Technical note 17—December 2006

Guidelines for the validation and verification of chemical test methods

Distribution: A copy of this Technical Note can be found on the NATA website at www.nata.asn.au in the Technical Publications section under *NATA Technical Notes* or by contacting NATA Communications on (02)9736 8222

Amendment History: **Issued** August 2004, **Reviewed**: December 2006

1. Introduction

A test method must be shown to be fit for purpose before use by a laboratory to ensure clients will be able to use the results generated by application of the method with confidence. **Method Validation** provides objective evidence that the particular requirements for a specific intended use are fulfilled.

For these reasons, method validation is an essential requirement of laboratory accreditation to ISO/IEC 17025. Accordingly, laboratories accredited to this Standard must demonstrate the validity of all methods used, by validating all in-house and modified standard methods and verifying standard methods. The ISO/IEC Standard recognises that validation is always a balance between costs, risks and technical possibilities (Note 3 to 5.4.5.3). The extent of validation required will depend on the status of the method under consideration and the needs relating to its intended application. A newly developed in-house method will normally require rigorous validation whereas minor modifications to a procedure previously validated in the laboratory may require only a comparative bias check before implementation for its intended scope.

If a laboratory wishes to apply a standard method that has been extensively validated via collaborative studies, eg. ASTM, or Australian Standard methods, consideration should be given to the extent of Method Verification that is required. **Method Verification** studies are typically less extensive than those required for method validation. Nevertheless the laboratory should demonstrate ability to achieve the published performance characteristics of the Standard method under their own test conditions.

This Technical Note describes the aspects of a method that should be considered when undertaking method validation or method verification, and provides guidance on how they may be investigated and evaluated. It is intended to be applicable to all fields of testing that employ chemical methods of analysis. An IUPAC Technical Report (Thompson et al., 2002) and recent publications by the Laboratory of the Government Chemist, UK, (LGC, 2003) and B. Hibbert (Hibbert, 2004) are acknowledged as key sources for the information and guidance provided herein.

2. Validation parameters

The first step in measurement and method validation is specifying what you intend to measure; both qualitatively describing the entity to be measured and the quantity. A method is then validated against this specification and client requirements.

The second step in validation is to determine certain selected performance characteristics, also called performance parameters. These are described below.

2.1 Selectivity

The selectivity of a method is the accuracy of its measurement in the presence of interferences. Methods that employ highly specific determinative procedures, such as chromatography/mass spectrometry, have the capability to be very selective. However, methods based on colorimetric measurements may be affected by the presence of coloured sample co-extracts or compounds with chemical properties similar to the analyte. While it is impractical to consider every potential interferent, analysts should call on their knowledge and experience to consider likely, worst case scenarios.

The effect of potential interferents may be checked by analysing samples to which known concentrations of the suspected interferents have been added. This should be carried out on samples containing the analyte over the concentration range expected in practice. In principle, methods should be developed to provide a level of selectivity without significant interferences.

2.2 Linearity

For an analytical method for which a measurement model defines a relation between the instrumental response and the concentration, the range of applicability of this model must be established.

For many analytical methods, the instrumental response as a function of concentration is linear within a stated range. This is normally demonstrated by graphical methods.

The following protocols (Thompson et al., 2002; LGC, 2003) are recommended for establishing the validity of the calibration model as part of method validation:

- there should be six or more calibration standards;
- the calibration standards should be evenly spaced over the concentration range of interest and should be independently prepared (i.e. calibration standards should not be prepared by serial dilution of a stock solution);
- the range should encompass 0-150 % or 50-150 % of the concentration likely to be encountered, depending on which of these is the more suitable; and
- the calibration standards should be run at least in duplicate, and preferably triplicate or more, in a random order.

A simple plot of the data will provide a quick indication of the nature of the relationship between response and concentration. Classical least squares regression, usually implemented in a spreadsheet program, is used to establish the equation of the relation between the instrumental response (y) and the concentration (x) which for a linear model is $y = a + bx$. The standard error of the regression ($s_{y/x}$) is a measure of the goodness of fit. The use of the correlation coefficient derived from regression analysis as a test for linearity may be misleading (Mulholland and Hibbert, 1997). The residuals should also be examined for evidence of non-linear behaviour (Miller and Miller, 2000). Graphs of the fitted data and residuals should always be plotted and inspected to confirm linearity and check for outliers.

If the relationship is not linear over the range of investigation it is necessary to either eliminate the cause of non-linearity, or restrict the concentration range covered by the method to ensure linearity. In some cases it may be appropriate to use a non-linear function, but care must be exercised to properly validate the chosen model.

Matrix effects

For some analytical methods, the measured response to a given amount of substance (most often measured as a concentration) will vary with the sample matrix. For example, matrix enhancement is a well-recognised occasional phenomenon in pesticide residue analysis using gas-liquid chromatography.

If no matrix effects are apparent, it is preferable to prepare calibration standards as simple solutions of the analyte. If matrix effects are suspected, they may be investigated by making standard additions of the analyte to a typical sample extract solution. The range of concentrations by standard addition should be the same as that of the matrix-free calibration so that the slopes of both calibration plots can be compared for significant difference. If the slopes are not significantly different, there is no need to compensate for matrix effects. However it must be noted that standard addition does not compensate for additive matrix effects.

2.3 Sensitivity

The sensitivity of a method is the rate of change of the measured response with change in the concentration (or amount) of analyte. The greater the sensitivity (slope/gradient of the calibration graph), the better a method is able to distinguish small changes in analyte concentration. If the sensitivity changes with day-to-day operating conditions the consideration of sensitivity during method validation may be restricted to ensuring a satisfactory, linear response is regularly achievable within the required concentration range. Sensitivity should be checked as part of a laboratory's ongoing quality assurance and quality control procedures.

2.4 Accuracy

Accuracy is a measure of the quality of a result, i.e. how useful the result is to the client. It has two components, precision and trueness.

Precision is a measure of aspects of random error.

The trueness of a method describes how close a test result is to the accepted reference value for the quantity measured. Lack of trueness indicates systematic error. Bias, strictly the term given to systematic error attributable to instrumental results, is commonly used in reference to all systematic errors. Bias is a quantitative expression of trueness. The trueness of a result improves as bias decreases.

Analytical recovery is a bias usually associated with sample preparation, extraction of the analyte from a sample and other analytical procedures prior to determination. The best way to estimate recovery is to analyse a matrix certified reference material (CRM) containing a stated concentration (or amount) of the analyte. Recovery is reported as a fraction of the determined concentration divided by the stated concentration. If suitable CRMs are not available, recovery may be estimated by analyzing a sample before and after the addition of a known amount of analyte (a process commonly referred to as 'spiking'). In this instance the recovery is calculated from the difference between the results obtained before and after spiking as a fraction of the added amount.

It is not possible in single laboratory validation exercises to separate recovery from any other contributions to the overall bias likely to effect test results. The experiments described below will give an estimate of overall bias.

2.4.1 Precision

The precision of a method is a measure of the closeness expected between independent replicate test results conducted under specified conditions. Precision is usually stated in terms of the standard deviation (s), or relative standard deviation (RSD) of replicate results. Two measures of precision, termed **repeatability** and **reproducibility** are commonly quoted. Australian Standard 2850 (1986) provides guidance on this aspect of method validation.

Repeatability refers to tests performed on identical test items under conditions that are as constant as possible, with the tests performed during a short interval of time in one laboratory by one operator using the same equipment. Repeatability is a useful indicator of method performance, but it underestimates the spread of results that can be expected under normal operating conditions over the longer term.

Reproducibility is the precision relating to a series of measurements made under more variable conditions, i.e. the same method on identical test items used by different operators with different equipment in different laboratories at different times. As such, reproducibility is not a component of the method validation performed by a single laboratory, but it is an important consideration when a laboratory seeks to compare its performance using a particular method with that achieved by laboratories participating in inter-laboratory studies. 'Intra-laboratory reproducibility', 'within-laboratory reproducibility' or 'intermediate precision' are terms used to describe the precision relating to reproducibility conditions restricted to a single laboratory.

In order for the stated precision to truly reflect the performance of the method under normal operating conditions, it must be

determined under such conditions. Test materials should be typical of samples normally analysed. Sample preparation should be consistent with normal practice and variations in reagents, test equipment, analysts and instrumentation should be representative of those normally encountered.

Precision may vary with analyte concentration. This should be investigated if analyte concentration is expected to vary by more than 50% of an average value. For some tests, it may be appropriate to determine precision at only one or two concentrations of particular significance to the users of test data, eg. a production QC specification or regulatory limit.

For single-laboratory validation, the best measure of precision is obtained by replicate analyses of independently prepared test portions of a laboratory sample, CRM or RM, under normal longer term operating conditions. Usually this will involve the determination of intra-laboratory reproducibility as described above.

If data are available from precision experiments carried out on different samples, possibly at different times and there is no significant difference between the variances from each data set, the data may be combined to calculate a pooled standard deviation.

2.4.2 Trueness

The bias of a measurement result may be seen as the combination of the bias of the method itself, laboratory bias and the bias attributable to a particular analytical run.

A reference material, containing an analyte of known concentration may be used to estimate the bias of a test result. If bias is not determined in each run, an estimate of the average bias is best achieved by comparing test results, obtained in different runs over several days, with the known value. Reference materials should match the matrices and analytes of the samples to be tested by the method.

Certified Reference Materials (CRMs)

CRMs contain measurands with assigned values, traceable to international standards with stated uncertainties. When CRMs are available to match the matrices and values of laboratory samples, they present the best option for estimating bias. Ideally, several CRMs with appropriate matrices and analyte concentrations should be measured. However, for most test methods, suitable CRMs are not available, and alternatives are necessarily employed to estimate bias.

Certified reference materials are also used to establish the traceability of calibrations.

Reference Materials (RMs)

If CRMs are not available, other reference materials may be used to estimate bias, provided they are matrix matched with the samples to be tested and sufficiently characterised with respect to the analytes of interest. Materials characterised by restricted collaborative testing may be suitable for the purpose. Laboratories may use RMs characterised against CRMs for routine quality control as an acceptable, cost-effective alternative to the regular analysis of CRMs.

Spiked samples

If neither suitable CRMs nor RMs are available, bias may be investigated by the analysis of spiked samples, i.e. samples to which a known concentration of analyte has been added. For some tests, eg. pesticide residue analysis, laboratories may be able to spike samples that have been determined not to contain detectable residues of the analyte(s) of interest. However, for many tests, it will be necessary to spike samples that contain natural concentrations of analyte(s).

In such cases, bias is estimated from the difference between results obtained for analysis of the sample in its spiked and original states. Caution is advised when evaluating bias from the analysis of spiked samples since the recovery may be better for spiked analyte compared to 'native' analyte, or incurred residues/contaminants. For example, whilst spiking drinking water with fluoride would allow a reliable estimate of recovery the same may not be true for spiking a soil with organochlorine pesticides. This is largely due to different extraction efficiencies for 'added' and 'native' analytes. If possible, spiked recovery data should be substantiated by some means; for example, participation in Proficiency Trials involving natural samples or samples with incurred residues/contamination.

In some cases, laboratories will have to rely solely on spiked recovery data to estimate bias. In such instances, it should be noted that while a 100% recovery does not necessarily indicate trueness, a poor recovery definitely indicates bias, albeit a possible underestimate of the total bias.

Reference methods

A reference method with a known bias may be used to investigate the bias of another method. Typical samples covering the range of matrices and analyte concentrations relevant to proposed testing programs are analysed by both methods. The significance of the bias of the test method may be estimated by statistical analysis (a t-test) of the results obtained.

2.5 Limit of detection and limit of quantitation

The limit of detection (LOD) of a method is the smallest amount or concentration of an analyte that can be reliably distinguished from zero. In other words, the LOD is the lowest value measured by a method that is greater than the uncertainty associated with it. (Taylor, 1989)

It is a NATA requirement that trace organic analytes must be positively identified by an appropriate confirmatory technique. In this context, for trace organic analyses, the LOD is the smallest amount or concentration that can be readily distinguished from zero and be positively identified according to predetermined criteria and/or levels of confidence.

The limit of quantitation (LOQ) of a method is often defined as the lowest concentration of analyte that can be determined with an acceptable level of uncertainty. Various conventions have been applied to estimating the LOQ. Perhaps the most common recommendation is to quote the LOQ as 3 times the LOD.

There is no need to estimate the LOD or LOQ for methods that will always be applied to determine analyte concentrations much greater than the LOQ. However, the estimates often have great importance for trace and ultra-trace methods where concentrations of concern are often close to the LOD or LOQ and results reported as 'not detected' may nevertheless have significant impact on risk assessments or regulatory decisions.

The LOD of a method should not be confused with the lowest instrumental response. The use of a signal to noise ratio for an analytical standard introduced to an instrument is a useful indicator of instrument performance but an inappropriate means of estimating the LOD of a method.

In order to estimate the LOD of a method, analyses should be performed on samples, including all steps of the analytical procedure. The LOD may be determined by analysing 7 replicate samples at each of 3 concentrations, the lowest concentration

being reasonably close to zero. A plot of standard deviation vs concentration is then extrapolated to estimate the standard deviation at zero concentration (s_0). The LOD of the method is taken as $3s_0$, which gives 95% confidence that the method would detect an analyte present in a sample at that concentration.

Alternatively, 7 replicate analyses may be performed at a single concentration equal to about twice the LOQ. (The analyst will need to apply informed judgement in selecting the appropriate concentration). In such circumstances, the standard deviation of these replicates can be assumed to approximate s_0, and the LOD may be calculated as described above.

2.6 Range

The working range of a method is defined as the concentration range within which results will have an acceptable level of uncertainty. In terms of the parameters discussed above, this could be taken to equate to the concentration range between the LOQ and the upper limit of the linear calibration. In practice, acceptable uncertainties may be achieved at concentrations greater than this upper limit (beyond the extent of the determined linear range). However, it is more prudent to consider the validated range, i.e. the range between the LOQ and the highest concentration studied during validation.

2.7 Ruggedness

The ruggedness (a measure of robustness) of a method is the degree to which results are unaffected by minor changes from the experimental conditions described in the procedure; for example, small changes in temperature, pH, reagent concentration, flow rates, extraction times, composition of mobile phase. Ruggedness is investigated by measuring the effects on the results of small, planned changes to the method conditions. In some cases, information may be available from studies conducted during in-house method development. Intra-laboratory reproducibility investigations, by their nature, take into account some aspects of a method's ruggedness.

The simplest tests of ruggedness consider only one method variable at a time. Youden and Steiner (1975) describe a Plackett-Burman designed experiment that provides an economical and efficient approach whereby seven variables are evaluated by conducting only eight analyses. Both approaches assume independence of effects.

In practice, an experienced analyst will be able to identify those method parameters with the potential to affect results and introduce controls, eg. specified limits for temperature, time or pH ranges, to guard against such effects.

2.8 Measurement Uncertainty

Measurement Uncertainty (MU) is defined as 'a parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand (ISO, 1993). Knowledge of MU is necessary for the effective comparison of measurements and for comparison of measurements with specification limits. Furthermore, MU is inexorably linked to **traceability**. Test results must be traceable to stated references, usually national or international standards, through an unbroken chain of comparisons, **all having stated uncertainties** (ISO, 1993; Eurachem/CITAC, 2003). The ISO/IEC 17025 Standard requires that laboratories estimate MU for their non-standard analytical methods and, where applicable, report the MU associated with results. Therefore the estimation of MU is an essential requirement of method validation.

Numerous references are available that present different approaches for the estimation of MU. ISO has published guidelines on the estimation of MU (ISO, 1995) and Eurachem/CITAC interpretation on how they may be applied to analytical measurements (Eurochem/CITAC, 2000). These documents have now been supplemented by guidelines and examples from a number of other sources (ILAC, 2002; APLAC, 2003; UKAS, 2000; ISO/TS, 2004; Magnusson et al., 2003) aiming to provide laboratories with more practical examples and simpler approaches which may be used to calculate reasonable estimates of MU. Excellent examples are also available from the website www.measurementuncertainty.org/

Technical Note 33 entitled *Guidelines for estimating and reporting measurement uncertainty of chemical test results* is available on the NATA website (www.nata.asn.au). The site also provides some worked examples as well as links to other informative web sites.

The information gained from other aspects of method validation, as described above, will be sufficient to produce a reasonable estimate of MU. These data can be supplemented with data from regular QC checks once the method is operational and data resulting from participation in relevant Proficiency Trials. Estimates may also be based on, or partly based on, published data and professional judgement. As with all aspects of method validation, estimates of measurement uncertainty should be fit-for-purpose. The required rigour for estimates will vary according to the rationale for testing; the principle being that estimates should be reasonable for the intended purpose. A reasonable estimate of MU may be obtained from consideration of long-term precision (intra-laboratory reproducibility) and bias. In some instances, other significant contributors to MU, eg. purity of standards, which may not be covered by these parameters may need to be included in the estimation.

It is recommended that a test result and its associated MU be expressed in the same units (Eurachem/CITAC, 2000). It is desirable to estimate MU at the values most important to the users of the results produced by the method e.g. critical concentrations such as a QC specification or regulatory limit.

NATA's current policies with respect to MU require laboratories to provide examples of MU estimates using a documented procedure. The procedure may cite literature references and include alternative approaches. Examples of MU estimates that are submitted for review should include:

- A brief description of the method including the formulae used to calculate results
- Consideration of the approach to be used for estimating MU
- Consideration of the possible contributors to MU (a 'fishbone' cause and effect diagram may help)
- Step-wise calculations for estimating each contribution to MU as per the chosen approach
- Argument for disregarding the uncertainty associated with any test parameter originally identified as a potential contributor to uncertainty
- The equation used for combining standard uncertainties
- The calculation of expanded uncertainty
- An example to show how results would be reported
- A reality check; i.e. does the estimate make sense, based on the laboratory's experience or other relevant information?

The example should be descriptive enough to allow an independent reviewer to easily follow the process. Values

calculated by spreadsheets should be accompanied by a description of how they were derived. Generally a spreadsheet alone will not suffice.

Any judgemental decisions based on experience should be briefly justified. Estimates based solely on precision need to be supported by evidence to demonstrate that precision is the only significant contributor to MU; and in particular that bias, and the uncertainty associated with an estimate of bias, are not significant contributors to the uncertainty of results.

3. Verification of previously validated methods

Methods published by organisations such as Standards Australia, ASTM, USEPA, APHA and IP have already been subject to validation by collaborative studies and found to be fit for purpose for the scope defined by the method. Therefore, the rigour of testing required to introduce such a method into a laboratory is less than that required to validate an in-house method. Essentially the laboratory only needs to verify that their operators using their equipment in their laboratory environment can apply the method satisfactorily. Full validation is required if a laboratory has reason to significantly modify a standard method, for example, use a different extraction solvent or use HPLC instead of GLC for determination.

Additional validation should be considered if the validation data for a standard method is not available to the laboratory or the laboratory needs to apply specifications more stringent than those for which the standard method has been validated. Minor modifications to previously validated in-house methods, for example, using the same type of chromatographic column from a different manufacturer, should also be verified.

The key parameters to consider in the verification process will depend on the nature of the method and the range of sample matrices likely to be encountered. The determination of bias and precision are minimum requirements. Ideally the laboratory will be able to demonstrate performance in line with method specifications. If not, judgement should be exercised to determine whether the method can be applied to generate test results fit for purpose.

For trace analyses the laboratory should also confirm that the achievable LOD and LOQ are fit for purpose.

4. Summary

Table 1 (below) summarises the parameters that need consideration when planning method validation and method verification investigations. The table also includes brief notes on how each performance characteristic may be determined and the

Table 1: method validation
Recommended minimum number of analyses required

Characteristics to be evaluated	Procedures to be followed	Number of determinations
Linearity	Analysis of calibration standards	Duplicate measurements for 6 or more standards evenly spaced over expected concentration range of samples
Sensitivity	Analysis of spiked samples or standards prepared in sample extract solution	Initial check for satisfactory gradient for plot of response vs concentration. (More appropriately a QC issue following initial check)
Selectivity	Consideration of potential interferences, analysis of samples spiked with possible interferents. (Method Development may have overcome potential issues)	If required, one-off tests should suffice
Trueness; bias	Analysis of: CRMs Other RMs Sample Spikes Comparison with Standard Methods Results from Collaborative Studies	At least 7 replicates Reference samples should be matrix and concentration matched with samples
Precision; intra-laboratory reproducibility Limit of Detection. Limit of Quantitation	Replicate analysis of samples; if possible selected to contain analytes at concentrations most relevant to users of test results Analysis of samples containing low concentrations of analytes. Note: The determination of LOD and LOQ is normally only required for methods intended to determine analytes at about these concentrations	At least 7 replicates for each matrix At least 7 replicates at each of 3 concentrations including a concentration close to zero (graphical method), or at least 7 replicates at a concentration estimated to be equal to twice the LOQ (statistical method). Separate determinations may be required for different matrices
Working Range	Evaluation of data from bias and possibly LOQ determinations	
Ruggedness	Consider those steps of the method which if varied marginally, would possibly affect results. Investigate if necessary (i) single variable test (ii) multi variable test	Introduce appropriate limits to method parameters likely to impact results if not carefully controlled Test and re-test with small change to one method parameter Plackett-Burman designed experiment. (Ref, Youden and Steiner, 1975)
Measurement Uncertainty	Utilise other validation data, combined with any other complementary data available, eg. results from collaborative studies, proficiency tests, round-robin tests, in-house QC data	Calculate a reasonable, fit-for-purpose estimate of MU. Ensure estimates are aligned with the concentration(s) most relevant to the users of results

recommended minimum number of replicate tests required for each determination.

Not all parameters need to be assessed for all methods. The rigour of validation should be sufficient to ensure that test results produced by a method are technically sound and will satisfy the client's needs. Well-planned method validation studies will be based on a clear understanding of the specific requirements for the method in use. Within this framework, carefully designed experiments will provide information to satisfy more than one of the parameters in Table 1. For example, information on precision and bias could be obtained from replicate analysis of a CRM, and precision data would also be generated via the determination of the LOD.

It is good practice for laboratories to keep comprehensive records of method validation, including the procedures used for validation, the results obtained and a statement as to whether the method is fit for its intended use.

5. References

APLAC (2003) *Interpretation and Guidance on the Estimation of Uncertainty of Measurement in Testing*, APLAC TC 005, APLAC, March 2003, www.ianz.govt.nz/aplac.

AS 2850:1996 *Chemical Analysis—Interlaboratory Test Programs for Determining Precision of Analytical Method(s)—Guide to the Planning and Conduct*.

Eurachem/CITAC (2000) *Quantifying Uncertainty in Analytical Measurement Eurachem/CITAC Guide CG4*, 2nd Edition, 2000, ISBN 0-948926-15-5, www.eurachem.ul.pt.

Eurachem/CITAC (2003) *Traceability in Chemical Measurement* Eurachem/CITAC Guide, 2003, www.eurachem.ul.pt

Hibbert (2004) *Method Validation*, in Encyclopedia of Analytical Science, 2nd Edition, Quality Assurance, Elsevier Ltd. 2004 (in print)

ILAC (2002) *Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of the Standard* ISO/IEC 17025, ILAC G17, 2002, www.ilac.org.

ISO (1993) *International Vocabulary of Basic and General Terms in Metrology*, 2nd Edition, ISO, Geneva, 1993, ISBN 92-67-01075-1.

ISO (1995) *Guide to the Expression of Uncertainty in Measurement*, 1st Edition, ISO, Geneva, 1995, ISBN 92-67-10188-9.

ISO/TS 2004 *Guidance for the Use of Repeatability, Reproduceability and Trueness Estimates in Measurement Uncertainty Estimation* ISO/TS 21748:2004

LGC (2003) *In-House Method Validation—A Guide for Chemical Laboratories*, Laboratory of the Government Chemist (LGC), UK, 2003.

B. Magnusson, T. Naykki, H. Hovind and M. Krysell, *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories*, NORDTEST Report TR537, 2003

J. C. Miller and J. N Miller, *Statistics and Chemometrics for Analytical Chemistry* 4th Edition, Prentice Hall, 2000, ISBN 0-13-022888-5, and/or VAMSTAT II Statistics Training for Valid Analytical Measurements CD-ROM, www.vam.org.uk.

M. I. Mulholland and D. B. Hibbert, *Linearity and the Limitation of Least Squares Calibration*, J. Chromatography A, 76273-82, 1997

NATA (2006) *Guidelines for estimating and reporting measurement uncertainty of chemical test results* Technical Note No. 33, June, 2006

Taylor, J.K., *Quality Assurance of Chemical Measurements*, sixth edition, Lewis Publishers, ISBN 0-87371-097-5, 1989, page 79.

Thompson M., Ellison S.L.R. and Wood R., *Harmonised Guidelines for Single-Laboratory Validation of Methods of Analysis* (IUPAC Technical Report), Pure Appl. Chem., 74(5), 835-855, 2002.

UKAS (2000) *The Expression of Uncertainty in Testing*, UKAS LAB12, 1st Edition, October 2000, www.ukas.com.

W. J. Youden and E. H. Steiner, *Statistical Manual of the Association of Official Analytical Chemists*, AOAC, 1975, ISBN 0-935584-15-3

6. Additional reading

F. M. Garfield, E. Klestan and J. Hirsch, *Quality Assurance Principles for Analytical Laboratories*, 3rd Edition, AOAC International, 2000, ISBN-0-935584-70-6.

Validation of Analytical Methods for Food Control, Report of a Joint FAO/IAEA Expert Consultation, December 1997, FAO Food and Nutrition Paper No. 68, FAO, Rome (1998).

M. Thompson and R. Wood. *Harmonised guidelines for internal quality control in analytical chemistry laboratories*, Pure Appl. Chem. 67 (4), 49-56 (1995).

M. Thompson, S. Ellison, A. Fajgelj, P. Willetts, R. Wood. *Harmonised guidelines for the use of recovery information in analytical measurement*, Pure Appl. Chem. 71 (2), 337-348 (1999).

Validation of Chemical Analytical Methods, NMKL Secretariat, Finland, NMKL Procedure No. 4, (1996).

EURACHEM Guide: T*he fitness for purpose of analytical methods. A Laboratory Guide to method validation and related topics*, LGC, Teddington 1996. Also available from the EURACHEM Secretariat and website.