codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00100 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 8 **CX/MAS 01/9**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Twenty-Third Session
Budapest, Hungary, 26 February - 2 March 2001

IN-HOUSE METHOD VALIDATION

Background

The 22nd session of the Committee on Methods of Analysis and Sampling discussed the possible use of in-house validated methods for Codex purposes and proficiency data in the validation process. The Committee decided to request the Netherlands, together with France and the United States, to prepare a paper on the use of information from the proficiency testing studies for the elaboration of characteristics of in-house validated methods for consideration by the Committee at its next session. The Committee agreed that when the next draft of the IUPAC *Harmonized Guidelines for the In-House Validation of Methods of Analysis* became available, it would consider the text to determine if it would be appropriate to recommend it to the Commission for adoption by reference for Codex purposes (ALINORM 99/23, para.51). The revised draft of the IUPAC Guidelines is attached as Annex I for consideration by the Committee, as decided at the last session.

Since the last session of the CCMAS the following meetings relevant to method validation have been held:

- AOAC/FAO/IUPAC/IAEA Workshop on Principles and Practices of Validation Methods (4-6 November 1999, Budapest, Hungary)
- FAO/IAEA Expert Consultation on "Practical Procedures to Validate Methods Performance of Analysis of Pesticide and Veterinary Drugs Residues, and Trace Organic Contaminants in Food" (8-11 November 1999, Miskolc, Hungary). This meeting prepared "Guidelines for Single Laboratory Validation of Analytical Methods for Trace-Level Concentrations of Organic Chemicals".

The Committee on Pesticide Residues and the Committee on Residues of Veterinary Drugs in Foods were informed of the results of the Workshop and Consultation in general terms, as the final report was not available. These Committees are currently discussing single-laboratory validation concerning methods of analysis for residues of pesticides and veterinary drugs, and work on this area is still at a very early stage.

The CCRVDF agreed that the results of the Consultation and other work underway at the international level could provide the basis for the criteria on methods of analysis for residues of veterinary drugs to be developed by the Committee. It was agreed that proposals concerning the criteria would be prepared for consideration by the next session (ALINORM 01/31, paras. 98-101).

The CCPR was informed of the conclusions of the Consultation and agreed that the concepts developed should form the basis for a new set of criteria for the assessment of analytical methods for Codex purposes. The Committee agreed that a paper describing appropriate performance parameters and criteria for CCPR purposes would be prepared for consideration by the next session (ALINORM 01/24, paras. 152-153).

The attached paper on in-house validation was prepared by the Netherlands in collaboration with other countries, as agreed by the Committee.

THE USE OF IN-HOUSE VALIDATED METHODS FOR CODEX PURPOSES

(Prepared by the Netherlands with the assistance of other countries)

INTRODUCTION

1. To evaluate the possible role of in-house validated methods for CODEX purposes it is essential that CODEX methods can be used to settle conflicts on composition of foodstuffs. Therefore, mutual acceptance of results is a prerequisite. Traditionally this mutual acceptance was obtained by using the same inter-laboratory validated methods. Recently, the *criteria approach* towards methods of analysis is getting more and more adopted. In this approach more than one method can be used, provided they fulfil certain criteria. An important criterion is the availability of interlaboratory precision data.

2. In the field of residue analysis fully validated methods are not always available. These methods develop rapidly because of the on going marketing of new active substances. Governments must be able to enforce limit on these substances as soon as they come on the market and cannot wait until methods might be validated. Because of the wide scope of these methods both with respect to analytes and substrates it cannot be expected that all combinations will be ever validated in the classical interlaboratory way. Possibly characteristic combinations can be studied, as was done *e.g.* in the EU Standards, Measuring and Testing project on pesticide residues in fruit and vegetables. However, those projects tend to be obsolete by the time of publishing. In this case classical selective detectors were used, whereas now GC-MS is the state of the art.

3. Because of the lack of up to date interlaboratory validated methods a strong need exist to use in-house validated methods exists in the field of residue analysis. In recent years much effort has been mad to elaborate suitable protocols for this in-house validation. As a result a harmonized IUPAC protocol is about to be finalized.

4. In general the use of in-house validated methods will be restricted to multiresidue methods. In the case of single analytes of great trade importance it is expected that full validation will be the starting point of a CODEX specification.

5. In-house validation also plays a role in the management of acute food quality accidents. Usually these incidents are not directly related to CODEX specifications. Trade parties will have to negotiate in those cases to accept mutual analytical results.

REQUIREMENTS FOR USE OF IN-HOUSE VALIDATED METHODS FOR CODEX PURPOSES

6. The main issue in use of analytical methods for Codex purposes is the mutual acceptance of laboratory data. One aspect is the comparability of the methods. Another aspect is the proficiency of the laboratory. It is in this field that strong improvements have been made during recent decennia. The guidelines on quality assurance and proficiency testing (IUPAC) and the draft guideline on in-house validation give important safeguards to the reliability of analytical data in addition to collaboratively validated methods.

7. A main topic in the acceptability of results obtained by in-house validated methods is the presence of data giving an interlaboratory reference. Three methods can give such a reference:

a. calibration using reference materials;
b. comparison of results achieved with other methods;
c. systematic participation in proficiency tests.

8. Taking into account this external reference conditions can be set up for cases where in-house validated methods can be used:

a. the application of in-house validated methods is in general restricted to multi-residue methods;
b. no inter-laboratory validated method is appropriate.

The in-house validated methods must fulfil the following criteria:

a. the method is validated according to an internationally recognized protocol (e.g. the IUPAC/ISO/AOAC protocols);
b. the use of the method is embedded in a quality assurance system under accreditation;

c. external reference is given at least by systematic participation in proficiency schemes.

THE ROLE OF PROFICIENCY STUDIES

9. Proficiency testing can play a role in the validation process, where it gives an external reference in the in-house validation procedure. Another possibility is the use in the interlaboratory validation of methods. In instances where many laboratories use the same method in proficiency testing rounds, interlaboratory reproducibility results can be obtained. The UK Ministry of Agriculture, Fishery and Foodstuffs (MAFF) has drawn up a guideline setting criteria for such an approach (ANNEX II). MAFF also published a practical example (ANNEX III).

CONCLUSIONS

10. CCMAS is invited to:

a. accept the principle that in-house validated methods can be fit for CODEX purposes;
b. accept the conditions and criteria for those methods as indicated in paragraph 8;
c. accept the guidelines of the AOAC/FAO/IUPAC/IAEA expert consultation on d. single-laboratory validation and forward them into the CODEX system at step 3 for government comments;
d. communicate the criteria mentioned above to the CCRVDF and the CCPR and to forward the guidelines of the AOAC/FAO/IUPAC/IAEA expert consultation on single-laboratory validation to consider them as guideline for CODEX uses without the need of further detailing to the different levels at the CCPR and CCRVDF;
e. discuss the MAFF guideline on the use proficiency data for the validation of methods and to consider its appropriateness for CODEX purposes.

ANNEX I. Draft IUPAC Harmonized Guidelines on Single-Laboratory Validation of Analytical Methods.

ANNEX II. Procedure to be used for the validation of methods through the use of results from proficiency testing schemes (United Kingdom.)

ANNEX III. MAFF validated method V38: Method for the enumeration of Listeria Monocytogenes in meat and meat products (United Kingdom).

Secretariat Note

The reports of the AOAC/FAO/IUPAC/IAEA *Workshop on Principles and Practices of Validation Methods* and the FAO/IAEA *Expert Consultation on Practical Procedures to Validate Methods Performance of Analysis of Pesticide and Veterinary Drugs Residues, and Trace Organic Contaminants in Food* have not yet been published.

A Summary Report of the Workshop, relevant information concerning the Workshop and Consultation, and the Guidelines developed by the Consultation are available on the IAEA website:

IAEA: http://www.iaea.org/programmes/rifa/trc/pest-qa_val.htm

A summary report of the Workshop is also available on the IUPAC website:

IUPAC: http://www.iupac.org/symposia/conferences/method_validation_4nov99/report.htm

The decision of the Committee at its last session was to consider the revised draft of the *IUPAC Harmonized Guidelines* when it became available in order to decide whether it should be adopted by reference. The Committee is invited to consider the document in Annex I for this purpose.

The above discussion paper does not refer to the IUPAC Guidelines in its conclusions and proposes instead to consider the guidelines prepared by the FAO/IAEA Consultation for adoption for Codex purposes. These guidelines have not yet been published.

ANNEX I

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

ANALYTICAL, APPLIED, CLINICAL, INORGANIC AND PHYSICAL CHEMISTRY DIVISIONS INTERDIVISIONAL WORKING PARTY FOR HARMONIZATION OF QUALITY ASSURANCE SCHEMES FOR ANALYTICAL LABORATORIES*

HARMONISED GUIDELINES FOR SINGLE-LABORATORY VALIDATION OF METHODS OF ANALYSIS

(Technical Report)

Resulting from the Symposium on Harmonisation of Quality Assurance Systems for Analytical Laboratories, Budapest, Hungary, 4-5 November 1999 held under the sponsorship of IUPAC, ISO and AOAC INTERNATIONAL

Prepared for publication by

MICHAEL THOMPSON[1], STEVEN L R ELLISON[2] AND ROGER WOOD[3]

[1] Department of Chemistry, Birkbeck College (University of London), London WC1H 0PP, UK
[2] Laboratory of the Government Chemist, Queens Road, Teddington, Middlesex TW11 0LY, UK
[3] Food Standards Agency, c/o Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA, UK

*Membership of the Working Party during 1997-2000 was as follows:

Chairman: A. Fajgelj, 1997- , (IAEA, Austria); *Members*: T.B. Anglov (Denmark); K. Bergknut (Norway); K.G. Boroviczeny (Germany); Carmen Camara (Spain); K. Camman (Germany); Jyette Molin Christensen (Denmark); S. Coates (AOAC Int., USA); W.P. Cofino (The Netherlands); P. De Bievre (Belgium); T.D. Geary (Australia); T. Gills (USA); A.J. Head (UK); J. Hlavay (Hungary); D.G. Holcombe (UK); P.T. Holland (New Zealand); W. Horwitz (USA); A. Kallner (Sweden); H. Klich (Germany); J. Kristiansen (Denmark); Helen Liddy (Australia); E.A. Maier (Belgium); H. Muntau (Italy); C. Nieto De Castro (Portugal); E. Olsen (Denmark); Nancy Palmer (USA); S.D. Rasberry (USA); M. Thompson (UK); M.J. Vernengo (Argentina); R. Wood (UK).

Harmonised guidelines for single-laboratory validation of methods of analysis (technical report)

Synopsis

Method validation is one of the measures universally recognised as a necessary part of a comprehensive system of quality assurance in analytical chemistry. In the past ISO, IUPAC and AOAC INTERNATIONAL have co-operated to produce agreed protocols or guidelines on the "Design, Conduct and Interpretation of Method Performance Studies"[1] on the "Proficiency Testing of (Chemical) Analytical Laboratories"[2] on "Internal Quality Control in Analytical Chemistry Laboratories"[3] and on "The Use of Recovery Information in Analytical Measurement". [4] The Working Group that produced these protocols/guidelines has now been mandated by IUPAC to prepare guidelines on the Single-laboratory Validation of methods of analysis. These guidelines provide minimum recommendations on procedures that should be employed to ensure adequate validation of analytical methods.

A draft of the guidelines has been discussed at an International Symposium on the Harmonisation of Quality Assurance Systems in Chemical Laboratory, the Proceedings from which have been published by the UK Royal Society of Chemistry.

CONTENTS

1. INTRODUCTION

1.1 BACKGROUND

Reliable analytical methods are required for compliance with national and international regulations in all areas of analysis. It is accordingly internationally recognised that a laboratory must take appropriate measures to ensure that it is capable of providing and does provide data of the required quality. Such measures include:

- using validated methods of analysis;
- using internal quality control procedures;
- participating in proficiency testing schemes; and
- becoming accredited to an International Standard, normally ISO/IEC 17025.

It should be noted that accreditation to ISO/IEC 17025 specifically addresses the establishment of traceability for measurements, as well as requiring a range of other technical and management requirements including all those in the list above.

Method validation is therefore an essential component of the measures that a laboratory should implement to allow it to produce reliable analytical data. Other aspects of the above have been addressed previously by the IUPAC Interdivisional Working Party on Harmonisation of Quality Assurance Schemes for Analytical Laboratories, specifically by preparing Protocols/Guidelines on method performance (collaborative) studies,[1] proficiency testing,[2] and internal quality control.[3]

In some sectors, most notably in the analysis of food, the requirement for methods that have been "fully validated" is prescribed by legislation.[5,6] "Full" validation for an analytical method is usually taken to comprise an examination of the characteristics of the method in an inter-laboratory method performance study (also known as a collaborative study or collaborative trial). Internationally accepted protocols have been established for the "full" validation of a method of analysis by a collaborative trial, most notably the International Harmonised Protocol[1] and the ISO procedure.[7] These protocols/standards require a minimum number of laboratories and test materials to be included in the collaborative trial to validate fully the analytical method. However, it is not always practical or necessary to provide full validation of analytical methods. In such circumstances a "single-laboratory method validation" may be appropriate.

Single-laboratory method validation is appropriate in several circumstances including the following:

- to ensure the viability of the method before the costly exercise of a formal collaborative trial;
- to provide evidence of the reliability of analytical methods if collaborative trial data are not available or where the conduct of a formal collaborative trial is not practicable;
- to ensure that "off-the-shelf" validated methods are being used correctly.

When a method is to be characterised in-house, it is important that the laboratory determines and agrees with its customer exactly which characteristics are to be evaluated. However, in a number of situations these characteristics may be laid down by legislation (e.g. veterinary drug residues in food and pesticides in food sectors). The extent of the evaluation that a laboratory undertakes must meet the requirements of legislation.

Nevertheless in some analytical areas the same analytical method is used by a large number of laboratories to determine stable chemical compounds in defined matrices. It should be appreciated that if a suitable collaboratively studied method can be made available to these laboratories, then the costs of the collaborative trial to validate that method may well be justified. The use of a collaboratively studied method considerably reduces the efforts which a laboratory, before taking a method into routine use, must invest in extensive validation work. A laboratory using a collaboratively studied method, which has been found to be fit for the intended purpose, needs only to demonstrate that it can achieve the performance characteristics stated in the method. Such a verification of the correct use of a method is much less costly than a full single laboratory validation. The total cost to the Analytical Community of validating a specific method through a collaborative

trial and then verifying its performance attributes in the laboratories wishing to use it is frequently less than when many laboratories all independently undertake single laboratory validation of the same method.

1.2 EXISTING PROTOCOLS, STANDARDS AND GUIDES

A number of protocols and guidelines[8-19] on method validation and uncertainty have been prepared, most notably in AOAC INTERNATIONAL, International Conference on Harmonisation (ICH) and Eurachem documents:

- The Statistics manual of the AOAC, which includes guidance on single laboratory study prior to collaborative testing[13]
- The ICH text[15] and methodology,[16] which prescribe minimum validation study requirements for tests used to support drug approval submission.
- The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics (1998)[12]
- Quantifying Uncertainty in Analytical Measurement (2000)[9]

Method validation was also extensively discussed at a Joint FAO/IAEA Expert Consultation, December 1997, on the Validation of Analytical Methods for Food Controls, the Report of which is available[19].

The present 'Guidelines' bring together the essential scientific principles of the above documents to provide information which has been subjected to international acceptance and, more importantly, to point the way forward for best practice in single-laboratory method validation.

2 DEFINITIONS AND TERMINOLOGY

2.1 GENERAL

Terms used in this document respect ISO and IUPAC definitions where available. The following documents contain relevant definitions:

i) IUPAC: Compendium of chemical terminology, 1987

ii) International vocabulary of basic and general terms in metrology. ISO 1993

2.2 DEFINITIONS USED IN THIS GUIDE ONLY:

Relative uncertainty: Uncertainty expressed as a relative standard deviation.

Validated range: That part of the concentration range of an analytical method which has been subjected to validation.

3 METHOD VALIDATION, UNCERTAINTY, AND QUALITY ASSURANCE

Method validation makes use of a set of tests which both test any assumptions on which the analytical method is based and establish and document the performance characteristics of a method, thereby demonstrating whether the method is fit for a particular analytical purpose. Typical performance characteristics of analytical methods are: applicability; selectivity; calibration; trueness; precision; recovery; operating range; limit of quantification; limit of detection; sensitivity; and ruggedness. To these can be added measurement uncertainty and fitness-for-purpose.

Strictly speaking, validation should refer to an 'analytical system' rather than an 'analytical method', the analytical system comprising a defined method protocol, a defined concentration range for the analyte, and a specified type of test material. For the purposes of this document, a reference to 'method validation' will be

taken as referring to an analytical system as a whole. Where the analytical procedure as such is addressed, it will be referred to as 'the protocol'.

In this document method validation is regarded as distinct from ongoing activities such as internal quality control (IQC) or proficiency testing. Method validation is carried out once, or at relatively infrequent intervals during the working lifetime of a method; it tells us what performance we can expect the method to provide in the future. Internal quality control tells us about how the method has performed in the past. IQC is therefore treated as a separate activity in the IUPAC Harmonisation Programme.[3]

In method validation the quantitative characteristics of interest relate to the accuracy of the result likely to be obtained. Therefore it is generally true to say that method validation is tantamount to the task of estimating uncertainty of measurement. Over the years it has become traditional for validation purposes to represent different aspects of method performance by reference to the separate items listed above, and to a considerable extent these guidelines reflect that pattern. However, with an increasing reliance on measurement uncertainty as a key indicator of both fitness for purpose and reliability of results, analytical chemists will increasingly undertake measurement validation to support uncertainty estimation, and some practitioners will want to do so immediately. Accordingly, measurement uncertainty is treated briefly in Appendix A as a performance characteristic of an analytical method, while Appendix B provides additional guidance on some procedures not otherwise covered.

4 BASIC PRINCIPLES OF METHOD VALIDATION

4.1 SPECIFICATION AND SCOPE OF VALIDATION

Validation applies to a defined protocol, for the determination of a specified analyte and range of concentrations in a particular type of test material, used for a specified purpose. In general, validation should check that the method performs adequately for the purpose throughout the range of analyte concentrations and test materials to which it is applied. It follows that these features, together with a statement of any fitness-for-purpose criteria, should be completely specified before any validation takes place.

4.2 TESTING ASSUMPTIONS

In addition to the provision of performance figures which indicate fitness for purpose and have come to dominate the practical use of validation data, validation studies act as an objective test of any assumptions on which an analytical method is based. For example, if a result is to be calculated from a simple straight line calibration function, it is implicitly assumed that the analysis is free from significant bias, that the response is proportional to analyte concentration, and that the dispersion of random errors is constant throughout the range of interest. In most circumstances, such assumptions are made on the basis of experience accumulated during method development or over the longer term, and are consequently reasonably reliable. Nonetheless, good measurement science relies on *tested* hypotheses. This is the reason that so many validation studies are based on statistical hypothesis testing; the aim is to provide a basic check that the reasonable assumptions made about the principles of the method are not seriously flawed.

There is an important practical implication of this apparently abstruse note. It is easier to check for gross departure from a reliable assumption than to 'prove' that a particular assumption is correct. Thus, where there is long practice of the successful use of a particular analytical technique (such as gas chromatographic analysis, or acid digestion methods) across a range of analytes and matrices, validation checks justifiably take the form of relatively light precautionary tests. Conversely, where experience is slight, the validation study needs to provide strong evidence that the assumptions made are appropriate in the particular cases under study, and it will generally be necessary to study the full range of circumstances in detail. It follows that the extent of validation studies required in a given instance will depend, in part, on the accumulated experience of the analytical technique used.

In the following discussion, it will be taken for granted that the laboratory is well practised in the technique of interest, and that the purpose of any significance tests is to check that there is no strong evidence to discount the assumptions on which the particular protocol relies. The reader should bear in mind that more stringent checks may be necessary for unfamiliar or less established measurement techniques.

4.3 SOURCES OF ERROR IN ANALYSIS

Errors in analytical measurements arise from different sources* and at different levels of organisation. One useful way of representing these sources (for a specific concentration of analyte) is as follows[+24]:

- random error of measurement (repeatability);
- run bias ;
- laboratory bias;
- method bias;
- matrix variation effect.

Though these different sources may not necessarily be independent, this list provides a useful way of checking the extent to which a given validation study addresses the sources of error.

The repeatability (within-run) term includes contributions from any part of the procedure that varies within a run, including contributions from the familiar gravimetric and volumetric errors, heterogeneity of the test material, and variation in the chemical treatment stages of the analysis, and is easily seen in the dispersion of replicated analyses. The run effect accounts for additional day-to-day variations in the analytical system, such as changes of analyst, batches of reagents, recalibration of instruments, and the laboratory environment (*e.g.*, temperature changes). In single-laboratory validation, the run effect is typically estimated by conducting a designed experiment with replicated analysis of an appropriate material in a number of separate runs. Between-laboratory variation arises from factors such as variations in calibration standards, differences between local interpretations of a protocol, changes in equipment or reagent source or environmental factors, such as differences in average climatic conditions. Between-laboratory variation is clearly seen as a reality in the results of collaborative trials (method performance studies) and proficiency tests, and between-method variation can sometimes be discerned in the results of the latter.

Generally, the repeatability, run effect and laboratory effect are of comparable magnitude, so none can safely be ignored in validation. In the past there has been a tendency for aspects to be neglected, particularly when estimating and reporting uncertainty information. This results in uncertainty intervals that are too tight. For example, the collaborative trial as normally conducted does not give the complete picture because contributions to uncertainty from method bias and matrix variation are not estimated in collaborative trials and have to addressed separately (usually by prior single-laboratory study). In single-laboratory validation there is the particular danger that laboratory bias also may be overlooked, and that item is usually the largest single contributor to uncertainty from the above list. Therefore specific attention must be paid to laboratory bias in single-laboratory validation.

In addition to the above-mentioned problems, the validation of a method is limited to the scope of its application, that is, the method as applied to a particular class of test material. If there is a substantial variation of matrix types within the defined class, there will be an additional source of variation due to within-class matrix effects. Of course, if the method is subsequently used for materials outside the defined class (that is, outside the scope of the validation), the analytical system is cannot be considered validated: an extra error of unknown magnitude is introduced into the measurement process.

It is also important for analysts to take account of the way in which method performance varies as a function of the concentration of the analyte. In most instances the dispersion of results increases absolutely with concentration and recovery may differ substantially at high and low concentrations. The measurement

* Sampling uncertainty in the strict sense of uncertainty due to the preparation of the laboratory sample from the bulk target is excluded from consideration in this document. Uncertainty associated with taking a test portion from the laboratory sample is an inseparable part of measurement uncertainty and is automatically included at various levels of the following analysis.

+ Many alternative groupings or 'partitions of error' are possible and may be useful in studying particular sources of error in more detail or across a different range of situations. For example, the statistical model of ISO 5725 generally combines laboratory and run effects, while the uncertainty estimation procedure in the ISO GUM is well suited to assessing the effects of each separate and measurable influence on the result.

uncertainty associated with the results is therefore often dependent on both these effects and on other concentration-dependent factors. Fortunately, it is often reasonable to assume a simple relationship between performance and analyte concentration; most commonly that errors are proportional to analyte concentration.* However, where the performance of the method is of interest at substantially different concentrations, it is important to check the assumed relationship between performance and analyte concentration. This is typically done by checking performance at extremes of the likely range, or at a few selected levels. Linearity checks also provide information of the same kind.

4.4 METHOD AND LABORATORY EFFECTS

It is critically important in single-laboratory method validation to take account of method bias and laboratory bias. There are a few laboratories with special facilities where these biases can be regarded as negligible, but that circumstance is wholly exceptional. (However, that if there is only one laboratory carrying out a particular analysis, then method bias and laboratory bias take on a different perspective). Normally, method and laboratory effects have to be included in the uncertainty budget, but often they are more difficult to address than repeatability error and the run effect. In general, to assess the respective uncertainties it is necessary to use information gathered independently of the laboratory. The most generally useful sources of such information are (i) statistics from collaborative trials (not available in many situations of single-laboratory method validation), (ii) statistics from proficiency tests and (iii) results from the analysis of certified reference materials.

Collaborative trials directly estimate the variance of between-laboratory biases. While there may be theoretical shortcomings in the design of such trials, these variance estimates are appropriate for many practical purposes. Consequently it is always instructive to test single-laboratory validation by comparing the estimates of uncertainty with reproducibility estimates from collaborative trials. If the single-laboratory result is substantially the smaller, it is likely that important sources of uncertainty have been neglected. (Alternatively, it may be that a particular laboratory in fact works to a smaller uncertainty than found in collaborative trials: such a laboratory would have to take special measures to justify such a claim.) If no collaborative trial has been carried out on the particular method/test material combination, an estimate of the reproducibility standard deviation s_H at an analyte concentration c above about 120 ppb can usually be obtained from the Horwitz function, $s_H = 0.02c^{0.8495}$, with both variables expressed as mass fractions. (The Horwitz estimate is normally within a factor of about two of observed collaborative study results). It has been observed that the Horwitz function is incorrect at concentrations lower than about 120 ppb, and a modified function is more appropriate.[21, 25] All of this information may be carried into the single-laboratory area with minimum change.

Statistics from proficiency tests are particularly interesting because they provide information in general about the magnitude of laboratory and method biases combined and, for the participant, information about total error on specific occasions. Statistics such as the robust standard deviation of the participants results for an analyte in a round of the test can in principle be used in a way similar to reproducibility standard deviations from collaborative trials, *i.e.*, to obtain a benchmark for overall uncertainty for comparison with individual estimates from single-laboratory validation. In practice, statistics from proficiency tests may be more difficult to access, because they are not systematically tabulated and published like collaborative trials, but only made available to participants. Of course, if such statistics are to be used they must refer to the appropriate matrix and concentration of the analyte. Individual participants in proficiency testing schemes can also gauge the validity of their estimated uncertainty by comparing their reported results with the assigned values of successive rounds[26]. This, however, is an ongoing activity and therefore not strictly within the purview of single-laboratory validation (which is a one-off event).

If an appropriate certified reference material is available, a single-laboratory test allows a laboratory to assess laboratory bias and method bias in combination, by analysing the CRM a number of times. The estimate of the combined bias is the difference between the mean result and the certified value.

---

* This may not be applicable at concentrations less than 10 times the detection limit.

Appropriate certified reference materials are not always available, so other materials may perforce have to be used. Materials left over from proficiency tests sometimes serve this purpose and, although the assigned values of the materials may have questionable uncertainties, their use certainly provides a check on overall bias. Specifically, proficiency test assigned values are generally chosen to provide a minimally biased estimate, so a test for significant bias against such a material is a sensible practice. A further alternative is to use spiking and recovery information[4] to provide estimates of these biases, although there may be unmeasurable sources of uncertainty associated with these techniques.

Currently the least recognised effect in validation is that due to matrix variation within the defined class of test material. The theoretical requirement for the estimation of this uncertainty component is for a representative collection of test materials to be analysed in a single run, their individual biases estimated, and the variance of these biases calculated. (Analysis in a single run means that higher level biases have no effect on the variance. If there is a wide concentration range involved, then allowance for the change in bias with concentration must be made.) If the representative materials are certified reference materials, the biases can be estimated directly as the differences between the results and the reference values, and the whole procedure is straightforward. In the more likely event that insufficient number of certified reference materials are available, recovery tests with a range of typical test materials may be resorted to, with due caution. Currently there is very little quantitative information about the magnitude of uncertainties from this source, although in some instances they are suspected of being large.

5 Conduct of Validation Studies

The detailed design and execution of method validation studies is covered extensively elsewhere and will not be repeated here. However, the main principles are pertinent and are considered below:

It is essential that validation studies are representative. That is, studies should, as far as possible, be conducted to provide a realistic survey of the number and range of effects operating during normal use of the method, as well as to cover the concentration ranges and sample types within the scope of the method. Where a factor (such as ambient temperature) has varied representatively at random during the course of a precision experiment, for example, the effects of that factor appear directly in the observed variance and need no additional study unless further method optimisation is desirable.

In the context of method validation, “representative variation” means that the factor must take a distribution of values appropriate to the anticipated range of the parameter in question. For continuous measurable parameters, this may be a permitted range, stated uncertainty or expected range; for discontinuous factors, or factors with unpredictable effects such as sample matrix, a representative range corresponds to the variety of types or “factor levels” permitted or encountered in normal use of the method. Ideally, representativeness extends not only to the range of values, but to their distribution. Unfortunately, it is often uneconomic to arrange for full variation of many factors at many levels. For most practical purposes, however, tests based on extremes of the expected range, or on larger changes than anticipated, are an acceptable minimum.

In selecting factors for variation, it is important to ensure that the larger effects are ‘exercised’ as much as possible. For example, where day to day variation (perhaps arising from recalibration effects) is substantial compared to repeatability, two determinations on each of five days will provide a better estimate of intermediate precision than five determinations on each of two days. Ten single determinations on separate days will be better still, subject to sufficient control, though this will provide no additional information on within-day repeatability.

Clearly, in planning significance checks, any study should have sufficient power to detect such effects before they become practically important (that is, comparable to the largest component of uncertainty).

In addition, the following considerations may be important:

- Where factors are known or suspected to interact, it is important to ensure that the effect of interaction is accounted for. This may be achieved either by ensuring random selection from different levels of

interacting parameters, or by careful systematic design to obtain ‘interaction’ effects or covariance information.

- In carrying out studies of overall bias, it is important that the reference materials and values are relevant to the materials under routine test.

6 Extent of validation studies

The extent to which a laboratory has to undertake validation of a new, modified or unfamiliar method depends to a degree on the existing status of the method and the competence of the laboratory. Suggestions as to the extent of validation and verification measures for different circumstances are given below. Except where stated, it is assumed that the method is intended for routine use.

6.1 THE LABORATORY IS TO USE A "FULLY" VALIDATED METHOD

The method has been studied in a collaborative trial and so the laboratory has to verify that it is capable of achieving the published performance characteristics of the method (or is otherwise able to fulfil the requirements of the analytical task). The laboratory should undertake precision studies, bias studies (including matrix variation studies), and possibly linearity studies, although some tests such as that for ruggedness may be omitted.

6.2 THE LABORATORY IS TO USE A FULLY VALIDATED METHOD, BUT NEW MATRIX IS TO BE USED

The method has been studied in a collaborative trial and so the laboratory has to verify that the new matrix introduces no new sources of error into the system. The same range of validation as the previous is required.

6.3 THE LABORATORY IS TO USE A WELL-ESTABLISHED, BUT NOT COLLABORATIVELY STUDIED, METHOD

The same range of validation as the previous is required.

6.4 THE METHOD HAS BEEN PUBLISHED IN THE SCIENTIFIC LITERATURE TOGETHER WITH SOME ANALYTICAL CHARACTERISTICS

The laboratory should undertake precision studies, bias studies (including matrix variation studies), ruggedness and linearity studies.

6.5 THE METHOD HAS BEEN PUBLISHED IN THE SCIENTIFIC LITERATURE WITH NO CHARACTERISTICS GIVEN OR HAS BEEN DEVELOPED IN-HOUSE

The laboratory should undertake precision studies, bias studies (including matrix variation studies), ruggedness and linearity studies.

6.6 THE METHOD IS EMPIRICAL

An empirical method is one in which the quantity estimated is simply the result found on following the stated procedure. This differs from measurements intended to assess method-independent quantities such as the concentration of a particular analyte in a sample, in that the method bias is conventionally zero, and matrix variation (that is , within the defined class) is irrelevant. Laboratory bias cannot be ignored, but is likely to be difficult to estimate by single-laboratory experiment. Moreover, reference materials are unlikely to be available. In the absence of collaborative trial data some estimate of interlaboratory precision could be obtained from a specially designed ruggedness study or estimated by using the Horwitz function.

6.7 THE ANALYSIS IS "AD HOC"

"Ad hoc" analysis is occasionally necessary to establish the general range of a value, without great expenditure and with low criticality. The effort that can go into validation is accordingly strictly limited. Bias should be studied by methods such as recovery estimation or analyte additions, and precision by replication.

6.8 CHANGES IN STAFF AND EQUIPMENT

Important examples include: change in major instruments; new batches of very variable reagents (for example, polyclonal antibodies); changes made in the laboratory premises; methods used for the first time by new staff; or a validated method employed after a period of disuse. Here the essential action is to demonstrate that no deleterious changes have occurred. The minimum check is a single bias test; a "before and after" experiment on typical test materials or control materials. In general, the tests carried out should reflect the possible impact of the change on the analytical procedure.

7 RECOMMENDATIONS

The following recommendations are made regarding the use of single-laboratory method validation:

- Wherever possible and practical a laboratory should use a method of analysis that has had its performance characteristics evaluated through a collaborative trial conforming to an international protocol.
- Where such methods are not available, a method must be validated in-house before being used to generate analytical data for a customer.
- Single-laboratory validation requires the laboratory to select appropriate characteristics for evaluation from the following: applicability, selectivity, calibration, accuracy, precision, range, limit of quantification, limit of detection, sensitivity, ruggedness and practicability. The laboratory must take account of customer requirements in choosing which characteristics are to be determined.
- Evidence that these characteristics have been assessed must be made available to customers of the laboratory if required by the customer.

---

8 REFERENCES

1. *"Protocol for the Design, Conduct and Interpretation of Method Performance Studies"*, W Horwitz, *Pure Appl. Chem.*, 1988, **60**, 855 864, revised W. Horwitz, *Pure Appl. Chem.*, 1995, **67**, 331-343.

2. *"The International Harmonised Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories"*, M Thompson and R Wood, *Pure Appl. Chem.*, 1993, **65**, 2123-2144. (Also published in *J. AOAC International*, 1993, **76**, 926-940.

3. *"Harmonised Guidelines For Internal Quality Control in Analytical Chemistry Laboratories"*, Michael Thompson and Roger Wood, J. Pure & Applied Chemistry, 1995, 67(4), 49-56.

4. *"Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement"*, Michael Thompson, Stephen Ellison, Ales Fajgelj, Paul Willetts and Roger Wood, J. Pure & Applied Chemistry, 1999, 71(2), 337-348.

5. *"Council Directive 93/99/EEC on the Subject of Additional Measures Concerning the Official Control of Foodstuffs"*, O. J., 1993, L290.

6. *"Procedural Manual of the Codex Alimentarius Commission, 10th Edition"*, FAO, Rome, 1997.

7. "*Precision of Test Methods"*, Geneva, 1994, ISO 5725, Previous editions were issued in 1981 and 1986.

8. "*Guide to the Expression of Uncertainty in Measurement*", ISO, Geneva, 1993.

9. "*Quantifying Uncertainty in Analytical Measurement*", EURACHEM Secretariat, Laboratory of the Government Chemist, Teddington, UK, 1995, EURACHEM Guide (under revision).

10. *"International vocabulary of basic and general terms in metrology"* ISO, Geneva 1993

11. *"Validation of Chemical Analytical Methods"*, NMKL Secretariat, Finland, 1996, NMKL Procedure No. 4.

12. *"EURACHEM Guide: The 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.

13. *"Statistics manual of the AOAC"*, AOAC INTERNATIONAL, Gaithersburg, Maryland, USA, 1975

14. *"An Interlaboratory Analytical Method Validation Short Course developed by the AOAC INTERNATIONAL"*, AOAC INTERNATIONAL, Gaithersburg, Maryland, USA, 1996.

15. *"Text on validation of analytical procedures"* International Conference on Harmonisation. Federal Register, Vol. 60, March 1, 1995, pages 11260

16. *"Validation of analytical procedures: Methodology"* International Conference on Harmonisation. Federal Register, Vol. 62, No. 96, May 19, 1997, pages 27463-27467.

17. *"Validation of Methods"*, Inspectorate for Health Protection, Rijswijk, The Netherlands, Report 95-001.

18. *"A Protocol for Analytical Quality Assurance in Public Analysts' Laboratories"*, Association of Public Analysts, 342 Coleford Road, Sheffield S9 5PH, UK, 1986.

19. *"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

20. *"Estimation and Expression of Measurement Uncertainty in Chemical Analysis"*, NMKL Secretariat, Finland, 1997, NMKL Procedure No. 5.

21. M Thompson, PJ Lowthian, *J AOAC Int*, 1997, **80**, 676-679

22. IUPAC recommendation: *"Nomenclature in evaluation of analytical methods, including quantification and detection capabilities" Pure and Applied Chem."* 1995, **67** 1699-1723

23. ISO 11843. "*Capability of detection*." (Several parts). International Standards Organisation, Geneva.

24. M. Thompson, *Analyst*, 2000, **125**, 2020-2025

25. *"Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing"* M Thompson, *Analyst*, 2000, **125**, 385-386.

26. *"How to combine proficiency test results with your own uncertainty estimate - the zeta score"*, Analytical Methods Committee of the Royal Society of Chemistry, AMC Technical Briefs, editor M. Thompson, *AMC Technical Brief No. 2, www.rsc.org/lap/rsccom/amc*

APPENDIX A: Notes on the requirements for study of method performance characteristics.

The general requirements for the individual performance characteristics for a method are as follows.

A1 APPLICABILITY

After validation the documentation should provide, in addition to any performance specification, the following information:

- the identity of the analyte, including speciation where appropriate (Example: 'total arsenic';
- the concentration range covered by the validation (Example: '0-50 ppm');
- a specification of the range of matrices of the test material covered by the validation (Example: 'seafood');
- a protocol, describing the equipment, reagents, procedure (including permissible variation in specified instructions, e.g., 'heat at 100±5° for 30±5 minutes'), calibration and quality procedures, and any special safety precautions required;
- the intended application and its critical uncertainty requirements (Example: 'The analysis of food for screening purposes. The standard uncertainty $u(c)$ of the result c should be less than $0.1 \times c$.').

A2 SELECTIVITY

Selectivity is the degree to which a method can quantify the analyte accurately in the presence of interferents. Ideally, selectivity should be evaluated for any important interferent likely to be present. It is particularly important to check interferents which are likely, on chemical principles, to respond to the test. For example, colorimetric tests for ammonia might reasonably be expected to respond to primary aliphatic amines. It may be impracticable to consider or test every potential interferent; where that is the case, it is recommended that the likely worst cases are checked. As a general principle, selectivity should be sufficiently good for any interferences to be ignored.

In many types of analysis, selectivity is essentially a qualitative assessment based on the significance or otherwise of suitable tests for interference. However, there are useful quantitative measures. In particular, one quantitative measure is the selectivity index b_{an}/b_{int}, where b_{an} is the sensitivity of the method (slope of the calibration function) and b_{int} the slope of the response independently produced by a potential interferent, provides a quantitative measure of interference. b_{int} can be determined approximately by execution of the procedure on a matrix blank and the same blank spiked with the potential interferent at one appropriate concentration. If a matrix blank is unavailable, and a typical material used instead, b_{int} can be estimated from such a simple experiment only under the assumption that mutual matrix effects are absent. Note that b_{int} is more easily determined in the absence of the analyte because the effect might be confused with another type of interference when the sensitivity of the analyte is itself affected by the interferent (a matrix effect).

A3 CALIBRATION AND LINEARITY

With the exception of gross errors in preparation of calibration materials, calibration errors are usually (but not always) a minor component of the total uncertainty budget, and can usually be safely subsumed into various categories estimated by "top-down" methods. For example random errors resulting from calibration are part of the run bias, which is assessed as a whole, while systematic errors from that source may appear as laboratory bias, likewise assessed as a whole. Never-the-less, there are some characteristics of calibration that are useful to know at the outset of method validation, because they affect the strategy for the optimal development of the procedure. In this class are such questions as whether the calibration function plausibly (a) is linear, (b) passes through the origin and (c) is unaffected by the matrix of the test material. The procedures described here relate to calibration studies in validation, which are necessarily more exacting than calibration undertaken during routine analysis. For example, once it is established at validation that a calibration function is linear and passes through the origin, a much simpler calibration strategy can be used for routine use (for example, a two point repeated design). Errors from this simpler calibration strategy will normally be subsumed into higher level errors for validation purposes.

A3.1 Linearity and intercept

Linearity can be tested informally by examination of a plot of residuals produced by linear regression of the responses on the concentrations in an appropriate calibration set. Any curved pattern suggests lack of fit due to a non-linear calibration function. A test of significance can be undertaken by comparing the lack-of-fit variance with that due to pure error. However, there are causes of lack of fit other than nonlinearity that can arise in certain types of analytical calibration, so the significance test must be used in conjunction with a residual plot. Despite its current widespread use as an indication of quality of fit, the correlation coefficient is misleading and inappropriate as a test for linearity and should not be used.

Design is all-important in tests for lack of fit, because it is easy to confound nonlinearity with drift. Replicate measurements are needed to provide an estimate of pure error if there is no independent estimate. In the absence of specific guidance, the following should apply:

- there should be six or more calibrators;
- the calibrators should be evenly spaced over the concentration range of interest;
- 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;
- the calibrators should be run at least in duplicate, and preferably triplicate or more, in a random order.

After an exploratory fit with simple linear regression, the residuals should be examined for obvious patterns. Heteroscedasticity is quite common in analytical calibration and a pattern suggesting it means that the calibration data are best treated by weighted regression. Failure to use weighted regression in these circumstances could give rise to exaggerated errors at the low end of the calibration function.

The test for lack of fit can be carried out with either simple or weighted regression. A test for an intercept significantly different from zero can also be made on this data if there is no significant lack of fit.

A3.2 Test for general matrix effect

It simplifies calibration enormously if the calibrators can be prepared as a simple solution of the analyte. The effects of a possible general matrix mismatch must be assessed in validation if this strategy is adopted. A test for general matrix effect can be made by applying the method of analyte additions (also called “standard additions”) to a test solution derived from a typical test material. The test should be done in a way that provides the same final dilution as the normal procedure produces, and the range of additions should encompass the same range as the procedure-defined calibration validation. If the calibration is linear the slopes of the usual calibration function and the analyte additions plot can be compared for significant difference. A lack of significance means that there is no detectable general matrix effect. If the calibration is not linear a more complex method is needed for a significance test, but a visual comparison at equal concentrations will usually suffice. A lack of significance in this test will often mean that the matrix variation effect [Section A13] will also be absent.

A3.3 Final calibration procedure

The calibration strategy as specified in the procedure may also need to be separately validated, although the errors involved will contribute to jointly estimated uncertainties. The important point here is that evaluation uncertainty estimated from the specific designs for linearity etc., will be smaller than those derived from the simpler calibration defined in the procedure protocol.

A4 TRUENESS

A4.1 Estimation of trueness

Trueness is the closeness of agreement between a test result and the accepted reference value of the property being measured. Trueness is stated quantitatively in terms of "bias"; with smaller bias indicating greater trueness. Bias is typically determined by comparing the response of the method to a reference material with the known value assigned to the material. Significance testing is recommended. Where the uncertainty in the reference value is not negligible, evaluation of the results should consider the reference material uncertainty as well as the statistical variability.

A4.2 Conditions for trueness experiments

Bias can arise at different levels of organisation in an analytical system, for example, run bias, laboratory bias and method bias. It is important to remember which of these is being handled by the various methods of addressing bias. In particular:

- The mean of a series of analyses of a reference material, carried out wholly within a single run, gives information about the sum of method, laboratory and run effect for that particular run. Since the run effect is assumed to be random from run to run, the result will vary from run to run more than would be expected from the observable dispersion of the results, and this needs to be taken into account in the evaluation of the results (for example, by testing the measured bias against the among-runs standard deviation investigated separately).
- The mean of repeated analyses of a reference material in several runs, estimates the combined effect of method and laboratory bias in the particular laboratory (except where the value is assigned using the particular method).

A4.3 Reference values for trueness experiments

A4.3.1 Certified reference materials (CRMs)

CRMs are traceable to international standards with a known uncertainty and therefore can be used to address all aspects of bias (method, laboratory and within-laboratory) simultaneously, assuming that there is no matrix mismatch. CRMs should accordingly be used in validation of trueness where it is practicable to do so. It is important to ensure that the certified value uncertainties are sufficiently small to permit detection of a bias of important magnitude. Where they are not, the use of CRMs is still recommended, but additional checks should be carried out.

A typical trueness experiment generates a mean response on a reference material. In interpreting the result, the uncertainty associated with the certified value should be taken into account along with the uncertainty arising from statistical variation in the laboratory. The latter term may be based on the within-run, between-run, or an estimate of the between-laboratory standard deviation depending on the intent of the experiment. statistical or materials. Where the certified value uncertainty is small, a Student's *t* test is normally carried out, using the appropriate precision term.

Where necessary and practicable, a number of suitable CRMs, with appropriate matrices and analyte concentrations, should be examined. Where this is done, and the uncertainties on the certified values are smaller than those on the analytical results, it would be reasonably safe to use simple regression to evaluate the results. In this way bias could be expressed as a function of concentration, and might appear as a non-zero intercept ("transitional" or constant bias) or as a non-unity slope ("rotational" or proportional bias). Due caution should be applied in interpreting the results where the range of matrices is large.

4.3.2 Reference materials

Where CRMs are not available, or as an addition to CRMs, use may be made of any material sufficiently well characterised for the purpose (a reference material[10]), bearing in mind always that while insignificant bias may not be proof of zero bias, significant bias on any material remains a cause for investigation. Examples of reference materials include: Materials characterised by a reference material producer, but whose values are not accompanied by an uncertainty statement or are otherwise qualified; materials

characterised by a manufacturer of the material; materials characterised in the laboratory for use as reference materials; materials subjected to a restricted round-robin exercise, or distributed in a proficiency test. While the traceability of these materials may be questionable, it would be far better to use them than to conduct no assessment for bias at all. The materials would be used in much the same way as CRMs, though with no stated uncertainty any significance test relies wholly on the observable precision of results.

A4.3.3 Use of a reference method

A reference method can in principle be used to test for bias in another method under validation. This is a useful option when checking an alternative to, or modification of, an established standard method already validated and in use in the laboratory. Both methods are used to analyse a number of typical test materials, preferably covering a useful range of concentration fairly evenly. Comparison of the results over the range by a suitable statistical method (for example, a paired *t*-test, with due checks for homogeneity of variance and normality) would demonstrate any bias between the methods.

A4.3.4 Use of spiking/recovery

In the absence of reference materials, or to support reference material studies, bias can be investigated by spiking and recovery. A typical test material is analysed by the method under validation both in its original state and after the addition (spiking) of a known mass of the analyte to the test portion. The difference between the two results as a proportion of the mass added is called the surrogate recovery or sometimes the marginal recovery. Recoveries significantly different from unity indicate that a bias is affecting the method. Strictly, recovery studies as described here only assess bias due to effects operating on the added analyte; the same effects do not necessarily apply to the same extent to the native analyte, and additional effects may apply to the native analyte. Spiking/recovery studies are accordingly very strongly subject to the observation that while good recovery is not a guarantee of trueness, poor recovery is certainly an indication of lack of trueness. Methods of handling spiking/recovery data have been covered in detail elsewhere.[4]

A5 PRECISION

Precision is the closeness of agreement between independent test results obtained under stipulated conditions. It is usually specified in terms of standard deviation or relative standard deviation. The distinction between precision and bias is fundamental, but depends on the level at which the analytical system is viewed. Thus from the viewpoint of a single determination, any deviation affecting the calibration for the run would be seen as a bias. From the point of view of the analyst reviewing a year's work, the run bias will be different every day and act like a random variable with an associated precision. The stipulated conditions for the estimation of precision take account of this change in view point.

For single laboratory validation, two sets of conditions are relevant: (a) precision under repeatability conditions, describing variations observed during a single run as expectation 0 and standard deviation s_r, and (b) precision under run-to-run conditions, describing variations in run bias δ_{run} as expectation 0, standard deviation s_{run}. Usually both of these sources of error are operating on individual analytical results, which therefore have a combined precision $s_{tot} = \left(s_r^2 / n + s_{run}^2 \right)^{1/2}$, where n is the number of repeat results averaged within a run for the reported result. The two precision estimates can be obtained most simply by analysing the selected test material in duplicate in a number of successive runs. The separate variance components can then be calculated by the application of one-way analysis of variance. Each duplicate analysis must be an independent execution of the procedure applied to a separate test portion. Alternatively the combined precision s_{tot} can be estimated directly by the analysis of the test material once in successive runs, and estimating the standard deviation from the usual equation. (Note that observed standard deviations are generally given the symbol *s*, to distinguish them from standard deviations σ)

It is important that the precision values are representative of likely test conditions. First, the variation in conditions among the runs must represent what would normally happen in the laboratory under routine use of the method. For instance, variations in reagent batches, analysts and instruments should be representative.

Second, the test material used should be typical, in terms of matrix and (ideally) the state of comminution, of the materials likely to encountered in routine application. So actual test materials or, to a lesser degree, matrix-matched reference materials would be suitable, but standard solutions of the analyte would not. Note also that CRMs and prepared reference materials are frequently homogenised to a greater extent than typical test materials, and precision obtained from their analysis may accordingly under-estimate the variation that will be observed for test materials.

Precision very often varies with analyte concentration. Typical assumptions are i) that there is no change in precision with analyte level, or ii) that the standard deviation is proportional to, or linearly dependent on, analyte level. In both cases, the assumption needs to be checked if the analyte level is expected to vary substantially (that is, by more than about 30% from its central value). The most economical experiment is likely to be a simple assessment of precision at or near the extremes of the operating range, together with a suitable statistical test for difference in variance. The F-test is appropriate for normally distributed error.

Precision data may be obtained for a wide variety of different sets of conditions in addition to the minimum of repeatability and between-run conditions indicated here, and it may be appropriate to acquire additional information. For example, it may be useful to the assessment of results, or for improving the measurement, to have an indication of separate operator and run effects, between or within-day effects or the precision attainable using one or several instruments. A range of different designs and statistical analysis techniques is available, and careful experimental design is strongly recommended in all such studies.

A6 RECOVERY

Methods for estimating recovery are discussed in conjunction with methods of estimating trueness (above).

A7 RANGE

The validated range is the interval of analyte concentration within which the method can be regarded as validated. It is important to realise that this range is not necessarily identical to the useful range of the calibration. While the calibration may cover a wide concentration range, the remainder of the validation (and usually much more important part in terms of uncertainty) will cover a more restricted range. In practice, most methods will be validated at only one or two levels of concentration. The validated range may be taken as a reasonable extrapolation from these points on the concentration scale.

When the use of the method focuses on a concentration of interest well above the detection limit, validation near that one critical level would be appropriate. It is impossible to define a general safe extrapolation of this result to other concentrations of the analyte, because much depends on the individual analytical system. Therefore the validation study report should state the range around the critical value in which the person carrying out the validation, using professional judgement, regards the estimated uncertainty to hold true.

When the concentration range of interest approaches zero, or the detection limit, it is incorrect to assume either constant absolute uncertainty or constant relative uncertainty. A useful approximation in this common circumstance is to assume a linear functional relationship, with a positive intercept, between uncertainty u and concentration c, that is of the form

$$u(c) = u_0 + \theta c$$

where θ is the relative uncertainty estimated a some concentration well above the detection limit. u_0 is the standard uncertainty estimated for zero concentration and in some circumstances could be estimated as $c_L / 3$. In these circumstances it would be reasonable to regard the validated range as extending from zero to a small integer multiple of the upper validation point. Again this would depend on professional judgement.

A8 DETECTION LIMIT

In broad terms the detection limit (limit of detection) is the smallest amount or concentration of analyte in the test sample that can be reliably distinguished from zero[22,23]. For analytical systems where the validation range does not include or approach it, the detection limit does not need to be part of a validation.

Despite the apparent simplicity of the idea, the whole subject of the detection limit is beset with problems outlined below:

- There are several possible conceptual approaches to the subject, each providing a somewhat different definition of the limit. Attempts to clarify the issue seem ever more confusing.
- Although each of these approaches depends of an estimate of precision at or near zero concentration, it is not clear whether this should be taken as implying repeatability conditions or some other condition for the estimation.
- Unless an inordinate amount of data is collected, estimates of detection limit will be subject to quite large random variation.
- Estimates of detection limit are often biased on the low side because of operational factors.
- Statistical inferences relating to the detection limit depend on the assumption of normality, which is at least questionable at low concentrations.

For most practical purposes in method validation, it seems better to opt for a simple definition, leading to a quickly implemented estimation which is used only as a rough guide to the utility of the method. However, it must be recognised that the detection limit as estimated in method development, may not be identical in concept or numerical value to one used to characterise a complete analytical method. For instance the "instrumental detection limit", as quoted in the literature or in instrument brochures and then adjusted for dilution, is often far smaller than a "practical" detection limit and inappropriate for method validation.

It is accordingly recommended that for method validation, the precision estimate used ($\hat{s}_0$) should be based on at least 6 independent complete determinations of analyte concentration in a typical matrix blank or low-level material, with no censoring of zero or negative results, and the approximate detection limit calculated as $3\hat{s}_0$. Note that with the recommended minimum number of degrees of freedom, this value is quite uncertain, and may easily be in error by a factor of two. Where more rigorous estimates are required (for example to support decisions on based on detection or otherwise of a material), reference should be made to appropriate guidance (see, for example, references 22-23).

A9 LIMIT OF DETERMINATION OR LIMIT OF QUANTIFICATION

It is sometimes useful to state a concentration below which the analytical method cannot operate with an acceptable precision. Sometimes that precision is arbitrarily defined as 10 %RSD, sometimes the limit is equally arbitrarily taken as a fixed multiple (typically 2) of the detection limit. While it is to a degree reassuring to operate above such a limit, we must recognise that it is a quite artificial dichotomy of the concentration scale: measurements below such a limit are not devoid of information content and may well be fit for purpose. Hence the use of this type of limit in validation is not recommended here. It is preferable to try to express the uncertainty of measurement as a function of concentration and compare that function with a criterion of fitness for purpose agreed between the laboratory and the client or end-user of the data.

A10 SENSITIVITY

The sensitivity of a method is the gradient of the calibration function. As this is usually arbitrary, depending on instrumental settings, it is not useful in validation. (It may be useful in quality assurance procedures, however, to test whether an instrument is performing to a consistent and satisfactory standard.)

A11 RUGGEDNESS

The ruggedness of an analytical method is the resistance to change in the results produced by an analytical method when minor deviations are made from the experimental conditions described in the procedure. The

limits for experimental parameters should be prescribed in the method protocol (although this has not always been done in the past), and such permissible deviations, separately or in any combination, should produce no meaningful change in the results produced. (A "meaningful change" here would imply that the method could not operate within the agreed limits of uncertainty defining fitness for purpose.) The aspects of the method which are likely to affect results should be identified, and their influence on method performance evaluated by using ruggedness tests.

The ruggedness of a method is tested by deliberately introducing small changes to the procedure and examining the effect on the results. A number of aspects of the method may need to be considered, but because most of these will have a negligible effect it will normally be possible to vary several at once. An economical experiment based on fractional factorial designs has been described by Youden[13]. For instance, it is possible to formulate an approach utilising 8 combinations of 7 variable factors, that is to look at the effects of seven parameters with just eight analytical results. Univariate approaches are also feasible, where only one variable at a time is changed

Examples of the factors that a ruggedness test could address are: changes in the instrument, operator, or brand of reagent; concentration of a reagent; pH of a solution; temperature of a reaction; time allowed for completion of a process etc.

A12 FITNESS FOR PURPOSE

Fitness for purpose is the extent to which the performance of a method matches the criteria, agreed between the analyst and the end-user of the data, that describe the end-user's needs. For instance the errors in data should not be of a magnitude that would give rise to incorrect decisions more often than a defined small probability, but they should not be so small that the end-user is involved in unnecessary expenditure. Fitness for purpose criteria could be based on some of the characteristics described in this Appendix, but ultimately will be expressed in terms of acceptable total uncertainty.

A13 MATRIX VARIATION

Matrix variation is, in many sectors, one of the most important but least acknowledged sources of error in analytical measurements. When we define the analytical system to be validated by specifying, amongst other things, the matrix of the test material, there may be scope for considerable variation within the defined class. To cite an extreme example, a sample of the class "soil" could be composed of clay, sand, chalk, laterite (mainly Fe_2O_3 and Al_2O_3), peat, etc., or of mixtures of these. It is easy to imagine that each of these types would contribute a unique matrix effect on an analytical method such as atomic absorption spectrometry. If we have no information about the type of soils we are analysing, there will be an extra uncertainty in the results because of this variable matrix effect.

Matrix variation uncertainties need to be quantified separately, because they are not taken into account elsewhere in the process of validation. The information is acquired by collecting a representative set of the matrices likely to be encountered within the defined class, all with analyte concentrations in the appropriate range. The material are analysed according to the protocol, and the bias in the results estimated. Unless the test materials are CRMs, the bias estimate will usually have to be undertaken by means of spiking and recovery estimation. The uncertainty is estimated by the standard deviation of the biases. (Note: This estimate will also contain a variance contribution from the repeat analysis. This will have a magnitude $2s_r^2$ if spiking has been used. If a strict uncertainty budget is required, this term should be deducted from the matrix variation variance to avoid double accounting.)

A14 MEASUREMENT UNCERTAINTY

The formal approach to measurement uncertainty estimation calculates a measurement uncertainty estimate from an equation, or mathematical model. The procedures described as method validation are designed to ensure that the equation used to estimate the result, with due allowance for random errors of all kinds, is a valid expression embodying all recognised and significant effects upon the result. It follows that, with one

caveat elaborated further below, the equation or 'model' subjected to validation may be used directly to estimate measurement uncertainty. This is done by following established principles, based on the 'law of propagation of uncertainty' which, for independent input effects is

$$u(y(x_1,x_2,\ldots)) = \sqrt{\sum_{i=1,n} c_i^2 u(x_i)^2}$$

where $y(x_1,x_2,\ldots x_n)$ is a function of several independent variables $x_1,x_2\ldots$, and c_i is a sensitivity coefficient evaluated as $c_i=\partial y/\partial x_i$, the partial differential of y with respect to x_i. $u(x_i)$ and $u(y)$ are *standard uncertainties*, that is, measurement uncertainties expressed in the form of standard deviations. Since $u(y(x_1,x_2,\ldots))$ is a function of several separate uncertainty estimates, it is referred to as a *combined standard uncertainty*.

To estimate measurement uncertainty from the equation $y=f(x_1,x_2\ldots)$ used to calculate the result, therefore, it is necessary first, to establish the uncertainties $u(x_i)$ in each of the terms x_1, x_2 etc. and second, to combine these with the additional terms required to represent random effects as found in validation, and finally to take into account any additional effects. In the discussion of precision above, the implied statistical model is

$$y=f(x_1,x_2\ldots) + \delta_{run} + e$$

where e is the random error for a particular result. Since δ_{run} and e are known, from the precision experiments, to have standard deviations σ_{run} and σ_r respectively, these latter terms (or, strictly, their estimates s_{run} and s_r) are the uncertainties associated with these additional terms. Where the individual within-run results are averaged, the combined uncertainty associated with these two terms is (as given previously) $s_{tot} = \left(s_r^2/n + s_{run}^2\right)^{1/2}$. Note that where the precision terms are shown to vary with analyte level, the uncertainty estimate for a given result must employ the precision term appropriate to that level. The basis for the uncertainty estimate accordingly follows directly from the statistical model assumed and tested in validation. To this estimate must be added any further terms as necessary to account for (in particular) inhomogeneity and matrix effect (see section A13). Finally, the calculated standard uncertainty is multiplied by a 'coverage factor', k, to provide an expanded uncertainty, that is, "an interval expected to encompass a large fraction of the distribution of values that may be attributed to the measurand"[8]. Where the statistical model is well established, the distribution known to be normal, and the number of degrees of freedom associated with the estimate is high, k is generally chosen to be equal to 2. The expanded uncertainty then corresponds approximately to a 95% confidence interval.

There is one important caveat to be added here. In testing the assumed statistical model, imperfect tests are perforce used. It has already been noted that these tests can not prove that any effect is identically zero; they can only show that an effect is too small to detect within the uncertainty associated with the particular test for significance. A particularly important example is the test for significant laboratory bias. Clearly, if this is the only test performed to confirm trueness, there must be some residual uncertainty as to whether the method is indeed unbiased or not. It follows that where such uncertainties are significant with respect to the uncertainty calculated so far, additional allowance should be made.

In the case of an uncertain reference value, the simplest allowance is the stated uncertainty for the material, combined with the statistical uncertainty in the test applied. A full discussion is beyond the scope of this text; reference 9 provides further detail. It is, however, important to note that while the uncertainty estimated directly from the assumed statistical model is the *minimum* uncertainty that can be associated with an analytical result, it will almost certainly be an underestimate; similarly, an expanded uncertainty based on the same considerations and using k=2 will not provide sufficient confidence.

The ISO Guide[8] recommends that for increased confidence, rather than arbitrarily adding terms, the value of k should be increased as required. Practical experience suggests that for uncertainty estimates based on a validated statistical model, but with no evidence beyond the validation studies to provide additional confidence in the model, k should not be less than 3. Where there is strong reason to doubt that the validation study is comprehensive, k should be increased further as required.

APPENDIX B. Additional considerations for UNCERTAINTY ESTIMATION IN VALIDATION STUDIES

B1 SENSITIVITY ANALYSIS.

The basic expression used in uncertainty estimation

$$u(y(x_1,x_2,\ldots)) = \sqrt{\sum_{i=1,n} c_i^2 u(x_i)^2}$$

requires the 'sensitivity coefficients' c_i. It is common in uncertainty estimation to find that while a given influence factor x_i has a known uncertainty $u(x_i)$, the coefficient c_i is insufficiently characterised or not readily obtainable from the equation for the result. This is particularly common where an effect is not included in the measurement equation because it is not normally significant, or because the relationship is not sufficiently understood to justify a correction. For example, the effect of solution temperature T_{sol} on a room temperature extraction procedure is rarely established in detail.
Where it is desired to assess the uncertainty in a result associated with such an effect, it is possible to determine the coefficient experimentally. This is done most simply by changing x_i and observing the effect on the result, in a manner very similar to basic ruggedness tests. In most cases, it is sufficient in the first instance to choose at most two values of x_i other than the nominal value, and calculate an approximate gradient from the observed results. The gradient then gives an approximate value for c_i. The term $c_i.u(x_i)$ can then be determined. (Note that this is one practical method for demonstrating the significance or otherwise of a possible effect on the results).

In such an experiment, it is important that the change in result observed be sufficient for a reliable calculation of c_i. This is difficult to predict in advance. However, given a permitted range for the influence quantity x_i, or an expanded uncertainty for the quantity, that is expected to result in insignificant change, it is clearly important to assess ci from a larger range. It is accordingly recommended that for an influence quantity with an expected range of $\pm a$, (where $\pm a$ might be, for example, the permitted range, expanded uncertainty interval or 95% confidence interval) the sensitivity experiment employ, where possible, a change of at least $4a$ to ensure reliable results

B2 .JUDGEMENT

It is not uncommon to find that while an effect is recognised and may be significant, it is not always possible to obtain a reliable estimate of uncertainty. In such circumstances, the ISO Guide makes it quite clear that a professionally considered estimate of the uncertainty is to be preferred to neglect of the uncertainty. Thus, where no estimate of uncertainty is available for a potentially important effect, the analyst should make their own best judgement of the likely uncertainty and apply that in estimating the combined uncertainty. Reference 8 gives further guidance on the use of judgement in uncertainty estimation.

APPENDIX II

UK-PROCEDURE TO BE USED FOR THE VALIDATION OF METHODS THROUGH THE USE OF RESULTS FROM PROFICIENCY TESTING SCHEMES.

Introduction

The purpose of a proficiency testing scheme is to test the competence of the laboratory and not to validate a method of analysis. In most proficiency testing schemes participants have a free choice of method of analysis and so there is no opportunity to formally validate a method using a proficiency testing scheme, i.e. a multiplicity of methods may be used by participants. However, in some situations, there is the possibility of validating a method of analysis if:

1. there are sufficient participants in the proficiency testing scheme who choose to use the same defined method of analysis or
2. a method of analysis is prescribed by the scheme co-ordinators.

For most proficiency testing schemes it is the former situation which will predominate but in the case of microbiology, in particular, it is very likely that a method of analysis will be prescribed. The former situation will tend to occur when a very empirical determination is being assessed in the proficiency testing scheme.

Organisation of Proficiency Testing Scheme

The proficiency testing scheme who's results are to be used must be organised according to the AOAC/ISO/IUPAC International Protocol on the organisation of proficiency testing schemes for the results to be recognised by MAFF. That protocol stipulates the procedures that must be incorporated in any proficiency testing scheme, and in particular the work that must be carried out by the scheme co-ordinators to ensure that samples received by the participants in the scheme are homogeneous. At present there are no internationally agreed stipulations to ensure sufficient homogeneity of samples used for collaborative trial exercises whereas there are for proficiency test exercises.

Number of Samples

In the case of a collaborative trial conforming in design to the Harmonised protocol a minimum of 5 test materials are required to be prepared. However, in most proficiency testing schemes there are insufficient test materials sent out in any one round (i.e. the dispatch of test material at a specific time) to meet the minimum requirements for number of materials as specified in the Harmonised Guidelines for Collaborative Studies (see appendix 1). Because of that it is necessary to "build up" the number of samples used to validate a method over a period of time. This may mean that the time taken to build up sufficient results to ensure validation of a method may extend over one or two years depending upon the test materials which are being used in the proficiency testing scheme.

Replication of Results

In most cases the validation of a method in a collaborative trial results in both within, and between, laboratory precision characteristics (i.e. repeatability and reproducibility). In most collaborative trials are dispatched as either blind duplicates or as split level test materials. This means that one of the aims of a collaborative trial, that of determining the within-laboratory variability, is readily achievable. However, because the aim of a proficiency test is different from a collaborative trial, the results of replicate analyses of any particular test material are normally not reported to the proficiency test co-ordinator - i.e. it is only the single result, as reported to the customer, that is returned. Because of that, it is frequently the case that it is not possible to obtain the within-laboratory variability of the method. In such cases only the overall precision of the method i.e. the between-laboratory precision will be quoted and not the within-laboratory in the MAFF Validated Method Series.

Laboratories

Not necessarily the same each round.

ANNEX III

MAFF VALIDATED METHODS FOR THE ANALYSIS OF FOODSTUFFS

MAFF VALIDATED METHOD V38 **OCTOBER 1996**

METHOD FOR THE ENUMERATION OF *LISTERIA MONOCYTOGENES* IN MEAT AND MEAT PRODUCTS

The Ministry of Agriculture, Fisheries and Food, CSL Food Science Laboratory Microbiological Method for the Enumeration of *Listeria monocytogenes* in Meat and Meat Products

Also published in the Journal of the Association of Public Analysts 1997, 33, 67 - 85

Correspondence on the MAFF Validated Methods Series may be sent to Roger Wood, JFSSG, Food Contaminants Division, c/o Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA.

MAFF VALIDATED METHOD V38: METHOD FOR THE ENUMERATION OF *LISTERIA MONOCYTOGENES* IN MEAT AND MEAT PRODUCTS

The Ministry of Agriculture, Fisheries and Food CSL Food Science Laboratory Microbiological Method for the Enumeration of *Listeria monocytogenes* in Meat and Meat Products

Correspondence on this method may be sent to Roger Wood, JFSSG, Food Contaminants Division, c/o Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA.

COSHH AND SAFETY CONSIDERATIONS

Analysts are reminded that appropriate hazard and risk assessments required by the Control of Substances Hazardous to Health Regulations, 1988 (See "Control of Substances Hazardous to Health - Approved Codes of Practice, Control of Substances Hazardous to Health Regulations, 1988") must be made before using this method.

Each laboratory should follow its own safety rules and national regulations, particularly COSHH, with respect to the sample preparation.

The procedures specified in this method shall only be carried out in laboratories with suitable facilities and under control of a qualified microbiologist.

These procedures shall not be performed in quality control laboratories, or in food manufacturing or processing premises, where there is a risk of contamination of the environment.

Full bacteriological precautions shall be taken at all times whilst carrying out the procedure specified in this method. Particular attention shall be given to the sterilisation of used equipment and media after testing suspect samples prior to disposal or reuse.

Note - For further and more detailed safety precautions reference is made to ISO 7218, Microbiology - General Guidance for Microbiological Examinations, in general and the clauses 3,4 and 7 in particular (10.3).

1. SCOPE AND FIELD OF APPLICATION

This method specifies procedures recommended for the enumeration of *Listeria monocytogenes* in meat and meat products.

2. DEFINITIONS

For the purpose of this recommended method the following definitions apply:

2.1 *Listeria monocytogenes*:
Pathogenic bacteria that form typical colonies on the specified solid selective medium and which display the morphological, physiological and biochemical characteristics described, when tests are carried out in accordance with this method.

2.2 Enumeration of *Listeria monocytogenes*:
Determination of the number of viable and confirmed *L. monocytogenes* bacteria per gram of product when the examination is carried out in accordance with this method.

3. PRINCIPLE

In general, the enumeration of *L. monocytogenes* necessitates three successive stages as in 3.1 to 3.3. See also the diagram of procedure in Appendix I.

3.1 Preparation of the test sample

The test sample is homogenised in suspension medium and decimal dilutions are prepared as necessary.

3.2 Enumeration and presumptive identification

The selective agar is inoculated from the initial suspension (3.1) and dilutions thereof, incubated at 30°C and examined after 48 h to check for the presence of colonies which, from their appearance, are considered to be presumptive *Listeria* spp..

3.3 Confirmation of identity

Colonies of presumptive *Listeria* spp. (3.2) are sub-cultured onto a non-selective solid medium for confirmation of identity by means of appropriate morphological, physiological and biochemical tests.

4. CULTURE MEDIA AND REAGENTS

4.1 Basic materials

In order to improve the reproducibility of the results, it is recommended that, for the preparation of the culture media, dehydrated basic components or complete dehydrated media are used. The manufacturer's instructions shall be rigorously followed.

The chemical products used for the preparation of the culture media and reagents shall be of recognised analytical quality.

The water used shall be distilled or deionised water, free from substances that might inhibit the growth of microorganisms under the test conditions.

When agar is specified, the amount used should be varied according to the manufacturer's instructions to give media of suitable firmness.

Measurements of pH shall be made using a pH meter, measurements being referred to a temperature of 25°C. Adjustments, if necessary, are made by adding either 1 M hydrochloric acid or 1 M sodium hydroxide solution.

If the prepared culture media and reagents are not used immediately, they shall, unless otherwise stated, be stored in the dark at a temperature between 2 and 5°C for no longer than 1 month, conditions which do not produce any change in their composition.

4.2 Diluent

	g/L
Peptone	1.0
Sodium chloride	8.5

Preparation: Dissolve the components in 1000 mL of water by gently heating. Dispense into final containers such that after autoclaving each tube or bottle contains 9.0 ± 0.2 mL. Adjust the pH so that after autoclaving it is 7.0 ± 0.2 at 25°C. Autoclave at 121°C for 15 min.

4.3 Culture media

4.3.1 Sample suspension medium (UVM 1 formulation)

4.3.1.1 Base

	g/L
Protease peptone	5.0
Tryptone	5.0
Meat extract	5.0
Yeast extract	5.0
Sodium chloride	20.0
Di-sodium hydrogen phosphate	12.0
Potassium dihydrogen phosphate	1.35
Aesculin	1.0

Preparation: Dissolve the dehydrated components in 1000 mL of water by boiling. Autoclave at 121°C for 15 min. Cool to 50°C.

4.3.1.2 Supplement per Litre of medium

	mg
Nalidixic acid	20.0
Acriflavine HCL	12.0

Preparation: Dissolve the components in 4 mL of water. Sterilise by filtration through a filter of pore size 0.22 µm (5.1.13).

4.3.1.3 Preparation of the complete medium

Aseptically add the supplement to the basal medium. Invert gently to dissolve. Adjust the pH so that, after sterilisation and the addition of supplement, it is 7.4 ± 0.2 at 25°C. Aseptically distribute the

complete medium into 225 mL volumes. The complete medium may be stored for up to one week at 2 - 5°C before use.

4.3.2 Listeria selective agar Oxford formulation

4.3.2.1 Base

	g/L
Special peptone	23.0
Starch	1.0
Sodium chloride	5.0
Aesculin	1.0
Ferric ammonium citrate	0.5
Lithium chloride	15.0
Agar	10.0

Preparation: Dissolve the dehydrated components in 1000 mL of water by boiling. Autoclave at 121°C for 15 min. Cool to 50°C.

4.3.2.2 Supplement per Litre of medium

	mg
Cycloheximide	400.0
Colistin sulphate	20.0
Acriflavin	5.0
Cefotetan	2.0
Fosfomycin	10.0

Preparation: Dissolve the components in 10 mL of a 1:1 solution of ethanol:water. Sterilise by filtration through a 0.22 µm pore size filter (5.1.13).

4.3.2.3 Preparation of the complete medium

Aseptically add the supplement to the basal medium and mix thoroughly. Adjust the pH so that, after sterilisation and addition of supplement, it is 7.0 $\pm$ 0.2 at 25°C. Transfer the complete medium in quantities of about 15 mL to sterile Petri dishes and allow to solidify on a level surface (5.1.14). The complete medium may be stored for up to one week at 2 - 5°C before use.

4.3.3 Tryptone Soya Yeast Extract Agar (TSYEA)

	g/L
Tryptone soy broth	30.0
Yeast extract	16.0
Agar	12.0 to 18.0

Preparation: Dissolve the dehydrated components in 1000 mL of water by boiling. Adjust the pH so that after sterilisation it is 7.3 $\pm$ 0.1 at 25°C. Autoclave at 121°C for 15 min and allow to cool to 50°C. Transfer quantities of about 15 mL to sterile Petri dishes and allow to solidify on a level surface (5.1.14).

4.3.4 Tryptone Soya Yeast Extract Broth (TSYEB)

	g/L
Tryptone soy broth	30.0
Yeast extract	16.0

Preparation: Dissolve the dehydrated components in 1000 mL of water by heating gently. Adjust the pH so that after sterilisation it is 7.3 ± 0.1 at 25°C. Transfer the TSYEB in quantities of about 10 mL to tubes or bottles. Sterilise for 15 min at 121°C.

4.3.5 Blood agar (not required if microwell haemolysis test used)

4.3.5.1 Base

	g/L
Protease peptone	15.0
Liver digest	2.5
Yeast extract	5.0
Sodium chloride	5.0
Agar	12.0

Preparation: Dissolve the dehydrated components in 1000 mL of water by boiling. Sterilise the blood agar base for 15 min at 121°C. Cool the medium to 47 ± 1°C.

4.3.5.2 Supplement per Litre of medium

	mL
Washed sheep red blood cells	70.0

Preparation: Centrifuge defibrinated sheep blood at 900 x g for 30 min, aseptically removing the supernatant liquid and re suspend the pellet in sterile 0.85% saline solution to the original volume. If the centrifuged suspension has haemolysed, a fresh suspension must be prepared.

4.3.5.3 Preparation of the complete medium

Add the washed sheep red blood cells to the sterilised agar base and mix well. Adjust the pH so that, after sterilisation and addition of supplement, it is 7.0 ± 0.1 at 25°C. Transfer the medium in quantities of about 15 mL to sterile Petri dishes and allow to solidify on a level surface (5.1.14).

4.3.6 Brain heart infusion broth

	g/L
Calf brain infusion solids	12.5
Beef heart infusion solids	5.0
Protease peptone	10.0
Dextrose	2.0
Sodium chloride	5.0
Disodium phosphate	2.0

Preparation: Dissolve the dehydrated components in 1000 mL water by heating gently. Adjust the pH so that after autoclaving it is 7.4 ± 0.2 at 25°C. Distribute in 10 mL volumes in screw-capped containers and autoclave at 121°C for 15 min.

4.3.7 Phosphate Buffered Saline (PBS)

	g/L
Sodium chloride	8.0
Potassium chloride	0.2
Disodium hydrogen phosphate	1.15
Potassium dihydrogen phosphate	0.2

Preparation: Dissolve the dehydrated components in 1000 mL water by heating gently. Adjust the pH so that after autoclaving it is 7.3 ± 0.1 at 25°C. Dispense in 10 mL volumes in screw-capped containers and autoclave at 115°C for 10 min.

4.3.8 Carbohydrate utilisation broth

4.3.8.1 Base

	g/L
Protease peptone	10.0
Beef extract	1.0
Sodium chloride	5.0
Bromocresol purple	0.02

Preparation: Dissolve the dehydrated components in 1000 mL water by gentle heating. Distribute into tubes or bottles in quantities of 10 mL. Sterilise at 121°C for 15 min. Adjust the pH so that after sterilising it is 6.8 $\pm$ 0.2 at 25°C.

4.3.8.2 Carbohydrates

	g/L
Rhamnose	50.0
Xylose	50.0

Preparation: Dissolve each carbohydrate separately in 1000 mL water, do not heat to dissolve. Sterilise by filtration through a 0.22 µm pore size filter (5.1.13).

4.3.8.3 Preparation of the complete medium

For each carbohydrate, aseptically add 1 mL carbohydrate solution (4.3.8.2) to each tube or bottle of basal medium (4.3.8.1).

4.3.9 Motility medium

	g/L
Casein peptone	20.0
Meat peptone	6.1
Agar	3.5

Preparation: Dissolve the dehydrated components in 1000 mL water by boiling. Adjust the pH so that after sterilisation it is 7.3 $\pm$ 0.2 at 25°C. Dispense in tubes or bottles in quantities of about 10 mL. Sterilise for 15 min at 121°C.

4.3.10 CAMP (Christie/Atkins/Munch-Peterson) test agar

Very thin-layered sheep blood agar plates are required for this test.

4.3.10.1 Base

	g/L
Protease peptone	15.0
Liver digest	2.5
Yeast extract	5.0
Sodium chloride	5.0
Agar	12.0

Preparation: Dissolve the dehydrated components in 1000 mL of water by boiling. Adjust the pH so that after sterilisation it is 7.0 $\pm$ 0.1 at 25°C. Sterilise the blood agar base for 15 min at 121°C. Cool the medium to 50°C and transfer quantities of about 15 mL to sterile Petri dishes and allow to solidify on a level surface (5.1.14).

4.3.10.2 Sheep blood medium

	mL
Basal medium (4.3.10.1)	100.0
Washed sheep red blood cells (see 4.3.5.2)	7.0

Preparation: Add the washed cell suspension to the sterilised, molten base cooled to 47 ± 1°C.

4.3.10.3 Preparation of the complete medium

Pour a very thin layer of sheep blood medium (4.3.10.2) over the basal medium (4.3.10.1) using no greater than 3 mL per plate. Allow to solidify in an even layer. If the blood is added to dishes containing the basal medium which have been prepared in advance, it may be necessary to warm the dishes by placing them in an incubator at 37°C for 20 min before pouring the blood layer. Dry plates before use.

4.3.10.4 CAMP reaction cultures

A weakly β-haemolytic strain of *Staphylococcus aureus* (eg NCTC 1803) and a strain of *Rhodococcus equi* (eg NCTC 1621) are required to undertake the CAMP test. Not all strains of *Staphylococcus aureus* are suitable for the CAMP test.

Maintain stock cultures of *S. aureus*, *R. equi*, *L. monocytogenes*, *L.innocua* and *L. ivanovii* by inoculating TSYEA plates (4.3.3), incubating at 37°C for 24 - 48h, or until growth has occurred and storing at 4°C. Sub-culture at least once per month.

5. APPARATUS & GLASSWARE

Usual microbiological laboratory equipment, and in particular:

5.1 Apparatus

5.1.1 Apparatus for dry sterilisation (oven) or wet sterilisation (autoclave)

Apparatus that will enter into contact with the culture media, the dilution fluid or the sample, except for apparatus that is supplied sterile (particularly plastic apparatus), shall be sterilised either

by being kept at 170 to 175°C for not less than 1 h in an oven or

by being kept in contact with saturated steam at 121°C for not less than 15 min in an autoclave.

An autoclave is also necessary for the sterilisation of culture media and reagents. It shall be capable of being maintained at 121°C.

5.1.2 Incubator: capable of being maintained at 30°C ± 1°C.

5.1.3 Incubator: capable of being maintained at 37°C ± 1°C.

5.1.4 Incubator: capable of being maintained at 25°C ± 1°C.

5.1.5 Waterbath: capable of being maintained at 47°C ± 1°C.

5.1.6 Blending equipment

One of the following shall be used:

a) a rotary blender, operating at a rotational frequency between 8000 and 45000 min^{-1}, with glass or metal bowls fitted with lids, resistant to the conditions of sterilisation.
b) a peristaltic type blender (Stomacher Model 400), with sterile plastic bags.

Note - The bowls or plastic bags should have sufficient capacity to allow the sample to be properly mixed with the appropriate amount of diluent. In general, the volume of the container should be equal to about twice the volume of the sample plus diluent.

5.1.7 Loops: of platinum-iridium, nickel-chromium or plastic of diameter approximately 3 mm.

5.1.8 Inoculating needle: of platinum-iridium, nickel-chromium or plastic.

5.1.9 pH-meter: (for measuring the pH of prepared media and reagents), having an accuracy of calibration of 0.1 pH unit at 25°C.

5.1.10 Refrigerator: (for storage of prepared media and reagents), capable of being maintained at 2 to 5°C.

5.1.11 Sterile round-bottom microtitre plates, for microwell haemolysis.

5.1.12 Automatic pipette, capable of dispensing 100 µL volumes.

5.1.13 Filters, of 0.22 µm pore size suitable for the filtration of aqueous solutions and organic solvents.

5.1.14 Level surface, for drying agar plates.

5.2 Glassware: The glassware shall be resistant to repeated sterilisation.

5.2.1 Culture bottles or flasks, for sterilisation and storage of culture media and incubation of liquid media.

5.2.2 Test tubes, of dimensions approximately 16 mm x 125 mm fitted with lids.

5.2.3 Screw-capped bottles of approximately 25 mL capacity.

5.2.4 Flasks or bottles, of capacity 250 mL.

5.2.5 Measuring cylinders, for preparation of the complete media.

5.2.6 Graduated pipettes, of nominal capacity 1 mL graduated in divisions of 0.1 mL.

5.2.7 Sterile Petri dishes, of glass or plastic of diameter 90 to 100 mm.
5.2.8 Spreaders: of glass or plastic.
5.2.9 Microscope slides/coverslips

6. PROCEDURE

See the diagram of procedure in Appendix I.

6.1 Preparation of test sample, initial suspension and dilutions

Add 225 mL sample suspension medium (4.3.1) to 25 g test sample in a Stomacher bag or blender bowl. Blend for 2 min. Prepare dilutions from the initial suspension as necessary in the diluent (4.2).

6.2 Inoculation

Transfer by means of a sterile pipette, 0.1 mL of the initial suspension (10^{-1} dilution) to each of two selective agar plates (4.3.2). Repeat the procedure for 10^{-2} dilution and further dilutions as necessary. Carefully spread the inoculum as quickly as possible over the surface of the agar plate, trying not to touch the sides of the dish, using a glass or plastic spreader (5.2.8). Use a sterile spreader for each plate. Retain the plates at room temperature, on a level surface (5.1.14) for about 15 min with the lids uppermost to allow the inoculum to soak into the agar.

6.3 Incubation

Invert the plates prepared according to 6.2 and incubate them at 30 $\pm$ 1°C for 48 h.

6.4 Counting and selection of colonies

Select dishes at two consecutive dilutions containing less than 150 typical colonies, that is colonies surrounded by a dark brown or black halo. Count these suspect colonies.

6.5 Confirmation

6.5.1 Selection of colonies for confirmation

From each plate containing less than 150 typical colonies (6.4) select five typical or suspect colonies or, if there are fewer than five such colonies, select all for confirmation.

6.5.2 Subculturing

Streak the selected colonies onto the surface of TSYEA plates (4.3.3) in a manner which will allow well isolated colonies to develop. Incubate the plates at 30°C for 24 h or until growth is satisfactory.

6.6 Confirmation

6.6.1 Catalase reaction

From each TSYEA plate (6.5.2) pick a typical colony and place it on a coverslip (5.2.9). Add a drop of 3% hydrogen peroxide solution to a microscope slide (5.2.9). Invert the coverslip and place onto the slide. This technique is used to prevent aerosol formation. All *Listeria* spp. are catalase positive demonstrated by the formation of gas bubbles.

6.6.2 Morphology and staining properties

Test for Gram reaction. From each TSYEA plate (6.5.2) pick a typical colony and prepare a heat-fixed mount on a microscope slide (5.2.9). Gram stain and examine under oil immersion on a light microscope. All *Listeria* spp. are Gram-positive short rods.

6.6.3 Motility at 25°C

From each TSYEA plate (6.5.2) select a well isolated typical colony and remove by means of a sterile inoculating needle (5.1.8). For each isolate stab inoculate a tube of motility medium (4.3.9) and incubate at 25°C for 48 h. Examine for growth around the stab, if negative reincubate for a further 5 days. *Listeria* spp. are motile giving a typical umbrelliform growth pattern.

6.6.4 Haemolysis (see also 6.6.5)

If the morphological and physiological characteristics, and catalase reaction indicate the possibility of *Listeria* spp., inoculate blood agar plates (4.3.5) to determine the haemolytic reaction.

Dry the agar surface well before use. Select a typical colony from each TSYEA plate (6.5.2) and streak the colony onto the blood agar by means of a loop (5.1.7). Use one plate per isolate. Simultaneously inoculate blood agar plates (4.3.5) with positive and negative control cultures (*L. monocytogenes*, *L. ivanovii* and *L. innocua*).

After 48h incubation at 37°C, examine the test strains and controls. *L. monocytogenes* shows narrow, slight zones of clearing (β-haemolysis): *L.innocua* should show no clear zone. *L. ivanovii* usually shows wide, clearly delineated zones of β-haemolysis. Remove the colony to examine the haemolysis underneath the colony. Hold plates up to a bright light to compare test cultures with controls.

6.6.5 Haemolysis using microwell technique

As an alternative to the preparation of blood agar plates for the determination of haemolytic activity, a microwell method may be used.

From each TSYEA plate (6.5.2) select a well isolated typical colony and remove by means of a sterile loop (5.1.7). For each isolate inoculate a Brain Heart Infusion Broth (4.3.6) and incubate at 37°C for 48 h.

Prepare a 2% sheep erythrocyte suspension by washing i.e. centrifuging and resuspending, sterile defibrintated sheep blood three times in PBS (4.3.7). From this suspension pipette 100 µL in duplicate into wells of a round bottom microtitre plate (5.1.11).

To the erythrocyte suspension add 100 µL of Brain Heart Infusion broth culture. Incubate the microtitre plate for 45 min at 37°C followed by incubation for 2 h at 4°C. The presence of haemolysins are shown by a homogeneous red liquid. A clear supernatant with a layer of red blood cells on the bottom of the well indicate no haemolytic activity. Reference strains of *L. monocytogenes* and *L. innocua* should be run concurrently with this test.

6.6.6 Further biochemical confirmation

For these assays a culture in TSYEB (4.3.4) corresponding to the typical colony used for the haemolysis reaction (6.6.4 or 6.6.5) is required. Pick a typical colony from each TSYEA plate (6.5.2) and suspend in a tube containing TSYEB (4.3.4). Incubate for 24 h at 37°C.

6.6.6.1 Carbohydrate utilisation

Inoculate the carbohydrate fermentation broths (4.3.8) each with one loopful of the TSYEB culture (6.6.6). Incubate for up to 7 days at 37°C, although positive reactions (acid formation indicated by a yellow colour) occur mostly within 24 - 48 h. Reference strains of *L. monocytogenes, L. ivanovii* and *L. innocua* should be run concurrently with this test.

6.6.6.2 CAMP test

Streak the *S. aureus* and *R. equi* cultures in single lines across the blood agar plate (4.3.10) so that the two cultures are parallel and diametrically opposite. A thin, even inoculum is required. This can be obtained by using an inoculating needle (5.1.8) or a loop (5.1.7) held at right angles to the agar. Streak the test strain in a similar fashion at right angles to these cultures so that the test culture and reaction cultures do not touch but at their closest are about 1 - 2 mm apart. Several test strains may be streaked on the same plate.

Simultaneously, streak control cultures of *L. monocytogenes*, L. *innocua* and *L. ivanovii*. Incubate the plates at 37°C for 18 - 24 h.

Positive reactions are indicated by an enhanced zone of β-haemolysis at the intersection of the test strain with either the *S. aureus* or *R. equi* culture. However, the appearance of positive results varies with the reaction culture. A positive reaction with *R. equi* is seen as a wide (5 - 10 mm) 'arrow-head' of haemolysis. Small (about 1 mm) zones of weak haemolysis around the intersection of the test and *R. equi* cultures are negative reactions. A positive reaction with *S. aureus* is seen as a small rounded zone of enhanced haemolysis extending only about 2 mm from the test strain and within the weakly haemolytic zone due to growth of the *S. aureus* culture. Large zones of haemolysis around the *S. aureus* culture do not occur.

L. monocytogenes and *L. seeligeri* show a positive CAMP reaction with *S. aureus* but not *R. equi. L. ivanovii* reacts with *R. equi* but not with *S. aureus*. The other *Listeria* spp. show negative CAMP reactions with both *S. aureus* and *R. equi.*

6.7 Interpretation of morphological and physiological properties and biochemical reactions

All *Listeria* spp. are small Gram-positive rods (only with 24 h old cultures) that demonstrate an umbrelliform growth pattern in the motility medium. They are catalase positive. *L. monocytogenes* utilises rhamnose but not xylose.

L. monocytogenes, L. ivanovii and *L. seeligeri* (weak) produce β-haemolysis on blood agar plates and positive reactions in the microwell haemolysis test. Of the three haemolytic *Listeria* spp. only *L. monocytogenes* fails to utilise xylose and is positive for rhamnose utilisation.

L. monocytogenes and *L. seeligeri* show a positive CAMP reaction with *S. aureus* but not with *R. equi. L. ivanovii* reacts with *R. equi* but not with *S. aureus*. The other *Listeria* spp. show negative CAMP reactions with both reaction cultures.

7. CONTROL CULTURES

Control cultures of *L. monocytogenes, L. ivanovii* and *L. innocua* should be run concurrently with all confirmatory tests.

8. EXPRESSION OF RESULTS

8.1 General

If all of the selected typical colonies (6.4) confirm as *L. monocytogenes*, the number of organisms present will be the same as that given by the count in 6.4. In all other cases the number shall be calculated from the percentage of isolates confirmed positive in relation to the total number of selected colonies (6.4).

Round the result to a whole number of colonies.

8.2 Calculation of the weighted mean

Calculate the number, N, of *L. monocytogenes* per gram of product using the following equation:

$$N = \frac{\Sigma c}{(n_1 + 0.1n_2)0.1d}$$

where

Σc = the sum of confirmed colonies on all dishes retained
n_1 = the number of dishes retained at the first dilution
n_2 = the number of dishes retained at the second dilution

d = the dilution factor corresponding to the first dilution

Round the result calculated to two significant figures.

Take as the result the number of micro organisms per gram of product, expressed as a number between 1.0 and 9.9 multiplied by 10^x, where x is the appropriate power of 10.

8.3 Estimation of small numbers

If the two dishes corresponding to the initial suspension contain less than 15 colonies, calculate the arithmetic mean *m* of the colonies counted on both dishes.

Report the result as follows:

estimated number N_E of *L. monocytogenes* per gram:

$N_E = m \times d^{-1}$ where d is the dilution factor of the initial suspension.

8.4 No characteristic colonies

If the two dishes corresponding to the initial suspension contain no characteristic colonies report the result as follows:

less than $1 \times d^{-1}$ *L. monocytogenes* per gram, where d is the dilution factor of the initial suspension.

9. VALIDATION

The procedure as described in this protocol has been used in an on-going proficiency test exercise organised by the Ministry of Agriculture, Fisheries and Food, CSL Food Science Laboratory, Norwich (10.5). A summary of results obtained is given in Appendix II. Test materials were distributed on four occasions. These comprised freeze-dried minced beef test materials, artificially inoculated with the target organism and a simulated autochthonous flora; in order to simulate, as closely as possible, a natural foodstuff. On each occasion analysts received duplicate blind test materials and were asked to use the method prescribed in this protocol.

All test materials used in the proficiency testing exercise were assessed for homogeneity (Appendix II) using the recommended procedures described in the ISO/IUPAC/AOAC International Harmonised Protocol for Proficiency Testing of (Chemical) Analytical Laboratories (10.6). Homogeneity was assessed immediately following preparation of the test materials (day 0) and again on the date of examination by laboratories (test day). For distribution rounds 1 to 3, test day was 12 days after preparation of the test materials. For round 4, test day was 19 days after preparation.

Statistical analyses of the results for any one test material are as described in the ISO/IUPAC/AOAC International Protocol for the Design, Conduct and Interpretation of Collaborative Studies (10.4).

10. REFERENCES

10.1 McLain, D., Lee, W., FS&S method for the isolation and identification of *Listeria monocytogenes* from processed meat and poultry products. USDA Laboratory communication, NU57, Washington, D.C..

10.2 Anon. 1993. British Standard Methods for Microbiological Examination for Dairy Purposes. Part 3 Methods for detection and/ or enumeration of specific groups of microorganisms. Section 3.15 Detection of *Listeria monocytogenes*. BS 4285: Section 3.15: 1993. ISO 10560: 1993.

10.3 Anon. 1996. British Standard Methods for Microbiological Examination of Food and Animal Feeding Stuffs. Part 0 General laboratory practices (ISO title - General guidance for microbiological examinations). BS 5763: Part 0: 1996. ISO 7218-1996.

10.4 Horwitz, W. 1988. Protocol for the design, conduct and interpretation of collaborative studies. **Pure & Applied Chemistry, 60**, 6, 855-864.

10.5 The QAS Secretariat, CSL Food Science Laboratory, Norwich Research Park, Colney, Norwich NR4 7UQ.

10.6 Thompson, M. and Wood, R. 1993. International harmonised protocol for proficiency testing of (chemical) analytical laboratories. **Journal of the Association of Analytical Chemists International, 76**, 4, 926-940.

10.7 Scotter S. L., Wood R., Williams A.P. 1993. Methods for the detection of *Listeria monocytogenes* in foods, J. Assoc. Publ. Analysts, **29,** 221-251.

APPENDIX I: METHOD FOR THE ENUMERATION OF *L. MONOCYTOGENES*

25 g test portion + 225 mL UVM I

Homogenise

Decimal dilutions as required

Enumerate by direct plating

Incubate for 48 h at 30°C

Count typical colonies on retained dishes of less than 150 colony forming units

Confirm 5 typical colonies from each plate retained

Calculate the weighted mean from numbers of confirmed *L. monocytogenes*

APPENDIX II: STATISTICAL ANALYSIS OF THE RESULTS FROM AN ON-GOING PROFICIENCY TESTING SCHEME

A.1 Matrix

All test materials were prepared on a minced beef matrix

A.2 Laboratories

Data points from participating laboratories were used in statistical analyses after the removal of aberrant results

A.3 Statistical outliers

Entries among tables derived from the original test results that deviate so much from comparable entries that they are considered to be irreconcilable with other data (10.4).

A.4 Assigned value

The robust mean calculated from data returned by all participants who carried out the method as prescribed

A.5 $\mathbf{S_r}$ The standard deviation of the repeatability

A.6 **r** repeatability (within laboratory variation) - the value below which the absolute difference between two single test results obtained with the same method on an identical test material under the same conditions may be expected to lie within a 95% probability

A.7 $\mathbf{S_R}$ the standard deviation of the reproducibility

A.8 **R** reproducibility (between laboratory variation) - the value below which the absolute difference between two single test results obtained with the same method on an identical test material under the same conditions may be expected to lie within a 95% probability

Table 1: Precision characteristics ($\mathbf{Log_{10}}$ colony forming units per gram) of the method derived from the results of an on-going proficiency testing scheme

Date of Testing	No. of laboratories	Homogeneity	No. of Statistical outliers	Assigned value	S_r	r	S_R	R
May 1994	12	satisfactory	0	4.70	0.08	0.22	0.15	0.42
November 1994	15	satisfactory	0	3.87	0.08	0.22	0.15	0.42
November 1995	18	satisfactory	2	5.19	0.09	0.24	0.16	0.44
November 1996	20	satisfactory	1	4.46	0.09	0.24	0.23	0.64
January 1997	20	satisfactory	1	3.53	0.07	0.19	0.11	0.31