methods and/or laboratories and other sources.

A1.3 Tabulation of Uncertainty Estimates

Compile the values estimated from the repeated experimental data with that for each of the potential sources of uncertainty identified as not being reflected in the repeated data variability (if any) and rank them in decreasing numerical order. Those sources that have a SD less than 1/3 of the largest SD can be ignored in the subsequent calculation of the combined uncertainty since their contribution to the combined uncertainty will be negligible.

A1.4 Calculation of the Combined Uncertainty

SDs cannot be manipulated to calculate the combined standard uncertainty. Instead, the SDs are converted to variances by squaring them and the variances are used for the calculation of the combined standard uncertainty. The combined standard uncertainty is the square root of the sum of the squares of the SDs (known as the Root Sum of Squares).

If RSDs have been calculated, the SD at a specific concentration C should be calculated by:

$$SD = RSD \times C$$

This allows for taking the concentration dependence of the SD into account. (NMKL Procedure No. 5 (1997) Estimation and expression of measurement uncertainty in chemical analysis). Precautions must be taken to not count the contribution of a source of uncertainty more than once in the calculation of the combined standard uncertainty. The between run SD calculated from daily spike recoveries for example, will include the variability found in the entire analytical process if the spike was inserted at the beginning of the analytical process (i.e., spiking done to the sample prior to any filtration, extraction, digestion, etc.). This is also true however, of the SD calculated from the routine

inclusion of any reference sample that is inserted at the very beginning of the analytical process (i.e., the reference sample undergoes the same filtration, extraction, digestion, etc. as the samples). Calculating the combined standard uncertainty by using the SDs from both of these sets of data would double count all of the contributing sources and result in an estimate of the measurement uncertainty that is too large. The established procedure in such an instance is to use the larger of the two SDs in order to give a "worst case" estimate.

As an example, if we have established the between run standard deviation from historical spike recovery data to be s_{spike}, the bias uncertainty to be s_{bias} and that no other sources of uncertainty have a standard deviation larger than 1/3 of the largest of these, the combined standard uncertainty u_c is given as:

$$u_c = \sqrt{{s_{spike}}^2 + {s_{bias}}^2}$$

A1.5 Reporting the Result

Apply the coverage factor k. The final concentration result C is then reported as $C \pm k \times u_c$ with a description of how the measurement uncertainty was calculated.

A1.6 Example Table to compile MU information

Example Table to compile MU information

Description of Uncertainty Source	Value x	Uncertainty measured or found	u(x) as Standard Deviation	u(x)/x	Source of u(x) information

A1.6.1 Steps to using this Table:

1. Define the measurand(s), the analyte, the measurement objectives required for data to be "fit-for-purpose" (includes LOD, precision, accuracy, analytical range, selectivity etc.);
2. List the anticipated sources of uncertainty (including parameters found in the equation used to calculate the final result to be reported);
3. List the repeated data sources (spikes, certified reference materials, in-house reference materials. duplicates, method validation files) both short term (one day or one run for example) and long term (over several months or longer);
4. Match the sources of uncertainty with repeat data that was collected while the sources of uncertainty may have varied. Long term spike recovery data may include changes in analysts, calibration sets, and laboratory environment;
5. Identify those sources of uncertainty that are included in more than one repeat data set. Both long-term spike and reference material standard deviation values will include uncertainty due to different analysts, calibration sets etc.; if these were varied while the spike and reference material data were being collected in routine runs. Use only one of these two standard deviation values to estimate the contribution to measurement uncertainty from the sources identified as being varied, usually the larger to be conservative. Alternatively, the two standard deviations can be pooled and the pooled value included for compilation into the overall estimate of measurement uncertainty;
6. Estimate the uncertainty due to those sources that have not varied during the collection of repeat data, either during method validation or routine analysis. This may involve using certificates for balances and masses or some other source of uncertainty information;
7. Compile the information into the table above and check to ensure that a source of uncertainty has not been counted more than once;
8. Remove those sources of uncertainty that have a standard deviation less than 1/3 the largest standard deviation;
9. Combine the remaining standard deviations using root sum of squares (RSS) technique

10. Multiply this combined standard deviation by the appropriate expansion factor to determine the expanded uncertainty;
11. Ensure the data meets the fit-for-purpose criteria; and,
12. If applicable, report the result with the expanded uncertainty. Indicate the expansion factor (k) and the confidence interval (usually 95%).

A 1.7 Example of MU determination using approach described in section 6.0 -Analysis of 1,1 Dichloroethane in soil

For the determination of 1,1 Dichloroethane in soil, a CRM was analysed in 10 batches over two weeks.

A1.7.1 Calculation of uncertainty for within laboratory reproducibility

Batch no	1,1 Dichloroethane concentration(mg/kg)
1	0.71
2	0.68
3	0.72
4	0.81
5	0.77
6	0.68
7	0.73
8	0.82
9	0.69
10	0.71

The mean conc = 0.732 mg/Kg
The standard deviation s_{rw}= 0.051
$u_{rw} = s_{rw}/x$ = 0.070

A1.7.2 Calculation of uncertainty for method and laboratory bias

From the reference material certificate, the certified value is given as 0.91± 0.04 mg/kg

The uncertainty of the reference value is therefore given by u_{cref} = 0.04/3 = 0.0133

The bias *b* is, b = x-Cref =0.732-0.91 = -0.178

The uncertainty component for method and laboratory bias is given by,

$$u_{bias} = \sqrt{(\boldsymbol{bias})^2 + \left(\frac{s_{bias}}{\sqrt{n}}\right)^2 + u_{cref}^2}$$

$$\sqrt{-0.178^2 + \left(\frac{0.051}{\sqrt{10}}\right)^2 + 0.0133^2}$$

= 0.0179

Calculate the combined standard uncertainty, Uc

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

$u_c = \sqrt{0.07^2 + 0.0179^2}$ = 0.072

The expanded uncertainty U using a coverage factor of k =2 is,

U = u_c x k = 0.072 x 2 = 0.144 mg/kg

APPENDIX 2: MEASUREMENT UNCERTAINTY FOR MICROBIOLOGICAL TESTING

A2.1 Aim

This Appendix applies to microbiological testing methods that are quantitative and whose uncertainties are based on Type A estimates. This appendix explains and expands on one approach and includes the use of experimental data such as that from routine laboratory QC work (duplicates, reference material usage, method validation studies, and proficiency testing (PT) and other inter-laboratory programs, etc.). Other approaches are available to evaluate measurement uncertainty and can be used by laboratories. It is the responsibility of the laboratory to demonstrate that the approach meets the requirements of ISO/IEC 17025, the CALA policy on uncertainty, and that the estimated measurement uncertainty is reasonable.

A2.2 Components of Uncertainty

The References Section of this guide lists organizations that have published documents which present possible sources of uncertainty for a microbiological method. In particular, Niemela (1996) provides a good discussion of the sources of uncertainty for microbiological methods. In addition, close examination of the steps in the laboratory method SOP and of the parameters found in the final concentration calculation, will usually help to identify the likely sources of uncertainty. The following factors have been shown to influence the precision of microbiological results and require appropriate QC procedures to minimize variation:

- Source of sample
- Method of sampling/subsampling
- Transportation
- Method of Analysis
- Level of performance verification or validation
- Culture Media and Reagents
- Preparation protocols
- Water quality
- Storage conditions and shelf-life
- Sample homogenization/mixing
- Inoculation procedure (e.g., Filtration technique)
- Incubation conditions
- Reading, interpreting and reporting results
- Microbial density
- Equipment
- Personnel

The uncertainty that may be associated with sample holding time, if all tests are run within the allowable holding time, will not be considered in this guide.

It can also be assumed that the uncertainty for colony counts may be derived from an examination of the variances associated with filtering or plating and colony counting among analysts.

CAUTION: This only applies if quality control results show that all other critical factors (e.g. incubator temperatures, refrigerator temperatures, media, within analyst repeatability, etc.) are in control.

A2.3 Laboratory Repeat Data Sets

The following is a list of sources of repeated measurements in a typical laboratory that may be used when estimating measurement uncertainty.

Sample Duplicates;

- In drinking water analysis, ground water or treated water samples are not very useful for capturing duplicate data because most results are 0/100mL.
- The raw water from rivers or lakes is a better source for duplicate testing.

Quality Control Data;

- Quantitative data generated as part of ongoing QC programs, e.g. for method performance validation, media QC etc. can be included for the calculation of expanded uncertainty.

Spike recovery data;

- Time consuming, but a reasonable method to measure within-analyst repeatability and among-analyst reproducibility over time;
- Similar in approach to and could be combined with split PT results as a source of data for within analyst repeatability and between analyst reproducibility;
- Should be done over several orders of magnitude for dilutions.

Reference sample insertion;

- There is a general lack of reference materials available for routine use in microbiological methods. However, where available reference samples provide a good source of uncertainty data.

Proficiency testing programs;

- Pooling of data derived by different methods will result in a larger MU than for a single laboratory but may still be useful;
- PT samples may provide material for within-analyst and between-analyst duplicate testing and results, which can be included in the data pool when determining the within analyst repeatability and between analyst reproducibility.

A2.4 Reproducibility Calculations for Estimating Combined (u_c) and Expanded Uncertainty (U)

A2.4.1 Combined Uncertainty

The combined uncertainty (u_c) is defined as the square root of the sum of the squares of the standard deviations (SD) or relative standard deviations (RSD) of the independent components, which comprise a method. This is known as the root sum of squares.

The traditional approach is to combine all data from the different sources described previously and determine the single analyst relative standard deviation (RSD_r) and the between analyst (e.g. counting, technique) or laboratory relative standard deviation (RSD_L) and calculate the combined uncertainty (u_c). (NOTE: Data used to determine single variance cannot be used to also determine between analyst or laboratory variance). However, it should be kept in mind that use of this traditional approach could lead to double-counting of the within-analyst variation.

$$u_C = \sqrt{(RSD_r^2 + RSD_L^2)}$$

Therefore, duplicate analyses using the same SOP by different analysts over an extended period of time (e.g. 1 year) would encompass the influence of all elements that impact on measurement uncertainty.

The combined uncertainty associated with each procedure can be determined more easily by combining results obtained when different analysts process samples and/or count colonies on plates or MF filters and calculating the Relative Standard Deviation of Reproducibility (RSD_R) directly.

In this situation, combined uncertainty may be reduced to:

$$u_C = \sqrt{RSD_R^2}$$

A2.4.2 Expanded uncertainty (U)

U= k (coverage factor for 95% confidence) x u_c (RSD_R)

The expanded uncertainty (U) is 2 x the combined uncertainty (u_c) if 30 or more values are used to calculate the SD or RSD.

All calculations can be easily handled using Microsoft Excel Spreadsheets.

A2.5 Data Handling

There are diverging opinions on whether data needs to be log transformed or not. When bacterial populations in different samples vary significantly, pooling of CFU counts may result in some high or low values that can skew the mean and result in an unreasonably large variance. Under these circumstances, it would be more appropriate to convert the data to $\log_{10}$ before doing any statistical analyses. However, if the raw data (counts per plate/filter) is separated into ranges, the data per range is assumed to be approximately normally distributed, which allows the use of the arithmetic values for statistical evaluation. A quick check is to plot the raw data!

The expanded uncertainty determined from data over the entire counting range of colonies per filter or plate may overestimate or underestimate uncertainty depending upon whether the data is weighted to high or low counts. Therefore, data should be separated into ranges (as indicated below) and, the combined uncertainty (u_c) determined for each range.

The following colony forming unit (CFU) ranges are suggested for MF techniques:

1-19 Colonies/Filter;

20-80 Colonies/Filter;

81-100 Colonies/Filter.

The following CFU ranges are suggested for plating (e.g. spread plate) procedures:

1-29 Colonies/Plate;

30-99 Colonies/Plate;

100-300 Colonies/Plate.

Note: The suggested ranges are arbitrary and based on experience in water testing laboratories. A laboratory can choose different ranges that better reflect the data being generated in the laboratory.

A2.6 Most Probable Number Methods (MPN)

MPN analyses typically refer to McCrady's tables to obtain a result as well as the 95% confidence limits. For the purposes of CALA's Policy, these tables can be used as estimates of uncertainty for a

test, provided the laboratory has reviewed the resulting data and identified any unusual combinations of results.

Any unusual combinations in excess of 1% of all MPN results are to be treated as non-conformances.

A2.7 Qualitative Methods (e.g. Presence-Absence)

There is no precision associated with presence/absence or qualitative methods and therefore no statistical estimate of uncertainty can be calculated. However, not being able to calculate MU does not mean that uncertainty is "not applicable"! The possible sources of variability that impact all microbiological methods (outlined above in Section A2.2) need to be controlled. These sources are not necessarily independent but can contribute to the overall uncertainty of a method. The variability of these sources needs to be considered with analysis of replicate samples, the use of control samples, inter-analyst sample testing and the participation in Proficiency Programs. Appropriate corrective actions when there is a non-conformance must be described in related documents and referenced in the methods. QC records must be maintained. For Qualitative Microbiological methods, a summary statement with the method verification report should include the following:

- A list of possible sources of uncertainty (See A.2.2).
- A statement regarding the consistency of performance indicated by method validation and PT testing.
- Statement of performance claims by Manufacturer or Method Literature.
- Certificate of test strains.
- Approval by Technical Management.

Laboratories are also to be aware of False Positive/False Negative Rates; e.g.,

- False Positive/False Negative Rates provided by the manufacturer (e.g. from IDEXX for Colilert), if available;
- False Positive/False Negative Rates provided for the method in the literature, if available;
- Laboratory may run confirmation tests on all or a percentage of positive and negative samples to determine False Positive/False Negative Rates for the method within the laboratory (this can be very time consuming);
- False Positive / False Negative rates in excess of published specification are to be treated by the laboratory as a non-conformance and root causes identified for corrective action.

A2.8 Addendum 1

The following information shows how to calculate uncertainty (e.g. variance, SD, RSD, RSD_2) based upon duplicate testing. (Sample Data from actual laboratory results)

Table A2.8-1: The Results of Duplicate Total Coliform (TC) Tests on a Series of Different Samples

Range 20 - 80 TC Colonies per Filter					
Sample	TC/Filter Duplicate 1	TC/Filter Duplicate 2	Absolute Difference (D)	Difference Squared (D_2)	Variance
1	46	45	1	1	0.5
2	55	45	10	100	50
3	47	41	6	36	18
4	23	18	5	25	12.5
5	23	23	0	0	0
6	34	38	4	16	8
7	50	54	4	16	8
8	14	21	7	49	24.5
9	33	43	10	100	50
10	69	61	8	64	32
11	77	78	1	1	0.5
12	26	24	2	4	2
13	63	62	1	1	0.5
14	42	38	4	16	8
15	42	48	6	36	18
16	36	41	5	25	12.5

Range 20 - 80 TC Colonies per Filter					
17	21	21	0	0	0
18	25	21	4	16	8
19	22	32	10	100	50
20	20	21	1	1	0.5
21	52	61	9	81	40.5
22	22	24	2	4	2
23	29	23	6	36	18
24	22	26	4	16	8
25	31	30	1	1	0.5
26	53	42	11	121	60.5
27	66	51	15	225	112.5
28	66	50	16	256	128
29	39	22	17	289	144.5
30	55	40	15	225	112.5

The mean of all duplicate values (counts) is 39 and n = 30 so 2n = 60.

The variance based upon duplicate counts from a series of samples can be determined in two ways. The same variance, SD, RSD and/or RSD_2 will be obtained either way.

In the first case, variance = $\sum D_2/2n$.

In the second case,

variance = $\frac{[\sum(variance\ pair\ 1 + variance\ pair\ 2\ .....+ variance\ pair\ n)]}{n}$

Table A2.8-2 shows that both methods for analyzing duplicate data will give the same results for uncertainty when we apply the methods to duplicate data from Table A2.8-1.

Table A2.8-2**:** Statistics and Uncertainty for Duplicate Total Coliform (TC) Counts in Table A2.8-1

Based on Variance for Duplicates $= \sum D_2/2n$		**Based on Variance for Duplicates** $= [\sum$ **(var pair 1 + var pair 2 ……+ var pair n)]/n**	
Statistic	Value	Statistic	Value
Number of data pairs (n)	30	Number of data pairs (n)	30
2n	60		
Mean duplicate count	39	Mean duplicate count	39
$\sum D_2$	1861		
Variance ($\sum D_2/2n$)	1861/60 = 31	Mean Variance	31
SD	$\sqrt{31}$ = 5.6	SD	$\sqrt{31}$ = 5.6
RSD (SD/mean count)	5.6/39 = 0.14	RSD (SD/mean count)	5.6/39 = 0.14
RSD_2	0.0196	RSD_2	0.0196

A2.9 Addendum 4

A2.9.1 Worked Examples

The following information presents 2 ways of collecting membrane filtration (MF) data for the range 20 - 80 colonies per filter and determining combined uncertainty (u_c).

This assumes that quality control results show that all equipment (e.g. incubators) and materials (e.g. media) are in control so that we can determine combined uncertainty (u_c) from only the uncertainty for filtering plus the uncertainty for counting among analysts.

A2.9.2 Method 1 (Testing Among all Analysts)

On 5 or more separate occasions, instruct all analysts to test the same sample but have one analyst count the colonies on all filters. This will give the variation predominantly associated with differences in filtering technique among analysts, keeping in mind that it will include within-analyst variation associated with counting.

In addition, on 5 or more separate occasions, instruct all analysts to count target colonies on the same filter. This will provide the variation associated with differences in target colony recognition and counting among analysts.

Repeat this procedure for each analyte (e.g. total coliform, faecal coliform, E.coli, HPC, etc.) and for colony counts in each range (i.e. 0 - 19, 20 - 80 and 81 - 150 target colonies per filter).

Tables A2.9-1 and A2.9- 2 provide examples for total coliform (TC) in the range of 20 - 80 colonies per filter, show how to organize the data and determine the RSD_2. This is followed by a calculation of combined uncertainty (u_c).

Table A2.9-1:Uncertainty for the Filtration Component Among Analysts

Total Coliform (TC) in the Range 20 - 80 TC/Filter **(all analysts filtered the same sample each time but one analyst counted colonies on all filters)**						
Analyst	Sample 1 TC/Filter	Sample 2 TC/Filter	Sample 3 TC/Filter	Sample 4 TC/Filter	Sample 5 TC/Filter	Sample 6 TC/Filter
1	38	46	50	50	68	74
2	41	28	58	54	81	70
3	31	26	42	50	65	69
4	33	34	50	33	73	64
5	23	30	58	52	68	71
Variance	**48**	**63**	**45**	**71**	**40**	**13**

Overall Mean Count = 51
Mean Variance = 47
SD = $\sqrt{47}$ = 6.9
RSD = 6.9/51 = 0.135
RSD_2 = 0.018

Table A2.9-2: Uncertainty for the Colony Counting Component Among Analysts

Total Coliform (TC) in the Range 20 - 80 TC/Filter (all analysts counted the colonies on the filter each time)					
Count from Analyst	Sample 1 TC/Filter	Sample 2 TC/Filter	Sample 3 TC/Filter	Sample 4 TC/Filter	Sample 5 TC/Filter
1	55	71	43	61	20
2	57	68	46	57	25
3	61	72	33	58	22
4	57	75	56	61	21
5	60	71	34	67	22
Variance	**6**	**6.3**	**89**	**15**	**3.5**

Overall Mean Count = 51
Mean Variance = 24
SD = $\sqrt{24}$ = 4.9
RSD = 4.9/51 = 0.096
RSD_2 = 0.0092

Note: The variation in counts for total coliforms (TC) is often large because TC colonies may show considerable variation in reaction and not all analysts recognize subtle positive reactions.

In this case, use the following formula to calculate combined uncertainty (u_c).

$$u_c = \sqrt{(RSD^2_{(filtration\ among\ analysts)} + RSD^2_{(counting\ among\ analysts)})}$$

So, in this case, the combined uncertainty (u_c) for the range 20 - 80 TC/Filter can be expressed as:

$$u_c = \sqrt{(0.018^2 + 0.0092^2)}$$

$$= 0.165$$

Remember to repeat the above process per range for each analyte (i.e. total coliforms, faecal coliforms, E.coli, plate counts, etc.).

A2.9.3 Method 2 (Between-Analyst Duplicate Testing)

Method 2 uses duplicate data between analysts to determine combined uncertainty. However, collecting duplicate data becomes complicated when there are 3 or more analysts. Nevertheless, the following procedure may be used and we will assume that there are 5 analysts in the laboratory.

Give each analyst an analyst number. In this example, there are 5 analysts numbered 1 to 5. Organize the analysts to perform duplicate tests between analysts on a regular basis but rotate the analyst pairs so that they perform duplicate testing in the following or similar manner.

Sample 1	(Analyst 1 and Analyst 2)
Sample 2	(Analyst 1 and Analyst 3)
Sample 3	(Analyst 1 and Analyst 4)
Sample 4	(Analyst 1 and Analyst 5)
Sample 5	(Analyst 2 and Analyst 3)
Sample 6	(Analyst 2 and Analyst 4)
Sample 7	(Analyst 2 and Analyst 5)
Sample 8	(Analyst 3 and Analyst 4)
Sample 9	(Analyst 3 and Analyst 5)
Sample 10	(Analyst 4 and Analyst 5)
Etc.	Etc.

When the rotation is complete start over.

Each time the analysts run duplicate tests get the analysts to run the filtrations on the sample and then count the colonies on their own filters.

Use the same procedure for each analyte (i.e. total coliform, faecal coliform, E.coli, HPC, etc.)

Continue the process throughout the year and analyze the data per range. Analyze the data when there are at least 30 duplicate counts per range (i.e. in the ranges 0 - 19, 20 - 80 and 81 - 150 target colonies per filter). To get a more reliable estimate of uncertainty, analyze the data each year (assuming that this will provide more than 30 duplicates per range).

Table A2.9-3 shows how to organize the duplicate data and calculate the RSD_2 for a range. This is followed by a calculation of combined (u_c) and expanded uncertainty (U).

Table A2.9-3: Uncertainty Among Analysts

Total Coliform (TC) in the Range of 20 - 80 TC/Filter **(5 analysts tested samples in duplicate in rotation and counted target colonies on their own filters)**					
Sample	Analyst Pair		TC/Filter Duplicate A	TC/Filter Duplicate B	Variance
	A	B			
1	1	2	50	60	50
2	1	3	41	28	84.5
3	1	4	25	34	40.5
4	1	5	36	44	32
5	2	3	40	31	40.5
6	2	4	66	74	32
7	2	5	53	35	162
8	3	4	35	42	24.5
9	3	5	64	51	84.5
10	4	5	49	57	32
11	(Start over) 1	2	Etc.	Etc.	Etc.
12	1	3			
13	1	4			
14	Etc.	Etc.			

Overall Mean Count = 46

Mean Variance = 58
SD = √58 = 7.6
RSD = 7.6/46 = 0.165
RSD_2 =0.027225

Use the following formula to calculate combined uncertainty (u_c), when Method 2 is used for collecting between analyst duplicate data, because the uncertainties for filtering and counting among analysts are combined in the duplicate testing procedure.

$$u_c = \sqrt{RSD^2} \text{ (BETWEEN ANALYST DUPLICATES)}$$

Therefore, the combined uncertainty (u_c) for the range 20 - 80 TC/Filter can be expressed as

$$u_c = \sqrt{0.027225}$$
$$= 0.165$$

Remember to repeat the above process per range for each analyte (i.e. total coliforms, fecal coliforms, E.coli, plate counts, etc.).

Note: If there are more than 2 analysts, laboratories should rotate analyst pairs to gather between-analyst duplicate data when using method 2 for determining combined uncertainty. Otherwise, they may not capture all the variation, which might occur among analysts in the laboratory.

In this case, the expanded uncertainty (U) will be 2 x (u_c)

$$U = 2 \times 0.165$$
$$= 0.33$$

Express the expanded uncertainty (U) as an RSD%

$$U = 0.33 \times 100$$
$$= 33\%$$

In this example, the count ± the expanded uncertainty for any count within the range of 20 - 80 colonies per filter will be the Count/Filter ± 33% of the Count/Filter.

So, if the TC count was 60 colonies per filter, the count ± its expanded uncertainty would be 60 ± 33% of 60 or 60 ± 20 (rounded) colonies per filter.

To obtain the final result per 100mL, multiply the result ± the expanded uncertainty by the dilution factor.

For example, if an analyst filtered 10mL of sample and the TC count on the filter was 60 colonies, the count ± expanded uncertainty per filter would be 60 ± 20. So, the final result to the client would be (60 ± 20) x 10 = 600 ± 200 TC/100mL at the 95% level of confidence.

Note: Laboratories will have to decide which of the above methods is best suited for their style of laboratory operation.

APPENDIX 3: MEASUREMENT UNCERTAINTY FOR ENVIRONMENTAL TOXICOLOGY TESTING

A3.1 Aim

This appendix considers and expands the CALA Policy on measurement uncertainty as it applies to environmental toxicology testing.

A3.2 Test Type

Most toxicology tests used by Canadian laboratories, for which CALA offers accreditation, require estimation of statistical endpoints (i.e., lethal concentration (LC_x), effective concentration (EC_x) and inhibition concentration (IC_x) where 'x' represents a specific effect level like 25% or 50% and/or the calculation of percent mortality (Environment Canada, 1999). Environment and Climate Change Canada or provincial environment ministries frequently require single concentration and LC_{50} acute lethality tests for the monitoring and control of industrial or municipal effluents. Accredited toxicology tests generally follow published standardized methods of Environment and Climate Change Canada and the USEPA, many of which are mandated under Canadian regulatory programs for monitoring and control of contaminants in effluents and sediments.

A3.3 Specification

All aquatic, sediment, and soil toxicity testing involve biological organisms, such as fish, invertebrate, bacteria, algae, and higher-level plants. The test result (statistical endpoint, e.g. LC_x, IC_x, EC_x or % mortality estimated for a given toxicity test) is specified in terms of a dilution of an environmental sample or concentration of a chemical and is based on observed effects on the exposed biological organisms as compared to the negative control response. The quantification of the endpoint, and its related uncertainty is, therefore, predominately associated with the test organism response.

A3.4 Quantitative and Semi-quantitative Assessments

Observed effects of the toxicant or toxicant mixture on test organisms (e.g. % mortality or inhibition) are used to assess the toxicity of the sample. Depending on the test design, different types of statistical endpoints are estimated based on one or more test observations. Single concentration tests involve the exposure of organisms to a single sample and a negative control. If these tests are conducted with replication, the data generated are suitable for quantitative analysis such as hypothesis testing. However, if the tests are conducted without replication, the available data are analysed in a semi-quantitative manner.

Tests conducted using a range of concentrations, such as dilutions of an environmental sample in an LC_{50} test, are commonly associated with endpoint estimates such as EC_x and IC_x, which are point estimates. Point estimates may also include the no-observed-effect concentration (NOEC) and the

lowest-observed-effect concentration (LOEC) for hypothesis testing, which are derived from quantal or quantitative analyses. Where there is limited response or mortality (e.g. little or no response in the test organisms at the highest concentration tested), the response data produced are suitable for a semi-quantitative assessment.

The data from quantitative tests can be analysed to derive an associated uncertainty much more readily than data from screening and semi-quantitative tests. Toxicology test methods often require statistical analysis of the result and require reporting such results with confidence limits or standard deviations of the mean result. This provides an indication of the uncertainty of an individual test but should not be confused with the uncertainty of the overall method which encompasses variability over time (e.g. different batches of organisms and different analysts).

A3.5 Type A and B Uncertainty Evaluations

As stated previously, there are two approaches that may be taken in estimating uncertainty, Type A and Type B. The Type A approach uses data from QA/QC work such as duplicate testing, reference toxicant testing, method validation studies and proficiency testing to estimate uncertainty. For example, cumulative reference toxicant data using a single species and toxicant can be used to show that the biological detector (test organism) is operating relatively consistently on a day-to-day basis. Proficiency tests are useful in showing that the biological detector is relatively constant between laboratories but show nothing about how the organisms will react to test samples containing different toxicants or toxicant mixtures.

Routine environmental toxicology testing (e.g. effluent monitoring) is not amenable to the Type A approach. The toxicant mixture is effectively unknown (e.g. a pulp-mill effluent containing hundreds of components and varying day-to-day) and there are no useful internal controls as in chemical analyses. Data from toxicological testing of unknown mixtures of toxicants cannot be accumulated and Type A evaluations are generally not applicable. A Type B evaluation, however, can still be used. By this approach, the contribution of individual factors is assessed and estimated, or data from individual tests are used to give an uncertainty estimate for the method. However, Type B evaluations on toxicology tests are not well covered in the toxicology literature and estimation of uncertainty is a best effort approach. A lab is required to estimate uncertainty for each of the endpoints in a test method (e.g. a survival and growth test will require uncertainty of survival results and growth results).

The testing laboratory must have in place a policy and procedure specifying the approach for estimating uncertainty as well as the circumstances under which they are applied. Measurement uncertainty calculations for each test must be updated with any change in test conditions or procedures (e.g., test is moved to a new water bath or shelf, a new supplier of organisms, a new balance, new technician, etc).

A3.6 Sources of Uncertainty

The possible sources of uncertainty for an environmental toxicology method are tabulated in many of the sources listed in this guide. Close examination of the steps in the laboratory methods and procedures will usually help to identify the likely sources of uncertainty in the method.

The toxicology laboratory must identify the sources of error in their laboratory (such as those listed below) and come up with an estimate of uncertainty for each of these components. The laboratory shall determine if any of these uncertainties is greater than 1/3rd of the major uncertainty (most likely to be the biological response, see Section 19 calculations).

If any estimated uncertainties exceed 1/3rd the value of the major uncertainty, the combined uncertainty must be given as described below. In other words, the uncertainty that is estimated must be a combined uncertainty of the biological response as described in Section 19 and other major sources of uncertainties listed below.

The toxicology laboratory must demonstrate that the other factors contributing to the uncertainty of a specific type of test are less than 1/3 of the biological response uncertainty. Only then can a lab claim that the uncertainty of the biological response as the major source of test uncertainty.

Some sources of uncertainty in toxicity tests may include:

- response of the biological detector (i.e., organism variability);
- sampling (at sample source and sub-sampling in the laboratory);
- transportation, storage and handling of samples;
- preparation of samples (e.g., glassware, balances);
- environmental and measurement conditions (e.g., temperature, light);
- preparation of standard materials;
- different technicians/analysts; and,
- maintenance of the test organism (culturing or holding).

Since a Type B evaluation is used, all sources of uncertainty should be considered, and their contribution to the expanded uncertainty evaluated. However, the major uncertainty is likely to be in the measurement step itself and, provided care is taken in the other steps in the process, the major uncertainty to estimate is that associated with the biological detector or test organism (i.e. the actual measurement).

The uncertainty associated with some processes is relatively easy to determine. For example, uncertainty in a dilution step may be about 0.1 to 0.5% (depending on variation in reading a pipette or measuring 25 litres of water etc.). Similarly, uncertainty associated with weighing is of the order of 0.1% or less depending on the balance (Eurachem CITAC, 2012).

Some sources of uncertainty, such as transportation of samples, are outside the control of the laboratory and cannot be included. Other processes are more dependent on the experience of the analyst; for example, the uncertainty associated with temperature measurement (within the allowable range) and the effects on the test animal during culturing and testing. What might be the

uncertainty associated with sampling given sediment and how this might affect the mortality of the test animal? What is the uncertainty that may result in selecting fish for tests - the uncertainty associated with all smaller vs. all larger fish (within limits) or how healthy the fish may be?

In comparison, the toxicity tests with known reference toxicant usually have a coefficient of variation in the range of 10% to 40%. Unknown sample results will likely have uncertainties exceeding this range. As a consequence, smaller contributors have much smaller significance. Variations in reference toxicant results may cover some of these factors (e.g. temperature control, health of the test animal, feeding the test animal) but not others. In any case, reference toxicants are not always run with every unknown sample and confidence intervals may vary depending on the degree of replication and number of test concentrations. Reference toxicant results should not be used to estimate uncertainty of uncontrolled factors.

If other factors are significant (more than 1/3$_{rd}$ of largest contributor), they have to be included in the final estimate to give a combined or expanded uncertainty.

A3.7 Approaches to Estimating Uncertainty of the Biological Response in Different Toxicity Test Types

Generally speaking, toxicology tests are integrative tests employing a biological detector. Tests are generally of two designs, those performed with undiluted samples, and those performed on diluted samples, both with or without replication.

A3.7.1 Estimating Uncertainty of Biological Response of Tests with Replicates

A3.7.1.1 TESTS WITH REPLICATES:

When the test is run with replicates, it is possible to use the standard deviation of these individual tests to estimate the uncertainty of the biological response for the method and the relative standard deviation for the test can be expressed as:

$$RSD_{test} = SD_{test}/X_{test}$$

Where RSD is the calculated relative standard deviation, SD is the standard deviation and X is the mean

The results from multiple tests can then be combined to estimate the overall uncertainty in the biological response for the method.

$$u_{biol\ resp} = \sqrt{\frac{\left((N_1 - 1)RSD_1^2\right) + \left((N_2 - 1)RSD_2^2\right) + \ldots}{(N_1 - 1) + (N_2 - 1) + \ldots..}}$$

Where RSD_i is the relative standard deviation of the toxicology result from the individual tests

N is the number of replicates in the test

The results from replicates may be identical and the resulting standard deviation is zero. Because of the nature of quantal testing, an uncertainty of zero is not an uncommon result. When pooling toxicology data for estimation of uncertainty for the method, every effort should be made to choose samples that reflect all types of samples processed in the lab; in addition, these results should include the full range of toxic response possible (e.g. no or low mortality, partial mortality and high or total mortality).

A3.7.1.2 TESTS WITHOUT REPLICATES:

When the test is always run singly, without replication, the uncertainty cannot be estimated in this way. It is recognised that in many situations, it is impracticable to run replicates and estimating uncertainty for individual tests is not possible. There is still an uncertainty associated with the method. Tests from duplicate testing can be used to provide an estimate of test variability. When no duplicate tests are available, the variation in results of the reference toxicant test can indicate some uncertainty (as discussed above) and may be the best effort available, but effort should not be made to include other data for estimates of uncertainty in the biological response. The reference toxicant data is only suitable if it has the same endpoints, duration and matrix as the samples conducted for laboratory clients.

A3.8 Combined and Expanded Uncertainty

If any contribution to the uncertainty (e.g. u_2) is greater than one third of the major contributor (usually the biological response) the uncertainties should be combined into a combined uncertainty as shown using relative uncertainties:

$$u_C = \sqrt{u_i^2 + u_2^2 + u_3^2 \ldots}$$

Since the method to combine the uncertainties involves summing the squares, any small contribution becomes much less important and can be disregarded.

Expanded uncertainty can be calculated in several ways. It can be calculated directly from the relative standard deviation (RSD) information by multiplying by a coverage factor (i.e., k = 2) to give the expanded uncertainty. In the case where a combined uncertainty has been calculated, the expanded uncertainty is determined using formula below:

$$U = k \times u_c$$

Where U is the expanded uncertainty, u_c is the combined uncertainty and k is the coverage factor. At this time, the appropriate value of k of toxicology tests is 2. If required by the client, the report must contain the result and the expanded uncertainty associated with that particular result with a description of how the measurement uncertainty was calculated.

A 3.9 Example: Uncertainty for *Daphnia magna* acute toxicity test

Step 1: List sources of uncertainties in the method:

- Storage of samples – allowable variation in temperature and holding time
- Sub-sampling in the laboratory
- Measurements of volume
- Environmental conditions during a test (such as temperature, lighting, etc)
- Receipt, holding and culturing of test organisms
- Analyst variability (experience, etc)
- Response of the biological detector (i.e. test organisms)

Step 2: Estimate the magnitude of each of these uncertainties using Type A or Type B evaluation.

Volumetric uncertainty:

Class A graduated cylinders 500 mL ± 4 mL as per catalogue information . The standard uncertainty was estimated assuming a rectangular distribution of the error.

u_{cyl} = (4 ml/√3)/500 ml = 0.0046

Adjustable pipettes data available from catalogues or laboratory's calibration.

10 ml pipette from calibration average volume 10mL and SD 0.05 ml

u_{pip}= 0.05mL / 10mL = 0.005

Pooled volumetric uncertainty = $\sqrt{(0.0046^2 + 0.005^2)}$ = 0.007

Storage of samples

May be able to look at the results of samples that were re-tested to see if variability with different storage times and temperatures within the holding times of the method. May be difficult to find this evidence and best judgement may be required.

Sub-sampling in the laboratory

This value may be small for aqueous samples but may be more significant for sediment or soil samples were non-homogeneity is more prevalent.

Some form of chemical testing may serve as a surrogate for the sub-sampling if the sub-sampling procedure is the same. For solid tests that use replicate data to estimate biological response, this may already be included in the biological response uncertainty and should not be considered twice.

Environmental Conditions During Testing

This is the influence of variability within the allowable range of a test. For example, you may look at tests for a reference toxicant (or some other sample) conducted at 18 versus 22°C – both which are within the acceptable range. If they are both within expected range (2SD of the mean) this may be negligible.

Culturing & Holding of Test Organisms

Estimate the uncertainty around culture health and holding. This may be best scientific judgement. Some evidence may include testing, such as the reference toxicant, when different criteria are observed. For example when the Daphnia magna culture mortality was 0% versus when it was 20% - both are acceptable according to the test method.

Analyst

The laboratory may look at analyst variability when side-by-side testing was conducted during training or alternate to PT testing.

The laboratory can also look at variability in reference toxicants for different analysts (i.e., is there a statistical difference in mean value from one analyst to the other).

Biological Response

Table 1. Mortality data from Daphnia magna single concentration test (triplicate analysis)

Test no	Mortality Rep 1	Mortality Rep 2	Mortality Rep 3	Mean Mortality	Std. dev.	N-1	RSD	(N-1) * RSD_2
1	0	0	0	0	0	2	0	0
2	0	0	0	0	0	2	0	0
3	90	100	100	96.66667	5.7735	2	0.059726	0.00713
4	100	100	100	100	0	2	0	0
5	100	100	100	100	0	2	0	0
6	70	50	60	60	10	2	0.166	0.05556
7	0	0	0	0	0	2	0	0
8	0	0	0	0	0	2	0	0
9	0	0	0	0	0	2	0	0

Test no	Mortality Rep 1	Mortality Rep 2	Mortality Rep 3	Mean Mortality	Std. dev.	N-1	RSD	(N-1) * RSD_2
10	20	10	10	13.33333	5.7735	2	0.43301	0.375
11	40	30	30	33.33333	5.7735	2	0.17320	0.06
12	10	0	10	6.66667	5.7735	2	0.86602	1.5
13	60	70	50	60	10	2	0.16667	0.05556
14	10	20	10	13.33333	5.7735	2	0.43301	0.375
15	100	100	100	100	0	2	0	0
16	10	10	10	10	0	2	0	0
17	0	0	0	0	0	2	0	0
18	10	10	20	13.33333	5.7735	2	0.433013	0.375
19	80	80	90	83.33333	5.7735	2	0.069282	0.0096
20	90	90	100	93.33333	5.7735	2	0.06186	0.00765
							Sum (Ni-1)RSD_2	2.82050
							sum Ni-1	40
							u_{biol}	0.26554

Note for tests with all 3 replicates of zero the RSD was entered as zero

Step 3:

Determine the largest contributor to uncertainty and identify any of the other factors that are more than 1/3rd of this largest contributor (major contributors). For toxicity tests the largest contributor will probably be the biological response.

Volumetric Combined value = 0.007

Biological response (may include sub-sampling, analyst variability)

u_{biol} = 0.265

The largest contributor to the uncertainty is the biological response. The data used to calculate this uncertainty also includes factors for subsampling and analyst variability. Other factors that are very small are not included.

Step 4:

Look at all of the major contributors to uncertainty to see if they are already addressed by one of the other major contributors. For example, if the biological response and analyst variability were both found to be significant but the data used to calculate the biological response includes reference toxicants conducted by different analysts or side-by-side testing with different analysts, then the analyst variability is incorporated in the estimate for biological response and will not be added in the combined uncertainty in step 5.

The largest contributor is biological response (which includes components from analyst variability, sub-sampling and environmental conditions during testing).

No other factors are greater than 1/3rd of the standard uncertainty for biological response therefore.

u_c = 0.265

Step 5:

Calculate the combined uncertainty u_c.

$u_c = \sqrt{(u_1^2 + u_2^2 + u_3^2 + ..)}$ where u_n are the major contributors to uncertainty if there are more contributors than biological response.

In this example there are none.

Step 6:

Calculate the expanded uncertainty U

$U = k \times u_c$

$U = 2 \times 0.265 = 0.530$ using a coverage factor of k =2

8.0 REFERENCES

1. ISO/IEC 17025:2017 - General Requirements for the Competence of Testing and Calibration Laboratories.
2. Ellison, S.L.R., M. Rosslein, and A. Williams, Editors, Quantifying Uncertainty in analytical Measurement, 3rd Edition, Eurachem/CITAC, available on internet at www.measurementuncertainty.org/mu/quam2.pdf, 2012.
3. Evaluation of measurement data — Guide to the expression of uncertainty in measurement, JCGM 100:2008 GUM 1995 with minor corrections, 2008.
4. ILAC Guide 17:2002, Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025. 2002, International Laboratory Accreditation Cooperation (ILAC).
5. International Vocabulary of Metrology - Basic and general concepts and associated terms (VIM), JCGM 200:12 3rd edition, 2008 version with minor corrections
6. CALA P07-2017 – CALA Application of Requirements in ISO/IEC 17025.
   https://cala.ca/wp-content/uploads/P07-CALA_Application-2017.pdf
7. CALA P19-01 - Measurement Uncertainty Policy.
8. ISO 11352:2012, Water quality – Estimation of measurement uncertainty based on validation and quality control data.
9. Nordtest Report TR 537, Handbook for the calculation of measurement uncertainty in environmental laboratories, 2nd Edition, 2003.
10. Environment Canada, Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology, Environmental Protection Series, EPS 1/RM/34 - December 1999.
11. European Cooperation for Accreditation EA-4/02. Evaluation of the Uncertainty of Measurement in Calibration September 2013 rev01.
    http://www.european-accreditation.org/publication/ea-4-02-m
12. European Cooperation for Accreditation (EA), Accreditation in Microbiological Laboratories. 2002.
    http://www.euroan-accreditation.org/publication/ea-4-10-g.
13. ILAC P10:01/2013 Policy on the Traceability of Measurement Results
    https://www.ilac.org/documents/ILAC_P10_01_2013.pdf.
14. NMKL: Norway. Measurement of uncertainty in microbiological examination of foods. 2nd. Ed. 2002.
15. NMKL Procedure No. 5 Estimation and expression of measurement uncertainty in chemical analysis 1997.

16. Barwick, V.J. and S.L.R. Ellison, VAM Project 3.2.1. Development and harmonisation of measurement uncertainty principles. 2000, LGC, UK, www.vam.org.uk, http://www.cala.ca/VAM_uncertainty.pdf, 2000.
17. Bell, Stephanie. A Beginner's Guide to Uncertainty of Measurement. Measurement Good Practice Guide No. 11. August 1999, Issue 2 with amendments March 2001.
18. Birch, K. Estimating Uncertainties in Testing, An Intermediate Guide to Estimating and Reporting Uncertainty of Measurement in Testing. 2001. Reprinted with minor amendments March 2003.
19. Niemela, S.I., A semi-empirical precision control criterion for duplicate microbiology colony counts. Letters in Applied Microbiology. 22(4): p. 315-319.1996.
20. Niemi, R.M. and S.I. Niemela, Measurement Uncertainty in Microbiological Cultivation Methods. Accred. Qual. Assur. 6: p. 372-375.2001.
21. EUROLAB Technical report 1/2007, Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation.
22. Voysey, P.A. and K. Jewell, Uncertainty Associated with Microbiological Measurement, Campden & Chorleyword Food Research Association 1999.
23. EUROLAB Technical Report 1/2006 – Guide to the Evaluation of Measurement Uncertainty for Quantitative Test Results.

9.0 REVISION HISTORY

Revision number	Revision date	Nature of revision
2.0	February 12, 2020	Created a separate guidance document from the information originally found in the annexes of the CALA uncertainty policy.