technical guide

Uncertainty of Measurement Precision and Limits of Detection in Chemical and Microbiological Testing Laboratories

Published by:
International Accreditation New Zealand
626 Great South Road, Greenlane, Auckland 1005
Private Bag 28 908, Remuera, Auckland 1136, New Zealand
Telephone 64 9 525 6655
Facsimile 64 9 525 2266
Email: info@ianz.govt.nz
Internet: http://www.ianz.govt.nz

AS TG5, May 2004

ISBN: 09086 11 60 9

Contents

1 Introduction

Decisions relating to industrial production and also legal decisions are often based on test results. They can involve large sums of money or even a person's liberty. It is therefore important, not only for the analyst, but also for their clients or the decision makers to know how reliable test results are. What confidence can be placed on results, which indicate that a product is unsatisfactory or that a person has committed an offence? Some New Zealand chemical and microbiological testing laboratories do not know what confidence they can place in their results and therefore misinterpretation of insignificant differences is resulting in incorrect business decisions.

When analysts look at the results, they are faced with a large list of possible sources of variation, some causing a systematic bias to the results and some being random in nature. The possible combined extent of these variations needs to be known for each test method at typical analyte levels before the analyst can know whether their method and results are "fit for purpose" or not.

Known biases should be corrected (unless that is not the convention for the method). The remaining components of variation should be evaluated so that an overall uncertainty estimate may be made.

The general accreditation criteria for laboratories ISO/IEC 17025: 1999, states that, "*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 results 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*."

This International Accreditation New Zealand (IANZ) guide explains the steps, which may be taken to identify uncertainty components and to estimate the uncertainty of measurements in the fields of Chemistry and Microbiology. It also explains the use of precision in such estimates. Limits of detection and compliance with specification are also discussed. Later sections contain some useful formulae and examples for these fields.

Readers are referred to the Eurachem / CITAC Guide for more details and further examples on this subject. This document is obtainable from the Eurachem web site.

The Eurachem document suggests two approaches that may be used for the estimation of uncertainty in chemical measurements. The first is to identify each source of uncertainty, estimate the uncertainty of each source separately and combine them in the correct way to arrive at an overall uncertainty. The second approach is to look for available precision data or obtain suitable such data, then to identify the sources of uncertainty which are not included in this data, estimate the uncertainty of each additional source and then combine these with the precision standard deviation.

Telarc / IANZ since the mid 1980s has been encouraging chemical testing laboratories to obtain precision data and to estimate standard deviations, confidence limits and limits of detection from these data. It is therefore assumed that most such laboratories in New Zealand will have estimates of precision and preferably intermediate precision for their tests. Many will also have precision data from inter-laboratory comparison programmes in which they participated or from published data.

Microbiology laboratories will have or can obtain test results for replicate samples analysed preferably by different people, media, incubators, etc but usually on the same day (morning then afternoon). Different days and inter-laboratory comparisons are a problem in microbiology because of the lack of stability of samples.

Because of the availability of these data, the main approach described here for the estimation of uncertainty of measurement in chemical and microbiological testing is the second Eurachem one, starting with precision data and combining this with any missing components. Where chemistry laboratories are involved in characterising reference materials or require a more rigorous approach, the estimation of uncertainty for each contributing item and then the combining of these results will be the preferred approach.

Many New Zealand chemical and microbiological testing laboratories are also involved in monitoring environmental samples. In most cases, either the Health Department of a regional council will have set limits for levels of pollutants that samples may contain. Frequently, laboratory results indicate that none of a particular pollutant was detected. Such information is, of course, useless if the limit of detection for the method is above the limit set by the authority. This situation may occur when a laboratory has not correctly calculated the limits of detection for its methods.

The uncertainty estimates for low levels may be extended readily for the calculation of limits of detection and this is presented in the later sections of this guide.

2 Some Definitions

2.1 Uncertainty of Measurement (ISO / VIM 1993)

"A parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

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

2. 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 characterized by experimental standard deviations. The other components, which can also be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information.

3. It is understood that the result of the measurement is the best estimate of the value of the measurand and that all components of uncertainty, including those arising from systematic effects such as components associated with corrections and reference standards, contribute to the dispersion."

2.2 Measurement Traceability (ISO / VIM 1993)

The property of the result of a measurement whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.

ISO/IEC 17025 requires that
"*For testing laboratories the requirements given for calibration laboratories apply for measuring and test equipment with measuring functions used, unless it has been established that the associated contribution from the calibration contributes little to the total uncertainty of the test result. When this situation arises, the laboratory shall ensure that the equipment used can provide the uncertainty of measurement needed.*

Where traceability of measurements to SI units is not possible and/or relevant, the same requirements for traceability to, for example, certified reference materials, agreed methods and/or consensus standards are required as for calibration laboratories."

2.3 Measurement

The set of operations having the objective of determining a value of a quantity with a stated level of uncertainty

Where a test yields, or is based on, a result, which is expressed in numerical terms, the testing process will be regarded as a measurement for the purposes of this document and for the application of the uncertainty clause in ISO/IEC 17025.

2.4 Measurand

The particular quantity that is subject to the measurement.

2.5 Empirical Test Method

A test method where the result is not traceable to an SI unit, but depends on the defined steps of the method. A different test method (often for the same named analyte) may give a different result.

An empirical test method is a test method intended to measure a property, which is dependent on the test method used to measure it. Different methods may return different results, which may not be related. In many cases, the method cannot be verified using another test method. Examples of empirical test methods are "the leachable concentration of chemicals" and "the hardness of a material". For the former, different extractant and leaching conditions will produce different results.

For the latter, different indenter shapes and applied forces will produce different results.

A rational test method is a test method intended to measure a property, which is defined independently of any test methods. There is an objective true value to that property and the method can be verified using other test methods. Examples of rational test methods are those where the total concentration of a compound in a sample is measured. It is recognised that although there is an objective true value, it may be very difficult to measure that value.

All microbiology test methods are empirical as are many (most) chemical test methods (particularly where there is a digestion, extraction, chemical reaction or clean-up step or where steps in the method involve incomplete recovery of an analyte and there is no correction for bias). Corrections for spike recovery results or use of similar compound internal standards are attempts at correcting for bias and making the method closer to a rational method.

2.6 True Value

The value that would be obtained by a perfect measurement. This is indeterminate.

2.7 Conventional True Value

The value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose.

Frequently, a number of results of measurements of a quantity are used to establish a conventional true value.

2.8 Accuracy or Trueness of a Measurement

This is the closeness of the agreement between the result of a measurement and a true value of the measurand.

Accuracy, and particularly the extent of interferences must be assessed primarily during the selection and, if necessary, development of a test method. This requires detailed consideration of the ways in which various sample types may affect the measurement.

For well-researched methods, these aspects should have been covered for the range of sample types specified and therefore specific checks within each laboratory using that method may not be required. However, it is prudent to include accuracy checks such as running spiked samples and reference samples, or split samples in inter-laboratory trials, for which the conventional true values are well established. When introducing a well-researched method, the laboratory must still check that it gives reliable results, usually by use of certified reference materials and replicates.

Where a new test method is developed or a method is applied to sample types for which its accuracy has not been assessed, detailed interference and accuracy checks are necessary. These may include;

- Detailed consideration of the chemistry and physics of the method in relation to various sample types
- Cross-checking of results with completely independent methods
- Checking the effect of adding interfering substances to “known” samples
- Spike recovery tests
- Reanalysing samples from which the measurand has been removed.
- Inter-laboratory comparisons
- Certified Reference Materials

2.9 Bias

The systematic difference between results of measurements and the true value of the measurand.

2.10 Precision

Closeness of agreement amongst results of successive measurements of the same measurand.

Precision increases as random variations decrease. It is possible to have results, which are precise but not accurate.

They may be all close to a mean or average value but because of some systematic affect there is a bias and all are much higher (or lower) than the true value.

2.11 Repeatability “r”

Closeness of the agreement between the results of successive measurements of the same measurand carried out under the same measurement conditions (same laboratory, same sample, same method, same equipment, same materials, same staff, same time) e.g. from duplicates within the same batch.

2.12 Reproducibility "R"
Closeness of agreement between the results of measurements of the same measurand carried out under changed conditions of measurement (different laboratory, same sample, same method, different equipment, different materials, different staff, different time) e.g. from inter-laboratory comparison results.

2.13 Intermediate Precision
Closeness of agreement between the results of measurements of the same measurand carried out under changed conditions of measurement but within the one laboratory (same laboratory, same sample, same method, different equipment, different materials, different staff, different time) e.g. from replicate analysis of Quality Control samples or matrix reference materials over time.

2.14 Criterion of Detection
This is the lowest result at which the analyst may have (say 95%) confidence that some analyte has been detected rather than none. This is about 1.7 times the standard deviation of low-level results (the single sided statistic is used).

2.15 Limit of Detection
If a test result is just below the criterion of detection then the analyst cannot say that he has found nothing. The true value will be expected (say with 95% confidence) to lie between zero and two times the criteria of detection (the confidence interval around the criteria of detection). Therefore all that can be said is that the true result is less than this upper confidence limit viz. twice the criteria of detection.

Therefore the Limit of Detection is twice the Criteria of Detection or about 3.4 times the standard deviation of low-level results.

2.16 Limit for Quantification
For a numerical value to be placed on a low-level test result, the lower confidence limit associated with this value should be significantly above zero.

For a result at the limit of detection (say 10 units), for a normal distribution and reasonable degrees of freedom, the 95% confidence limits will be plus and minus 2/3.4 of the result (e.g. 10 units +/- 5.9 units). Such a result may not be helpful to the client.

It may therefore be reasonable to specify that the confidence limits for a reported numerical test result should be no more than say one quarter of that result.

Therefore to achieve this the limit for quantification should be about 8 times the standard deviation of low-level results.

2.17 Certified Reference Material
Reference material accompanied by a certificate, one or more of whose property values are certified by a procedure which establishes traceability to an accurate realisation of the unit in which the property values are expresses, and for which each certified value is accompanied by an uncertainty at a stated level of confidence.

3 IANZ Policy

It is the policy of IANZ that accredited testing laboratories shall comply with the requirements of ISO/IEC 17025 in relation to the estimation and reporting of uncertainty of measurement in testing.

The requirements are stated mainly in Clauses 5.4.6 and 5.10.3.1 c) of ISO/IEC 17025.

Procedures and recommendations stated in the documents referred to in the notes following Clause 5.4.6 are not requirements.

The measurement uncertainty definition in ISO – VIM 1993[2] applies.

4 General Interpretation And Guidance

The following guidance and interpretation are based on the APLAC TC 005 Guide.

4.1 Tests for which Uncertainty Applies
Where a test produces numerical results, or the reported result is based on a numerical result, the uncertainty of measurement for those numerical results shall be estimated. In cases where the nature of the test method precludes rigorous, metrologically and statistically valid estimation of the measurement uncertainty, a testing laboratory shall identify all significant components of uncertainty and make a reasonable attempt to estimate the overall uncertainties of those measurements. This applies whether the test methods are rational or empirical.

Where results of tests are not numerical (e.g., pass/fail, positive/negative or other qualitative expressions) estimates of uncertainty or other variability are not required. However, laboratories are encouraged to have an understanding of the variability of the results where possible and especially the possibility of false negatives or false positives.

The significance of the uncertainty of qualitative test results is recognised and so is the fact that the statistical machinery required to handle the calculation of such uncertainty exists. However, in view of the complexity of the issue and the lack of agreed approaches, IANZ does not require laboratories to estimate the uncertainty of qualitative test results at this time but will keep this under review.

4.2 Defining the Measurand

It is recognised that in chemical and microbiological testing, the measurand is often defined in terms of the method (empirical method) and is not directly traceable to SI. Care is need in defining the measurand to ensure that all uncertainty components will be identified and accounted for.

Traceability for chemical test methods often has several components. Balances, thermometers and volumetric glassware are traceable to SI units and full uncertainty budgets should be available if such measurements make a significant contribution to the overall uncertainty. Traceability of the final result calculation is usually to a reference material which, although often not complying with the definition (VIM) of a certified reference material, should be the best available. Where specified sample preparations, extractions, digestions, chemical reactions, clean-ups, temperatures, etc are included in the method and no correction for method bias (e.g. recoveries) is specified, the method is regarded as empirical and is traceable to its specified instructions. Different methods would give different results.

4.3 Identifying the Components of Uncertainty

The laboratory shall identify all the significant components of uncertainty for each test. An individual component contributing less than 1/5 to 1/3 of the total uncertainty of the measurement will not have much impact on the total uncertainty of the measurement. However, it would become significant if the method included a number of components of uncertainty of this size.

Even where reliance is to be made on overall precision data, the laboratory shall at least attempt to identify all significant components. This will provide information to confirm that the approach taken is reasonable and all significant components have been accounted for.

Flowcharting the steps of the test method and using fish-bone diagrams to present the uncertainty components provide useful approaches.

In some cases, groups of test method steps (similar sample type, same procedure, same weights and same volumes) may be common to several different test methods and once an estimate of uncertainty has been obtained for that group of steps, it may be used in the estimates of uncertainties for all methods where the group applies e.g. sub-sampling and sample preparation components applied to a specific sample type.

4.4 Approaches to the Estimation of Uncertainty

There are various published approaches to the estimation of uncertainty and/or variability in testing. ISO/IEC 17025 does not specify any particular approach. Laboratories are encouraged to use statistically valid approaches. All approaches, which give a reasonable estimate and are considered valid within the relevant technical discipline, are equally acceptable and no one approach is favoured over the others. The following paragraphs summarise of two approaches, which are described in the Eurachem document “Quantifying Uncertainty in Analytical Measurement.

Both the intermediate precision and reproducibility (from inter-laboratory comparisons) described in ISO 5725 (see clause 5.4.6.3 Note 3 of ISO/IEC 17025) may be used in estimating testing uncertainty. However, these alone may omit some uncertainty sources, which, if significant, should also be estimated and combined.

Method validation and verification data from repeat analysis of matrix reference materials, in-house standards, replicate analyses, inter-laboratory comparison programmes, etc., will be useful in establishing method precision. For chemical testing, a well-designed intermediate precision result or the precision from an inter-laboratory comparison will often incorporate the major components of uncertainty.

For microbiological testing, in most practical cases, precision will be the only significant component and the only one, which in practice can be readily estimated.

The uncertainty of physical measurements, the purity of calibration reference materials and their uncertainties, the uncertainties associated with recovery (bias) trials (when recovery factors are applied to results), as well as precision data shall all be considered in the evaluation of measurement uncertainty for chemical testing.
The ISO Guide to the Expression of Uncertainty in Measurement (GUM) (see clause 5.4.6.3 Note 3 of ISO/IEC 17025) is regarded as the more rigorous approach to the estimation of uncertainty. However, in certain cases, the validity of GUM estimates from a particular mathematical model may need to be verified, e.g., through inter-laboratory comparisons.

The "Number of Significant Figures" approach and Note 2 of ISO/IEC 17025 are not considered suitable for the evaluation of measurement uncertainty in chemical or microbiological testing.

4.5 Degree of Rigour

The degree of rigour and the method to be used for estimating uncertainty shall be determined by the laboratory in accordance with Note 1 of Clause 5.4.6.1 of ISO/IEC 17025.

To do this, the laboratory shall:

- Consider the requirements and limitations of the test method and the need to comply with "good practice" in the particular sector
- Ensure that it understands the requirements of the client (see clause 4.4.1a of ISO/IEC 17025). It is often the case that the client understands its problem but does not know what tests it requires or their uncertainty and needs guidance on the tests required for solving its problem
- Use methods, including methods for estimating uncertainty, which meet the needs of the client (see clause 5.4.2 of ISO/IEC 17025). It should be noted that what clients want may not be what they need
- Consider the narrowness of limits on which decisions on conformance with specification are to be made
- Consider the cost effectiveness of the approach adopted.

In general, the degree of rigour needed is related to the level of risk that can be tolerated.

Rigorous consideration of individual sources of uncertainty, combined with mathematical combination to produce a measurement uncertainty is considered appropriate for the most critical work, including the characterisation of reference materials. However, if an inappropriate model is used, this approach will provide an inadequate measurement uncertainty and is not necessarily better than the following approach. Estimation of measurement uncertainty based on the overall estimate of precision through inter-laboratory studies, method validation or other quality control data, taking into consideration additional uncertainty sources will be commonly used for chemical and microbiological testing. Additional sources that need to be considered may include sample homogeneity and stability, calibration/reference material, bias/recovery, equipment uncertainty (where only one item of equipment was used in obtaining the precision data).

In general, if less rigour is exercised in estimating measurement uncertainty, the estimated measurement uncertainty value should be larger than an estimate obtained from a more rigorous approach. Semi-quantitative measurements require less rigorous treatment of measurement uncertainty.

When a compliance decision is clear, then a less rigorous approach to measurement uncertainty estimation may be justified.

If the estimated uncertainty in reported results means that results will be unacceptable to the laboratory's client or decision maker, or will be too large for determination of compliance with the specification, the laboratory should endeavour to reduce the uncertainty, e.g., through identification of the largest contributors to uncertainty and working on reducing these.

4.6 Collaborative Trials and Proficiency Testing

Proficiency testing precision may not always provide complete measurement uncertainty data if significant aspects (e.g. sampling or sample homogeneity) have not been taken into account.

Three examples are:

- Matrix differences may occur between the proficiency test samples and the samples routinely tested by a laboratory.
- Result levels may not be the levels routinely tested in a laboratory and/or may not cover the full range of levels encountered in routine laboratory work.
- Participating laboratories may use a variety of empirical methods (different measurand) to produce the PT results.

Statistical analysis of PT results may give an indication of the precision obtainable from a particular method.

4.7 Uncertainty Arising from Sampling

Measurement uncertainty strictly applies only to the result of a specific measurement on an individual specimen.

During contract review, there shall be consideration and agreement with the client as to whether the test result and uncertainty are to be applied to the specific sample tested or to the bulk from which it came.

Where sampling (or sub-sampling) is to be treated as part of the test (measurand), the uncertainty arising from such sampling shall be considered by the laboratory. Estimating the representativeness of a sample or set of samples from a larger population requires additional statistical analysis.

Where a test method includes specific sampling procedures designed to characterise a batch, lot or larger population, the measurement uncertainties for individual measurements are often insignificant relative to the statistical variation of the batch, lot or larger population. In cases where the measurement uncertainty of individual measurements is significant in relation to the sampling variation, the measurement uncertainty shall be taken into consideration when characterising the batch, lot or larger population.

Where the test method includes a specific sub-sampling procedure, it is necessary to analyse the representativeness of the sub-sample as part of the measurement uncertainty estimation. Where there is doubt about the representativeness of a sub-sample, it is recommended that multiple sub-samples be taken and tested to evaluate the representativeness of each.

Where only one sample is available and is destroyed during the test, the precision of sampling cannot be determined directly. However, the precision of the measurement system shall be considered. A possible method for estimation of the precision of sampling is to test a batch of “homogeneous” samples for a highly repeatable measurand and calculate the sampling standard deviation from the results obtained.

4.8 Reporting Measurement Uncertainty

If the laboratory has not estimated the uncertainties of its measurements then it will not know whether its test results are valid for the intended use of the client. This is why ISO/IEC 17025 makes it a requirement for laboratories to estimate uncertainties for all measurements. However, there will be occasions when the laboratory decides not to report its uncertainties to its clients. ISO/IEC 17025, however, requires reporting of uncertainties in some specific circumstances.

For quantitative test results, measurement uncertainty shall be reported where required by clause 5.10.3 b) of ISO/IEC 17025, which includes the following circumstances:

- When it is relevant to the validity or application of the result
- When a client’s instructions so require
- When the uncertainty affects compliance with a specification limit

When measurement uncertainty is not reported under the provision of the third paragraph of clause 5.10.1 of ISO/IEC 17025, its absence shall not affect the accuracy of the conclusion, clarity of the reported information nor introduce any ambiguity in the information provided to the client.

Measurement uncertainty associated with results below the limit of quantification shall not be reported, as currently no reliable convention for this purpose exists.

The requirement to report measurement uncertainty when it is relevant to the validity or application of the test result will often need to be interpreted. In such cases, the client’s needs and the ability of the client to use the information may be taken into consideration. Although in the short term, some clients will not be in a position to make use of measurement uncertainty data, this situation can be expected to improve.

In this case, the laboratory may need to provide an interpretation of the results if the client could be misled by the numerical results alone.

When reporting measurement uncertainty the reporting format described in the GUM is recommended. The results of the uncertainty estimations would normally be reported based on a level of confidence of 95%. The indiscriminate use of a coverage factor of 2 is not recommended. Not all combined uncertainties are normally distributed and, where practicable, the uncertainty appropriate to the 95% confidence level for the appropriate distribution should be used. The coverage factor used for calculating the expanded uncertainty should also be reported.

When reporting the test result and its uncertainty, the use of excessive numbers of significant figures shall be avoided. Unless otherwise specified, the primary result shall be rounded to the number of significant figures consistent with the measurement uncertainty. When the test method prescribes rounding to a level that implies greater uncertainty than the actual measurement uncertainty, the uncertainty implied by this rounding should be reported as the measurement uncertainty of the reported result.

Laboratories shall have the competence to interpret measurement results and their associated measurement uncertainty to their clients.

At present, microbiological laboratories may decide not to report uncertainty unless required by clients. This is because uncertainties are very large and clients are not yet prepared to understand and accept these.

When reporting microbiological uncertainties, a description of the procedure used to estimate the uncertainty should also be included because, of the various methods used around the world, some give significant underestimates of the actual uncertainty.

4.9 Determining Compliance with Specification

Decisions on when and how to report compliance or non-compliance vary according to requirements of the client and other interested parties. However, the laboratory shall take its measurement uncertainty into consideration appropriately when making compliance decisions / statements and clients shall not be misled in relation to the reliability of such decisions.

In general, the principles described in APLAC TC 004 shall be followed.

Difficulties may arise where:

- Specifications do not quote the empirical method to be used and a number of methods for the analyte but giving different results are available.

- Where results are close to a specification limit one laboratory may state that a sample complies with the specification (their confidence interval for the result excludes the specification limit) whereas another laboratory may state that there is doubt as to whether or not that same sample complies with the specification (their confidence interval includes the specification limit).

4.10 Assessment for Accreditation

During assessment and surveillance of IANZ accredited laboratories, the assessment team shall evaluate the capability of the laboratory to estimate the uncertainty of measurements for all tests included in its accreditation scope. They shall check that the estimation methods applied are valid, all significant uncertainty components have been included, and all the IANZ criteria are met. The assessment team shall also ensure that the laboratory can achieve the claimed limits of detection.

4.11 Additional Interpretations for Chemical Testing

For empirical methods, the method bias is by definition zero and only individual laboratory and measurement standards bias effects need to be considered. Where results are corrected for recovery (from spiking experiments), the uncertainty associated with these spiking results should be included.

It is important to ensure that all appropriate effects are covered but not double accounted. Where appropriate, effects of matrix type and changes in concentration should be included in the measurement uncertainty.

Where proficiency testing precision data are used, it is important to ensure that the data used to estimate the measurement uncertainty are relevant. In particular, they should relate to the same measurand (same test method and matrix).

There may be issues if the proficiency programme includes laboratories with a wide range of backgrounds and skill levels.

In chemical testing, it is customary to evaluate the uncertainty at various selected levels of the analyte. However, when a measurement is being made to test for compliance with limit values, it is necessary to use an uncertainty value for test results close to the compliance limit. Therefore, it is useful to select the limit values as the levels at which the uncertainty is evaluated. This approach is most likely to provide the best estimate of the measurement uncertainty at levels adjacent to the limit values.

Professional judgement may be used for estimating the magnitude of uncertainty attributed to certain components where better estimates are not available or readily obtainable. In such cases, at least a short-term precision estimate of the component should be included in the evaluation. Professional judgement should not normally be used for significant uncertainty components. Where professional judgement has to be used for significant components, it must be based on objective evidence or previous experience. Measurement uncertainty estimates containing significant components evaluated by professional judgement shall not be used for applications demanding the most rigorous evaluations of uncertainty.

4.12 Additional Interpretations for Microbiological Testing

There are four main types of microbiological tests;

- General quantitative procedures
- MPN procedures
- Qualitative procedures
- Specialist tests, e.g. pharmaceuticals.

Various approaches to estimating uncertainty are available for general quantitative testing.

Quantitative microbiological methods are empirical because results are dependant on method defined conditions such as temperature, incubation period and media. Therefore, they do not have a method bias contributor to their measurement uncertainty.

A few certified reference materials are available for quantitative microbiological tests. Where they are available, their certified results are mostly obtained from collaborative studies. Therefore only consensus values relating to the specified method are available and, as with all other microbiological test methods, these methods are also empirical. These reference materials are useful in demonstrating competence of the laboratory but should not be used to quantify bias.

4.12.1 Use of Precision Data

The use of precision data as described in the Eurachem / CITAC Guide is applicable for microbiology. "Intermediate precision" as set out in ISO 5725 is considered to incorporate most if not all of the significant uncertainty components of microbiological tests.

Any components of uncertainty not included in intermediate precision, such as performance of different batches of media, variation in incubator conditions (where only one incubator is available), etc., may be examined for significance by other statistical means or by redesign of the precision exercise.

The distribution of results from plate count tests is not normal but skewed (long right tail). Such data may first be transformed by taking the logarithm_{10} of each result, to obtain a close to normal distribution. The log standard deviation and confidence limits may then be calculated before anti-logging each confidence limit separately.

Where other components of uncertainty such as sampling or equipment variations need to be combined with a precision result that was estimated using logs and anti-logs, sophisticated mathematical calculations may be necessary.

For those plate count tests where test results are less than about 20 CFU, the result may well be below the limit of quantification (see Section 12) and laboratories are cautioned about reporting numerical values for such results.

At this stage, very little method performance data are available from collaborative trials or proficiency testing, although this situation may change in the future. Proficiency testing data may not always provide suitable measurement uncertainty data because significant aspects may not have been taken into account:

- Matrix differences may occur between the proficiency test samples and the samples routinely tested by a laboratory. Samples may have been specially homogenised and stabilised.

- The population levels may not be the levels routinely tested in a laboratory and/or may not cover the full range of population levels encountered in routine laboratory work.
- Participating laboratories may use a variety of empirical methods (different measurand) to produce the PT results.

However, statistical analysis of PT results may give an indication of the precision obtainable from a particular method.

4.12.2 Most Probable Number (MPN) Procedures

For MPN procedures it is traditional to refer to McCrady's Tables to obtain the test result as well as the 95% confidence limits. These data have been established statistically but possibly without considering all sources of uncertainty. Some versions of these tables also contain significant errors arising from the rounding of data during their preparation.

Laboratories are encouraged to identify unusual combinations of positive tubes and to reject such results. If this is done effectively, then the uncertainties quoted in the tables will, in the mean time, be regarded as a reasonable estimate of uncertainty for these methods. This may be regarded as an application of Note 2 of clause 5.4.6.2 of ISO/IEC 17025.

4.12.3 Note 2 of Clause 5.2.6.2 of ISO/IEC 17025

Application of Note 2 for some areas of specialist testing e.g., pharmaceutical microbiological assays, may also be applicable, as the methods concerned include validation of the assay parameters, specify limits to the values of the major sources of uncertainty of measurement and define the form of presentation of calculated results.

4.12.4 Poisson Distribution

The Poisson distribution and confidence limit approaches as described in BS 5763 and ISO 7218 may significantly underestimate uncertainty as not all sources of uncertainty are taken into account. The distribution of particles / organisms in a liquid may be described using the Poisson distribution but other components of uncertainty associated with the test procedure are not included. Indeed some components of uncertainty such as dilutions and consistency of reading plates will not be described by a Poisson distribution.

For one-off analyses, the Poisson distribution approach will give a quick estimate of the uncertainty (see Note to 10.1.6 of BS 5763). However, this will likely give a significant underestimate of the actual uncertainty.

4.12.5 Negative Binomial Model

The negative binomial model described in ISO/TR 13843 may be more appropriate than Poisson alone as it covers the Poisson distribution plus "over-dispersion" factors.

5 Estimating Uncertainty Of Measurement Step By Step

5.1 Specify the measurand and decide if it is empirical or traceable to SI

A clear and unambiguous statement of what is being measured must be prepared.

Where the method is an empirical method (many chemical tests and all microbiological tests), some careful thought is needed because what is actually being measured is usually not what the name of the method implies.

For example, "fat in milk by gravimetric determination (Roese Gottlieb)" is actually "all the substances which extract from the milk sub-sample using the specified solvent, time, temperature and equipment and which remain after the drying stage". If some oil or grease had spilled into the sample these would also be measured as "fat" as would any oily contaminations from reagents or equipment. Some volatile "fat" components may also be lost during the drying stage. Maybe the client wants to know the "fat in the sample and not the sub-sample". This is a different measurand and the uncertainty of sub-sampling will need to be taken into consideration. If the client wants to know the "fat in the shipment" then this a different measurand again and homogeneity of the shipment will need to be taken into consideration. It is also noted that not all countries define fat in terms of lipid content. They may also define fat in terms of fatty acid content.

Another example is "total petroleum hydrocarbons in soil". Perhaps the measurand is "the material in the sub-sample, which is extracted by the specified solvent and which absorbs infra-red light at the specified wavelength and under the specified conditions of dilution, cell length etc as compared with a specified mixture of hydrocarbons".

One should not forget the material, which is removed (presumably fats) if the clean-up step

is invoked. But wasn't the client interested in the whole sample and not the sub-sample?

A microbiological example may be "Listeria in Milk". The measurand may be " the result of the specified calculation from the colonies which are counted when the sub-sample is resuscitated, diluted, plated, incubated (temperature and time) and otherwise treated as specified in the method"
Further examples of typical tests done in chemical testing laboratories, which are empirical and which require careful specification of the measurand are:

- Permanganate value
- Chemical Oxygen Demand
- Biochemical Oxygen Demand of Water
- Total Suspended Solids of Water
- "Weak Acid Extractable" or Lead in Soil
- Available Phosphorus in Soil
- Crude or Dietary Fibre in Food
- Lean Meat in Sausages
- Total Organic Nitrogen in Sausages
- Free Sulphur Dioxide in Wine
- Mercury in Fish
- Aflatoxins in Peanuts
- Arsenic, Antimony, Cadmium in Ceramics
- Flash Point of Avgas
- Acid and Base Number of Oil
- Cholesterol in Blood
- Cyanide in Stomach Contents
- Accelerants in Fire Debris.

The list goes on covering the large majority of all chemical test methods accredited by IANZ, which are empirical and are not traceable to the mole. Change the method and one has a different measurand giving a different answer.

All require careful definition before one can decide which components of uncertainty should be included in uncertainty estimates.

Questions also arise about whether the blank is pre-determined and subtracted and whether spike experiments give recovery results which are used to correct for method bias in an attempt to provide some traceability to the mole.

5.2 Write the formula for calculating the results

An example equation for a pesticide in bread being analysed by solvent extraction followed by gas liquid chromatography using an external standard is,

$$R = \frac{I_{samp} \cdot c_{ref} \cdot V_{extr}}{I_{ref} \cdot Rec \cdot m_{samp}} \mu g/g \quad \textbf{Formula 1}$$

Where
R = Pesticide result in sample (mg/kg)

I_{samp} = Peak intensity of sample extract

I_{ref} = Peak intensity of reference standard

c_{ref} = Reference standard concentration (ug/ml)

V_{extr} = Final volume of extract (ml)
Rec = Recovery

M_{samp} = Mass of investigated sub-sample (g)

5.3 Identify and list all possible sources of uncertainty

To assist with this process, it may be helpful to flowchart the entire method from sampling or sub-sampling to the final result including listing all materials and equipment used.

Items to note are:

- Start at the beginning – is sampling included in the measurand?
- Consider all inputs such as set-up of equipment and preparation of reagents
- Note control items such as masses, volumes, temperatures, times, pressures concentrations
- Note sub-sampling and sample preparation / dilution / digestion / extraction steps
- Note reference materials and their stated purities / values
- Are you using standard additions or spikes? Are you correcting for recoveries or using these only as quality controls?
- Are you doing duplicates and reporting their means?
- Are you correcting for an average blank or is a blank run with each sample or batch?
- Are dilutions made prior to presentation of samples to the instrument?
- What are the plate counting processes?

Each step of the method should be examined to identify possible sources of uncertainty.

Variations that relate to possible bias:

- Instrument calibration errors or drift
- Calibration curve fitting errors
- Balances, glassware or thermometer calibrations
- Reference material impurities
- Recoveries and blanks.

Random variations:

- Inadequate sample storage

- Sample homogeneity, sampling, sub-sampling, clumping, dilutions
- Sample storage and transport variations
- Sub-sampling
- Random sample matrix effects, stability, variability, form of binding of analyte (different to spike)
- Day to day differences
- Reagent variations
- Reaction stoichiometry – departures from expected chemistry or incomplete reactions
- Measuring and weighing
- Measurement conditions – temperature and humidity effects on measurements and sample stability, etc
- Errors in equivalence point detection
- Variations in volumetric glassware calibrations (tolerances)
- Repeatability of the method
- Instrument settings; readings; other effects
- Electronic bleeps on instruments and computers
- Computational effects – straight line or curved calibration and rounding too soon
- Blank correction – blank uncertainty and appropriateness to subtract blank
- Operator effects, colour changes, plate readings
- Other random effects (always include this).

Errors which are not considered for uncertainty estimates but which should be avoided or corrected:

- Incorrect sample
- Wrong client information
- Containers or exhibits mislabelled
- Samples preserved incorrectly
- Wrong method choice
- Errors on method card
- Wrong dilutions or dilution factors
- Basic chemistry wrong
- Determining incorrect species
- Matrix interference
- Wrong recovery factor applied
- Wrong chemicals or low quality chemicals
- Wrong units used or reported
- Incorrect standard solutions
- Calculation programme errors
- Computer programme errors
- Reporting wrong species
- Number transfer errors
- Calculation of data entry errors
- Sample tube mix-ups
- Typing errors.

5.4 List available data

Such data as calibration certificate uncertainties, inter-laboratory, in-house or published precision data, purity of reference materials, method validation data, etc should be identified.

Examples are:

- Method validation / verification data such as replicates of various samples or matrix reference materials under various conditions (intermediate precision)
- QC replicates or house standard repeat analyses under various conditions (intermediate precision) or same conditions (repeatability “r”)
- Inter-laboratory comparison statistics (reproducibility “R”)
- Published precision data from method validations (r and R)
- Certificates of instrument or reference material calibrations
- In-house equipment calibration data
- Published experimental research data.

5.5 Select suitable precision data

Select available precision data covering the maximum number of sources of uncertainty and express these as a standard deviation.

ILCP precision (R) will probably include variations from reference materials and instrument calibrations. However, it may not include real sample and sub-sampling variations as samples are usually specially prepared, homogenised and stabilised. Results from outlier laboratories may distort the R value.

Intermediate precision will probably not include variations in reference materials (same used for long periods) or larger equipment (only one in the lab) or uncertainty in the calibration of equipment. If intermediate precision is determined on real samples and samples are split before the sub-sampling procedure then sub-sampling variation will be included.

Precision data may not include the full variations in control points that are permissible within the defined method, because most labs tend to set these points to the centre of the permissible range (except during robustness experiments for method validation). Some measure of this may be desirable. When relying on published precision data (e.g. published R), the laboratory must demonstrate that its own precision (r) is comparable with the published “r”.

If results are to be corrected for method bias / recovery (using results from spiking studies) then the uncertainty of these spike tests should be estimated and incorporated.

Precision frequently varies significantly with the level of the result. The precision data should be adjusted for concentration of the analyte or a relationship established between precision and concentration.

For microbiological testing, precision data derived from the skewed distribution will have the upper confidence limit significantly further from the result than the lower one.

5.6 Calculate intermediate precision

5.6.1 Data quality

The design of the data collection procedure is the most important aspect of any estimation of intermediate precision. Data must include variations caused by different days, operators, equipment, laboratories, reagents etc if realistic estimations of intermediate precision are to be obtained.

Replicates analysed by the same analyst at the same time are likely to give an unrealistically small precision results (repeatability (r) only).

For estimating reproducibility, results obtained from several different laboratories, each analysing a sub-sample from the same sample or batch, are required.

Intermediate precision may be obtained from analyses by different analysts on different days using where possible different equipment, materials, standards, etc and using standard procedures usually applied in the laboratory.

Use of a statistical formula without careful thought is undesirable. There needs to be careful attention to the sources of variability that should be considered in getting a realistic design for the collection of data to estimate intermediate precision.

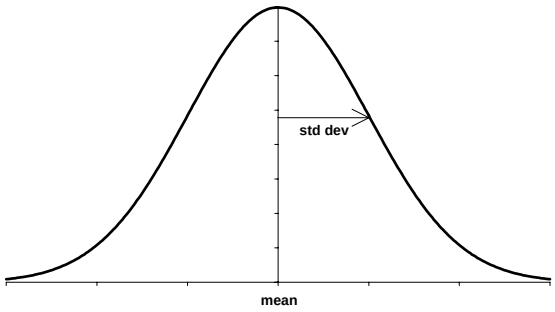
Because R and intermediate precision include r in addition to other significant components of precision, an estimation of one or both of these will be the most useful.

In some circumstances a laboratory may be asked to certify the composition of a batch of material. Here, sampling variation and product uniformity within the batch become important and these variations may be so large as to render analytical variations insignificant.

5.6.2 Assumptions of normality

It is assumed throughout this paper that replicate results form a normal distribution about a mean. There should, however, be coarse checks that this assumption is reasonable. If there are few results, they should be plotted out along a line. If many, a histogram or similar display may be drawn (see Figure 1). Alternatively, the use of normal probability plots can demonstrate normality of a data set.

Figure 1



Indications that the assumption is not valid are:

- several results lie well away from the rest
- a scatter of values that is quite asymmetric, e.g. values may be closely bunched at the high or low end of the scale.

Microbiological replicate results are often found to be bunched at the low end (long tail at the upper end) of the scale. Such results may be transformed by taking $\log_{10}$ values of each test result to give a log distribution which is closer to normal. Normal statistics may then be applied for the calculation of precision.

5.6.3 Replicate analyses in one laboratory

In a good routine testing laboratory, it is normal for at least a selection of analyses to carried out in duplicate or triplicate (n = 2 or 3). This allows only a very inaccurate estimate of standard deviation to be calculated for each sample. However, where a material such as a matrix reference material or a house standard is analysed repeatedly, and especially under varying conditions, the estimation of a standard deviation is more reliable.

The formula is,

$$S^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$$ **Formula 2**

where S^2 = the variance
S = estimated standard deviation
n = number of results
y = an individual test result
$\bar{y}$ = "mean" of individual results
$= \frac{\sum y}{n}$
$y - \bar{y}$ = the deviation of each result from the mean

The calculation of the standard deviation using formula 2 is illustrated with the following examples.

Example 1
The determination of nickel in an alloy
The sample gave results of 4.26, 4.18, 4.23 and 4.27% by mass.

Table 1

	Results	Deviation from average	Deviation squared
	4.26	0.025	0.000625
	4.18	0.055	0.003025
	4.23	0.005	0.000025
	4.27	0.035	0.001225
Sum	16.94	0.12	0.0049

Sample size (n) = 4
Mean ($\bar{y}$) = 4.235

$$S^2 = \frac{0.0049}{3}$$

S = 0.0404

More replicates will give a more reliable estimate of precision.

Example 2
The determination of coliform bacteria in ground beef
The sample gave the following CFU per gram results:
3.0×10^5, 2.7×10^5, 3.5×10^5, 1.0×10^6, 3.3×10^5, 2.5×10^5, 3.1×10^5, 3.3×10^5

Table 2

Results	Log_{10} Results	Log_{10} difference	Log_{10} difference^2
300000	5.477	-0.070	0.00493
270000	5.431	-0.116	0.01346
350000	5.544	-0.003	0.00001
1000000	6.000	0.453	0.20488
330000	5.519	-0.029	0.00083
250000	5.398	-0.149	0.02233
310000	5.491	-0.056	0.00314
330000	5.519	-0.029	0.00083
Log_{10} mean	5.547	Sum Log Diff^2	0.25041
	Log Std	Deviation	0.18914
	Log Conf	Limits 95% +/-	0.37827
	Log	Upper Limit	5.926
	Log	Lower Limit	5.169
		Upper Limit	**842624**
		Lower Limit	**147600**
		Mean	**352663**

Thus the mean is 3.5×10^5 and the 95% confidence interval is from 1.5×10^5 to 8.4×10^5

5.6.4 **Alternative data collection**
To avoid the need to analyse one sample many times for improving standard deviation estimates, data from replicate (e.g. duplicate) analyses may be used. Replicates may include those from samples with (slightly) different results or those from a same sample analysed in different laboratories. Such routine replicates, which encompass all variations, can then be used to calculate much better estimates of precision than where only a few replicates on one sample are available.

An example model for data collection for intermediate precision, which incorporates many of the components of uncertainty for a test method is presented in Table 3.

The sample results are paired up to form 15 sets of duplicates (30 test results).

Table 3

LABORATORY 1	ANALYST 1	ANALYST 2
Day 1	Sample 1	Sample 4
Day 2	Sample 2	Sample 3
Day 3	Sample 3	Sample 8
Day 4	Sample 4	Sample 2
Day 5	Sample 8	Sample 1
LABORATORY 2	ANALYST 3	ANALYST 4
Day 3	Sample 3	Sample 5
Day 7	Sample 5	Sample 3
Day 8	Sample 6	Sample 1
Day 9	Sample 1	Sample 2
Day 10	Sample 2	Sample 6
LABORATORY 3	ANALYST 5	ANALYST 6
Day 5	Sample 1	Sample 8
Day 6	Sample 5	Sample 7
Day 9	Sample 7	Sample 1
Day 10	Sample 8	Sample 9
Day 11	Sample 9	Sample 5

In some analyses there is a relationship between standard deviation and the level of the determinant. Standard deviations should, therefore, be calculated at various test result levels to cover the full range of results expected in real samples. A relationship between standard deviations and test results may be checked, by plotting these. Where a proportional relationship is found, the standard deviation for the particular level may be quoted or used in subsequent calculations. Failure to do this may lead to a nonsensical confidence interval with a negative lower bound for low levels and for higher test result levels, confidence intervals may be too small.

Data collections should continue at regular intervals as standard deviations vary over the years.

5.6.5 Calculation of standard deviations from a series of replicate results

From a series of estimates of standard deviation using formula 2 on replicate results of similar samples, an estimate of overall method standard deviation may be calculated using the formula,

$$S_R^2 = \frac{\sum_{i=1}^{p}(n_i - 1)S_i^2}{\left[\sum_{i=1}^{p} n_i\right] - p} \qquad \textbf{Formula 3}$$

where S_R = overall standard deviation for the method

S_I = estimated standard deviation for replicates of the i th sample using formula 2

n_i = number of replicates for the i th sample

$\sum_{i=1}^{p} n_i$ = total number of test results

p = number of samples

$\left[\sum_{i=1}^{p} n_1\right] - p$ = the degrees of freedom

See Text by Davies, O L; Section 3.341.

Replicates should be chosen to cover the maximum possible number of components of variation (see Table 3).

For a set of duplicates, n always equals 2 for all i samples, and Formula 2 simplifies to,

$$S_R^2 = \frac{1}{2p}\sum_{i=1}^{p}\left|y_{i1} - y_{i2}\right|^2 \qquad \textbf{Formula 4}$$

Where $\left|y_{i1} - y_{i2}\right|$ is the difference between each duplicate.

The maximum number of degrees of freedom to justify these calculations should be sought and triplicates provide more degrees of freedom than duplicates per sample set. One should aim for at least 10 degrees of freedom.

Formula 3 or 4 should be applied at several levels of determinant throughout the expected range of results to provide several S_R values.

Calculation of standard deviation using Formula 4 is illustrated by the following examples:

Example 3
Duplicates for free sulphur dioxide in wine

Table 4

Sample	Result 1	Result 2	$y_{i1} - y_{i2}$	$\|y_{i1} - y_{i2}\|^2$
1	42	45	-3	9
2	68	63	5	25
3	50	51	-1	1
4	38	41	-3	9
8	43	39	4	16
3	48	47	1	1
5	29	25	4	16
6	60	57	3	9
1	44	46	-2	4
2	66	67	-1	1
1	47	46	1	1
5	27	28	-1	1
7	18	20	-2	4
8	44	43	1	1
9	71	62	9	81

$$\sum \left| y_{i1} - y_{i2} \right|^2 = 179$$

Thus: p = 15

S_R^2 = 5.97 and

S_R = 2.44

Example 4
Duplicates for Aerobic Plate Count at 35^{O}C on ready-to-eat meals:

Table 5

Duplicates		Logs of duplicates		Difference	Difference2
1300000	1900000	6.114	6.279	-0.165	0.0272
16000000	14000000	7.204	7.146	0.058	0.0034
4200000	5000000	6.623	6.699	-0.076	0.0057
1700000	1600000	6.230	6.204	0.026	0.0007
13000	9600	4.114	3.982	0.132	0.0173
2800	4100	3.447	3.613	-0.166	0.0274
4700000	3200000	6.672	6.505	0.167	0.0279
1300000	1000000	6.114	6.000	0.114	0.0130
360000	270000	5.556	5.431	0.125	0.0156
2900000	1400000	6.462	6.146	0.316	0.1000
2110000	1600000	6.324	6.204	0.120	0.0144
100000	120000	5.000	5.079	-0.079	0.0063
2700000	2000000	6.431	6.301	0.130	0.0170
730000	1400000	5.863	6.146	-0.283	0.0800
640000	440000	5.806	5.643	0.163	0.0265
2200000	1500000	6.342	6.176	0.166	0.0277
280000	450000	5.447	5.653	-0.206	0.0425
54000	64000	4.732	4.806	-0.074	0.0054
15000	11000	4.176	4.041	0.135	0.0181
81000	100000	4.908	5.000	-0.092	0.0084
25000	17000	4.398	4.230	0.167	0.0281

P = 20

Sum Differences^2	0.5125
Log S_R^2 (divide by 2p)	0.0128
Log S_R	0.1132
Log CL = +/- (95% t = 2)	0.226
Upper Limit (log)	5.828
Overall Mean (log)	5.601
Lower Limit (log)	5.375

Upper Limit **672474**

Mean **399291**

Lower Limit **237084**

5.6.6 Outlying results

Extreme outlying results may be identified by statistical means, for example the Cochran's maximum variance test described in ISO 5725.

Where outliers or stragglers are identified, steps should be taken to explain them by identifying some technical error, computational error, clerical error, wrong sample, etc and to correct them. Action should then be taken to prevent their recurrence.

Result should only be rejected when there is a valid explanation that they are outliers. Otherwise all results should remain in the data to be used.

5.7 List the sources of uncertainty not covered by selected precision data

By studying the design of the project used to obtain the data for the precision calculations and considering the list of all sources of uncertainty, the analyst can decide which components have been covered by the precision data and which have not.

Assuming a well designed precision experiment where replicate samples are carried through the whole procedure including sub-sampling, then some common items, which may not have been included, are:

- Balance uncertainty: Where the laboratory always uses the same balance for weighing say a small quantity of the reference material (used for calibrating the equipment), this may be a significant item not included in the precision estimate. For example the lab may weigh out 2 mg of material on a five-place balance with an uncertainty (on the calibration certificate) of +/- 0.08 mg, to make up a standard. This reference material may be used for a few weeks or months before it is replaced.
- Purity of reference material: The certificate with the reference material may state that it is better than (say) 90% pure. This can be taken to mean that the purity is somewhere (anywhere) between 90 and 100% but not outside these limits.
- Equipment: Where the laboratory always uses only the same item of equipment for the particular tests, the uncertainty (unknown bias) associated with that equipment will not be included in the precision estimate.
- Recoveries and blanks: Where test results are corrected for a fixed recovery result or for a fixed blank value, the uncertainty associated with these values will not be included in the precision estimate. These uncertainties may be as large as the precision for the actual test and will need to be estimated.
- Sample preparation and sub-sampling: If replicates for the precision project are taken after sample preparation and/or sub-sampling, then the uncertainty associated with sample preparation and/or sub-sampling will not be included in the precision estimate.

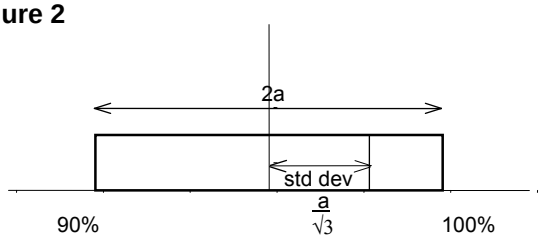
There may be other significant items on the particular list, which also need to be identified and their uncertainties estimated separately.

5.8 Estimate or locate information on the uncertainties of each of these identified additional sources and express each as a standard deviation.

Taking the above items:

- Balance: The uncertainty for the balance will be on the balance calibration certificate. It should be as a standard deviation or may be converted to one by dividing by the coverage factor, which should also be on the certificate if the expanded uncertainty has been quoted.
  The best accuracy (95% confidence limit) of 0.00016 g is divided by the coverage factor of 2 to obtain the standard deviation for this component of 0.00008g or 0.08 mg.
- Reference material: If the certificate on the reference material indicates that it is better than 90% pure and the laboratory always uses this same reference material to calibrate its equipment then the uncertainty allowance would be based on treating this as a rectangular distribution (see figure 2) from 90 to 100%.

Figure 2



The rule for rectangular distributions for calculating the standard deviation is:

- o divide the interval by 2 and then by the square route of 3,
- o i.e. (100-90) / 2 / √3 = 2.89%
  for the mean value of 95% purity.

- Recovery: From a series of spiking trials, the recovery may have been found to be 85% +/- 6% at 95% confidence.

Assuming the coverage factor (often referred to by the symbol k) was 2 (that is the reported confidence limits are about 95% confidence) then the standard deviation for this contribution will be 3% for the 85% recovery result.

Predetermined blank values may be similarly evaluated.

- Homogeneity - Previous special trials to assess the homogeneity of a sample may have been conducted by analysing a component of the sample, which is stable and gives results with very small test method uncertainty values.

  For example, the precision (as a standard deviation) for the selected stable component, in a reference solution, may be 1% of the test result while its precision (as a standard deviation) for a series of sub-samples from the same primary sample may be 5% of the result. In this case, the standard deviation attributable to the sub-sampling would be $\sqrt{(5^2 - 1^2)} = 4.9\%$ of the result (relative standard deviation).

5.9 Identify how each component of uncertainty fits into the formula for calculating the test results.

An example of a formula for calculating the test results is extended from in formula 1 to

$$R = \frac{I_{sample} . C_{ref} . V_{ext} . P . H}{I_{ref} . Rec . m_{sample}} \mu g / g.$$

Where components of uncertainty are not in the equation, for example overall precision P and sample homogeneity H (sub-sampling) these need to be included appropriately. In this case they are both included as multipliers. Other components are already included within the equation.

Once the precision function is included as a multiplier (result x 1 as it does not change the result) any uncertainty components in the equation, which are already included in the precision experiment need to be deleted.

In this case the two "I" components (peak heights) and the "V" (volume) component will be included in precision.

The mass of sample component will be insignificant as (say) a four-place balance was used to weigh 2 grams of sample. However, the uncertainty of the mass of the reference material used to make up the standard solution will be significant (see section 5.7 above). This would come under the C_{ref} component.

The uncertainty associated with the purity of the reference material would also come under the C_{ref} component.

5.10 Estimate the overall uncertainty for the method by combining each source value using the appropriate formulae. The rules for combining uncertainties are as follows. All uncertainty components are expressed as standard deviations.

Rule 1: Where the calculation of the measurand (test result) involves only adding or subtracting quantities, (e.g. y = p + q - r) the combined uncertainty is calculated using Formula 5.

$$u_c = \sqrt{u(p)^2 + u(q)^2 + u(r)^2 .....} \quad \textbf{Formula 5}$$

Rule 2: Where the calculation of the measurand involves multiplying or dividing quantities (e.g. y = p x q / r) the combined uncertainty is calculated using formula 6 to combine the relative uncertainties.

$$uc = y\sqrt{[u(p)/p]^2 + [u(q)/q]^2 + [u(r)/r]^2 .....} \quad \textbf{Formula 6}$$

Where there is a mixture of additions and multiplications in the measurand formula, the combined uncertainties of the components added or subtracted are first calculated (Formula 5) and then the overall combined uncertainty is calculated for the remaining (combined or single) multiplied or divided components using their relative uncertainties (Formula 6).

For combining precision and homogeneity components, use rule 2 (Formula 6).

From the measurand equation above as the example, it is noted that all components are either multiplied or divided and therefore rule 2 (Formula 6) will be used.

The calculation for overall uncertainty is

$$U_R = R\sqrt{\left[\frac{U(mass)}{mass}\right]^2 + \left[\frac{U(ref)}{ref}\right]^2 + \left[\frac{U(Prec)}{Prec}\right]^2 + \left[\frac{U(Homog)}{Homog}\right]^2 + \left[\frac{U(Rec)}{Rec}\right]^2}$$

This includes relative uncertainties for the mass of reference material weighed, the purity of the reference material, the precision, the homogeneity of the sample (ability to sub-sample) and the recovery (from spike

experiments). The uncertainty of the balance used for weighing the sample is usually not significant but this should be checked.

Using some results from above (section 5.8):

Balance Calibration	+/- 0.08 mg
Mass of Reference Material (say)	2.0 mg
Uncertainty of RM Purity	+/- 2.89%
Reference Material Purity	95 %
Precision as SD (say)	+/- 0.1 µg/g
Typical Test Result (say)	2.0 µg/g
Homogeneity (2 ug/g x 4.9%)	+/- 0.098µg/g
Recovery Precision as SD	+/- 3 %
Recovery	85 %

$$U_R = 2\sqrt{\left[\frac{0.08}{2.0}\right]^2 + \left[\frac{2.89}{95}\right]^2 + \left[\frac{0.1}{2}\right]^2 + \left[\frac{0.098)}{2}\right]^2 + \left[\frac{3}{85}\right]^2}$$

$$u_R = 2\sqrt{0.0016 + 0.00093 + 0.0025 + 0.0024 + 0.0013} \text{ ug/g}$$

u_R = 0.18 ug/g

The largest components here are the precision and the sample homogeneity (sub-sampling). If these were the only components incorporated, then the uncertainty result would have been 0.14 ug/g.

Components that are less than about 1/3 of the major components have little effect on the overall uncertainty.

A well-designed precision experiment could have included sub-sampling in the precision component and thus the precision alone would have been a reasonable (but a little small) estimate of uncertainty. For many chemical tests, this will be the case but until the laboratory has gone through the exercise they are not in a position to know that.

5.11 Uncertainty for Microbiological Tests

For microbiological tests, because of the complications of the skewed distribution, the estimate of uncertainty from a well-designed intermediate precision experiment will be regarded as a reasonable estimate of the uncertainty of results.

Where MPN tables are used to obtain microbiological results, the 95% confidence limits quoted in the tables will be regarded as a reasonable estimate of the uncertainty of these results in the mean time.

6 Expanded Uncertainty and Confidence Limits

The “confidence interval” is the range within which the true or best possible result is likely to lie at the stated confidence (e.g. 95%).

The “confidence limits” are the bounds of the confidence interval. The confidence limits may be referred to as “expanded uncertainty”.

If the typical “method” uncertainty as a standard deviation $S_{R,}$ for a single result (or mean of replicates), has been calculated as in the previous sections using Formulae 2 or 3 for microbiological testing and Formulae 5 and/or 6 for chemical testing, then the confidence limits can be determined from Formula 7.

$$\text{Confidence Limits} = \bar{y} \pm \frac{t.S_R}{\sqrt{n}} \qquad \textbf{Formula 7}$$

where:

$\bar{y}$ = mean of replicates (e.g. duplicates) done on this new sample

n = number of replicates done on this new sample

t = “coverage” factor from statistical *t*-tables

This formula applies where the overall uncertainty relates to a normal distribution of test results and the major component of uncertainty is precision, which is the case for chemical and logged microbiological results.

Formula 7 shows that the mean of a duplicate is $\sqrt{2}$ times more precise than an individual test result. ($\sqrt{n}$ for n replicates). However, this only applies to the precision component of the uncertainty and therefore replicate testing cannot be used to reduce uncertainty to negligible proportions.

At this stage the analyst must decide what level of confidence is required, i.e. how likely is it that the confidence interval will contain the true result.

For most purposes a 95% confidence is acceptable. This means that the confidence interval for a particular result would contain the true result for 95 out of 100 tests. The 95% level is the confidence level recommended for use in the ISO 5725 standard, and by the UK Water Research Centre TR 66.

This method of calculating the confidence interval gives a 95% probability of capturing the true mean result, which for this purpose is assumed to have some particular fixed but unknown value. The 95% probability is conditional on this.

Having chosen the confidence level required the coverage factor (*t*) can now be found in statistical table such as Table 6.

Table 6
Probability Points Of The *t*- Distribution
Double–sided Test

	Pr%			
Double Sided	0.5	1	5	10
Degrees of freedom	Coverage Factors (t)			
1	127	63.7	12.7	6.31
2	14.1	9.92	4.30	2.92
3	7.45	5.84	3.18	2.35
4	5.60	4.60	2.78	2.13
5	4.77	4.03	2.57	2.01
6	4.32	3.71	2.45	1.94
7	4.03	3.50	2.36	1.89
8	3.83	3.36	2.31	1.86
9	3.69	3.25	2.26	1.83
10	3.58	3.17	2.23	1.81
11	3.50	3.11	2.20	1.80
12	3.43	3.05	2.18	1.78
13	3.37	3.01	2.16	1.77
14	3.33	2.98	2.14	1.76
15	3.29	2.95	2.13	1.75
16	3.25	2.92	2.12	1.75
17	3.22	2.90	2.11	1.74
18	3.20	2.88	2.10	1.73
19	3.17	2.86	2.09	1.73
20	3.15	2.85	2.09	1.72
21	3.14	2.83	2.08	1.72
22	3.12	2.82	2.07	1.72
23	3.10	2.81	2.07	1.71
24	3.09	2.80	2.06	1.71
25	3.08	2.79	2.06	1.71
26	3.07	2.78	2.06	1.71
27	3.06	2.77	2.05	1.70
28	3.05	2.76	2.05	1.70
29	3.04	2.76	2.05	1.70
30	3.03	2.75	2.04	1.70
40	2.97	2.70	2.02	1.68
60	2.91	2.66	2.00	1.67
120	2.86	2.62	1.98	1.66
Infinity	2.81	2.58	1.96	1.64

Note that Pr % is the probability or risk of error for this statistical method and the degrees of freedom relate to the numbers of replicates for the original calculation of S_R.

Continuing the example from Section 5.10 consider 5% risk (95% confidence).

S_R was 0.18

Degrees of freedom = 30 – 15 = 15
(for say 30 duplicates used to obtain the precision)

Coverage factor = 2.13
(double sided 5% from Table 6)

Thus confidence limits (CL_{95}) which bound the 95% confidence interval in which the true result is expected to lie, for a new sample analysed in duplicate giving say 1.94 and 2.00 ppm of pesticide are found as follows:

$$\bar{y} = \frac{|y_1 + y_2|}{2} = 1.97$$

$$n = 2$$

$$CL_{95} = y \pm \frac{t.S_R}{\sqrt{n}}$$

$$= 1.97 \pm \frac{2.13 \times 0.18}{\sqrt{2}}$$

$$= 1.97 \pm 0.27$$

Thus the 95% confidence interval is from 1.70 to 2.24and the true result is expected to lie between 1.70 and 2.24 at the 95% a confidence level.

This may create a problem for decision makers if the legal limit for pesticide is 2.00 ppm.

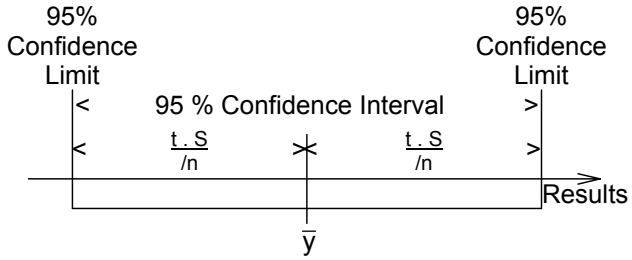
Note from Table 6, that:

- for a 5% risk of error, $t \approx 2$ and
- for 0.5 % risk of error, $t \approx 3$,

where there are a reasonable number of degrees of freedom.

Figure 3

Confidence Interval



The normal distribution curve is about the true or best possible result, which is unknown. However, the true result is expected to lie within the confidence interval for 95% of tests. One cannot view the centre of a confidence interval as the best point estimate of the true result.

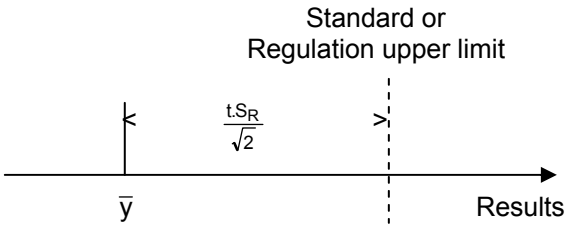
The interval simply defines the range of plausible values (see Bhattacharyya and Johnston; “Statistical Concepts and Methods”; Section 8.3).

7 Comparing a Test Result with a Fixed Value

7.1 General

To approach this situation, we must consider where the observed test result (or mean of a replicate) $\bar{y}$ is in relation to the true result and the fixed value. The diagram for this can be drawn as in figure 4.

Figure 4
Comparing a Result With a Standard



The true result is to be below standard value on 95% of occasions.

For the sample to comply, the true result must be somewhere below the standard / regulation upper limit at a confidence of say 95%. For this to be the case, there should be only a 5% risk that the true result could be above the standard or regulation limit and 95% probability that it may be anywhere below that limit. This means that the single sided statistic (Table 7) should be used to calculate the upper confidence limit for the test result.

We calculate the upper confidence limit thus:

$$\bar{y} + \frac{tS_R}{\sqrt{n}}$$

where *t* is taken from a single sided statistical table (Table 7).

Note from Table 7 that for 95% confidence, *t* is about 1.7 for a reasonable number of degrees of freedom.

From the previous example, pesticide must not exceed 2 ppm. What is the mean of a duplicate below which an analyst may accept the sample as compliant, with 95% confidence?

Degrees of freedom = 15 (say)
Pr = 5% single sided
t = 1.75 single-sided
S_R = 0.18
n = 2

Therefore

2.00 ppm = $\bar{y} + \frac{1.75 \times 0.18}{\sqrt{2}}$ or

2.00 = $\bar{y}$ + 0.22 and

$\bar{y}$ = 1.78 ppm

Thus for the duplicate of 1.94 and 2.00 (mean 1.97), the laboratory could not make a decision about compliance or non-compliance. Only if the mean was 1.78 or less could the laboratory say with 95% confidence that the sample was in compliance with the requirement limit of 2.00 ppm.

Note that if a regulator wanted to be 95% confident that a sample of bread contained greater that 2.00 ppm of pesticide before taking regulatory action, then the single sided statistic is still used but the lower confidence limit is calculated.

2.00 ppm = $\bar{y} - \frac{tS_R}{\sqrt{n}}$ and

$\bar{y}$ = 2.22

At values above 2.22 ppm it would be 95% safe for the regulator to take action.

Table 7
Probability Points Of The *t*-Distribution
Single-sided Test

	Pr%			
	0.5	1	5	10
Degrees of freedom	Coverage factors (t)			
1	63.7	31.8	6.31	3.08
2	9.92	6.96	2.92	1.89
3	5.84	4.54	2.35	1.64
4	4.60	3.75	2.13	1.53
5	4.03	3.36	2.01	1.48
6	3.71	3.14	1.94	1.44
7	3.50	3.00	1.89	1.42
8	3.36	2.90	1.86	1.40
9	3.25	2.82	1.83	1.38
10	3.17	2.76	1.81	1.37
11	3.11	2.72	1.80	1.36
12	3.05	2.68	1.78	1.36
13	3.01	2.65	1.77	1.35
14	2.98	2.62	1.76	1.34
15	2.95	2.60	1.75	1.34
16	2.92	2.58	1.75	1.34
17	2.90	2.57	1.74	1.33
18	2.88	2.55	1.73	1.33
19	2.86	2.54	1.73	1.33
20	2.85	2.53	1.72	1.32
21	2.83	2.52	1.72	1.32
22	2.82	2.51	1.72	1.32
23	2.80	2.50	1.71	1.32
24	2.80	2.49	1.71	1.32
25	2.79	2.48	1.71	1.32
26	2.78	2.48	1.71	1.32
27	2.77	2.47	1.70	1.31
28	2.76	2.47	1.70	1.31
29	2.76	2.46	1.70	1.31
30	2.75	2.46	1.70	1.31
40	2.70	2.42	1.68	1.30
60	2.66	2.39	1.67	1.30
120	2.62	2.36	1.66	1.29
Infinity	2.58	2.33	1.64	1.28

7.2 Results where Compliance or Non-compliance Cannot be Stated with 95% Confidence

In the above example, the test result was 1.97 ± 0.27 at 95% confidence. With this result, one cannot say, with 95% confidence, if the sample complied or non-complied with the legal limit of 2.00 even though the result was less than 2.00.

However, it is possible to state that the sample is likely to comply with the limit but at a lesser percentage confidence. The APLAC TR 004 report allows for this procedure. However, if this result is to be reported as “likely to comply” with the standard, the APLAC report requires that the laboratory must also state the actual percentage confidence it has in this compliance statement.

The procedure for calculating the actual confidence in the compliance statement is based on working backwards through the *t*-distribution tables as follows:

From Formula 7 the *t* value may be calculated because all other information is known

$$\text{Confidence Limits (CL)} = \pm \frac{t\, S_R}{\sqrt{n}} \quad \text{or}$$

$$t = \frac{CL \,.\, \sqrt{n}}{S_R}$$

$$S_R = 0.18$$

Here the confidence limit (CL) is the difference between the result and the legal limit.

$$CL = 2.00 - 1.97 = 0.03 \quad \text{and}$$

$$n = 2$$

Degrees of Freedom (from original standard deviation estimate) were 15.

$$\text{Therefore } t = \frac{0.03 \,.\, \sqrt{2}}{0.18} = 0.24$$

Now from a more extensive single sided *t* -distribution table, a value closest to *t* = 0.24 for 15 degrees of freedom is found in the column headed 40.

This means that the laboratory is only 60 percent confident that the test result indicates compliance with the legal limit.

Decision makers who are not given this information but simply receive the test result of 1.97 ppm may unknowingly be making a very risky decision because of a lack of this vital information.

7.3 Suitability of Test Method for Stating that a Product Test Result is Within Upper and Lower Product Specification Limits

Where a product is being tested to determine if the result is within upper and lower product specification limits, the laboratory must choose a test method with a suitable uncertainty, capable of giving a compliance statement for a reasonable percentage of results which fall within the specification limits.

To do this, the uncertainty of results must be significantly smaller than the product’s

specified allowable range (upper minus lower product limits).

APLAC TC 004 suggests that "for an expanded uncertainty of measurement U and a specified product allowable range 2T (2T = upper limit – lower limit), the ratio U : T is a measure of the ability of the test method to distinguishing compliance from non-compliance."

For example, consider a product limit interval of 2T and the expanded uncertainty for the test result of U. For a compliance statement to be made, the test result would need to be U (single sided) above the lower limit and U (single sided) below the upper limit. If U : T were 1 : 3 then the interval between the lower limit plus U and the upper limit minus U would be 66.7% of the total product allowable range interval 2T. In such cases, 66.7% of possible results between the upper and lower specification limits could be stated with 95% confidence as complying.

The APLAC document therefore concludes that a ratio of 1 : 3 or better is reasonable for a laboratory to use in deciding on the suitability of a test method in these circumstances.

8 Comparing a Test Result with a Natural Population

If a single test result (or the mean of replicates for a single sample) is being compared with previously obtained results that make up a population of "acceptable" results for a product such as fruit juice the analyst must answer the question, "does the result belong inside or outside the acceptable population?" and, for instance, can they conclude that the juice has been watered?

The fruit juice population will have (say) a normal distribution such as figure 1 from which uncertainty as a standard deviation, S_J has been calculated using Formula 2.

The test sample replicates will have a standard deviation S_R calculated using Formulae 5 and / or 6.

When the difference between the normal fruit juice population mean $\bar{x}$ and the sample test result (or mean of replicates) $\bar{y}$ is greater than

half their combined confidence intervals, then there is less than Pr% probability that the sample belongs to the population of fruit juices and the analyst may well say, "the sample is not fruit juice".

The combined standard deviation S_M can be calculated from Formula 8.

$$S_M^2 = S_J^2 + \frac{S_R^2}{n} \qquad \textbf{Formula 8}$$

Therefore for a 5% or less risk of rejecting the sample when it is actually fruit juice,

$$\left|\bar{x} - \bar{y}\right| \geq t.S_M \qquad \textbf{Formula 9}$$

Where *t* equals 1.7 for a single-sided statistic and degrees of freedom are taken as the smallest of those used for the S_J or S_R calculations.

For example, magnesium in fruit juice is typically at a level of about 80 ppm with a standard deviation of 5 ppm. Tap water contains negligible magnesium. The lowest level of magnesium an unknown sample may have, before it is considered with 95% confidence to have been adulterated, may be calculated.

The method for magnesium determination has a standard deviation of (say) 2.5 ppm and duplicate results are determined on each sample.

$$S_M^2 = 5^2 + \frac{2.5^2}{2}$$

$$= 25 + 3.1$$

$$S_M = 5.3$$

Thus for $\bar{x} - \bar{y} \geq tS_M$ or

$80 - \bar{y} \geq 1.7 \times 5.3$ or

(5% risk single sided)

$80 - \bar{y} \geq 9$ then

$\bar{y} \leq 71$

Therefore, a sample with a magnesium result of 71 ppm or less, is not a pure fruit juice because it is outside the plausible values for fruit juice at 95% confidence.

Note that for a sample $\bar{y}$ result just above 71 ppm of magnesium, the sample may be judged to be not inconsistent with fruit juice with 95% confidence. The S_J is the standard deviation of the fruit juice population and must have been derived from samples representing the likely geographical origins and varieties of the fruits in the suspect juice.

9 Comparing Two Test Results

This situation is similar to the previous one except that usually the two results have the same uncertainty and thus,

what was S_J becomes $\frac{S_R}{\sqrt{n}}$

where each sample has n replicates.

Now formula 8 simplifies to

$$S_M = \frac{\sqrt{2}S_R}{\sqrt{n}}$$

or for duplicate test results $S_M = S_R$.

Therefore, for the two test (R) results $\bar{x}$ and $\bar{y}$,

if $\left|\bar{x} - \bar{y}\right| \geq \frac{t\sqrt{2}S_R}{\sqrt{n}}$ (or *t*.S_R for duplicates)

then the two test (or mean of replicate) results will be judged as different because the true result for each is outside the plausible range for the other, at the confidence level (e.g. 95%) chosen for the *t* value.

10 Criterion of Detection

This is the minimum value that a single test result (or mean of replicates) may have for the analyst to say that something is present with 95% confidence. The term is defined and discussed in the UK Water Research Centre Technical Report TR 66.

It can often be assumed that the "blank" results for a method will vary and form a population distribution identical to that of results for a sample with the measurand close to the blank (low level). This will occur whether the blank contains traces of the measurand or not.

Therefore, the population of the blank and the population of a low level sample can be compared using the methods described in the previous section for comparing two test results.

If $\bar{y} - \bar{b} \geq \frac{t\sqrt{2}S_R}{\sqrt{n}}$ where

$\bar{b}$ = mean of replicate blanks and

S_R = uncertainty as a standard deviation for a low level sample (or blank) calculated from Formulae 5 and/or 6

Then the analyst may conclude that some measurand is present in the sample with 95% confidence (the sample result is different from the blank result). On 5% of occasions this conclusion will be wrong.

Criterion of Detection is when

$$\bar{y} - \bar{b} = \frac{t\sqrt{2}S_R}{\sqrt{n}} \qquad \textbf{Formula 10}$$

Because we are interested only in positive results $\left(\bar{y} > \bar{b}\right)$ (use single sided statistic) and 95% confidence is adequate, and duplicates for samples and blanks are run,

Criterion of Detection is when

$$\bar{y} - \bar{b} = \frac{1.7 \times \sqrt{2} \times S_R}{\sqrt{2}} \qquad \text{or}$$

Criteria for Detection = 1.7 S_R
(for means of duplicates)

Criteria for detection = 2.4 S_R
(for a single test result).

For some analyses, a paired blank or average blank (base line of instrument chart joining several blank responses) is subtracted during the calculation of each result, $\bar{y}$. Here, variation of the blank is already incorporated into the S_R calculation.

In these cases,

$$S_M = \frac{S_R}{\sqrt{2}}$$

and the criterion of detection is when

$$\bar{y} = \frac{1.7S_R}{\sqrt{2}}$$

= 1.2 S_R for duplicates

In this case, when a mean of a duplicate $\bar{y}$ is greater than or equal to 1.2 S_R then there is 95% confidence that at least some of the measurand is present in the sample.

For example, a water sample is analysed in duplicate for lead and a very small result for the method is obtained.

Y_1	=	0.01
Y_2	=	0.02
S_R	=	0.007
$\bar{b}$	=	0.003
Now $\bar{y}$	=	0.015
$\bar{y}$ - $\bar{b}$	=	0.0120 and

$$\frac{t\sqrt{2}S_R}{\sqrt{2}} = 1.7 \times 0.007$$

$$= .0119$$

Therefore the water contains lead (with better than 95% "confidence").

Note that the 95% confidence limits (double sided) for this result are

$$\bar{y} - \bar{b} = 0.012 \pm \frac{t\sqrt{2} \times 0.007}{\sqrt{2}}$$

$$= 0.012 \pm 0.014 \quad !!!$$

(For 95% confidence limits $t \approx 2$ (double sided)).

11 Limit of Detection

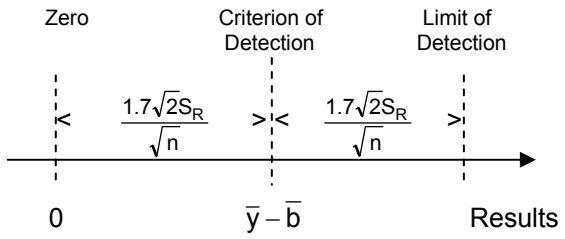
The UK Water Research Centre Technical Report discusses Limit of Detection.

We have just determined the minimum result $\bar{y} - \bar{b}$ for which the analyst will conclude with 95% confidence that something has been found.

However, if the result is only just below this value, the analyst cannot say that nothing has been found. Looking at Figure 4, for a test result (or mean of n replicates) which is just below the criterion of detection, the true sample result may lie anywhere within the confidence interval for 90% of results (or replicate means). Therefore all that can be said with 95% confidence is that the true result is less than the upper confidence limit *viz*. twice the criterion of detection. The true result is just as likely to be zero as any other value in this range (90% between the limits of detection plus 5% below zero).

Figure 5

Limit of Detection



For a test result (or mean of replicates) exactly on the criterion of detection, the true result is below the "Limit of Detection" on 95% of occasions and the true result is above zero on 95% of occasions.

The Limit of Detection is defined as this upper confidence limit for a result that is exactly on the criteria of detection. It is used when reporting all "less than" results.

Limit of Detection = $\frac{2t\sqrt{2}S_R}{\sqrt{n}}$ **Formula 11**

or for duplicate results at 95% confidence using the single sided statistic,

Limit of Detection = $\frac{2 \times 1.7 \times \sqrt{2} \times S_R}{\sqrt{2}}$

Limit of Detection = 3.4 S_R
(for duplicate test results) or

Limit of Detection ≈ 4.8 S_R
(for a single test result).

Now when a result $\bar{y} - \bar{b}$ is in the region below the criterion of detection $\frac{t\sqrt{2}S_R}{\sqrt{n}}$ then it should be reported as "Less Than $\frac{2t\sqrt{2}S_R}{\sqrt{n}}$".

For example, if the water sample result for lead had been

$Y_1 = 0.010$
$Y_2 = 0.015$
$S_R = 0.007$
$\bar{b} = 0.003$

then $\bar{y} = 0.0125$ and
$\bar{y} - \bar{b} = 0.0095$ and
$\frac{t\sqrt{2}S_R}{\sqrt{n}} = 0.0119$

This result must be reported as "Less than 0.024" (3.4 x 0.007).

12 Limit of Quantification

It has already been shown that for a result close to the criterion of detection, the lower confidence limit is zero (or negative). Obviously it would be nonsense to report, "The sample contained 0.01 ppm ± 0.01 ppm of lead".

The UK Water Research Centre Report suggests that a confidence interval, which is about 20% of the result being reported, may be acceptable for such trace constituents.

That means that a result would be for example 10 ± 1 (confidence interval is 2).

This may be used as a guide to determine a "limit of quantification", or the lowest result for which a numerical value should be reported.

Thus $\bar{y} - \bar{b} = \frac{10.t.S_R}{\sqrt{2}}$

For the usual double sided 95% confidence limits ($t \approx 2$) and duplicate sample and blank results n = 2,

$$\bar{y} - \bar{b} = \frac{10 \times 2 \times S_R}{\sqrt{2}} \quad \text{or}$$

the limit of quantification for means of duplicates is = $14\ S_R$ or

for a single result less blank, (n = 1),

$$y - b = 10 \times 2 \times S_R$$

Limit of Quantification = $20\ S_R$
(ratio 10 : 1)

It may well be that a smaller ratio (result : uncertainty) is acceptable for some circumstances. For example, 10 ± 5 may be an acceptable result depending on the risks.

In this case, for a single result less blank (95% confidence),

$$y - b = 2 \times 2 \times S_R \quad \text{or}$$

for a single test result the

Limit of Quantification = $4\ S_R$.
(ratio 2 : 1)

For the above lead in water example

S_R = 0.007 and

$4\ S_R$ = 0.028 ppm

Therefore the method can be recommended for reporting numerical results only if they are above 0.028 ppm. They will have confidence limits of $\pm$ 0.014 ppm at 95% confidence (double sided).

For results of y – b below this (and above the criterion of detection), a statement could be used such as,
"Lead was detected but at a level below 0.028 ppm".

Where the main contribution to uncertainty is the precision component, the level of quantification may be lowered by analysing more replicates and reporting the mean of these.

Example –
For quadruplicates where S_R = 0.007,

Limit of Quantification (ratio 2:1) is

$$= \frac{2 \times 2 \times S_R}{\sqrt{4}}$$

$= 2.S_R$ and

Limit of Quantification = 0.014 ppm

Therefore a numerical result of 0.014 ppm could be reported.

However, as the uncertainty arising from the precision component is thus reduced, other components become more significant. These other components are not reduced by analysing more replicates.

With strict application of this procedure using the ratio of 10:1, valuable information may not be reported. In some cases, a report may be of value using ratios as low as 2:1 as in the above example.

13 Acknowledgement

This technical guide is based on Telarc Technical Guide No 5 prepared for Telarc by Robertson and Maindonald of the Department of Scientific and Industrial Research and on the AOAC paper by Robertson and Chan.

14 References

NZS / ISO / IEC 17025, *General Requirements for the Competence of Testing and Calibration Laboratories*, 1999

Eurachem / CITAC Guide, *Quantifying Uncertainty in Analytical Measurement*, Second Edition, 2000

ISO / VIM, *International Vocabulary of Basic and General Terms in Metrology*, 1993

ISO 5725, *Accuracy (Trueness and Precision) of Measurement Methods and Results*, 1994

ISO, *Guide to the Expression of Uncertainty in Measurement* (GUM), 1995

APLAC TC 005, *Interpretation and Guidance on the Estimation of Uncertainty of Measurement in Testing*, 2003

APLAC TC 004, *Method of Stating Test Results and Compliance With Specification*, 2001

BS 5763, *Methods for Microbiological Examination of Food and Anumal Feeding Stuffs, Part 5 – Enumeration of Micro-organisms – Colony Count at* 30^0C, 1981

ISO 7218, *Microbiology of Food and Animal Feeding Stuffs – General Rules for Microbiological Examinations*, 1966

ISO TR 13843, *Water Quality – Guidance on Validation of Microbiological Methods*, 2000

Davies, O.L. (ed); 1957; *Statistical Methods in Research and Production*, Third Edition; Oliver and Boyd, London

International Standard ISO 5725; 1981; *Precision of test methods – Determination of repeatability and reproducibility by inter-laboratory tests*, First Standardisation, Switzerland.

Cheesman, R.V. and Wilson, A.L. (revised by M.J. Gardner); 1989; *Manual on Analytical Quality Control for the Water Industry*, Technical Report TR 66; UK Water Research Centre, England.

Bhattacharyya, G.K. and Johnston, R.A.; 1977; *Statistical Concepts and Methods*, John Wiley & Sons, New York.

Robertson, J.M. and Maindonald, J; *Precision and Limits of Detection for Analytical Methods,* Telarc technical guide Number 5, February 1989

Robertson, Max and Chan, Terence; *APLAC Interpretation and Guidance on the Estimation of Uncertainty of Measurement in Testing*; Journal of AOAC International, Volume 86, No 5, 2003