P19 – CALA Measurement Uncertainty Policy
Revision 1.10 – May 2010

CALA
Laboratory Accreditation

TABLE OF CONTENTS

CALA MEASUREMENT UNCERTAINTY POLICY

1.0 SCOPE

This policy is to be implemented by all accredited laboratories. Uncertainty is to be treated as one of the considerations examined during method validation

2.0 BACKGROUND

When ISO Guide 25 was re-written as ISO/IEC 17025 the requirement to estimate measurement uncertainty was added,

> *Testing laboratories shall have and shall apply procedures for estimating uncertainty of measurement. In certain cases the nature of the test method may preclude rigorous, metrologically and statistically valid, calculation of uncertainty of measurement. In these cases the laboratory shall at least attempt to identify all the components of uncertainty and make a reasonable estimation, and shall ensure that the form of reporting of the result does not give a wrong impression of the uncertainty. Reasonable estimation shall be based on knowledge of the performance of the method and on the measurement scope and shall make use of, for example, previous experience and validation data.*
>
> *ISO/IEC 17025:2005 clause 5.4.6.2*

> *When estimating the uncertainty of measurement, all uncertainty components which are of importance in the given situation shall be taken into account using appropriate methods of analysis.*
>
> *ISO/IEC 17025:2005 clause 5.4.6.3*

The only exception to the requirement to estimate uncertainty for each test is explained in a subsequent note,

> *In those cases where a well-recognized test method specifies limits to the values of the major sources of uncertainty of measurement and specifies the form of presentation of calculated results, the laboratory is considered to have satisfied this clause by following the test method and reporting instructions (see 5.10)*
>
> *ISO/IEC 17025:2005 clause 5.4.6.2 Note 2*

3.0 POLICY

Laboratories accredited under the CALA Accreditation Program for Environmental Laboratories shall fulfil the requirements of ISO/IEC 17025 with respect to the estimation of measurement uncertainty associated with testing for those tests which produce numerical results. This applies whether the test methods are rational or empirical.

Laboratories shall report the expanded uncertainty estimate as part of the reported result when the reporting of the estimate of measurement uncertainty is:
Required by the customer;
Required to establish that the data is fit-for-purpose; or,
Required because the data is being used to establish compliance (of the body being represented by the analysed sample) with a requirement.

The requirement which underlies this policy is that given in ISO/IEC 17025, Clause 5.4.6. Other documents and Guides may be used by laboratories to develop methods in meeting this requirement.

3.1 CALA Requirements

There are a few tasks that CALA requires for all estimates of measurement uncertainty. Further guidance is provided in section 4.0. As well, the appendices provide much greater detail for specific fields of testing. The guidance in sections 4.0 and in the appendices is intended to provide information, not as a prescriptive, step-by-step, procedure. The required steps are as follows:
Inventory all components of uncertainty in the test (e.g., sampling, sub-sampling, calibration, etc.);
Determine the significance of each component, eliminating any component that is insignificant;
Identifying all available data that can be used in the uncertainty estimate and identifying the component that it applies to (e.g., duplicate data, spike recovery data, etc.);
Identify any gaps in data; and,
Use the available data, and logically derived estimates where gaps exist, to calculate the expanded uncertainty. The coverage factor, k, is 2 when n is >29 or the appropriate (95% confidence level) Student distribution 't' (two tailed) factor for n<30.

3.1.1 Multi analyte tests

Where test methods generate multi-analyte data of 10 analytes or more (e.g., ICP, or GC techniques), laboratories shall select 3 analytes that represent each of three levels of uncertainty of the results small, medium and large levels of uncertainty - for which to estimate the measurement uncertainty.

3.1.2 Large Analytical Range

In cases where the analyte is expected to occur over a wide concentration range (more than a factor of 10), the estimation of uncertainty should be done at low, medium and high concentrations within that range. This will reflect the increase in the uncertainty with concentration. The other analytes will be classified within these three categories. If the relationship between SD and concentration is shown to be linear, the laboratory can estimate an expanded RSD.

3.1.3 Re-estimating Uncertainty

Laboratories will be required to re-estimate measurement uncertainty only when changes to their operations are made that may affect sources of uncertainty and these sources have not been shown to be unaffected through method validation or other studies.

3.1.4 Different Matrices

Laboratories shall make independent estimates of measurement uncertainty for tests performed on samples with significantly different matrices.

3.1.5 Use of Duplicate Data in Estimate

When laboratories select an analytical portion from a sample that may not be homogeneous, the laboratory shall include sub-sample uncertainty as part of the combined standard uncertainty calculation (e.g., soil, sediment, wastes, etc.).

4.0 GUIDANCE ON THE IMPLEMENTATION OF THE CALA MEASUREMENT UNCERTAINTY POLICY

There are two approaches that can be taken in estimating the uncertainty of measurement associated with testing. The first method, termed Type A, estimates uncertainties through the use of experimental data such as that from routine laboratory QA/QC work (duplicates, reference material usage, method validation studies, and proficiency testing (PT) and other inter-laboratory programs, for example).

The second method, Type B, involves the cause and effect-based metrological estimation of specific uncertainties from each identified source of uncertainty. It is similar to the approach used by calibration laboratories.

Several organisations and groups have published advice on the estimation of measurement uncertainty to the analytical laboratory community (see references).

The first approach (Type A) is the one used by most specifier agencies when requiring estimations of the uncertainty of measurement for analytical laboratories, and is the one that will be described below.

4.1 Using the Type A Approach

The following steps involve the use of experimental data to estimate the uncertainty of measurement for environmental laboratories. (See the Appendix for more detail):
Using the method SOP and the final result-calculation equation, identify and list all potential sources of uncertainty;
Identify and compile recent laboratory repeat analysis and PT data that is available;
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;
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);
Apply the appropriate coverage factor 'k'. (see 3.1.1 above); and,
Report the result with the expanded uncertainty and with a description of how the uncertainty was calculated.

4.1.1 Reporting the Uncertainty to Customers

As noted in the policy above, laboratories must report the measurement uncertainty:
When it is relevant to the validity or application of the test result,
When so instructed by a client, or
When the uncertainty affects compliance to a specification limit.

Estimates of measurement uncertainty quoted in reports shall reflect conservative "worst case" scenarios of variability incorporating long term effects on significant sources of uncertainty such as different analysts, instrument drift and other factors that reflect routine laboratory operations. A short description of how the estimate of uncertainty was determined should also be included. This description would include information on the source(s) of the data used to estimate the SDs included in the calculation of the combined standard uncertainty.

5.0 REFERENCES

ISO/IEC 17025:2005 - General Requirements for the Competence of Testing and Calibration Laboratories.
ILAC Guide 17: Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025. 2002, ILAC: Rhodes, NSW, Australia, http://www.ilac.org, 2002.
APLAC Policy, Interpretation and Guidance on the Estimation of Uncertainty of Measurement in Testing, Asia-Pacific Laboratory Cooperation, (APLAC) 2002.
Ellison, S.L.R., M. Rosslein, and A. Williams, Editors, *Quantifying Uncertainty in analytical Measurement*, 2nd Edition, Eurachem/CITAC, available on internet at www.measurementuncertainty.org/mu/quam2.pdf, 2000.

APPENDIX 1: MEASUREMENT UNCERTAINTY FOR ANALYTICAL CHEMISTRY

A1.1 Aim

This appendix considers and expands each of the CALA Protocol steps, as they apply to analytical chemistry, in more detail. It also explains what is meant and how to perform each task required.

A1.2 Sources of uncertainty

The possible sources of uncertainty for an analytical method are tabulated in many of the sources listed in paragraph 5 of the Policy. 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 in the method. ILAC Guide 17 lists these as:
definition of the measurand;
sampling;
transportation, storage and handling of samples;
preparation of samples;
environmental and measurement conditions;
the personnel carrying out the tests;
variations in the test procedure;
the measuring instruments;
calibration standards or reference materials;
software and/or, in general, methods associated with the measurement; and,
uncertainty arising from correction of the measurement results for systematic effects.

Spike or reference material recovery efficiency is an example of a source of variation in the test procedure category above:
Inter-laboratory bias and standard deviations derived from proficiency testing programs are examples of uncertainty in the correction for systematic effects. Data from reference material analyses can also be used for this purpose.
Recovery, as measured by the use of a reference material that is matrix matched to the samples or from a spike is not reported, its uncertainty should be measured (as part of method validation for example) but it need not be reported. If a recovery-corrected analytical result is to be corrected, the uncertainty of the corrected result reported must include an estimate of the uncertainty associated with the recovery value.

A1.3 Laboratory Repeat Data Sets

These are sources of repeated measurements from which SDs and RSDs can be calculated. The laboratory can vary one or more of the above sources of uncertainty during the collection of the repeat data and the SD calculated will include uncertainty attributed to the varied source(s). The various repeat data sets include:

Proficiency testing programs: These 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 provided the laboratory can demonstrate that their bias is within certain bounds, consistent with the collaborative study estimate of between-laboratory SD. It is, however, very likely to be an over estimate of what the intra-laboratory uncertainty actually is (by a factor of 2 according to conventional wisdom). Note the possibility of using PT results for bias detection and correction.

Reference sample insertion: both certified reference samples and in-house reference samples inserted into routine runs for control charting applications are a source of long term uncertainty data. Sources of uncertainty that can vary during repeated insertion of these samples over time are analysts, calibration sets, calibration solution sources, environmental conditions, instrument drift and many more. As a consequence, the standard deviation calculated from this data will reflect the uncertainty contributions from these sources. Sources of variability that are not included however, are factors that can change from sample to sample such as matrix effects and sample non-homogeneity (or heterogeneity). If a reference sample is to be used to estimate a bias, the uncertainty in the bias estimate must include the uncertainty in the certified value of the reference material. The combined standard uncertainty must include the bias uncertainty if the result is corrected for bias.

Spike recovery data: Can give the same information as reference sample insertion and in some cases can reflect variability due to different sample matrices. This type of interpretation should be made with caution however since the spike portion may be 100% recovered but the analyte portion may not be (due to speciation differences for example).

Method validation replicate data: This 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. A thorough discussion of the use of method validation data in the estimation of uncertainty is *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. This can be downloaded as a pdf file from the VAM web site.

Sample duplicate insertion: This can be a valuable source of uncertainty data, known as replicability $SD_{dupl.}$, 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).

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 accounted for. More than 20 duplicate pairs should be run of samples of a similar concentration.

$$SD_{dupl} = \sqrt{(R^2/2N)}$$

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 (see paragraph 2 of A1.8 below).

A1.4 Match Repeat Data with Uncertainty Sources

The objective of this step is to select laboratory Quality Control and validation data that includes as many sources of variability as possible so that these do not have to be estimated using the more difficult (and time consuming) Type B approach. The most effective means of achieving this is to design the analytical method to ensure spikes, reference samples and duplicates are inserted as early as possible into the analytical run. In addition, from the "Method Validation Replicate Data" section in A1.3 above, the method validation program should include the variation of as many potentially significant sources of uncertainty as possible.

A1.5 Estimate the Uncertainty for any Sources not Accommodated 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. Examples of these calculations are found in the EURACHEM/CITAC guide *Quantifying Uncertainty in Analytical Measurement* available as a pdf file from their web page.

A1.6 Tabulate 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.

A1.7 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).*

If no actual data is available, a first approximation of the inter-laboratory reproducibility RSD is given by:

$$RSD_R = 2 \times C^{(-0.15)}$$

The intra-laboratory RSD is one half of that (Official Journal of the European Communities L 77, 16.3.2001, p. 14). The formula is matrix and analyte independent but is unreliable at low and high concentration extremes.

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 provided the spike was inserted at the very beginning. 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. 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 SD from historical spike recovery data to be SD_{spike}, the bias uncertainty from a Proficiency Testing Program to be SD_{PT} and the sample non-homogeneity SD from sample duplicate insertions to be SD_{hom} and that no other sources of uncertainty have an SD larger than 1/3 of the largest of these three, the combined standard uncertainty SD_C is given as:

$$SD_c = \sqrt{(SD_{spike}{}^2 + SD_{PT}{}^2 + SD_{\text{hom}}{}^2)}$$

A1.8 Applying the Coverage Factor "k"

The Expanded Uncertainty is derived by multiplying the Combined Standard Uncertainty 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. If $n \geq 30$, $k = 2$. If $n < 30$, k is the appropriate Student's t factor for n-1 degrees of freedom and a 95% confidence level.

A1.9 Reporting the Result

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

A1.10 Uncertainty at the Limit of Detection and at the Limit of Quantitation

Only when a measured value is larger than the uncertainty with which it can be measured does it have any credibility. This point is known as the Limit of Detection (LOD). The lowest concentration at which a result can have a meaningful uncertainty assigned to it is the Limit of Quantitation (LOQ). The LOD has been most commonly set at a concentration that gives a signal that is 3 times the standard deviation of the measurement process at zero concentration or $3s_o$. Similarly, the LOQ has been set at $10s_o$.

The value for s_o is the Method Detection Limit determined as described in the CALA document *Quality Control for Environmental Laboratories*. This gives a relative uncertainty at

the LOD and LOQ of ± 100% and ± 30% respectively, both with a 95% confidence level. (J.K. Taylor, *Quality Assurance of Chemical Measurements* Lewis Publishers Inc., pages 79-82 (1987).

A1.11 Hierarchy of Data Selection for Estimation of Uncertainty

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 (IV) Least Suitable:

Uncertainty Specified within the Method: In those cases where a well-recognized test method (such as a peer-reviewed AOAC method or one published by agencies such as the Ontario MOE, the US EPA or ASTM) specifies limits to the values of the major sources of uncertainty of measurement and specifies the form of presentation of calculated results, the laboratory should follow the reporting instructions (see Note 2 to Clause 5.4.6.2 of CAN-P-4D (ISO/IEC 17025).

> Note: The laboratory would be expected to demonstrate that their results obtained when using this method have the reliability specified in the method in order for this clause to apply.

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 50 points collected from their insertion into routine analytical runs. See paragraph 4 of this Appendix above.

Pooled Sample Replicate Data: In cases where sample replicates are analysed and there is sufficient data above the limit of quantitation, include pooled sample replicate data to estimate uncertainty that incorporates sub-sample uncertainty as a source. See A1.3 above.

Proficiency Testing Sample Data: In cases where the previous options are not available and where Proficiency Testing samples are analysed with sufficient data above the limit of quantitation, pooled Proficiency Testing sample data can be used to estimate uncertainty. See A1.3 above.

A1.12 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

A1.12.1 Steps to using this Table:

Define the measureand(s), the analyte, the measurement objectives required for data to be "fit-for-purpose" (includes LOD, precision, accuracy, analytical range, selectivity etc.);
List the anticipated sources of uncertainty (including parameters found in the equation used to calculate the final result to be reported);
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);
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;
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;
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;
Compile the information into the table above and check to ensure that a source of uncertainty has not been counted more than once;
Remove those sources of uncertainty that have a standard deviation less than 1/3 the largest standard deviation;
Combine the remaining standard deviations using root sum of squares (RSS) technique (See paragraph 1 of A1.8 above);

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

APPENDIX 2: MEASUREMENT UNCERTAINTY FOR MICROBIOLOGICAL TESTING

A2.1 Aim

This Appendix applies to environmental microbiological testing methods that are quantitative and whose uncertainties are based on Type A estimates. These include 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.).

This appendix considers and expands each of the CALA Protocol steps in more detail. It also explains each step and how to perform each task required.

Dr. William Mills developed the first version of this appendix. CALA also acknowledges the work of Mr. Michael Brodsky, the author of this revision, and the assistance of the CALA Microbiology Technical Assessors for their comments as this revised draft evolved. Garry Horsnell provided all of the real data used in the addenda to this Appendix.

A2.2 Components of Uncertainty

Paragraph 8 of this CALA Policy lists organizations that have published documents, which present possible sources of uncertainty for a microbiological method. 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:

Sampling;

- Source of sample;
- Method of sampling;
- Transportation time and temperature of sample;
- Storage time and temperature of sample after receipt until analysis.

Method of Analysis;

- Source (SMEWW, AOAC, ASTM, in-house);
- Level of performance verification or validation.

Culture Media and Reagents;

- Formulation specifications;
- Preparation protocols;
- Water quality ;
- Performance verification criteria;
- Storage conditions and shelf-life.

Analytical Procedure;

- Sample homogenization/mixing;
- Subsampling;
- Preparing and dispensing dilutions;
- Inoculation procedure e.g. Filtration technique;
- Incubation conditions;
- Reading, interpreting and reporting results;
- Microbial density.

Equipment;

- Maintenance;
- Calibration;
- Repair.

Personnel;

- Hiring;
- Validating & Maintaining Competency.

The uncertainty which may be associated with sample holding time, if all tests are run within the allowable holding time, (e.g. often within 30 hours of sample collection) will not be considered in this policy;

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; and,

Niemelä [22] and [19] has provided a good discussion of these sources of uncertainty for microbiological methods.

A2.3 Measures of Spread or Dispersion (Precision)

Variance,S^2 is estimated as:

$$S^2 = \frac{\sum (x_i - \overline{X})^2}{N-1}$$

Where x_i = data point

$\overline{X}$ = the mean or average of all data points

Standard Deviation, SD is estimated as:

$$SD = \sqrt{S^2}$$

Relative Standard Deviation, RSD, is estimated as:

$$RSD = \frac{SD}{\overline{X}}$$

95% confidence interval for population mean is estimated as:

$$\overline{X} \pm (Students\ t \times SD / \sqrt{N})$$

Precision can be measured by:

i) Repeatability or Replicability: measures random error of the method for duplicate tests performed under identical conditions (within analyst variation)

For individual analysts repeatability:

$$S_r^{\ 2} = \frac{\sum (x_i - \overline{X})^2}{N-1}$$

Repeatability Variance (S_r^2):

$$S_r^2 = \frac{\sum Between\ duplicate\ \text{var}\ iances}{N}$$

or

$$S_r^2 = \frac{\sum (Differences\ between\ duplicates)^2}{2N}$$

Where N= number of samples or duplicate pairs

Repeatability Standard Deviation $SD_r = \sqrt{S_r^{\ 2}}$

Relative Standard Deviation of Repeatability (RSD_r) = $SD_r / \overline{X}$

ii) Reproducibility: measures random error under changed conditions of measurement ("ruggedness of the test") (between analyst or laboratory variation):

$$S_R^2 = \frac{\sum (x_i - \overline{X})^2}{N-1}$$

If between-analyst duplicates are used to measure reproducibility, reproducibility variance (S_R^2) may be determined in the following manner (see Addendum 1 to this Appendix):

$$S_R^2 = \frac{\sum (\text{var}\, iances\ of\ \ between\ analyst\ duplicates)}{N}$$

or

$$S_R^2 = \frac{\sum (differences\ of\ \ between\ analyst\ duplicates)^2}{2N}$$

Where, n= number of samples or duplicate pairs

Reproducibility Standard Deviation =SD_R= $\sqrt{S_R^2}$

Relative Standard Deviation of Reproducibility (RSD_R) = $SD_R / \overline{X}$

A2.4 Laboratory Repeat Data Sets

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:

Proficiency testing programs;

- Pooling of data derived by different methods diminishes the usefulness of PT information for estimating measurement of uncertainty;
- 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.

Reference sample insertion;

- There is a general lack of reference materials available for routine use in microbiological methods. At least one vendor claims to be able to provide freeze-dried pellets that will provide a guaranteed range of colony forming units (CFU).

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.

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.

A2.5 Reproducibility Calculations for Estimating Combined (Uc) and Expanded Uncertainty (Ue)

The combined uncertainty (Uc) 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 within analyst relative standard deviation (RSD_r) and the between analyst or laboratory relative standard deviation (RSD_L) and calculate the combined uncertainty (Uc). (NOTE: Data used to determine within analyst variance cannot be used to also determine between analyst or laboratory variance).

A2.5.1 Combined uncertainty

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

Alternatively, 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 uncertainty of measurement.

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. (Addendum 4 presents examples).

In this situation, combined uncertainty may be reduced to:

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

A2.5.2 Expanded uncertainty (Ue)

Ue= *k* (coverage factor for 95% confidence) x Uc (RSD_R)
The expanded uncertainty (Ue) is 2 x the combined uncertainty (Uc) if 30 or more values are used to calculate the SD or RSD.
All calculations can be easily handled using Microsoft Excel Spreadsheets.

A2.6 Data Handling

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 will be approximately normally distributed, which allows the use of the arithmetic values for statistical evaluation. (See Addendum 2)

Furthermore, 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 (See Addendum 3). Therefore, data should be separated into ranges (as indicated below) and, the combined uncertainty (Uc) 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.

A2.7 Evaluation Of Results Against A Microbiological Guideline

Such a guideline might be expressed as:

HPC ≥ 500 CFU/mL

HPC = Heterotrophic Plate Count and is the same, in some instances as, "Total Viable Count," "Total Plate Count," "Aerobic Plate Count," or "Standard Plate Count." HPC is the preferred designation in water microbiology.

For a result to be considered as having exceeded a guideline, the lower limit for the confidence interval of measurement uncertainty is required to be above this value. This represents the most conservative estimation of uncertainty.

For example, if the limit is 500 CFUs/mL, NMKL [17] states that a result of 500 CFUs/mL +/- 45 CFUs/mL is NOT considered to exceed the guideline. 545 CFUs/mL +/- 45 CFUs/mL is the **smallest** count that can be reported as exceeding the same guideline.

Alternatively a χ^2 (chi-square test) can be used:

$$X^2 = \frac{(C - L)^2}{L}$$

$\chi^2 \geq 4$ **OR** $C \geq \sqrt{L}\ (+/-(2/\sqrt{L}))$ (95% confidence limits)

Where C = colony count
And Where L = Limit value

The microbial guideline will be exceeded if either $\chi^2 > 4$ or if the count, C, is > L (+ $2/\sqrt{L}$)

For a guideline of 500, the # of CFUs in the sample (of the example given above) would have to be ≥ 545 to be statistically in excess of the guideline

A2.8 Most Probable Number Methods (MPN)

The Draft APLAC Uncertainty Guideline accepts the data in the McCrady's tables [30, 31] as reasonable estimates of uncertainty for MPN results.

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 and root causes identified - then corrected.

A2.9 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 taken into account 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 A2.2).
- 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.

A2.10 Hierarchy of Data Selection for Estimation of Uncertainty

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 Most Suitable, to Least Suitable.

Uncertainty Specified within the Method: In those cases where a well recognized test method (such as a peer-reviewed AOAC method or one published by agencies such as the Ontario MOE, the US EPA or ASTM) specifies limits to the values of the major sources of uncertainty of measurement and specifies the form of presentation of calculated results, the laboratory should follow the reporting instructions (see Note 2 to Clause 5.4.6.2 of ISO/IEC 17025).

e.g. Pour Plate counting (SMEDP)

Relative Standard Deviation of Repeatability, RSD_r

RSD_r r ≤ 7.7% (0.077)

Relative Standard Deviation of Reproducibility, RSD_R

RSD_R R ≤ 18.2% (0.182)

Calculation of Combined uncertainty, Uc:

Sum of Squares: $(0.077)^2 + (0.182)^2 = 0.0371 = 3.7\%$
Combined uncertainty = $\sqrt{0.0371} = 0.193 = 19.3\%$

Expanded uncertainty, Ue:

(Use coverage factor k=2 for 95% confidence)

Ue = k x Uc
= 2 x 19.3%
= 38.6%

Note: The laboratory would be expected to demonstrate that their results obtained when using this method have the reliability specified in the method in order for this clause to apply.

Quality Control Samples (QCS) and Spikes: In cases where matrix specific QCS and/or matrix spike data are available, include uncertainty estimated from the standard deviation of the LCS or matrix spikes of more than 30 points

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

Pooled Sample Replicate Data: In cases where sample replicates are analysed and there is sufficient data above the limit of quantitation, include pooled sample replicate data to estimate uncertainty.

A2.11 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 A2.11-1:
The Results of Duplicate Total Coilform (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
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
n = 30	Mean = 39	Mean D = 6.2	$\sum D^2 = 1861$	Mean Var = 31	n = 30

In Table A2.11-1, the number (n) of duplicate pairs is 30. So, 2n is 60. The mean of all duplicate values (counts) is 39.

It was mentioned earlier that the variance based upon duplicate counts from a series of samples could 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,

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

Table A2.11-2 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 A1-1.

Table A2.11-2:

Statistics and Uncertainty for Duplicate Total Coliform (TC) Counts in Table A1-1

Based on Variance for Duplicates $= \sum D^2/2n$		Based on Variance for Duplicates $= [\sum (\text{var pair 1} + \text{var pair 2} \ldots\ldots + \text{var pair n})] / n$	
Statistic	**Value**	**Statistic**	**Value**
Number of data pairs (n)	30	Number of data pairs (n)	30
2n	60		
Mean count	39	Mean count	39
$\sum D^2$	1861		
Variance ($\sum D^2/2n$)	1861/60 = 31	Mean Variance	31
SD	√31 = 5.6	SD	√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

So, either method is acceptable for calculating uncertainty based on duplicates.

A2.12 Addendum 2

Many microbiologists suggest that bacterial colony counts should be transformed or converted to the logarithm (base 10) of the counts before performing statistical analyses. However, this is not necessary if the untransformed data is already approximately normally distributed. Furthermore, it is not necessary if duplicate data, within a range of counts per filter, is analyzed separately.

Table A2.12-1 and A2.12-2 show that, the uncertainty based upon an analysis of duplicates per range will be similar regardless of whether the counts are transformed to their logarithm.

Table A2.12-1 shows, untransformed and log transformed, duplicate data in the range of 20 - 80 colonies per filter.

Table A2.12-1:

Untransformed and Log Transformed Data for Duplicate Total Coliform (TC) Colony Counts
Range 20 – 80 TC Colonies per Filter

TC/Filter Duplicate 1	TC/Filter Duplicate 2	D^2	Log TC/Filter Duplicate 1	Log TC/Filter Duplicate 2	D^2
46	45	1	1.662758	1.653213	0.000911
55	45	100	1.740363	1.653213	0.007585
47	41	36	1.672098	1.612784	0.003518
23	18	25	1.361728	1.255273	0.011333
23	23	0	1.361728	1.361728	0
34	38	16	1.531478	1.579784	0.002333
50	54	16	1.69897	1.732394	0.001117
14	21	49	1.146128	1.322219	0.031008
33	43	100	1.518514	1.633468	0.013215
69	61	64	1.838849	1.78553	0.002864
77	78	1	1.886491	1.892095	0.000314
26	24	4	1.414973	1.380211	0.001208
63	62	1	1.799341	1.792392	0.000483
42	38	16	1.623249	1.579784	0.001889
42	48	36	1.623249	1.681241	0.003363
36	41	25	1.556303	1.612784	0.00319
21	21	0	1.322219	1.322219	0
25	21	16	1.39794	1.322219	0.005734
22	32	100	1.342423	1.50515	0.02648

Table A2.12-1:

Untransformed and Log Transformed Data for Duplicate Total Coliform (TC) Colony Counts
Range 20 – 80 TC Colonies per Filter

TC/Filter Duplicate 1	TC/Filter Duplicate 2	D^2	Log TC/Filter Duplicate 1	Log TC/Filter Duplicate 2	D^2
20	21	1	1.30103	1.322219	0.000449
52	61	81	1.716003	1.78533	0.004806
22	24	4	1.342423	1.380211	0.001428
29	23	36	1.462398	1.361728	0.010134
22	26	16	1.342423	1.414973	0.005264
31	30	1	1.491362	1.477121	0.000203
53	42	121	1.724276	1.623249	0.010206
66	51	225	1.819544	1.70757	0.012538
66	50	256	1.819544	1.69897	0.014538
39	22	289	1.591065	1.342423	0.061823
55	40	225	1.740363	1.60206	0.019128

Table A2.12-2 presents an analysis of uncertainty based on the duplicate data in Table A2-1.

Table A2.12-2:

Comparison of Statistics and Uncertainty for Untransformed Versus Log Transformed Duplicate Total Coliform (TC) Colony Counts from Table A2.12-1
Range 20 - 80 TC Colonies per Filter

Statistics and Uncertainty Based on Untransformed Data		Statistics and Uncertainty Based on Log Transformed Data	
Statistic	Value	Statistic	Value
n	30	n	30
2n	60	2n	60
Mean Count (C)	39	Mean Log Count	1.554043
$\sum D^2$	1861	$\sum D^2$	0.256
SD (dups)	5.6	SD (dups)	0.065
RSD (SD/mean)	5.6/39 = 0.14	RSD (SD/mean)	0.065/1.554 = 0.04
RSD% (RSD x 100)	14%	RSD% (RSD x 100)	4%
2RSD%	28%	2RSD%	8%
Uncertainty Range	C ± 28% C	Uncertainty Range	Log C ± 8% log C
Uncertainty At count = 39	39 ± 28% or 28 to 50	Uncertainty At count = 39 Where, log 39 = 1.591	1.591 ± 8% = 1.591 ± 0.127 or 1.464 to 1.718 Antilog = 29 to 52

The analysis shows that, when the duplicate data is analyzed per range, it will not make much difference if the uncertainty is determined with or without converting the duplicate counts per filter to logarithms.

A2.13 Addendum 3

Tables A2.13-1 and A2.13-2 show that, if analysts wish to convert colony counts to their logarithm, they should not combine all duplicate data from the entire acceptable colony counting range of 0 - 150 colonies per filter before analysis. They should still analyze duplicate data within ranges. Otherwise, they may obtain unrealistic estimates of uncertainty.

In Tables A2.13-1 and A2.13-2, the mean variance is used to estimate uncertainty. However, as mentioned earlier, $\sum D^2/2n$ can also be used to calculate variance and uncertainty.

Table A2.13-1 shows, untransformed and log transformed, duplicate data, which covers the range from 0 - 150 colonies per filter.

Table A2.13 -1:
Untransformed and Log Transformed Data for Duplicate Total Coliform (TC) Colony Counts
Data Lumped Together for the Entire Range 0 - 150 TC Colonies per Filter

TC/Filter Duplicate 1	TC/Filter Duplicate 2	Variance	Log TC/Filter Duplicate 1	Log TC/Filter Duplicate 2	Variance
2	1	0.5	0.30103	0	0.045310
2	4	2	0.30103	0.60206	0.045310
1	2	0.5	0	0.30103	0.45310
4	3	0.5	0.60206	0.477121	0.007805
6	8	2	0.778151	0.90309	0.007805
8	5	4.5	0.90309	0.69897	0.020832
15	7	32	1.176091	0.845098	0.054778
5	3	2	0.69897	0.477121	0.024608
2	4	2	0.30103	0.60206	0.045310
12	16	8	1.079181	1.20412	0.007805
8	14	18	0.90309	1.146128	0.029534
6	4	2	0.778151	0.60206	0.015504
8	12	8	0.90309	1.079181	0.015504
1	2	0.5	0	0.30103	0.045310
9	2	24.5	0.954243	0.30103	0.213343
4	7	4.5	0.60206	0.845098	0.029534
7	4	4.5	0.845098	0.60206	0.029534
1	3	2	0	0.477121	0.113822
3	6	4.5	0.477121	0.778151	0.045310
1	5	8	0	0.69897	0.244280
36	39	4.5	1.556303	1.591065	0.000604

Table A2.13 -1:
Untransformed and Log Transformed Data for Duplicate Total Coliform (TC) Colony Counts
Data Lumped Together for the Entire Range 0 - 150 TC Colonies per Filter

TC/Filter Duplicate 1	TC/Filter Duplicate 2	Variance	Log TC/Filter Duplicate 1	Log TC/Filter Duplicate 2	Variance
49	57	32	1.690196	1.755875	0.002157
74	61	84.5	1.869232	1.78533	0.003520
56	58	2	1.748188	1.763428	0.000116
100	101	0.5	2	2.004321	0.000009
123	110	84.5	2.089905	2.041393	0.001177
112	91	220.5	2.049218	1.959041	0.004066
103	108	12.5	2.012837	2.033424	0.000212
93	88	12.5	1.968483	1.944483	0.000288
96	93	4.5	1.982271	1.968483	0.00095

Table A2.13-2 shows the uncertainty, which will be obtained from the log-transformed data in Table A2.13-1.

Table A2.13-2:

Statistics and Uncertainty Based on Log-Transformed Duplicate TC Colony Counts in Table A2.13-1
When Data is Lumped Together for the Entire Range 0 - 150 Total Coliform (TC) Colonies per Filter

Statistic	Value
n	30
Mean Log Count (Log C)	1.039308
Mean variance	0.036626
SD	0.19
RSD (SD/mean)	0.19/1.039308 = 0.18
RSD% (RSD x 100)	18%
2RSD%	36%
Uncertainty Range	Log C ± 36% log C
Uncertainty for a count of 102 Where, log 102 = 2.009	2.009 ± 36% = 2.009 ± 0.72 or from 1.289 to 2.729 (as logs) Antilog 19 to 536

When duplicate data over the entire range from 0 - 150 colonies per filter was lumped together and log transformed, analysis indicated that the uncertainty surrounding a count of 102 would be from 19 to 536 colonies per filter.

However, if an analyst gets 102 colonies on duplicate 1, it is highly unlikely that the analyst will get either 19 or 536 colonies on duplicate 2 unless the analyst has made a serious blunder during filtration.

Because the duplicate data was heavily weighted to low counts (i.e. 0 - 19 colonies per filter) and because the data was lumped together rather than separated into ranges, the precision or uncertainty of counts in the high range was overestimated even though the duplicate counts were converted to their logarithm.

Table A2.13-3 shows that the estimate of uncertainty will be more realistic, regardless of whether the data is log-transformed, if the data from Table A3-1 is analyzed per range.

Table A2.13-3:
Statistics and Uncertainty Based on Duplicate MF Colony Counts from Table A2.13-1 When Data from the Range of 81 - 150 TC Colonies per Filter was Analyzed Separately

Untransformed		Log Transformed	
Statistic	Value	Statistic	Value
n	6	n	6
Mean count (C)	102	Mean log count	2.00449
Mean variance	55.8	Mean variance	0.00097
SD	7.5	SD	0.03
RSD (SD/mean)	7.5/102 = 0.07	RSD (SD/mean)	0.03/2.00449 = 0.15
RSD% (RSD x 100)	7%	RSD% (RSD x 100)	1.5%
2RSD%	14%	2RSD%	3%
Uncertainty Range	C ± 14% C	Uncertainty Range	Log C ± 3% log C
Uncertainty for a count of 102	102 ± 14% = 102 ± 15 or from 87 to 117	Uncertainty for a count of 102 Where, log 102 = 2.009	2.009 ± 3% = 1.949 to 2.069 Antilog 88 to 118

The analysis shows that, when untransformed data for the range 81- 150 is analyzed separately, the estimate of precision or uncertainty for a count of 102 will range from 87 to 117 colonies per filter. Using a log transformation, the estimate of precision or uncertainty for a count of 102 will range from 88 to 118 colonies per filter.

Now, the estimates of uncertainty are similar, more in line with the 95% confidence limits based on Poisson scatter and more realistic. Once again, a log transformation is not necessary.

A2.14 Addendum 4

A2.14.1 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 (Uc).

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 (Uc) from only the uncertainty for filtering plus the uncertainty for counting among analysts.

A2.14.1.1 Method 1 (Testing Among all Analysts)

On 5 or more separate occasions, get all analysts to test the same sample but get one analyst to count the colonies on all filters. This will eliminate any variation associated with differences in counting among analysts and give the variation associated only with differences in filtering technique among analysts.

In addition, on 5 or more separate occasions, get all analysts to count target colonies on the same filter. This will provide the variation associated only 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 A2.14-1 and A2.14- 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 RSDs[2]. This is followed by a calculation of combined uncertainty (Uc).

Table A2.14-1:
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	TC/Filter Sample 1	TC/Filter Sample 2	TC/Filter Sample 3	TC/Filter Sample 4	TC/Filter Sample 5	TC/Filter Sample 6
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 = √47 = 6.9				
		RSD = 6.9/51 = 0.135				
		RSD^2 = 0.018				

Table A2.14-2:

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 = √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 (Uc).

$$Uc = \sqrt{RSD^2_{\text{(FILTRATION AMONG ANALYSTS)}} + RSD^2_{\text{(COUNTING AMONG ANALYSTS)}}}$$

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

$$Uc = \sqrt{(0.018 + 0.0092)}$$
$$= 0.165$$

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

A2.14.1.2 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 A2.14-3 shows how to organize the duplicate data and calculate the RSD^2 for a range. This is followed by a calculation of combined (Uc) and expanded uncertainty (Ue).

Table A2.14-3:
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.

Table A2.14-3:
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 A	Analyst Pair B	TC/Filter Duplicate A	TC/Filter Duplicate B	Variance
12	1	3			
13	1	4			
14	Etc.	Etc.			
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
Etc.	Etc.	Etc.	Etc.	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 (Uc), 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.

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

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

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

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 (Ue) will be 2 x (Uc)

$$Ue = 2 \times 0.165$$
$$= 0.33.$$

The expanded uncertainty (Ue) as an RSD%

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

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 of 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 endpoint estimates (i.e., a specific effect level such as lethal concentration (LC_X), effective concentration (EC_X) and inhibition concentration (IC_X) and/or calculation of percent mortality (Environment Canada, 1999). Environment 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 methods of Environment Canada and the USEPA, many of which are mandated under Canadian regulatory programs for monitoring and control of contaminants in effluents and sediments. The environmental toxicity tests that are offered within the CALA accreditation program are listed in Section 24 of this appendix.

A3.3 Specification

All aquatic, sediment, and soil toxicity testing involves 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. The quantification of the endpoint, and its related uncertainty is, therefore, 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 quanta1 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 limited 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.

A3.5 Type A and B Uncertainty Evaluations

As stated in this Policy, there are two approaches that may be taken in estimating uncertainty, Type A and Type B. CALA has used the Type A approach in developing its Policy. 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 an individual test is used to give an uncertainty estimate. However, Type B evaluations on toxicology tests are not well covered in the toxicology literature and estimation of uncertainty is a best effort approach.

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 Policy. Close examination of the steps in the laboratory methods and procedures will usually help to identify the likely sources of uncertainty in the method.

Basically, 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 in Sections 20 through 23 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;
sampling (at sample source and sub-sampling in the laboratory);
transportation, storage and handling of samples;
preparation of samples;
environmental and measurement conditions;
preparation of standard materials; 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 (and probably only) 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, 1990).

Some sources of uncertainty, such as transportation of samples, are outside the control of the laboratory and cannot be accounted for. 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 1/3rd of largest contributor), they have to be included in the final estimate to give a combined or expanded uncertainty (see Sections 20 through 23).

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

Generally speaking, toxicology tests are a broad-spectrum monitoring test employing a biological detector. Tests are generally of two types, those performed only with undiluted samples, with or without replication and those performed on a series of diluted samples.

A3.7.1 Single Concentration or Percent Mortality Tests on Undiluted Samples

Tests with replicates: When the test is run with replicates, it is possible to attach criteria for acceptability of replication and to calculate the mean and standard deviation of the results. This standard deviation may be used to estimate the uncertainty of the measurement and can be expressed as:

$$u = SD$$

Where u is the uncertainty and SD is the calculated standard deviation. For "combined uncertainty" (u_c), refer to sections 9-16, 20 and 21; for "expanded uncertainty" (U), refer to section 22.

Negative controls are run at the same time as test samples, with criteria for acceptability (e.g., 10% allowable mortality) as a measure of test validity. The results from replicates may be identical and the resulting calculated uncertainty is zero for that test and should be reported as zero. If the control results show more variation than the sample results, then the uncertainty associated with the control results is to be used. Because of the nature of quantal testing, an uncertainty of zero is not an uncommon result.

Tests without replicates: When the test is run singly, without replication, the standard deviation cannot be calculated and the uncertainty cannot be estimated in this way. It is recognised that in many situations, it is impracticable to run replicates and estimating uncertainty in these cases is not possible. There is still an uncertainty associated with the result (i.e., the uncertainty is not zero) but it cannot be readily estimated. The variation in results of the reference toxicant test can indicate some uncertainty (as discussed above) and may be the best effort available, but should not be used to estimate uncertainties.

> Under IS0/IEC 17025, the uncertainty estimate associated with a reference toxicant test is not applicable to the data obtained from a test on an unknown sample.

A3.7.2 LC_X, EC_X and IC_X Tests: Acute/Chronic with Lethal or Sublethal Responses

Toxicity tests where point estimate endpoints are calculated require the collection of data points from multiple dilutions (generally 5 or more test concentrations a negative control). Depending on the test, there may be replicates for each concentration in the dilution series. Different tests mandate different test designs.

If the testing results in no response, no uncertainty is attached to the result.

When responses are observed, point estimates may be calculated by using computer software to fit the data to a response curve (e.g. effect on the y-axis vs. log concentration on the x-axis), such as the LCx/Ecx/ICx endpoints and the associated confidence level (probit analysis). The software calculates a best fit for the response line and the variation in the actual data from the calculated straight line provides the values for calculating the confidence interval. For point estimates, the 95% confidence interval can be used as an acceptable estimate of the uncertainty.

For non-quantal (continuous) data, non-linear regression may also be used, and with sufficient data points, can be used to generate a best-fit line with the associated confidence interval or uncertainty.

Some tests, however, do not generate sufficient points to calculate a reliable confidence interval. An example of this is an LC_{50} test of an effluent, in which complete mortality is observed in the 100% concentration (undiluted effluent) and no mortality is observed in the 50% concentration (second highest concentration). In this situation, using the Binomial method or Spearman - Karber method gives a statistically conservative confidence interval that is an acceptable estimate of the true 95% confidence interval.

The testing laboratory should have in place a policy and procedure specifying the approach for estimating uncertainty as well as the circumstances under which they are applied.

A3.7.3 Toxicant Controls

Reference toxicants indicate nothing about the toxicity of an unknown contaminant or mixture. However, they do indicate if the detector (such as fish or daphnia) is behaving within specification. That is, they can be used to check the calibration of the bio-detector (such as a particular batch of fish). If the contaminant or contaminant mixture is known and constant, data can be accumulated. This data can be analysed by Type A evaluations and the estimate of the associated uncertainty can be made as discussed earlier, based on the standard deviation in a normal distribution of more than 30 measurements. The same is true of Proficiency Test results.

A3.7.4 LC50 Determinations for Known Substances

The testing approach for the LC_{50} uncertainty for known compounds is similar. Again, each determination is a specific toxicity test and uncertainty cannot be expressed in terms of data from other toxicants. However, sufficient data can and should be collected (either by multiple replicates at the one time or even by successive determinations over time) to generate a reliable estimate of the LC_{50} and the associated uncertainty. The LC_{50} uncertainty for known compounds can use a Type A or B approach.

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 (e.g. u_1) the uncertainties should be combined into a combined uncertainty as shown:

$$U_C = \sqrt{U_i^2 + U_2^2 + U_3^2 ....}$$

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 or SD_R) information by multiplying by a coverage factor (i.e., **k** = 2) to give the expanded uncertainty. For further detail, consult the main CALA policy document on estimation of uncertainty of measurement in environmental testing. 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, $\mathbf{u_c}$ is the combined uncertainty and **k** is the coverage factor. At this time, the appropriate value of **k** of toxicology tests is 2.

A3.9 Reporting the results

The report should contain the result and the expanded uncertainty associated with that particular result. If it has been established in the laboratory that biological response is the only major contributing factor (refer to sections 9-16, 20, and 21), the expanded uncertainty should be reported as follows:

For single concentration, or percent mortality tests on undiluted samples with replicates, the uncertainty associated with replicate results will be:

$$U_C = 2 \times SD$$

For single concentration or percent mortality tests on undiluted samples without replicates: no uncertainty will be attached.

For LC_X, EC_X and IC_X tests, the uncertainty estimation of the result will be the confidence interval calculated by the software used.

An indication of the major source(s) of the uncertainty and how it was estimated should be included where applicable, or as required.

The reference toxicant data result and its related uncertainty should also be included to indicate the reliability of the test. This result may also indicate some of the other contributors to the uncertainty (i.e. the u_2 factor above).

APPENDIX 4: DEFINITIONS OF TERMS USED IN THIS POLICY (REPRINTED FROM A2LA GUIDE[8])AND REFERENCES

4.1 Definitions

Accuracy (of measurement): (VIM 3.5): closeness of the agreement between the result of a measurement and a true value of the measureand

> 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

Correlation: (ISO 3534-1): the relationship between two or several random variables within a distribution of two or more random variables
NOTE: Most statistical measures of correlation measure only the degree of linear relationship.

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 3.10): result of a measurement minus a true value of the measureand

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 measureand.

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.

Influence quantity: (VIM 2.7): quantity that is not the measureand but that affects the result of the measurement

Examples: temperature of a micrometer used to measure length; frequency in the measurement of the amplitude of an alternating electric potential difference; bilirubin concentration in the measurement of haemoglobin concentration in a sample of human blood plasma.

Level of confidence: (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.

Measureand: (VIM 2.6): particular quantity subject to measurement
EXAMPLE: Vapor pressure of a given sample of water at 20°*C*.
NOTE: The specification of a measureand may require statements about quantities such as time, temperature, and pressure.

Measurement: (VIM 2.1): set of operations having the object of determining a value of 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 3.6): closeness of the agreement between the results of successive measurements of the same measureand 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.

Reproducibility: (VIM 3.7): closeness of the agreement between the results of measurements of the same measureand 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; observer; measuring instrument; reference standard; location; conditions of use; and, time.

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

Results are here usually understood to be corrected 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: (VIM 3.9): parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measureand

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 measureand, 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 VIM).

4.2 Bibliography

1. ISO/IEC 17025:2005 - General Requirements for the Competence of Testing and Calibration Laboratories
2. CALA P07 – CALA Application of Requirements in ISO/IEC 17025, www.cala.ca/P07-CALA_Interpretations.pdf.
3. Ellison, S.L.R., M. Rosslein, and A. Williams, Editors, Quantifying Uncertainty in Analytical Measurement, 2nd Edition, Eurachem/CITAC, available on internet at www.measurementuncertainty.org/mu/quam2.pdf, 2000.

4. ILAC Guide 17: Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025. 2002, ILAC: Rhodes, NSW, Australia, http://www.ilac.org, 2002.
5. Albert, R. and W. Horwitz, A Heuristic Derivation of the Horwitz Curve. Anal. Chem., 1997. 69(4): p. 789-790.
6. Uncertainties in qualitative testing and analysis. Accreditation and Quality Assurance, 2000. 5(8.): p. 346-348.
7. APLAC Policy, Interpretation and Guidance on the Estimation of Uncertainty of Measurement in Testing, Asia-Pacific Laboratory Cooperation, (APLAC) 2002.
8. Adams, T.M., A2LA Guide for the Estimation of Measurement Uncertainty In Testing. 2002, American Association for Laboratory Accreditation (A2LA): Frederick, MD. p. 42.
9. 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.
10. Estimation and Expression of Measurement Uncertainty in Chemical Analysis. 1997, NMKL. p. 15.
11. McQuaker, N., Quality Control for Environmental Laboratories. Revision 4.5 October 2001, CALA: Ottawa, ON.
12. Taylor., J.K., Quality Assurance of Chemical Measurements. 1987, Boca Raton, FL: Lewis Publishers Inc.
13. A2LA Interim Policy on Measurement Uncertainty for Testing Laboratories. 2000, American Association for Laboratory Accreditation (A2LA): Frederick, MD.
14. Excel spreadsheet used on the Course in Measurement of uncertainty in microbiological examination of food. 2002, NMKL, www.nmkl.org/Engelsk/publications.htm,2002.
15. Measurement of uncertainty in microbiological examination of foods. 2nd. Ed. 2002, NMKL: Norway, http://www.nmkl.org/Engelsk/reports.htm,2002.
16. NMKL, Measurement of Uncertainty in Microbiological Examination of Foods. 1999, NKML (Nordic Committee on Food Analysis. p. 22, www.nmkl.org,1999.
17. Accreditation in Microbiological Laboratories. 2002, European Cooperation for Accreditation (EA), http://www.europeanaccreditation.org/,2002.
18. Mills, W.J. Uncertainty in Microbiological Analysis of Environmental Samples. in CALA Uncertainty Workshop. 2001. Edmonton, AB: CALA.
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. Voysey, P.A. and K. Jewell, Uncertainty Associated with Microbiological Measurement. 1999, Campden & Chorleyword Food Research Association. p. 271999.
21. Niemi, R.M. and S.I. Niemela, Measurement Uncertainty in Microbiological Cultivation Methods. Accred. Qual. Assur. 6: p. 372-375.2001.
22. Niemela, S.I., Uncertainty of Quantitative Determinations Derived by Cultivation of Microorganisms. 2002, Centre for Metrology and Accreditation: Helsinki, Finland. p. 75, http://www.mikes.fi/documents/upload/Publication%20J3%202002_1.pdf,2002.

23. Norli, H.S. NMKL Procedure no 8, 2nd Ed., 2002: Measurement of uncertainty in microbiological examination of foodsProf. Eystein Skjerve. in AOAC Annual Meeting. 2002. Los Angeles, CA: AOAC.
24. Schop, R., Personal Communication, D.W.J. Mills, Editor. 2002: Toronto, ON2002.
25. USEPA, Membrane Filter Method for the Simultaneous Detection of Total Coliforms and Escherichia coli in Drinking Water. 2000, USEPA, Office of Research Environmental Protection and Development Cincinnati OH 45268: Washington, DC. p. 21, http://www.epa.gov/nerlcwww/MI_emmc.pdf, 2000.
26. USEPA, Improved Enumeration Methods for the Recreational Water Quality Indicators: Enterococci and Escherichia coli. 2000, United States Environmental Protection Agency, Office of Science and Technology , Washington DC 20460, http://www.epa.gov/ost/beaches/rvsdman.pdf,2000.
27. McQuaker, N.R., Measurement Uncertainty for Environmental Laboratories. 2000, CALA: Ottawa, ON2000.
28. Mills, W.J., Uncertainty Estimate for a Microbiological Dataset. 2002, Unpublished Data 2002.
29. Tholen, D., Telephone Conversation, W.J. Mills, Editor. 2002: Chicago, IL2002.
30. American Public Health Association. 1998. Standard Methods for the Examination of Water and Wastewater, 20th ed. (L.S. Clesceri, A.E. Greenberg and A.D. Eaton, Eds.) [re. MPN tables Section IX]
31. American Public Health Association. 1993. Standard Methods for the Examination of Dairy Products. 16th ed. [Section XI (A).
32. Environment Canada, 1999. Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology, Environmental Protection Series, EPS 1/RM/34 -December 1999.