Eurachem

A focus for analytical chemistry in Europe

The Fitness for Purpose of Analytical Methods

A Laboratory Guide to Method Validation and Related Topics

Second Edition 2014

Eurachem

A focus for analytical chemistry in Europe

Eurachem Guide

The Fitness for Purpose of Analytical Methods

A Laboratory Guide to Method Validation and Related Topics

Second edition

Acknowledgements

This document has been produced by members of the Eurachem Method Validation Working Group and others co-opted for this task. Those who have contributed to this edition are listed below.

Project group

Vicki Barwick	LGC (UK)
Pedro P. Morillas Bravo	Canal de Isabel II Gestión (ES)
Stephen L. R. Ellison	LGC (UK)
Joakim Engman	National Food Agency (SE)
Elin L. F. Gjengedal	Norwegian University of Life Sciences (NO)
Ulla Oxenbøll Lund	Eurofins Miljø A/S (DK)
Bertil Magnusson (editor)	SP Technical Research Institute of Sweden (SE)
Hans-Thomas Müller	Mersin (TR)
Marina Patriarca	Istituto Superiore di Sanità (IT)
Barbara Pohl	Merck KGaA (DE)
Piotr Robouch	European Commission (EU)
Lorens P. Sibbesen (chairman)	Labquality International (DK)
Elvar Theodorsson	University Hospital in Linköping (SE)
Florent Vanstapel	University Hospital Leuven, Leuven (BE)
Isabelle Vercruysse	BELAB (BE)
Aysun Yilmaz	Cevre Food and Industrial Analysis Laboratory (TR)
Perihan Yolci Ömeroglu	Okan University (TR)
Ulf Örnemark (editor)	Emendo Dokumentgranskning (SE)

Recommended citation

This publication should be cited* as: "B. Magnusson and U. Örnemark (eds.) Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, (2nd ed. 2014). ISBN 978-91-87461-59-0. Available from www.eurachem.org."

**Subject to journal requirements*

Contents

Foreword to the second edition

Since the first edition of this Guide in 1998, a number of important developments in analytical quality have taken place. Firstly, the ISO 9000 series of standards, which is widely used to provide a basis for a quality management system, has been revised. Its philosophy forms an integral part of international conformity assessment standards and guides, which underpins competence requirements for laboratories, proficiency testing (PT) providers and reference material (RM) producers. These documents all stress the importance of using validated methods.

Secondly, several general or sector-specific guides on method validation have been revised or developed. EU legislation contains mandatory requirements for analytical measurements in many sectors.

Thirdly, much effort has been invested by the analytical community in implementing the uncertainty concept. For example, in its Harmonized guidelines for single-laboratory validation of methods of analysis (2002) IUPAC predicted that, "...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...". In the following years, accreditation bodies issued policies and guidance documents clearly recognising the use of method validation data in the measurement uncertainty estimation process.

Furthermore, the International vocabulary of metrology – Basic and general concepts and associated terms (VIM) has been substantially revised, taking into account chemical and biological measurements. Although terminology related to method validation is far from harmonised, the situation has improved. VIM is also a normative document for laboratories accredited to, e.g. ISO/IEC 17025 and ISO 15189.

The second edition of this Guide aims to reflect changes in international standards and guidance documents and puts less emphasis on terms and definitions. Instead the Guide refers to the VIM and other readily available sources. As a consequence, the list of terms and definitions has been omitted from the Annex. Literature cited in this edition of this Guide are listed in the Bibliography at the end. Additional sources and literature related to method development and validation is available as a 'Reading list' under the menu item 'Publications' on the Eurachem website at www.eurachem.org. Annex A is revised as a consequence of changes to ISO 78-2. This edition has also been extended to include information on the statistical basis of limit of detection calculations (Annex B), analysis of variance (Annex C) and qualitative analysis (Annex D).

It is becoming increasingly common among routine laboratories, especially in the clinical sector, to use commercially available measuring systems. This means that the responsibility for validation mainly lies with the manufacturer. The laboratory's work will focus on verifying the manufacturer's published performance data and demonstrate that the method works on the end-user's premises.

However, looking back to the foreword to the first edition, we conclude that the six principles stated there are still relevant, and are consistent with the requirements of international standards such as ISO/IEC 17025.

Foreword to the first edition*

An initiative in the UK to promote good practice in analytical measurement has identified six principles of analytical practice which, taken together, are considered to constitute best practice. The six principles which are described in more detail in a separate guide† are:

1. "Analytical measurements should be made to satisfy an agreed requirement." (i.e. to a defined objective).
2. "Analytical measurements should be made using methods and equipment which have been tested to ensure they are fit for purpose."
3. "Staff making analytical measurements should be both qualified and competent to undertake the task." (and demonstrate that they can perform the analysis properly).
4. "There should be a regular independent assessment of the technical performance of a laboratory."
5. "Analytical measurements made in one location should be consistent with those made elsewhere."
6. "Organisations making analytical measurements should have well defined quality control and quality assurance procedures."

These principles are equally relevant to laboratories whether they are working in isolation or producing results which need to be compared with those from other laboratories.

This document is principally intended to assist laboratories in implementing Principle 2, by giving guidance on the evaluation of testing methods to show that they are fit for purpose.

* The first edition (1998) of this Guide was developed by a Eurachem Working Group from a draft originally produced by LGC. The following persons were members of the Eurachem group at that time:
D. Holcombe, P. De Bièvre, D. Böttger, C. Eastwood, J. Hlavay, M. Holmgren, W. Horwitz, M. Lauwaars, B. Lundgren, L. Massart, J. Miller, J. Morkowski, B. te Nijenhuis, B. Nyeland, R. Philipp, P. Radvila, J. Smeyers-Verbeke, R. Stephany, M. Suchanek, C. Vandervoorst, H. Verplaetse, H. Wallien, M. Walsh, W. Wegscheider, D. Westwood, H. J. van de Wiel.

† The manager's guide to VAM, UK Department of Trade and Industry, Valid Analytical Measurement Programme. Published as VAM Principles M. Sargent. Anal. Proc., 1995, 32, 201-202.

Abbreviations and symbols

The following abbreviations, acronyms and symbols occur in this Guide.

AMC	Analytical Methods Committee
ANOVA	Analysis of variance
AOAC International	a globally recognized standards developing organization
ASTM International	a globally recognized standards developing organization
BIPM	International Bureau of Weights and Measures
CCQM	Consultative Committee for Amount of Substance – Metrology in Chemistry
CEN	European Committee for Standardization
CITAC	Cooperation on International Traceability in Analytical Chemistry
CLSI	Clinical and Laboratory Standards Institute
CRM	certified reference material
EA	European co-operation for Accreditation
EC	European Commission
EPA	Environmental Protection Agency
EQA	external quality assessment
EU	European Union
GUM	Evaluation of measurement data – Guide to the expression of uncertainty in measurement
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
JCGM	Joint Committee for Guides in Metrology
LOD	limit of detection
LOQ	limit of quantification
NATA	National Association of Testing Authorities
QA	quality assurance
QC	quality control
RSC	Royal Society of Chemistry
SANCO	European Commission's Directorate-General for Health and Consumers
SOP	standard operating procedure
PT	proficiency testing
RM	reference material
RSD	relative standard deviation
UV/VIS	ultraviolet/visible
VIM	International vocabulary of metrology – Basic and general concepts and associated terms

b	absolute bias
$b(\%)$	relative bias in %
k_Q	multiplier used in calculating limit of quantification
m	number of measurements
n	number of replicate observations averaged when reporting results
n_b	number of blank observations averaged when calculating the blank correction
r	repeatability limit
R	reproducibility limit
$R(\%)$	relative recovery (apparent recovery) in per cent
$R'(\%)$	relative spike recovery in per cent
s	standard deviation
s_0	estimated standard deviation of single results at or near zero concentration
s_0'	standard deviation used for calculating an LOD or LOQ
s_I	intermediate precision standard deviation
s_r	repeatability standard deviation
s_R	reproducibility standard deviation
u	standard uncertainty
$\overline{x}$	mean value (arithmetic average)
x_{ref}	reference value
$\bar{x}_{\text{ref}}$	mean value of measurements with an alternative method, e.g. a reference method
$\bar{x}'$	mean value of spiked sample in a recovery experiment
x_{spike}	added concentration in a recovery experiment

1 Introduction

1.1 Rationale and scope for this Guide

Method validation is an important requirement in the practice of chemical analysis. Most analytical chemists are aware of its importance, but why it should be done and when, and exactly what needs to be done, is not always clear to them. Some analysts used to see method validation as something that can only be done in collaboration with other laboratories and therefore refrained from it. Requirements in standards such as ISO/IEC 17025 [1], ISO 15189 [2] and ISO 15195 [3] have helped in clarifying this. For example, the need to demonstrate that methods are fit for purpose is stressed in Clause 5.4.2 of ISO/IEC 17025:

"*The laboratory shall use test and/or calibration methods, including methods for sampling, which meet the needs of the customer and which are appropriate for the tests and/or calibrations it undertakes...*" and further: "*When the customer does not specify the method to be used, the laboratory shall select appropriate methods...*".

The purpose of this Guide is to discuss the issues related to method validation and increase readers' understanding of what is involved, why it is important, and give some idea of how it can be accomplished.

The Guide is expected to be of most use to a) laboratory managers responsible for ensuring that the methods under their supervision are adequately validated and b) analysts responsible for planning and carrying out studies on methods for validation purposes. Other staff may find the guidance of use as a source of background information – senior staff from a management point of view and junior staff from a technical or educational point of view.

The Guide focuses on single-laboratory validation. It aims to direct the reader towards established protocols where these exist and where they do not, give a simple introduction to the processes involved in validation and provide some basic ideas to enable the reader to design their own validation strategies. It includes references to further material on particular technical aspects of validation.

This Guide is aimed at the validation of quantitative methods. However, some of the principles described here are also relevant for qualitative methods for determining the presence of one or more analytes, e.g. the concepts of selectivity and limit of detection (LOD).

The Guide avoids emphasis on the use of statistics although undoubtedly those with a working knowledge of elementary statistics will find the method validation process easier to understand and implement. Several references are made to publications on basic statistics for chemists [4, 5, 6].

The analyst's understanding of method validation is inhibited by the fact that many of the metrological and technical terms used to describe processes for evaluating methods vary in different sectors of analytical measurement, both in their meaning and the way they are determined. This Guide cannot say where a term is used correctly or incorrectly although it is intended to provide some clarification. The best advice when using a term that may be misunderstood, is to state the source and which convention has been used.

It is implicit in the method validation process that the studies to determine method performance characteristics[*] are carried out using equipment that is within specification, working correctly, and adequately calibrated. Therefore, this Guide does not cover specifically the concepts of 'equipment qualification' or 'instrument qualification'. Likewise the analyst carrying out the studies must be competent in the field of work under study, and have sufficient knowledge related to the work to be able to make appropriate decisions from the observations made as the study progresses.

1.2 Notes on the use of this Guide

1.2.1 Terminology

In the revision of this Guide the main focus has been on updating the terminology and literature references to reflect developments since the Guide was first published fifteen years ago. With regards to terminology we have, where possible, followed the 3rd edition of the VIM first published in 2007 [7, 8]. This has been supplemented, where necessary, with

[*] Commonly used synonyms for method performance characteristics are 'method performance parameters', 'metrological characteristics' and 'performance properties'.

terminology used in ISO/IEC 17025:2005 [1], other ISO documents [9, 10, 11] and the IUPAC Harmonized Guidelines for Single-Laboratory Validation from 2002 [12] to reflect terms commonly used in analytical laboratories.

In some cases it may be difficult to decide which term to use when several similar terms are in use. For clarity it has been considered important to use a term consistently throughout the Guide. One example is the term used to describe the document that gives a detailed description of the method to be validated using personnel and equipment in a particular laboratory. For quantitative analysis VIM refers to the *measurement procedure*, in ISO/IEC 17025 this is the *method*, in ISO 15189 [2] it is the *examination procedure* and many laboratories refer to their *standard operating procedure* (*SOP*). The working group has decided to adhere to ISO/IEC 17025 and use the generic term *method*. Consequently, this Guide uses the commonly recognised term 'method validation' although 'procedure validation' would be more correct.

The terms 'ruggedness' and 'selectivity' are preferred to 'robustness' and 'specificity' [13] since the former are used by IUPAC [12].

Various terms, e.g. 'calibration', 'measurement', 'testing', 'analysis' and 'examination' are used to describe laboratory work. This Guide uses 'analysis' in a general sense and specifies, where necessary, the circumstances. Similarly, this Guide often refers to a measured concentration although several other quantities are regularly investigated in the chemistry laboratory [14].

In the processes of sampling, sample preparation and analysis terms such as 'sampling target', 'primary sample', 'increment', 'composite sample', 'subsample', 'laboratory sample', 'test sample', 'test portion' and 'test solution' may be used [15, 16]. In this Guide we normally use the general term 'sample' or 'test sample' [17].[*] The most important terms used in the Guide are defined in the text. Definitions in VIM, ISO 9000 [9] and IUPAC [17, 18] have been provided wherever possible. The terms in VIM related to analytical chemistry are further explained in the Eurachem Guide "Terminology in analytical measurement" [8]. Users should note that there is still no universal agreement on the definition of some of the terms used in method validation.

1.2.2 Quick References

In Section 6, the shaded boxes provide '*Quick Reference*' advice related to the specific performance characteristic of a method. However, it is recognised that in many cases laboratories will not have the time and resources to carry out experiments in the detail described here. Carrying out the operations described in the boxes, using less replication than suggested, will still yield useful information and is certainly better than no work at all. However, the information provided will be less reliable than if full replication had been utilised.

[*] Test sample: Sample, prepared from the laboratory sample, from which test portions are removed for testing or for analysis [17].

2 What is method validation?

2.1 Definitions

Definitions of ***validation*** from three international documents are given in Table 1. ***Method validation*** is basically the process of defining an analytical requirement, and confirming that the method under consideration has capabilities consistent with what the application requires. Inherent in this is the need to evaluate the method's performance. The judgement of method suitability is important; in the past method validation tended to concentrate only on evaluating the performance characteristics.

Method validation is usually considered to be very closely tied to method development. Many of the method performance characteristics (Table 2) that are associated with method validation are usually evaluated, at least approximately, as part of method development. However, it is important to remember that formal validation of the final version of the method (the documented procedure) should be carried out.

Some sectors use the concepts of 'primary validation' and 'secondary validation', the latter in the sense of verification [19]. The concepts 'qualification' and 'metrological confirmation' [20] also seem to cover verification (Table 1).

2.2 What is the difference between validation and verification?

ISO 9000 [9] defines verification as "confirmation, through provision of objective evidence, that specified requirements have been fulfilled". This is very similar to the definition of validation in Table 1. The VIM [7] states that verification is "provision of objective evidence that a given item fulfils specified requirements" and that validation is a "verification, where the specified requirements are adequate for an intended use".

A laboratory may adopt a validated procedure which, e.g. has been published as a standard, or buy a complete measuring system to be used for a specific application from a commercial manufacturer. In both these cases, basic validation work has already been carried out but the laboratory will still need to confirm its ability to apply the method. This is **verification**. It means that some experimental work must be done to demonstrate that the method works in the end-user's laboratory. However, the workload is likely to be considerably less compared to validation of a method that has been developed in-house.

The terms validation and verification are further discussed in the Eurachem Guide on terminology in analytical measurement [8].

Table 1 – Definitions of the concept 'validation' in ISO 9000, ISO/IEC 17025 and VIM

Definition	Reference
confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled	ISO 9000 [9][a]
confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled	ISO/IEC 17025 [1]
verification, where the specified requirements are adequate for an intended use	VIM [7][b]

[a] ISO 9000 defines 'qualification process' as "process to demonstrate the ability to fulfil specified requirements".

[b] VIM defines 'verification' as "provision of objective evidence that a given item fulfils specified requirements"

Table 2 – Overview of performance characteristics commonly evaluated during method validation

Performance characteristic
Selectivity
Limit of detection (LOD) and limit of quantification (LOQ)
Working range
Analytical sensitivity
Trueness • bias, recovery
Precision • repeatability, intermediate precision and reproducibility
Measurement uncertainty[a]
Ruggedness (robustness)
[a] Strictly, measurement uncertainty is not a performance characteristic of a particular measurement procedure but a property of the results obtained using that measurement procedure.

3 Why is method validation necessary?

3.1 Importance of analytical measurement

Millions of tests, measurements and examinations are made every day in thousands of laboratories around the world. There are innumerable reasons underpinning them, for example: as a way of valuing goods for trade purposes; supporting healthcare; checking the quality of drinking water, food and feed; analysing the elemental composition of an alloy to confirm its suitability for use in aircraft construction; forensic analysis of body fluids in criminal investigations. Virtually every aspect of society is supported in some way by analytical work.

The cost of carrying out these measurements is high and additional costs may arise from decisions made on the basis of the results. For example, tests showing food to be unfit for consumption may result in compensation claims; tests confirming the presence of banned drugs could result in fines, imprisonment or even, in some countries, execution. Clearly it is important to make a correct measurement and be able to show that the result is correct.

3.2 The professional duty of the analytical chemist

If the result of an analysis cannot be trusted then it has little value and the analysis might as well have not been carried out. When customers commission analytical work from a laboratory, it is assumed that the laboratory has a degree of expert knowledge that the customers do not have themselves. The customer expects to be able to trust results reported and usually only challenges them when a dispute arises. Thus the laboratory and its staff have an obvious responsibility to justify the customer's trust by providing the right answer to the analytical part of the problem, in other words results that have demonstrable 'fitness for purpose'. Implicit in this is that the tests carried out are appropriate for the analytical part of the problem that the customer wishes solved, and that the final report presents the analytical data in such a way that the customer can readily understand it and draw appropriate conclusions. Method validation enables chemists to demonstrate that a method is 'fit for purpose'.

For an analytical result to be fit for its intended use it must be sufficiently reliable that any decision based on it can be taken with confidence. Thus the method performance must be validated and the uncertainty on the result, at a given level of confidence, estimated. Uncertainty should be evaluated and quoted in a way that is widely recognised, internally consistent and easy to interpret [21]. Most of the information required to evaluate uncertainty can be obtained during validation of the method. This topic is dealt with briefly in Section 6.7 and in more detail in the Eurachem/CITAC Guide Quantifying Uncertainty in Analytical Measurement [22].

Regardless of how good a method is and how skilfully it is used, an analytical problem can be solved by the analysis of samples only if those samples are appropriate to the problem. Taking appropriate samples is a skilled job, requiring an understanding of the problem and its related chemistry. A laboratory should, wherever possible, offer advice to the customer on the taking of samples as part of its customer care. Clearly there will be occasions when the laboratory cannot themselves take or influence the taking of the samples. On these occasions results of analysis will need to be reported on the basis of the samples as received, and the report should make this distinction clear.

We have mostly (and rightly) focused on the overall objective of performing method validation, i.e. demonstrating that methods are 'fit for purpose'. However, it should be recognised that a method validation study gives additional benefits to the laboratory undertaking the validation. It provides a solid knowledge and experience of the practical details of performing the method, including awareness of any critical steps in the process. Validation gives the laboratory and its employees a greater confidence in their own results.

3.3 Method development

The validation work is preceded by a development phase which may involve different staff and which can take a number of forms.

At one extreme, it may involve adapting an existing method by making minor changes so that it is suitable for a new application. For example, a method required to determine toluene in water might be adapted from an established method for benzene in water. The matrix is the same, and the two analytes have broadly similar

properties. It is likely that the same principles of isolation, identification, and quantification that are applied to benzene can also be applied to toluene. If, on the other hand, a method is required to determine benzene in soil, adaptation of the benzene in water method may not be the best option. Adaptation of some other method for determining organics in soil may be a better starting point.

At the other extreme, the analytical chemist may start out with a few sketchy ideas and apply expertise and experience to devise a suitable method. This clearly involves a great deal more work and a degree of doubt as to whether the final method will be successful. It is not unusual for method development to involve work on a number of different ideas simultaneously before eventually choosing one winner.

Regardless of how much effort has been invested during method development, there is no guarantee the method will perform adequately during validation (or under routine conditions in a particular laboratory). When different staff are involved in the development and validation phase this offers the possibility of checking that the instructions (the measurement procedure) can be understood and implemented.

4 When should methods be validated or verified?

4.1 Method validation

A method should be validated when it is necessary to demonstrate that its performance characteristics are adequate for use for a particular purpose. For example, it is stated in Clause 5.4.5.2 of ISO/IEC 17025 [1] that the laboratory shall validate:

- non-standard methods;
- laboratory-designed/developed methods;
- standard methods used outside their intended scope;
- amplifications and modifications of standard methods.

Validation must be as extensive as necessary to meet the requirements in connection with the given use or the given application [23]. The extent ('scale', 'scope') of validation will depend on the application, the nature of the changes made, and the circumstances in which the method is going to be used.

Validation is also required when it is necessary to demonstrate the equivalence of results obtained by two methods, e.g. a newly developed method and an existing standard/regulatory method.

4.2 Method verification

For standard(ised) methods, such as those published by, e.g. ISO or ASTM, validation by the laboratory using the method is not necessary. However, the laboratory needs to verify the performance of the method as detailed in ISO/IEC 17025 Clause 5.4.2:

…The laboratory shall confirm that it can properly operate standard methods before introducing the tests or calibrations.

Verification is also required when there is an important change such as a new but similar instrument, relocation of equipment etc.

In laboratory medicine a majority of measurements and tests are performed with commercial procedures which have already been validated by the manufacturer, but which need to be verified by the end-user [24]. ISO 15189 [2] stresses that *examination procedures used without modification shall be subject to independent verification by the laboratory before being introduced into routine use.* This could also include when an instrument is updated with new software, or when quality control indicates that the performance of an established method is changing with time.

5 How should methods be validated?

5.1 Who carries out method validation?

5.1.1 Approaches to method validation

Once the initial method development is finished, the laboratory should document the measurement procedure in detail (see Annex A). It is this documented procedure that is taken forward for the formal validation.

There are two main approaches to method validation; the interlaboratory comparison approach and the single-laboratory approach. Regardless of the approach, it is the laboratory using a method which is responsible for ensuring that it is fit for the intended use and, if necessary, for carrying out further work to supplement existing validation data.

5.1.2 Interlaboratory approach

Much has been published in the literature concerning method validation by dedicated interlaboratory comparisons often referred to as 'collaborative studies' or 'cooperative studies'. There are a number of protocols relating to this type of validation [25, 26, 27, 28], as well as the ISO 5725 standards [29] which can be regarded as the most generally applicable. If a method is being developed which will have wide-ranging use, perhaps as a published standardised procedure, then a collaborative study involving a group of laboratories is probably the preferred way of carrying out the validation. A published method validated in this way is demonstrated to be robust. Published information normally contains precision (repeatability, reproducibility and/or corresponding precision limits) and, sometimes, bias estimates. Where a method has been validated by a standards approving organisation, such as ISO, CEN or AOAC International, the user will normally need only to verify published performance data and/or establish performance data for their own use of the method. This approach, therefore, reduces the workload for the laboratory using the method.

5.1.3 Single-laboratory approach

Laboratories will from time to time find that a method is needed but not available as a published standard. If the method is developed for use in one laboratory, for example because there is no general interest in the method or because other laboratories are competitors, the single-laboratory approach is appropriate [12].

Whether or not methods validated in a single laboratory will be acceptable for regulatory purposes depends on any guidelines covering the area of measurement concerned. It should normally be possible to get a clear policy statement from the appropriate regulatory body.

5.2 Extent of validation studies

The laboratory has to decide which performance characteristics (see Table 2 and Section 6) need to be investigated in order to validate the method and, in some cases, how detailed the investigation of a single performance characteristic should be. The IUPAC protocol [12] lists a number of situations, which takes into account, among other things, the status of the method and the competence of the laboratory.

Where the scope of the analytical work is well defined and applications are similar over time, it may be possible for an organisation or sector to issue general guidelines for the extent of validation studies. An example from the pharmaceutical sector is shown in Table 3.

Starting with a carefully considered analytical specification given in the scope of the documented procedure (see A.5 in Annex A) provides a good base on which to plan the validation process, but it is recognised that in practice this is not always possible. The assessment of method performance may be constrained. This is acknowledged in ISO/IEC 17025, clause 5.4.5.3 as *Validation is always a balance between costs, risks and technical possibilities*. The laboratory should do its best within the constraints imposed, taking into account customer and regulatory requirements, existing experience of the method, available tools (Section 5.4), and the need for metrological compatibility [7] with other similar methods already in use within the laboratory or used by other laboratories. Some performance characteristics may have been determined approximately during the method development or method implementation stage. Often a particular set of experiments will yield information on several performance characteristics, so with careful planning the effort required to get the necessary information can be minimised.

Table 3 – Extent of validation work for four types of analytical applications. Example from the pharmaceutical sector [13]. 'x' signifies a performance characteristic which is normally validated.

Performance characteristic	Type of analytical application			
	Identification test	Quantitative test for impurity	Limit test for impurity	Quantification of main component
Selectivity	x	x	x	x
Limit of detection			x	
Limit of quantification		x		
Working range including linearity		x		x
Trueness (bias)		x		x
Precision (repeatability and intermediate precision)		x		x
NOTE The table is simplified and has been adapted to the structure and terminology used in this Guide.				

The implications of the constraints discussed above are particularly critical where the method is not going to be used on a routine basis. The process of validating methods which are going to be used on a routine basis is comparatively well-defined. Clearly the same principles apply for ad hoc analysis as for routine testing. It is necessary to have an adequate level of confidence in the results produced. Establishing the balance between time and cost constraints and the need to validate the method is difficult. In some circumstances it may be more appropriate to subcontract the analyses to another laboratory where they can be performed on a routine basis.

5.3 Validation plan and report

The validation work shall be performed, and the results reported, according to a documented procedure.

The outline of a validation plan ('validation protocol') and validation report may be stated in sectoral guidelines (see Section 5.5). National accreditation bodies may point to minimum requirements for this documentation [23]. However, a simple template for a combined validation plan and validation report could, e.g. consist of the following sections.

- Title: This section should identify the method and when and who is performing the work. Brief information about the method scope and a short description of the method should be given, as well as details of the status of the method (e.g. an international standard, a method developed in-house etc.), the analyte, measurand, measurement unit, types of sample and the intended use. Sampling and subsampling can be part of the measurement procedure and must, in those cases, be validated. Even if these steps are performed elsewhere, it is useful to include information about them in the validation plan/report.
- Planning: This section should outline the purpose, e.g. full validation of a new method, verification of performance of a standardised method, extension to method scope, etc. The extent of the validation work should be indicated, i.e. the performance characteristics which will be investigated and any associated requirements.
- Performance characteristics: This section should give a brief explanation of the performance characteristic, repeat any specific requirements, outline the experiments which will be done and how the results are to be evaluated. Results and conclusions from the experiments should be stated. Separate sections are used for each performance characteristic.
- Summary: The last section should summarise the validation work and the results. Implications concerning routine use, and internal and external quality control, can be given. Most importantly, a concluding statement as to whether the method is fit for purpose shall be given. Note that this is a requirement in ISO/IEC 17025 [1].

5.4 Validation tools

5.4.1 Blanks

Use of various types of blanks enables assessment of how much of the measured signal is attributable to the analyte and how much to other causes. Various types of blank are available to the analyst:

- **Reagent blanks**[*]: Reagents used during the analytical process (including solvents used for extraction or dissolution) are analysed in order to determine whether they contribute to the measurement signal.
- **Sample blanks**. These are essentially sample matrices with no analyte present, e.g. a human urine sample without a specific drug of abuse, or a sample of meat without hormone residues. Sample blanks may be difficult to obtain but such materials are necessary to give a realistic estimate of interferences that would be encountered in the analysis of test samples.

5.4.2 Routine test samples

Routine test samples are useful because of the information they provide on precision, interferences etc. which could be realistically encountered in day-to-day work. If the analyte content of a test material is accurately known, it can be used to assess measurement bias. An accurate assessment of analyte content can be obtained using a reference method, although such methods are not always available.

5.4.3 Spiked materials/solutions

These are materials or solutions to which the analyte(s) of interest have been deliberately added. These materials or solutions may already contain the analyte of interest so care is needed to ensure the spiking does not lead to analyte levels outside of the working range of the method. Spiking with a known amount of analyte enables the increase in response to the analyte to be measured and calculated in terms of the amount added, even though the absolute amounts of analyte present before and after addition of the spike are not known. Note that most methods of spiking add the analyte in such a way that it will not be as closely bound to the sample matrix as it would be if it was present naturally. Therefore, bias estimates obtained by spiking can be expected to be over-optimistic.

Spiking does not necessarily have to be restricted to the analyte of interest. It could include anything added to the sample in order to gauge the effect of the addition. For example, the sample could be spiked with varying amounts of a particular interference in order to judge the concentration of the interferent at which determination of the analyte is adversely affected. The nature of the spike obviously needs to be identified.

5.4.4 Incurred materials

These are materials in which the analyte of interest may be essentially alien, but has been introduced to the bulk at some point prior to the material being sampled. The analyte is thus more closely bound in the matrix than it would be had it been added by spiking. The analyte value will depend on the amounts of analyte in contact with the material, the rate of take-up and loss by the matrix and any other losses through metabolism, spontaneous disintegration or other chemical or physical processes. The usefulness of incurred samples for validation purposes depends on how well the analyte value can be characterised. The following are examples of incurred materials:

1. Herbicides in flour from cereal sprayed with herbicides during its growth;
2. Active ingredients in pharmaceutical formulations added at the formulation stage.
3. Egg-white powder (known protein content) added to a cookie dough before baking when investigating allergens.

5.4.5 Measurement standards

Care must be taken when referring to ‘standards’ as the term also applies to written documents, such as ISO standards. Where the term refers to substances used for calibration or identification purposes it is convenient to refer to them as measurement standards or calibrants/calibrators [7]. These are traditionally thought of as solutions of single substances but in practice can be anything in which a particular parameter or property has been characterised to the extent it can serve as a metrological reference.

It is important to distinguish between reference materials (RMs) and certified reference materials (CRMs) [7, 30] because of the significant difference in how they can be used in the method validation process (6.5.2). RMs can be virtually any material used as a basis for reference, and could include laboratory reagents of known purity, industrial chemicals, or other artefacts. The property or analyte of interest needs to be

[*]A reagent blank taken through the entire analytical procedure is sometimes called a ‘procedural blank’.

stable and homogenous but the material does not need to have the high degree of characterisation, metrological traceability, uncertainty and documentation associated with CRMs.

The characterisation of the parameter of interest in a CRM is generally more strictly controlled than for an RM, and in addition the characterised value is certified with a documented metrological traceability and uncertainty. Characterisation is normally done using several different methods, or a single primary measurement procedure, so that as far as possible, any bias in the characterisation is reduced or even eliminated.

Assessment of bias requires a reliable reference point, preferably, a CRM with the same matrix and analyte concentrations as the test samples.

5.4.6 Statistics

Statistical methods are essential for summarising data and for making objective judgements on differences between sets of data (significance testing). Analysts should familiarise themselves with at least the more basic elements of statistical theory particularly as an aid to evaluation of precision, bias, linear range, LOD, LOQ and measurement uncertainty. A number of useful books introducing statistics for analytical chemistry are referenced [5, 6, 31, 32, 33, 34].

5.5 Validation requirements

Requirements for how to carry out method validation may be specified in guidelines within a particular sector relevant to the method [13, 25, 35 for example]. Where such requirements exist, it is recommended they are followed. This will ensure that particular validation terminology, together with the statistics used, is interpreted in a manner consistent within the relevant sector. Official recognition of a method may require characterisation using a collaborative study.

5.6 Method validation process

Faced with a particular customer problem, the laboratory must first set the analytical requirement which defines the performance characteristics that a method must have to solve that problem (Figure 1).

In response to these requirements, the laboratory needs to identify a suitable existing method, or if necessary develop/modify a method. Note that certain regulations may require a particular method to be followed. Table 4 shows the type of questions which might be posed in formalising an analytical requirement (column 1) and the corresponding performance characteristics of the method which may need to be evaluated (column 2). The laboratory will then identify and evaluate relevant performance characteristics and check them against the analytical requirement. The validation process ends with a conclusion and statement of whether or not the analytical requirement is met. If the analytical requirement is not met, further method development is necessary. This process of development and evaluation continues until the method is deemed capable of meeting the requirement.

In reality an analytical requirement is rarely agreed with the customer beforehand in such a formal way. Customers usually define their requirements in terms of cost and/or time and rarely know how well methods need to perform, although performance requirements for methods may be specified where the methods support a regulatory requirement or compliance with a specification. For example, the European Union (EU) have published requirements, e.g. for the analysis of drinking water [36], for analyses performed within the water framework directive [37], for the determination of the levels of veterinary drug residues in food of animal origin [38] and of pesticide residues in food and feed [39].

However, it will usually be left to the analyst's discretion to decide what performance is required. Very often this will mean setting an analytical requirement in line with the method's known capability (e.g. as published in standardised methods, as observed in proficiency testing (PT) schemes or estimated from mathematical models, such as the Horwitz function [40]).

Financial constraints may dictate that development of a method that satisfies a particular analytical requirement is not economically feasible, in which case the decision must be taken whether to relax the requirement to a more achievable level or rethink the justification for the analysis.

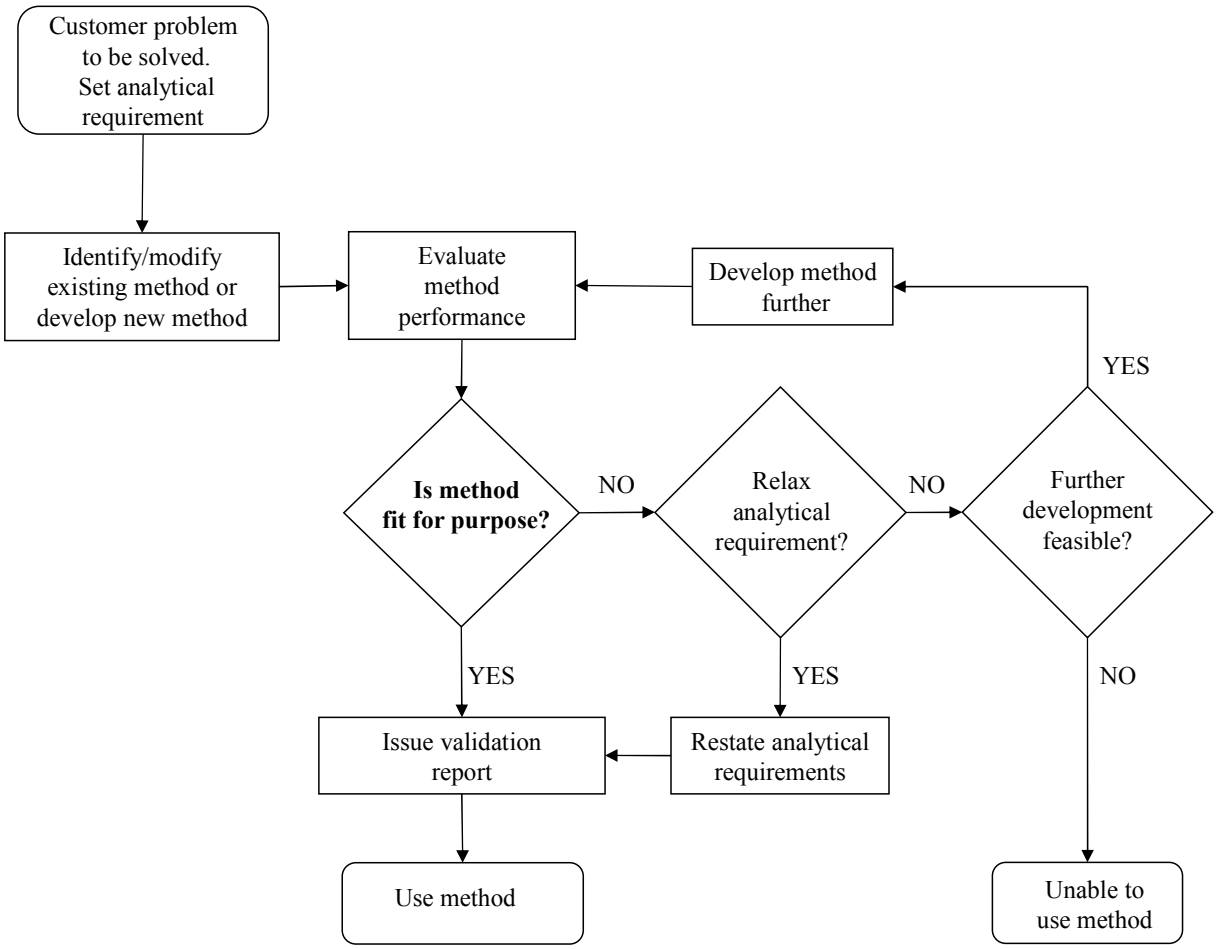


Figure 1 – The method validation process: from the customer problem to the laboratory decision on whether or not the customer request can be carried out with an identified method. Note: method validation consists of a stage where performance characteristics are evaluated and then compared with analytical requirements. Regardless of what existing performance data may be available for the method, fitness for purpose will be determined by how the method performs when used by the designated analyst with the available equipment/facilities.

Table 4 – Questions which might be posed in formalising an analytical requirement, and related performance characteristics with references to the appropriate sections in this Guide

Question	Performance characteristic	Section	Note
Do resource constraints apply and how – people, time, money, equipment and reagents, laboratory facilities?	-	-	a)
Is sampling and subsampling required (and will this be done within the laboratory)?			
Are there any restrictions on sample size/availability?			
What is the chemical, biological and physical nature of the matrix?			
Is the analyte dispersed or localised?			
Is a qualitative or quantitative answer required?	Selectivity LOD and LOQ	6.1 6.2	
What are the analytes of interest and the likely levels present (%, µg/g, ng/g, etc.....)? Are the analytes present in more than one chemical form (e.g. oxidation states, stereoisomers), and is it necessary to be able to distinguish between different forms?	Selectivity LOD and LOQ Working and linear ranges	6.1 6.2 6.3	
What quantity is intended to be measured ('the measurand')? Is it the 'total' concentration of the analyte present that is of interest, or the 'amount extracted' under specified conditions?	Recovery	6.5	
What trueness and precision are required? What is the target uncertainty and how is it to be expressed?	Trueness and recovery Repeatability, intermediate precision, reproducibility Uncertainty	6.5 6.6 6.7	b)
What are the likely interferences to the analyte(s)?	Selectivity	6.1	
Have tolerance limits been established for all parameters, critical for performing the analysis (e.g. time of extraction, incubation temperature)?	Ruggedness	6.8	c)
Do results need to be compared with results from other laboratories?	Uncertainty	6.7	b)
Do results need to be compared with external specifications?	Uncertainty	6.7	b)

a) Not all of the elements of the analytical requirement link directly to method validation requirements but dictate more generally as to whether particular techniques are applicable. For example, different techniques will be applicable according to whether the analyte is dispersed through the sample or isolated on the surface.

b) One essential element of the analytical requirement is that it should be possible to judge whether or not a method is suitable for its intended purpose and thus must include the required uncertainty expressed either as a standard uncertainty or an expanded uncertainty.

c) Published standardised procedures have normally been shown to be rugged within the scope of the procedure, i.e. matrix types and working range. Therefore single-laboratory verification for implementation of a published standardised procedure need not normally include ruggedness.

6 Method performance characteristics

6.1 Selectivity

6.1.1 Terms and definitions

Analytical selectivity relates to "*the extent to which the method can be used to determine particular analytes in mixtures or matrices without interferences from other components of similar behaviour"* [41].

Definitions in various documents [7, 18, 42] more or less agree with this interpretation. While IUPAC recommends the term 'selectivity', some areas, e.g. the pharmaceutical sector [13], use 'specificity' or 'analytical specificity'. The latter is recommended to avoid confusion with 'diagnostic specificity' as used in laboratory medicine [43].

6.1.2 Effects of interferences

In general, analytical methods can be said to consist of a measurement stage which may or may not be preceded by an isolation stage. In the measurement stage, the concentration of an analyte is normally not measured directly. Instead a specific property (e.g. intensity of light) is quantified. It is, therefore, crucial to establish that the measured property is only due to the analyte and not to something chemically or physically similar, or arising as a coincidence thus causing a bias in the measurement result. The measurement stage may need to be preceded by an isolation stage in order to improve the selectivity of the measuring system.

Interferences may cause a bias by increasing or decreasing the signal attributed to the measurand. The size of the effect for a given matrix is usually proportional to the signal and is therefore sometimes called a 'proportional' effect. It changes the slope of the calibration function, but not its intercept. This effect is also called 'rotational' [44].

A 'translational' or 'fixed effect' arises from a signal produced by interferences present in the test solution. It is therefore independent of the concentration of the analyte. It is often referred to as a 'background' or 'baseline' interference. It affects the intercept of a calibration function, but not its slope.

It is not unusual for both proportional and translational effects to be present simultaneously. The method of standard additions can only correct for proportional effects.

6.1.3 Assessment of selectivity

The selectivity of a procedure must be established for in-house developed methods, methods adapted from the scientific literature and methods published by standardisation bodies used outside the scope specified in the standard method. When methods published by standardisation bodies are used within their scope, selectivity will usually have been studied as part of the standardisation process.

The selectivity of a method is usually investigated by studying its ability to measure the analyte of interest in samples to which specific interferences have been deliberately introduced (those thought likely to be present in samples). Where it is unclear whether or not interferences are already present, the selectivity of the method can be investigated by studying its ability to measure the analyte compared to other independent methods. Example 1 and Example 2 below and Quick Reference 1 illustrate the practical considerations regarding selectivity.

Confirmatory techniques can be useful as a means of verifying identities. The more evidence one can gather, the better. Inevitably there is a trade-off between costs and time taken for analyte identification, and the confidence with which one can decide if the identification has been made correctly.

Whereas evaluation of repeatability requires the measurement to be repeated several times by one technique, confirmation of analyte identity requires the measurement to be performed by several, preferably independent, techniques. Confirmation increases confidence in the technique under examination and is especially useful when the confirmatory techniques operate on significantly different principles. In some applications, for example, the analysis of unknown organics by gas chromatography, the use of confirmatory techniques is essential. When the measurement method being evaluated is highly selective, the use of other confirmatory techniques may not be necessary.

An important aspect of selectivity which must be considered is where an analyte may exist in the sample in more than one form such as: bound or unbound; inorganic or organometallic; or different oxidation states. The definition of the measurand is hence critical to avoid confusion.

Example 1 – Chromatography. A peak in a chromatographic trace may be identified as being due to the analyte of interest on the basis that an RM containing the analyte generates a signal at the same point on the chromatogram. But, is the signal due to the analyte or to something else which coincidentally co-elutes, i.e. a fixed effect? It could be either or both. Identification of the analyte, by this means only, is unreliable and some form of supporting evidence is necessary. For example, the chromatography could be repeated using a column of different polarity, employing a different separation principle to establish whether the signal and the signal generated by the RM still appear at the same time. Where a peak is due to more than one compound, a different polarity column may be a good way of separating the compounds. In many cases modern mass spectrometric instruments can offer a high selectivity, e.g. gas or liquid chromatography with mass spectrometric detection.

Example 2 – Spectroscopy. In infrared spectroscopy, identification of unknown compounds may be made by matching absorbance signals (i.e. 'peaks') in the analyte spectrum with those of reference spectra stored in a spectral library. Once it is believed the correct identification has been made, a spectrum of an RM of the analyte should be recorded under exactly the same conditions as for the test portion. The larger the number of peaks which match between analyte and RM, the better the confidence that can be placed on the identification being correct. It would also be worthwhile examining how dependant the shape of the spectrum was with respect to how the analyte was isolated and prepared for infrared analysis. For example, if the spectrum was recorded as a salt disc, the particle size distribution of the test portion in the disc might influence the shape of the spectrum.

Quick Reference 1 – Selectivity

What to do	How many times	What to calculate/determine from the data	Comments
Analyse test samples, and RMs by candidate and other independent methods.	1	Use the results from the confirmatory techniques to assess the ability of the method to confirm analyte identity and its ability to measure the analyte in isolation from other interferences.	Decide how much supporting evidence is reasonably required to give sufficient reliability.
Analyse test samples containing various suspected interferences in the presence of the analytes of interest.	1	Examine effect of interferences. Does the presence of the interferent inhibit detection or quantification of the analytes?	If detection or quantification is inhibited by the interferences, further method development will be required.

6.2 Limit of detection and limit of quantification

6.2.1 Terms and definitions

Where measurements are made at low concentrations, there are three general concepts to consider. First, it may be necessary to establish a value of the result which is considered to indicate an analyte level that is significantly different from zero. Often some action is required at this level, such as declaring a material contaminated. This level is known as the 'critical value', 'decision limit' or, in EU directives, CCα [38].

Second, it is important to know the lowest concentration of the analyte that can be detected by the method at a specified level of confidence. That is, at what true concentration will we confidently exceed the critical value described above? Terms such as 'limit of detection' (LOD), 'minimum detectable value', 'detection limit', or, in EU directives, CCβ [38] are used for this concept.

Third, it is also important to establish the lowest level at which the performance is acceptable for a typical application. This third concept is usually referred to as the limit of quantification (LOQ)*.

* Synonyms used include 'quantification limit', 'quantitation limit', 'limit of quantitation', 'limit of

Terminology relating to all these concepts is very diverse and varies between sectors. For example, the terms 'limit of detection' (LOD) or 'detection limit' (DL) were previously not generally accepted, although used in some sectoral documents [13, 38]. However, they are now incorporated into the VIM [7] and IUPAC Gold Book [17]. ISO uses as a general term 'minimum detectable value of the net state variable' which for chemistry translates as 'minimum detectable net concentration' [45, 46, 47, 48]. In this Guide the terms 'critical value', 'limit of detection (LOD)' and 'limit of quantification' (LOQ) are used for the three concepts above. In method validation, it is the LOD and LOQ that are most commonly determined.

It is also necessary to distinguish between the instrument detection limit and the method detection limit. The instrument detection limit can be based on the analysis of a sample, often a reagent blank, presented directly to the instrument (i.e. omitting any sample preparation steps), or on the signal-to-noise ratio in, e.g. a chromatogram. To obtain a method detection limit, the LOD must be based on the analysis of samples that have been taken through the whole measurement procedure using results calculated with the same equation as for the test samples. It is the method detection limit that is most useful for method validation and is therefore the focus of this Guide.

The following paragraphs describe the experimental estimation of LOD and LOQ. The statistical basis for the calculation of the LOD is given in Annex B. Because the LOD and LOQ both depend on the precision at or near zero, Section 6.2.2 first describes the experimental estimation of the standard deviation of results near zero.

6.2.2 Determination of the standard deviation at low levels

Both LOD and LOQ are normally calculated by multiplying a standard deviation (s_0') by a suitable factor. It is important that this standard deviation is representative of the precision obtained for typical test samples, and that sufficient replicate measurements are made to give a reliable estimate. In this section, the standard deviation s_0' is based on a standard deviation s_0 for single results near zero, adjusted for any averaging or blank correction used in practice (see below). Alternative approaches are discussed in Section 6.2.5

The following issues should be considered in determining LOD and LOQ from an experiment using simple replication.

Suitable samples for estimating LOD and LOQ: The samples used should preferably be either a) blank samples, i.e. matrices containing no detectable analyte, or b) test samples with concentrations of analyte close to or below the expected LOD. Blank samples work well for methods where a measurable signal is obtained for a blank, such as spectrophotometry and atomic spectroscopy. However for techniques such as chromatography, which rely on detecting a peak above the noise, samples with concentration levels close to or above the LOD are required. These can be prepared by, for example, spiking a blank sample (see Section 5.4).

When blank samples or test samples at low concentrations are not available, reagent blanks* can often be used. When these reagent blanks do not go through the whole measurement procedure, and are presented directly to the instrument, the calculation based on these measurements will give the instrument LOQ/LOD.

Covering the scope of the method: For methods with a scope covering very different matrices it may be necessary to determine the standard deviation for each matrix separately.

Ensuring representative replication: The standard deviation should be representative of the performance of the method as used in the laboratory, i.e. the standard deviation is to be calculated based on test results where analyses are performed exactly according to the whole documented measurement procedure, including any sample preparation steps. The values used for calculating the standard deviation s_0 should be in the measurement units specified in the procedure.

Conditions of measurement: The standard deviation is normally obtained under repeatability conditions and this is the procedure described in this section. However, a more reliable estimate can be obtained from the use of

determination', 'reporting limit', 'limit of reporting' and 'application limit'.

* There is confusion regarding the terminology relating to blanks – for further discussion see Section 5.4.1.

intermediate precision conditions. This approach is discussed further in Section 6.2.5.

Number of observations: The number of replicates (m) should be sufficient to obtain an adequate estimate of the standard deviation. Typically between 6 and 15 replicates are considered necessary; 10 replicates are often recommended in validation procedures/protocols (see Section 6.2.5.1).

Allowing for averaging: In many measurement procedures the mean of replicates is reported in routine use of the method, where each replicate is obtained by following the entire measurement procedure. In this case the standard deviation of single results s_0 should be corrected by dividing with the square root of n, where n is the number of replicates averaged in routine use.

Allowing for the effect of blank corrections: If blank corrections are specified in the measurement procedure, care needs to be taken when determining the standard deviation used to calculate the LOD or LOQ. If the results obtained during the validation study were all corrected by the same blank value – the approach recommended here for simplicity – the standard deviation of the results will be smaller than that seen in practice when results are corrected by different blank values obtained in different runs.

In this case s_0 should be corrected by multiplying by $\sqrt{\frac{1}{n}+\frac{1}{n_b}}$ where n is the number of replicate observations averaged when reporting results where each replicate is obtained following the entire measurement procedure, and n_b is the number of blank observations used to calculate the blank correction.

Note that under intermediate precision conditions results will be corrected by different blank values so no correction of the standard deviation is necessary (see Section 6.2.5).

Example 3 illustrates these calculations and the flow chart in Figure 2 summarises the corrections required for averaging and blank correction.

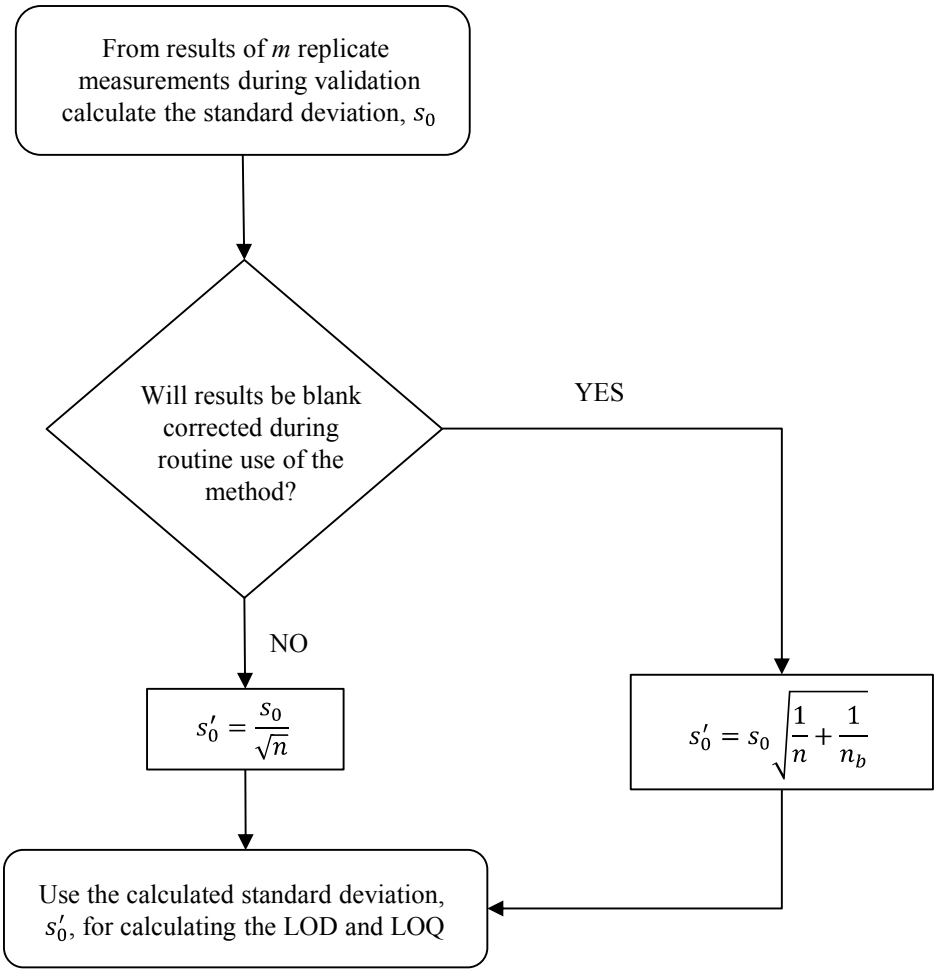
Example 3 – A validation exercise is based on the analysis of a sample blank. Ten (m) independent measurements of the sample blank are made under repeatability conditions. The results have a mean value of 2 mg/kg and a standard deviation s_0 of 1 mg/kg.

Case 1 – The measurement procedure states that test samples should be measured once (n=1) and the results corrected by the result for a single sample blank sample (n_b=1). In a series of measurements each run consists of single replicates of routine samples and one (n_b) blank sample. The standard deviation for calculating LOD/LOQ is then, according to Figure 2 equal to:

$$s_0' = s_0\sqrt{\frac{1}{n}+\frac{1}{n_b}} = 1\sqrt{\frac{1}{1}+\frac{1}{1}} = 1\sqrt{2} = 1.4 \text{ mg/kg}$$

Case 2 – The measurement procedure states that test samples should be analysed in duplicate (n=2) and also that the blank sample should be analysed in duplicate. In a series of measurements each run consists of duplicates (n=2) of routine samples and two (n_b) blank samples. The concentration obtained for routine samples is corrected by subtracting the mean value of the two blank samples. The standard deviation for calculating LOD/LOQ is then, according to Figure 2 equal to:

$$s_0' = s_0\sqrt{\frac{1}{n}+\frac{1}{n_b}} = 1\sqrt{\frac{1}{2}+\frac{1}{2}} = 1 \text{ mg/kg}$$



s_0 is the estimated standard deviation of *m* single results at or near zero concentration.

s_0' is the standard deviation used for calculating LOD and LOQ.

n is the number of replicate observations averaged when reporting results where each replicate is obtained following the entire measurement procedure.

n_b is the number of blank observations averaged when calculating the blank correction according to the measurement procedure.

Figure 2 – Calculation of the standard deviation, s_0' to be used for estimation of LOD and LOQ. The flow chart starts with an experimental standard deviation, s_0 calculated from the results of replicate measurements under repeatability conditions on a sample near zero concentration, either without blank correction or with a blank correction applied to all results as specified by the method. This blank correction may be based on a single blank observation or on a mean of several blank observations.

6.2.3 Estimating LOD

For validation purposes it is normally sufficient to provide an approximate value for the LOD, i.e. the level at which detection of the analyte becomes problematic. For this purpose the '3*s*' approach shown in Quick Reference 2 will usually suffice.

Where the work is in support of regulatory or specification compliance, a more exact approach is required, in particular taking into account the degrees of freedom associated with s_0. This is described in detail by IUPAC [49] and others [50, 51]. Where the critical value and/or LOD are used for making decisions, the precision should be monitored and the limits may need to be recalculated from time to time. Different sectors and/or regulations may use different approaches to LOD estimation. It is recommended that the convention used is stated when quoting a detection limit. In the absence of any sectoral guidance on LOD estimation, the approaches given in the Quick Reference 2 can be used as a general guidance.

6.2.4 Estimating LOQ

The LOQ is the lowest level of analyte that can be determined with acceptable performance. ('Acceptable performance' is variously considered by different guidelines to include precision, precision and trueness, or measurement uncertainty [52]. In practice, however, LOQ is calculated by most conventions to be the analyte concentration corresponding to the obtained standard deviation (s_0') at low levels multiplied by a factor, k_Q. The IUPAC default value for k_Q is 10 [49] and if the standard deviation is approximately constant at low concentrations this multiplier corresponds to a relative standard deviation (RSD) of 10 %. Multipliers of 5 and 6 have also sometimes been used which corresponds to RSD values of 20 % and 17 % respectively [53, 54]. See further Reference [8] and Quick Reference 3.

Quick Reference 2 – Limit of detection (LOD)

What to do	How many times	What to calculate from the data	Comments
a) Replicate measurements of blank samples, i.e. matrices containing no detectable analyte. or Replicate measurements of test samples with low concentrations of analyte.	10	Calculate the standard deviation, s_0 of the results. Calculate s_0' from s_0 following the flow chart in Figure 2. Calculate LOD as LOD = $3 \times s_0'$.	
b) Replicate measurements of reagent blanks. or Replicate measurements of reagent blanks spiked with low concentrations of analyte.	10	Calculate the standard deviation, s_0 of the results. Calculate s_0' from s_0 following the flow chart in Figure 2. Calculate LOD as LOD = $3 \times s_0'$.	Approach b) is acceptable, when it is not possible to obtain blank samples or test samples at low concentrations. When these reagent blanks are not taken through the whole measurement procedure, and are presented directly to the instrument, the calculation will give the instrument LOD.

NOTES

1) For some analytical techniques, e.g. chromatography, a test sample containing too low a concentration or a reagent blank might need to be spiked in order to get a non-zero standard deviation.
2) The entire measurement procedure should be repeated for each determination.
3) The standard deviation is expressed in concentration units. When the standard deviation is expressed in signal domain the LOD is the concentration corresponding to the blank signal $y_B + 3 \times s_0'$. A short example of LOD calculations in the signal domain is given also in Reference [5].

Quick Reference 3 – Limit of quantification (LOQ)

What to do	How many times	What to calculate from the data	Comments
a) Replicate measurements of blank samples, i.e. matrices containing no detectable analyte. or Replicate measurements of test samples with low concentrations of analyte.	10	Calculate the standard deviation, s_0 of the results. Calculate s_0' from s_0 following the flow chart in Figure 2. Calculate LOQ as LOQ = $k_Q \times s_0'$.	The value for the multiplier k_Q is usually 10, but other values such as 5 or 6 are commonly used (based on 'fitness for purpose' criteria).
b) Replicate measurements of reagent blanks. or Replicate measurements of reagent blanks spiked with low concentrations of analyte.	10	Calculate the standard deviation, s_0 of the results. Calculate s_0' from s_0 following the flow chart in Figure 2. Calculate LOQ as LOQ = $k_Q \times s_0'$.	Approach b) is acceptable, when it is not possible to obtain blank samples or test samples at low concentrations. When these reagent blanks are not taken through the whole measurement procedure and are presented directly to the instrument the calculation will give the instrument LOQ.
NOTES 1) For some analytical techniques, e.g. chromatography, a test sample containing too low a concentration or a reagent blank might need to be spiked in order to get a non-zero standard deviation. 2) The entire measurement procedure should be repeated for each determination. 3) The standard deviation is expressed in concentration units.			

6.2.5 Alternative procedures

The previous sections have described a general approach to estimating LOD and LOQ, based on the standard deviation of results at concentrations near zero, obtained under repeatability conditions. This approach is widely applied but alternative procedures are given in other standards and protocols.

In some cases, e.g. where blank values differ significantly from day-to-day, intermediate precision conditions are preferred to repeatability conditions. For example, if quality control results for test samples at low concentration levels are available, the standard deviation of these results can be used in the estimation of LOD and LOQ. Where the standard deviation used to calculate LOD and LOQ is obtained under intermediate precision conditions, the adjustment to take account of blank correction shown in Figure 2 is not required. Therefore the experimental standard deviation obtained from the internal quality control is equal to the standard deviation s_0' to be used for calculating LOD and LOQ. ISO 11843-2 [46] describes how the instrument LOD can be obtained directly from a calibration curve.

6.2.5.1 Reliability of estimates of LOD and LOQ

It should be noted that even with the 10 replicates indicated in Quick Reference 2 and Quick Reference 3, estimates of a standard deviation are inherently variable. Therefore, the estimate of LOD/LOQ obtained during validation should be taken as an indicative value. This will be sufficient if an estimate of LOD/LOQ is required simply to demonstrate that the concentrations of samples will be well above the LOD/LOQ. Where laboratory samples are expected to contain low concentrations of the analyte, the LOD/LOQ should be monitored on a regular basis.

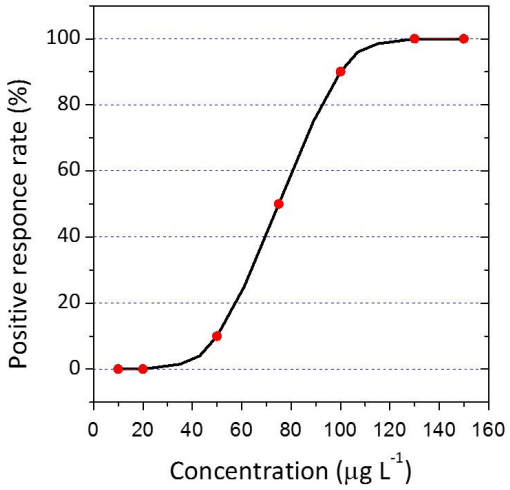
6.2.6 Capability of detection for qualitative analysis

A qualitative analysis (Annex D) involves identification or classification of substances and is effectively a 'yes'/'no' answer at a given cut-off concentration of an analyte [55]. For qualitative methods, precision cannot be expressed as a standard deviation or relative standard deviation, but may be expressed as true and false positive and negative rates.

In a validation study the cut-off concentration can be determined by establishing the false positive and negative rates at a number of levels below and above the expected cut-off concentration. The cut-off limit is where false negative rates for concentrations above the limit are low – with a stated probability, e.g. 5 %. During validation the proposed cut-off limit given in the documented procedure is assessed. (See Example 4 and Quick Reference 4).

Example 4 – Determination of cut-off concentration for a qualitative method with a stated cut-off equal to 100 µg L^{-1}. Ten observations were recorded at each level. A response curve with fraction (in %) of positive results versus concentration was constructed, from which it was possible to determine, by inspection, the threshold concentration at which the test becomes unreliable. With a criterion of < 5 % false negative results, the cut-off concentration is between 100 and 130 µg L^{-1}.

C (µg L^{-1})	No. of positive/negative results
150	10/0
130	10/0
100	9/1
75	5/5
50	1/9
20	0/10
10	0/10



Quick Reference 4 – Limit of detection (LOD) for qualitative analysis

What to do	How many times	What to calculate/determine from the data
Measure, in random order, sample blanks spiked with the analyte at a range of concentration levels.	10	A response curve of % positive or negative results versus concentration should be constructed, from which it will be possible to determine, by inspection, the threshold concentration at which the test becomes unreliable.

6.3 Working range

6.3.1 Definition

The 'working range'* is the interval over which the method provides results with an acceptable uncertainty. The lower end of the working range is bounded by the limit of quantification LOQ. The upper end of the working range is defined by concentrations at which significant anomalies in the analytical sensitivity are observed. An example of this is the plateauing effect at high absorbance values in UV/VIS spectroscopy.

6.3.2 Considerations for the validation study

The working range of the method to be validated should be stated in the scope of the documented procedure (see A.5 in Annex A). During validation it is necessary to confirm that the method can be used over this interval. In order to assess the working range, the laboratory needs to consider both the method linearity and the proposed calibration procedure given in the method.

6.3.3 Method and instrument working range

Many methods rely on the test sample received in the laboratory (the laboratory sample) being processed (digested, extracted, diluted) before it can be presented to the measuring instrument and a signal recorded. In these cases there are two working ranges. The method working range, given in the scope of the method (e.g. Section A.5 in Annex A), relates to the concentration in the laboratory sample. It is expressed, for example, in mg kg^{-1} for a solid test sample. The instrument working range is defined in terms of the concentration in a processed test sample presented to the instrument for measurement (e.g. mg L^{-1} in a solution after extracting the sample). An example of an instrument working range is given in Figure 3A where the concentrations in the calibration standards are plotted versus instrument signal. An example of a method working range is given in Figure 3B where the known test sample concentrations are plotted versus measured concentration. The measured concentration is the result obtained by applying the measurement procedure (including any sample preparation) using the instrument calibrated according to the written method.

In the course of the validation both the instrument working range and the method working range should be assessed. Data on the working range is often generated during method development. In such cases it will be sufficient to include this data in the validation report.

6.3.4 Assessing instrument working range

Between the LOQ and the upper end of the instrument working range, the response of the instrument obeys a known relationship, e.g. linear, curvilinear etc. During validation it is necessary to *i*) confirm this relationship, *ii*) demonstrate that the instrument working range is compatible with the interval stated in the method scope, and *iii*) verify that the proposed instrument calibration procedure (single point, bracketing, or multiple points) is adequate.

In order to assess the instrument working range and confirm its fitness for purpose, calibration standards with a concentration span that exceeds the expected concentration range by ± 10 % or even ± 20 % should be studied and the signals plotted (see Quick Reference 5 step 1). For a range 1 to 100 mg L^{-1}, ± 20 % indicates from 0.8 to 120 mg L^{-1}. The chosen concentrations should be evenly spaced across the range. The initial assessment of the working range is by a visual inspection of the response curve. The next step is to confirm the relationship between concentration and instrument response by examining the regression statistics and residual plot for the chosen model (e.g. linear, quadratic) (see Quick Reference 5 step 2). The assessment may also include special statistical measures, such as 'goodness of fit' tests [56, 57]. From the response curve and the supporting statistics obtained over the instrument working range, the analyst can assess if the suggested calibration procedure given in the method is appropriate. This is further assessed by evaluating the method working range.

* The VIM term [7] is 'measuring interval' or 'working interval'

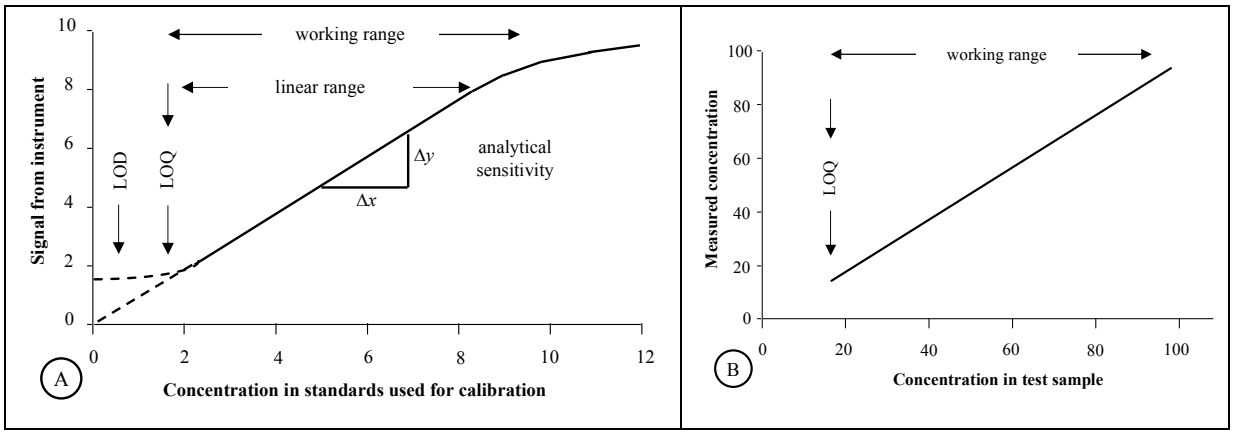


Figure 3 – A) Typical example of response curve obtained with an instrumental method. The performance characteristics 'working range', 'linear range', 'analytical sensitivity', 'LOD' and 'LOQ' are identified. B) Typical example of a curve obtained with a measurement procedure where the test sample concentration is plotted versus measured concentration.

6.3.5 Assessing method working range

In order to assess the method working range 1) samples with known concentrations and sample blanks should be available; 2) the samples used should be taken through the entire measurement procedure; 3) the concentrations of the different samples should preferably cover the whole range of interest and 4) the instrument should have been calibrated according to the suggested calibration procedure. The measurement result for each test sample is calculated according to the written procedure (see step 3 in Quick Reference 5). These values are plotted on the *y*-axis against the known concentrations of the samples (*x*-axis) as in Figure 3B. The method working range and linearity are assessed by visual inspection of the plot, supported by statistics and a residuals plot from a linear regression.

The assessment of the working range will be supported by data from precision and bias studies (see Sections 6.5.2 and 6.6.2.1), providing that these studies cover concentrations across the whole method working range.

The method working range needs to be established for each matrix covered in the method scope. This is because interferences can cause non-linear responses, and the ability of the method to extract/recover the analyte may vary with the sample matrix.

Quick Reference 5 – Working and linear range

What to do	How many times	What to calculate from the data	Comments
1) Measure blank plus calibration standards, at 6-10 concentrations evenly spaced across the *range of interest.*	1	Plot response (y axis) against concentration (x axis). Visually examine to identify approximate linear range and upper and lower boundaries of the working range for the instrument. then go to 2).	This will give visual confirmation of whether or not the instrument working range is linear. Note: When the signal is not directly proportional to concentration, e.g. when working with pH or other ion selective electrodes or immunometric methods, a transformation of the measured values is needed before linearity can be assessed.
2) Measure blank plus calibration standards, 2-3 times at 6-10 concentrations evenly spaced across the *linear range.*	1	Plot response (y axis) against concentration (x axis). Visually examine for outliers which may not be reflected in the regression. Calculate appropriate regression statistics. Calculate and plot residuals (difference between observed y value and calculated y value predicted by the straight line, for each x value). Random distribution of residuals about zero confirms linearity. Systematic trends indicate non-linearity or a change in variance with level.	This stage is necessary to test a working range, thought to be linear and especially where the method uses a two point calibration. If the standard deviation is proportional to concentration then consider using a weighted regression calculation rather than a simple non-weighted linear regression. It is unsafe to remove an outlier without first checking it using further measurements at nearby concentrations. In certain circumstances for instrument calibration it may be better to try to fit a non-linear curve to the data. The number of samples then needs to be increased. Functions higher than quadratic are generally not advised.
3) Calibrate the instrument according to the proposed calibration procedure. Measure, according to the written method, blank plus reference materials or spiked sample blanks 2-3 times at 6-10 concentrations evenly spaced across the range of interest.	1	Plot the measured concentration (*y*-axis) against the concentration of the test samples (*x*-axis). Visually examine to identify approximate linear range and upper and lower boundaries of the working range. Calculate appropriate regression statistics. Calculate and plot residuals (difference between observed y value and calculated y value predicted by the straight line, for each x value). Random distribution of residuals about zero confirms linearity. Systematic trends indicate non-linearity.	This step is required to assess whether the proposed instrument range and calibration procedure are fit for purpose. If data are available from bias and precision studies that cover the range of interest, a separate method working range study may not be required.

6.4 Analytical sensitivity

6.4.1 Definition

Analytical sensitivity is the change in instrument response which corresponds to a change in the measured quantity (for example an analyte concentration), i.e. the gradient of the response curve [7, 18]. The prefix 'analytical' is recommended to avoid confusion with 'diagnostic sensitivity' used in laboratory medicine [43]. The term 'sensitivity' is sometimes used to refer to limit of detection but this use is discouraged in the VIM.

6.4.2 Applications

Analytical sensitivity is not a particularly important performance characteristic. There are, however, at least two useful applications:

1. The theoretical analytical sensitivity is sometimes known. Many ion selective electrodes show a Nernstian behaviour, e.g. the signal from a well-functioning glass electrode is expected to change by 59 mV/pH.
2. In spectrophotometric measuring systems the absorbance can be predicted from the Beer-Lambert law. This can be used as a check of instrument performance and standards sometimes require such checks to be made [58].

6.5 Trueness

6.5.1 Terminology to describe measurement quality

In this Guide we use the three related performance characteristics *trueness*, *precision* and *uncertainty* to describe the quality of results obtained with a method. However, scientists frequently use different concepts, such as types of error (random, systematic and gross errors), accuracy (trueness and precision) and uncertainty. Some of these concepts have a qualitative meaning and some are quantitative. Over the years, terms as well as definitions have changed and new terms have been introduced. In addition, different sectors still favour different terms, all of which leads to a great deal of confusion. Figure 4 illustrates the links between the terms and further details are given in VIM [7] and the Eurachem Guide on terminology [8].

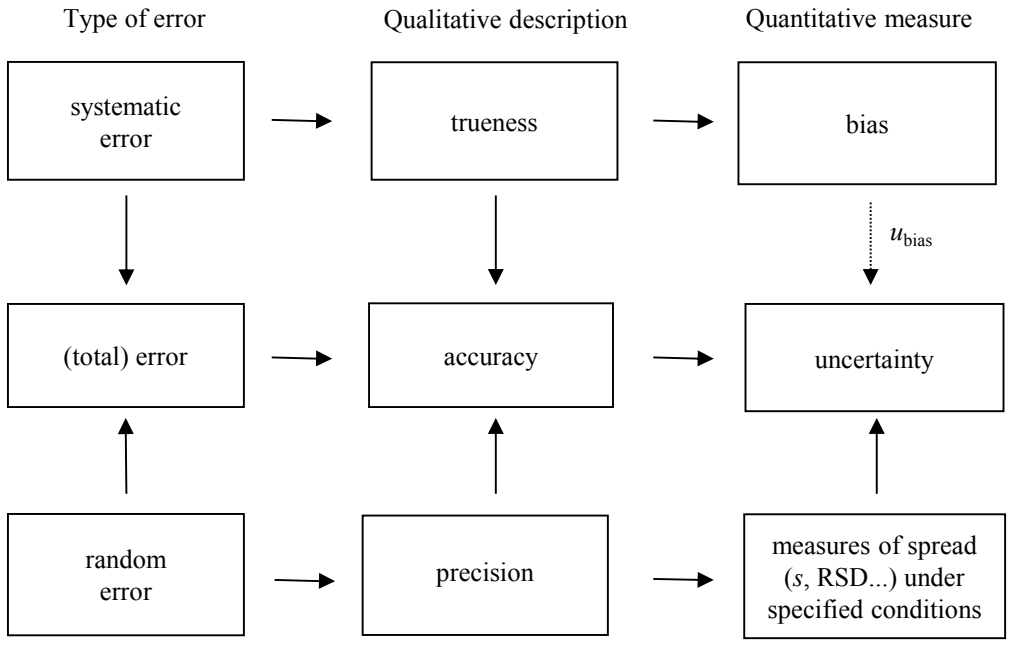


Figure 4 – Illustration of the links between some fundamental concepts used to describe quality of measurement results (based on the work of Menditto et al. [59]). An uncertainty evaluation according to GUM [21] assumes correction for known bias and that the uncertainty of the bias correction u_{bias} is included in the final uncertainty statement. This is implied by the dotted arrow below the box 'bias'. Both the accuracy concept and the uncertainty concept assume that measurements are performed according to the documented procedure and that effects of 'gross errors' (mistakes) are not included.

Measurement 'accuracy' expresses the closeness of a single result to a reference value* [29, 48]. (for the exact definition see VIM 2.13). Method validation seeks to investigate the accuracy of results by assessing both systematic and random effects on single results. Accuracy is, therefore, normally studied as two components: 'trueness' and 'precision'. In addition, an increasingly common expression of accuracy is 'measurement uncertainty', which provides a single figure. The evaluation of trueness is described below while precision is discussed in Section 6.6 and uncertainty in Section 6.7.

Measurement 'trueness' is an expression of how close the mean of an infinite number of results (produced by the method) is to a reference value. Since it is not possible to take an infinite number of measurements, trueness cannot be measured. We can, however, make a practical assessment of the trueness. This assessment is normally expressed quantitatively in terms of 'bias'.

6.5.2 Determination of bias

A practical determination of bias relies on comparison of the mean of the results ($\bar{x}$) from the candidate method with a suitable reference value (x_{ref}).* Three general approaches are available: a) analysis of reference materials, b) recovery experiments using spiked samples, and c) comparison with results obtained with another method – see Quick Reference 6. Bias studies should cover the method scope and may therefore require the analysis of different sample types and/or different analyte levels. To achieve this, a combination of these different approaches may be required.

The bias can be expressed in absolute terms

$$b = \bar{x} - x_{\text{ref}} \qquad \text{(Eq. 1)}$$

or relative in per cent

$$b(\%) = \frac{\bar{x} - x_{\text{ref}}}{x_{\text{ref}}} \times 100 \qquad \text{(Eq. 2)}$$

or as a relative spike recovery

$$R'(\%) = \frac{\bar{x}' - \bar{x}}{x_{\text{spike}}} \times 100 \qquad \text{(Eq. 3)}$$

where $\bar{x}'$ is the mean value of the spiked sample and x_{spike} is the added concentration.

However in some sectors of analytical measurement, the relative recovery ('apparent recovery') in per cent is also used [60].

$$R(\%) = \frac{\bar{x}}{x_{\text{ref}}} \times 100 \text{ (Eq. 4)}$$

To determine the bias using an RM, the mean and standard deviation of a series of replicate measurements are determined and the results compared with the assigned property value of the RM. The ideal RM is a certified matrix reference material with property values close to those of the test samples of interest. CRMs are generally accepted as providing traceable values [61, 62]. It is also important to remember that a particular RM should only be used for one purpose during a validation study. For example, an RM used for calibration shall not also be used to evaluate bias.

Compared to the wide range of sample types and analytes encountered by laboratories the availability of RMs is limited, but it is important that the chosen material is *appropriate to the use.* It may be necessary to consider how the RM was characterised, for example if the sample preparation procedure used during characterisation of the material is not intended to give the total analyte concentration but the amount extracted under certain conditions. For regulatory work, a relevant certified material, ideally matrix-matched if available, should be used. For methods used for long-term in-house work, a stable in-house material can be used to monitor bias but a CRM should be used in the initial assessment.

In the absence of suitable RMs, recovery studies (spiking experiments) may be used to give an indication of the likely level of bias. Analytes may be present in a variety of forms in the sample and sometimes only certain forms are of interest to the analyst. The method may thus be deliberately designed to determine only a particular form of the analyte. A failure to determine part of or all of the analyte present may reflect an inherent problem with the method. Hence, it is necessary to assess the efficiency of the method for detecting all of the analyte present [60, 63].

Because it is not usually known how much of a particular analyte is present in a test portion, it is difficult to be certain how successful the method has been at extracting it from the sample matrix. One way to determine the efficiency of extraction is to spike test portions with the analyte at various concentrations, then extract the spiked test portions and measure the analyte concentration. The inherent problem with this is that analyte introduced in such a way will probably not be bound as strongly as that which

* The reference value is sometimes referred to as a 'true value' or a 'conventional true value'.

is naturally present in the test portion matrix and so the technique will give an unrealistically high impression of the extraction efficiency.

It may be possible to assess bias by comparing results from the candidate method with those obtained from an alternative method. There are two general types of alternative method which may be encountered – a reference method or a method currently in routine use in the laboratory. A reference method is intended to provide an 'accepted reference value' for the property being measured and will generally give results with a smaller uncertainty than the candidate method. A particular type of reference method is a primary method.[*] The second case arises when the purpose of the validation is to demonstrate that the candidate method gives results that are equivalent to an existing method. Here the aim is to establish that there is no significant bias in relation to the results produced by the existing method (although this method may itself be biased).

In both cases the results from the candidate and alternative methods, for the same sample or samples, are compared. The sample(s) may be in-house RMs, or simply typical test samples. The advantage of this approach is that the materials do not need to be CRMs as the alternative method provides the reference value. The method can therefore be tested on 'real' samples that are representative of those that will be encountered routinely by the laboratory.

[*] 'Primary method': a method having the highest metrological qualities, whose operation is completely described and understood in terms of SI units and whose results are accepted without reference to a standard of the same quantity (CCQM). The corresponding VIM term (see 2.8 in [7]) is 'primary reference measurement procedure'.

Quick Reference 6 – Trueness

What to do	How many times	What to calculate/determine from the data	Comments
a) Measure RM using candidate method.	10	Compare mean value, $\bar{x}$ with reference value x_{ref} for the RM. Calculate bias, b, per cent relative bias, $b(\%)$ or the relative per cent recovery (apparent recovery). $b = \bar{x} - x_{\text{ref}}$ $b(\%) = \frac{\bar{x} - \text{x}_{\text{ref}}}{x_{\text{ref}}} \times 100$ $R(\%) = \frac{\bar{x}}{x_{\text{ref}}} \times 100$	Gives a measure of bias taking into account the effect of both method and laboratory bias.
b) Measure matrix blanks or test samples unspiked and spiked with the analyte of interest over a range of concentrations.	10	Compare the difference between mean spiked value $\bar{x}'$ and mean value $\bar{x}$ with the added concentration x_{spike}. Calculate the relative spike recovery $R'(\%)$ at the various concentrations: $R'(\%) = \frac{\bar{x}' - \bar{x}}{x_{\text{spike}}} \times 100$	Spiked samples should be compared with the same sample unspiked to assess the net recovery of the added spike. Recoveries from spiked samples or matrix blanks will usually be better than for routine samples in which the analyte is more tightly bound.
c) Measure RM/test sample using candidate method and alternative method.	10	Compare mean value $\bar{x}$ with mean value $\bar{x}_{\text{ref}}$ of measurements made using alternative method. Calculate bias b or per cent relative bias $b(\%)$ or the relative per cent recovery (apparent recovery). $b = \bar{x} - \bar{x}_{\text{ref}}$ $b(\%) = \frac{\bar{x} - \bar{\text{x}}_{\text{ref}}}{\bar{x}_{\text{ref}}} \times 100$ $R(\%) = \frac{\bar{x}}{x_{\text{ref}}} \times 100$	Gives a measure of the bias relative to the alternative method. The alternative method may be a reference method or, if the intention is to replace one method with another and there is a need to demonstrate equivalent performance, a method currently in use in the laboratory. The alternative method may itself be biased, in which case the experiment will not provide an absolute measure of trueness.
NOTE Bias may vary with matrix and concentration level which means that the number of matrices and concentration levels to be examined must be stated in the validation plan.			

6.5.3 Interpreting bias measurements

Figure 5 shows two components of bias, here referred to as 'method bias' and 'laboratory bias'.

The method bias arises from systematic errors inherent to the method, irrespective of which laboratory uses it. Laboratory bias arises from additional systematic errors specific to the laboratory and its interpretation of the method. In isolation, a laboratory can only estimate the combined (total) bias from these two sources. However, in checking bias, it is important to be aware of the conventions in force for the particular purpose. For example, for some food applications, regulatory limits are set in terms of the results obtained from the specified empirical ('operationally defined') standard method. Method bias for 'empirical' measurement procedures is by definition zero. Bias arising solely from the particular method (see Figure 5) is then ignored, and metrological comparability with other laboratories using the same method is the main concern. In this situation, the laboratory should ideally determine bias using a reference material certified by the particular regulatory or empirical method under investigation, in which case the usual guidance for checking and interpreting bias applies. Where no such material is available, or to add further information, the laboratory may use alternative materials, but should then take care to consider any known differences between the method under investigation and the method(s) used to obtain the reference value when they interpret the results.

To fulfil a particular analytical requirement, the same analyte may be measured using several different measuring instruments at many sites within the same organisation. In this case, numerous and complex sources of bias arise within the organisation. In this common and complex situation, the organisation may establish procedures for estimating a representative uncertainty covering all sites/instruments for each application. This should preferably use material having the same properties, including sample matrix, as the samples intended to be measured. Variance component analysis can be used to identify the main causes of variation contributing to the overall measurement uncertainty, allowing follow-up action to reduce differences across the organisation.

For most purposes, however, acceptability of bias should be decided on the basis of overall bias measured against appropriate RMs, spiked materials or reference methods, taking into account the precision of the method and any uncertainties in reference values, and the accuracy required by the end use. Statistical significance tests are recommended [64, 65].

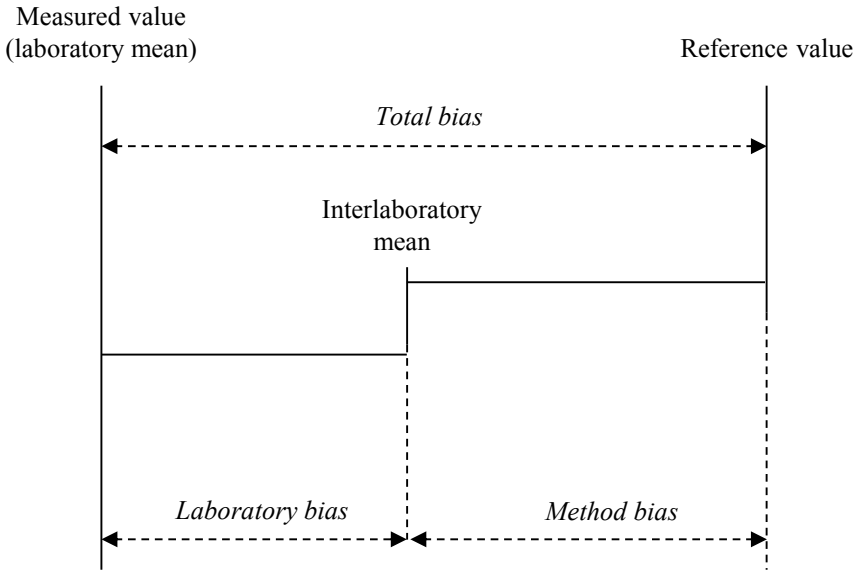


Figure 5 – The total measured bias consists of method bias and laboratory bias. Note: Laboratory and method biases are shown here acting in the same direction. In reality this is not always the case.

6.6 Precision

6.6.1 Replication

Replication is essential for obtaining reliable estimates of method performance characteristics such as precision and bias. Experiments involving replicate analysis should be designed to take into account all of the variations in operational conditions which can be expected during routine use of the method. The aim should be to determine typical variability and not minimum variability.

6.6.2 Precision conditions

Precision (measurement precision) is a measure of how close results are to one another [7, 29]. It is usually expressed by statistical parameters which describe the spread of results, typically the standard deviation (or relative standard deviation), calculated from results obtained by carrying out replicate measurements on a suitable material under specified conditions. Deciding on the 'specified conditions' is an important aspect of evaluating measurement precision – the conditions determine the type of precision estimate obtained.

'Measurement repeatability' and 'measurement reproducibility' represent the two extreme measures of precision which can be obtained. Documentation of standardised methods (e.g. from ISO) will normally include both repeatability and reproducibility data where applicable.

Repeatability, expected to give the smallest variation in results, is a measure of the variability in results when a measurement is performed by a single analyst using the same equipment over a short timescale.[*]

Reproducibility, expected to give the largest variation in results, is a measure of the variability in results between laboratories.[†]

Between these two extremes, 'intermediate (measurement) precision' gives an estimate of the variation in results when measurements are made in a single laboratory but under conditions that are more variable than repeatability conditions. The exact conditions used should be stated in each case. The aim is to obtain a precision estimate that reflects all sources of variation that will occur in a single laboratory under routine conditions (different analysts, extended timescale, different pieces of equipment etc.).[‡]

6.6.2.1 Estimates of precision – general aspects

Precision is generally dependent on analyte concentration, and so should be determined at a number of concentrations across the range of interest. This could include a particular concentration of interest (such as a regulatory limit) plus concentrations at the limits of the measuring interval. If relevant, the relationship between precision and analyte concentration should be established. In cases where the measured concentration is well above the detection limit, the precision is often found to be proportional to analyte concentration. In such cases it may be more appropriate to express precision as a relative standard deviation since this is approximately constant over the range of interest.

For qualitative methods, precision cannot be expressed as a standard deviation or relative standard deviation, but may be expressed as true and false positive (and negative) rates [55] (see Section 6.2.6).

Evaluation of precision requires sufficient replicate measurements to be made on suitable materials. The materials should be representative of test samples in terms of matrix and analyte concentration, homogeneity and stability, but do not need to be CRMs. The replicates should also be independent, i.e. the entire measurement process, including any sample preparation steps, should be repeated. The minimum number of replicates specified varies with different protocols, but is typically between 6 and 15 for each material used in the study.

It should be kept in mind that it is difficult to estimate a reliable standard deviation from data sets with few replicates. If admissible, the values calculated from several small sets of replicate measurements can be combined (pooled) to obtain estimates with sufficient degrees of freedom.

[*] Repeatability is sometimes referred to as 'within-run', 'within-batch' or 'intra-assay' precision.

[†] In validation reproducibility refers to the variation between laboratories using the same method. Reproducibility may also refer to the variation observed between laboratories using different methods but intending to measure the same quantity [7].

[‡] Intermediate precision is sometimes referred to as 'within-laboratory reproducibility', 'between-run variation', 'between batches variation' or 'inter-assay variation'.

Certain experimental designs, analysed using analysis of variance (ANOVA), are an efficient way of obtaining estimates of repeatability and intermediate precision with a suitable number of degrees of freedom (see Section 6.6.4 and Annex C for further explanation of this approach). See Quick Reference 7 for information on experiments to assess precision.

6.6.3 Precision limits

From the standard deviation s it is useful to calculate a 'precision limit' [29, 48]. This enables the analyst to decide whether there is a significant difference, at a specified level of confidence, between results from duplicate analyses of a sample obtained under specified conditions. The repeatability limit (r) is calculated as follows:

$$r = \sqrt{2} \times t \times s_r \quad \text{(Eq. 5)}$$

where the factor $\sqrt{2}$ reflects the difference between two measurements, t is the two-tailed Student t-value for a specified number of degrees of freedom (which relates to the estimate of s_r) and at the required level of confidence. For relatively large numbers of degrees of freedom, $t \approx 2$ at the 95 % confidence level, so the repeatability limit is often approximated as:

$$r = 2.8 \times s_r \quad \text{(Eq. 6)}$$

The intermediate precision limit and the reproducibility limit (R) are calculated in a similar way, replacing s_r with s_I and s_R, respectively.

Documentation of standardised methods (e.g. from ISO) will normally include data for both the repeatability limit and reproducibility limit where applicable.

6.6.4 Simultaneous determination of repeatability and intermediate precision

Approaches to simultaneous determination of repeatability and intermediate precision are described in ISO 5725-3 [29]. In addition, a design based on the Harmonized guidelines for single-laboratory validation of methods of analysis [12] offers the possibility to determine repeatability and intermediate precision from a single study. Subsamples of the selected test material are analysed in replicate under repeatability conditions across a number of different runs, with maximum variation in conditions between the runs (different days, different analysts, different equipment, etc.). Via one-way ANOVA [5, 6], repeatability can be calculated as the within-group precision, while the intermediate precision is obtained as the square root of the sum of squares of the within-group and between-group precision. This type of design can provide an efficient way of obtaining sufficient degrees of freedom for estimates of repeatability and between-group precision. For example, 8 groups of 2 replicates leads to 8 and 7 degrees of freedom for the estimates of repeatability and between run precision, respectively. See further Annex C.

Quick Reference 7 – Repeatability, intermediate precision and reproducibility

What to do	How many times	What to calculate/determine from the data	Comments
Measure RMs, surplus test samples or spiked sample blanks at various concentrations across working range. Repeatability and intermediate precision can be determined from separate studies (see a) and b) below) or simultaneously in a single study (see c) below.			
a) Same analyst and equipment, same laboratory, short timescale.	6-15 replicates for each material.	Determine standard deviation *(s)* of results for each material.	Estimates repeatability standard deviation s_r for each material.[a]
b) Different analysts and equipment, same laboratory, extended timescale.	6-15 replicates for each material.	Determine standard deviation *(s)* of results for each material.	Estimates intermediate precision standard deviation s_I for each material.
c) Different analysts and equipment, same laboratory, extended timescale.	6-15 groups of duplicate measurements[b] obtained under repeatability conditions on different days/equipment for each material.	Calculate repeatability standard deviation from ANOVA results for each material. Calculate between-group standard deviation from ANOVA and combine with repeatability standard deviation for each material.	Estimates repeatability standard deviation s_r for each material. Estimates intermediate precision standard deviation s_I for each material.
d) Different analysts and equipment, different laboratories, extended timescale.	6-15 groups of duplicate measurements[b] obtained under repeatability conditions in different laboratories for each material.	Calculate repeatability standard deviation from ANOVA results for each material. Calculate between-laboratory standard deviation from ANOVA results and combine with repeatability standard deviation for each material.	Estimates repeatability standard deviation s_r for each material. Estimates reproducibility standard deviation s_R for each material. This requires a special inter-laboratory comparison ('collaborative trial').

[a] A repeatability standard deviation can also be estimated by pooling of several small data sets, e.g. n =2, from different days.

[b] Duplicate measurements within each group will provide a balanced number of degrees of freedom for the estimates of the within- and between-group standard deviations. Increasing the number of replicates per group will increase the number of degrees of freedom associated with the estimate of the repeatability.

6.7 Measurement uncertainty

A full discussion of (measurement) uncertainty is beyond the scope of this Guide but detailed information can be found elsewhere [21, 22]. Uncertainty is an interval associated with a measurement result which expresses the range of values that can reasonably be attributed to the quantity being measured. An uncertainty estimate should take account of *all recognised effects* operating on the result. The uncertainties associated with each effect are combined according to well-established procedures.

Several approaches to obtaining an uncertainty estimate for the results from chemical measurements are described [22, 66, 67, 68]. These take into account:

- the overall, long-term precision of the method (i.e. the intermediate precision or reproducibility);
- bias and its uncertainty, including the statistical uncertainty involved in the bias measurements, and the uncertainty in the reference value [69, 70, 71, 72, 73];
- equipment calibration. Uncertainties associated with calibration of equipment such as balances, thermometers, pipettes and flasks are often negligibly small in comparison to the overall precision and the uncertainty in the bias. If this can be verified then calibration uncertainties do not need to be included in the uncertainty estimate;
- any significant effects operating in addition to the above. For example, temperature or time ranges permitted by the method may not be fully exercised in validation studies, and their effect may need to be added. Such effects can be usefully quantified by ruggedness studies (see Section 6.8), or related studies which establish the size of a given effect on the result.

Where the contribution of individual effects is important, for example in calibration laboratories, it will be necessary to consider the individual contributions from all individual effects separately.

Note that, subject to additional consideration of effects outside the scope of a collaborative study, the reproducibility standard deviation forms a working estimate of combined standard uncertainty provided that the laboratory's bias, measured on relevant materials, is small with respect to the reproducibility standard deviation, the in-house repeatability is comparable to the standard method repeatability, and the laboratory's intermediate precision is not larger than the published reproducibility standard deviation [67].

6.8 Ruggedness

6.8.1 Definition

The 'ruggedness' ('robustness') of an analytical procedure is "a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters. Ruggedness provides an indication of the method's reliability during normal usage" [13].

6.8.2 Ruggedness test

In any method there will be certain stages which, if not carried out sufficiently carefully, will have a significant effect on method performance and may even result in the method not working at all. These stages should be identified, usually as part of method development, and if possible, their influence on method performance evaluated using a 'ruggedness test' ('robustness test'). The AOAC has defined this term and describes an established technique for how to carry out such a test using a Plackett-Burman experimental design [74].

A 'ruggedness test' involves making deliberate changes to the method, and investigating the subsequent effect on performance.[*] It is then possible to identify the variables in the method which have the most significant effect and ensure that, when using the method, they are closely controlled. Where there is a need to refine the method further, improvements can probably be made by concentrating on those parts of the method known to be critical.

The ruggedness of a procedure must be established for in-house developed methods, methods adapted from the scientific literature and methods published by standardisation bodies used outside the scope specified in the standard method. When methods published by standardisation bodies are used within the scope of the method, ruggedness will usually have been studied as part of the standardisation process. Therefore a ruggedness study is in most cases not necessary at the single-laboratory level. Information about ruggedness should be indicated in the laboratory procedure in the form

[*] The effect on the measurand is normally studied but an alternative is to investigate the effect on an experimental parameter, e.g. the peak resolution in a chromatogram.

of stated tolerance limits for the critical experimental parameters (See Example 5 and Quick Reference 8).

Example 5 – Extracts from ISO 11732 [58]. The instructions indicate the criticality of some experimental parameters.

- NH_4Cl dried to constant mass at 105 ± 2 °C.
- The given quantities can be reduced (e.g. by one tenth).
- Being stored in a plastic bottle (polyethylene) at room temperature, the solution is stable for about 1 month.
- The absorbance of the solution should be should be 0.3 – 0.5.
- Degas and purify the solution..., fill it into the reagent reservoir and let it stand for at least 2 hours.
- This solution may be stored in a refrigerator for at most one week.
- Containers of glass, polyalkenes or polytetrafluoroethylene (PTFE) are suitable for sample collection.
- In exceptional cases, the sample may be stored up to two weeks, provided the sample has been membrane-filtered after acidification.

Quick Reference 8 – Ruggedness

What to do	How many times	What to calculate/determine from the data	Comments
Identify variables which could have a significant effect on method performance. Set up experiments (analysing RMs or test samples) to monitor the effect on measurement results of systematically changing the variables.	Most effectively evaluated using experimental designs. E.g. 7 parameters can be studied in 8 experiments using a Plackett-Burman experimental design [74].	Determine the effect of each change of condition on the measurement results. Rank the variables in order of the greatest effect on method performance. Carry out significance tests to determine whether observed effects are statistically significant.	Design quality control or modify the method in order to control the critical variables, e.g. by stating suitable tolerance limits in the standard operating procedure.

7 Using validated methods

When using someone else's method, whether it is a method developed elsewhere within the laboratory, a published method, or even a standard or regulatory method, there are two issues which need to be considered.

Firstly is the existing validation data adequate for the required purpose or is further validation necessary? It should be noted that, in addition to the amount of information provided on the method performance, the reliability of the validation data sources is also an issue. Data obtained in collaborative studies or by recognised standardisation organisations are generally considered reliable, less so data published only in the scientific literature or provided by manufacturers of equipment and/or reagents. Secondly, if the existing validation data is adequate, is the laboratory able to verify the performance claimed possible in the method? (see Section 2.2). Are the available equipment and facilities adequate? If the method has been validated by extensive testing under all extremes of operating conditions, then a new competent analyst will probably operate satisfactorily within the existing performance data. However, this should always at least be checked. It is usually sufficient to test the analyst's ability to achieve the stated repeatability and to check for any bias, provided that the standard method is used within its scope. This is covered more fully below.

Standardised methods are generally produced by some form of collaborative study and the standardisation bodies which produce them frequently have statistical experts to help ensure that validation studies are correctly designed, performed and evaluated. The standard ISO 5725 [29] describes a model on which interlaboratory comparisons of methods should be based in order to provide reliable information on the method's performance. This model is increasingly applied, but not all standard methods have been subjected to it. It would be dangerous to assume that all standard methods have been properly validated and it is the analyst's responsibility to check whether or not the information provided on the method's performance is adequate.

Similarly, it is often assumed that standard methods can be used straight off the shelf and the published performance data achieved straight away by whoever uses the method. This is not a safe assumption. Even those who are familiar or expert in the particular field of chemistry covered by the method will need to practice before becoming fully proficient.

When using validated methods (or for that matter any method) the following rules are recommended to ensure that acceptable performance is achieved.

1. Firstly, the analyst should be completely familiar with a new method before using it for the first time. Ideally the method will first be demonstrated to the analyst by someone already expert in its use. The analyst should then use it under initially close supervision. The level of supervision will be stepped down until the analyst is deemed sufficiently competent to 'go solo'. For example competence might be established in terms of the analyst's ability to achieve the levels of performance stated in the method, such as repeatability, limit of detection, etc. This is typical of the way someone might be trained to use a new method and laboratory training procedures will frequently be designed in this way with objective measures in place to test competence at intervals during the training. In any case, the analyst should have read through the method and familiarised themselves with the theory behind the measurement, mentally rehearsing the various stages, identifying points where breaks can be taken, and parts of the process where the analyst is committed to continuous work. Where reagents need to be prepared, how stable are they once prepared? Do they need to be prepared in advance? A classic pitfall is to spend several hours preparing a number of samples and then finding the preparation of the reagent needed for the next stage of the work involves a complicated synthesis, in the meantime the samples themselves may be degrading.
2. Secondly, an assessment needs to be made of how many samples can be conveniently handled at a time. It is better to analyse a few samples well than to try to analyse a large number and have to repeat most of them.
3. Finally, make sure everything needed for the method is available before work is started. This involves gathering together the right equipment, reagents and standards (with any

attendant preparation), perhaps reserving space in fume hoods, etc.

If it is necessary to adapt or change someone else's validated method then appropriate revalidation will be necessary. Depending on their nature, the changes may well render the original validation data irrelevant.

8 Using validation data to design quality control

8.1 Introduction

'Quality assurance' (QA) and 'quality control' (QC) are terms whose meanings are often varied according to the context. According to ISO, quality assurance addresses the activities the laboratory undertakes to provide confidence that quality requirements will be fulfilled, whereas quality control describes the individual measures which are used to actually fulfil the requirements [9].

Method validation gives an idea of a method's capabilities and limitations which may be experienced in routine use while the method is in control. Specific controls need to be applied to the method to verify that it remains in control, i.e. is performing in the way expected. During the validation stage the method was largely applied to samples of known content. Once the method is in routine use it is used for samples of unknown content. Suitable internal QC can be applied by continuing to measure stable test samples, thus allowing the analyst to decide whether the variety of answers obtained truly reflects the diversity of samples analysed or whether unexpected and unwanted changes are occurring in the method performance. In practice these known samples should be measured with every batch of samples as part of the quality control process. The checks made will depend on the nature, criticality and frequency of the analysis, batch size, degree of automation and test difficulty, and also on the lessons learnt during development and validation processes. Quality control can take a variety of forms, both inside the laboratory (internal) and between the laboratory and other laboratories (external).

8.2 Internal quality control

Internal QC refers to procedures undertaken by laboratory staff for the continuous monitoring of operations and measurement results in order to decide whether results are reliable enough to be released [18, 75]. This includes replicate analysis of stable test samples, blanks, standard solutions or materials similar to those used for the calibration, spiked samples, blind samples and QC samples [76]. The use of control charts is recommended for monitoring of QC results [76, 77]. The QC adopted must be demonstrably sufficient to ensure the validity of the results. Different kinds of quality control may be used to monitor different types of variation within the process. QC samples, analysed at intervals in the analytical batch will indicate drift in the system; use of various types of blank will indicate what the contributions to the instrument signal besides those from the analyte are; duplicate analyses give a check of repeatability.

QC samples are typical samples which over a given period of time are sufficiently stable and homogeneous to give the same result (subject to random variation in the performance of the method), and available in sufficient quantities to allow repeat analysis over time. Over this period the intermediate precision of the method can be checked by monitoring values obtained from analysis of the QC sample, usually by plotting them on a control chart. Limits are set for the values on the chart (conventionally 'warning limits' are set at $\pm 2s$ about the mean value, and 'action limits' are set at $\pm 3s$ about the mean value). Provided the plotted QC values conform to certain rules pertaining to the set limits, the QC is deemed to be satisfactory. As long as the QC sample value is acceptable it is likely that results from samples in the same batch as the QC sample can be taken as reliable. The acceptability of the value obtained with the QC sample should be verified as early as practicable in the analytical process so that in the event of a problem, as little effort as possible has been wasted on unreliable analysis of the samples themselves.

During method validation initial estimates of different precision measures are obtained. In order to set realistic limits on the control chart, the measurements must reflect the way the method is actually intended to be used on a day-to-day basis. Thus measurements during validation should mimic all possible variations in operating conditions: different analysts; variations in laboratory temperature etc. If this is not done, then the standard deviation will be unrealistically small, resulting in limits being set on the chart which cannot possibly be complied with in normal use. For this reason, it is generally advised to reassess the stated limits after one year or when a sufficient number of results have been collected [76].

The use of various types of blanks enables the analyst to ensure that calculations made for the analyte can be suitably corrected to remove any contributions to the response which are not

attributable to the analyte. Replicate analysis of routine test samples provides a means of checking for changes in precision in an analytical process, which could adversely affect the result [78]. Replicates can be adjacent in a batch to check repeatability.

Analysis of blind samples is effectively a form of repeat analysis and provides a means of checking precision. It consists of replicated test portions placed in the analytical batch, possibly by the laboratory supervisor, and is so-called because the analyst is not normally aware of the identity of the test portions or that they are replicates. Thus the analyst has no preconceived ideas that the particular results should be related.

Standards or materials similar to those used for calibration, placed at intervals in an analytical batch, enable checks to be made that the response of the analytical process to the analyte is stable.

It is the responsibility of the laboratory management to set and justify an appropriate level of quality control, based on risk assessment, taking into account the reliability of the method, the criticality of the work, and the feasibility of repeating the analysis if it doesn't work correctly first time. It is widely accepted that for routine analysis, a level of internal QC of 5 % is reasonable, i.e. 1 in every 20 samples analysed should be a QC sample. However, for robust, routine methods with high sample throughput, a lower level of QC may be reasonable. For more complex procedures, a level of 20 % is not unusual and on occasions even 50 % may be required. For analyses performed infrequently, a full system validation should be performed on each occasion. This may typically involve the use of an RM containing a certified or known concentration of analyte, followed by replicate analyses of the sample and a spiked sample (a sample to which a known amount of the analyte has been deliberately added). Those analyses undertaken more frequently should be subject to systematic QC procedures incorporating the use of control charts and check samples.

8.3 External quality control

Regular participation in proficiency testing (PT), also known as external quality assessment (EQA) is a recognised way for a laboratory to monitor its performance against both its own requirements and the norm of peer laboratories. PT helps to highlight variation between laboratories (reproducibility), and systematic errors (bias).

PT schemes and other types of interlaboratory comparison are accepted as being an important means of monitoring the degree of equivalence of analytical results at national and international level. Accreditation bodies recognise the benefit of these schemes and strongly encourage laboratories to participate in PT/EQA as an integral part of their quality management [79]. It is important to monitor PT results as part of the QC procedures and take action as necessary.

In certain instances, accreditation bodies may specify participation in a particular PT scheme as a requirement for accreditation. The value of PT is of course only as good as the schemes themselves. Requirements for the competence of PT providers are described in the standard ISO/IEC 17043 [80]. Practical information on how to select, use and interpret PT schemes is presented in a Eurachem Guide [81]. Information about a large number of schemes can be found in the EPTIS database (www.eptis.bam.de). However, for emerging fields of analysis or rare applications in particular, there may be no scheme that is fully appropriate. These and other limitations are now considered in a recent guidance document [82] that requires accredited laboratories to derive a strategy for their participation in PT.

9 Documentation of validated methods

9.1 From draft to final version

The method subject to validation, is performed using a documented procedure which should be considered a draft until the validation report is approved. Once the validation process is complete it is important to document the analytical procedure so that the method can be clearly and unambiguously implemented. There are a number of reasons for this.

- The various assessments of the method made during the validation process assume that, in use, the method will be used in the same way each time. If it is not, then the actual performance of the method will not correspond to the performance predicted by the validation data. Thus the documentation must limit the scope for introducing accidental variation to the method.
- Proper documentation is also necessary for auditing and evaluation purposes and may also be required for contractual or regulatory reasons.
- Appropriate documentation of the method will help to ensure that application of the method from one occasion to the next is consistent. Since the quality of documentation has a direct effect on how consistently the method can be applied, it is likely to have an influence on the precision and measurement uncertainty. In fact, the uncertainty contribution associated with inadequately documented methods could be so large that it effectively makes the method useless. Any anomalies in the documentation must be resolved before a sensible estimate of the uncertainty can be obtained.

9.2 Recommendations

9.2.1 Checking the instructions

It is not easy to document a method properly. Information should appear in roughly the order that the user will be expected to need it. A common trap is to assume that everyone will understand the mechanics of the method to the same extent as the person who has developed and documented it. This assumed knowledge can be dangerous. A useful way to test the documentation is for a competent colleague to work through the documentation exactly in the way described. If this corresponds to what was intended then the documented method should stand up well to use by a variety of analysts and deliver consistent results. If not then redrafting is necessary to describe the procedures in more detail and reduce ambiguity.

9.2.2 Recommendations in standards

A number of standards provide guidance on what type of information should be included when documenting a method. From the chemists' point of view probably the most useful are the ISO 78 series, which describe the documentation of a number of different types of chemical analysis methods (standardisation bodies produce, validate and of course document a large number of methods each year, and need as consistent an approach as possible and produce these standards principally for the benefit of their own technical committees). ISO 78-2 [83] advises on method documentation for general chemical methods. A layout based around this standard is included in Annex A. The standards indicate a logical order for material with recommended headings and advice on the information which should appear under each heading. When using these standards the reader should note the need to balance flexibility of approach against consistency. Whilst it is desirable that all methods should have the same document format, it should also be recognised that not all methods warrant the same degree of detail and frequently it will be appropriate to omit some of the recommended sections from the documentation.

9.2.3 Document control

A laboratory documenting its own methods may well benefit from developing a 'house style'. As well as presenting relevant information in a logical easy-to-use way, it also enables the burden of the documentation work to be spread across a number of authors. Drafts generated by a number of authors can be checked for consistency using a single checking authority.

Documented methods form an important part of a laboratory's quality management system and should be subject to an appropriate degree of document control. The purpose of this is to ensure that only methods and procedures which have been authorised as fit for use are actually used. Therefore, as part of the documentation process, methods should carry information which enables the user to judge whether the method has been authorised for use and whether it is complete. Other information should be available

regarding the version number and date of the method; the author; how many copies of the method exist; and any copying restrictions.

From time to time methods may require updating. The technology underpinning the procedure may have been improved, for example. Document control enables the smooth withdrawal of obsolete methods and issue of revised methods. These days the process of document control is greatly simplified using specific software. Changes should be made only by those so authorised. This may be controlled in the software where the relevant files may have widespread 'read-only' access and very limited 'write' access.

10 Implications of validation data for calculating and reporting results

It is important that the analyst is able to translate the data, generated during analysis of samples using the validated method, into results which directly contribute to solving the customer's problem. The performance characteristics established during the validation process help to do this. Data for repeatability, intermediate precision and reproducibility can be used to establish whether differences found when analysing samples are significant. Quality controls based on the validation data can be used to confirm that the method is in control and producing meaningful results. Estimation of the measurement uncertainty enables expression of the result as a range of values with an accepted level of confidence.

It is important that the analyst has access to validation data which can be used to support the validity of the results. Whether or not such information is passed to the customer is another matter. Very often the customer will not have the technical skills to appreciate the significance of the data. In such circumstances it is perhaps safer to make the data available on request.

Issues such as method validation, variability and measurement uncertainty need to be treated carefully in certain circumstances, for example in legal or forensic contexts. It may be better to be open about the existence of uncertainty attached to measurements and be prepared to justify decisions made in the light of knowing that uncertainty.

Care needs to be taken when trying to use an analytical result with its accompanying uncertainty to try to decide whether or not the original consignment from which the sample has been taken complies with a specification or limit [84]. Such a decision may not be the responsibility of the analyst, although the analyst may be required to provide technical advice to assist in the decision making process.

When reporting results, the analyst must decide whether to correct for any biases which may have been detected or to report results uncorrected but acknowledge the existence of the bias.

Care should be taken when reporting results as 'not detected'. On its own this statement is uninformative and should be accompanied by an explanation of what the limit of detection is in that instance. Sometimes it is appropriate to report a numerical value even though this may be below the apparent limit of detection. Authorities may sometimes request that the limit of quantification be stated.

Where a statement of uncertainty is required with the result, it may be appropriate to quote an expanded uncertainty by applying a suitable coverage factor. For example, a coverage factor of 2 corresponds to an interval with a level of confidence of approximately 95 %. For further guidance on how to report measurement uncertainty, see Section 9 in the Eurachem/CITAC Guide [22].

Annex A – Method documentation protocol

The adequate documentation of methods is discussed in Section 9 of the Guide. The following format is included for reference as a suitable layout. It is based on ISO 78-2 [83], but contains some additional advice on calibration, quality control, and document control. Annex A is for guidance only and should be adapted to suit any special requirements.

A.1 Foreword

A.1.1 Update and review summary

This section has a twofold purpose. Firstly, it is intended to enable minor changes to be made to the text of the method without the need for a full revision and reprint of the method. Secondly, it is recommended that every method should be reviewed for fitness-for-purpose periodically and the summary serves as a record that this has been done. The summary typically would be located at the front of the method, just inside the front cover.

A.1.2 Updates

Any hand written changes to the text of the method would be accepted provided the changes were also recorded in the table below (hand-written entries acceptable) and appropriately authorised. It would be implicit that the authorisation endorsed the fact that the effects of the changes on the method validation had been investigated and caused no problems, and that the changes had been made to all copies of the method.

#	Section	Nature of amendment	Date	Authorisation
1 (e.g.)	3.4	Change flow rate to 1.2 ml min^{-1}	8/2/96	DGH

A.1.3 Review

At any given time it would be expected that the date at which a method was seen to be in use would be between the *review* and *next review* dates, as shown in the table.

Review date	Outcome of review	Next review date	Authorisation

A.2 Introduction

The introduction is used, if necessary, to present information, such as comments concerning the technical content of the procedure or the reasons for its preparation. If background information on the method is required, it should preferably be included in this clause.

A.3 Title

The title shall express the sample types to which the test method applies, the analyte or the characteristic to be determined and the principle of the determination. It should be limited, wherever possible, to the following information. Preferred format:

Determination of A{*analyte or measurand*} (in the presence of B{*interference*}) in C {*matrix*} using D {*principle*}.

A.4 Warnings

Draw attention to any hazards and describe the precautions necessary to avoid them. Detailed precautions may be given in the relevant sections, but notice must be drawn to the existence of hazards and need for precautions here. Provide suitable warnings of any hazards involved with:

- handling the samples;
- handling or preparing solvents, reagents, standards, or other materials;
- operation of equipment;

- requirements for special handling environments, e.g. fume cupboards;
- consequences of scaling up experiment (explosion limits).

A.5 Scope

This section enables a potential user to see quickly whether the method is likely to be appropriate for the desired application, or whether limitations exist. The following details should be covered:

- a description of the underlying problem (why the method is needed);
- the analyte(s) or measurand(s) which can be determined by the method;
- the form in which analyte(s) is determined – speciation, total/available etc.;
- the sample matrix(es) within which those analyte(s) may be determined;
- a working range (measuring interval) over which the method may be used. This should refer to properties, e.g. concentrations, in the laboratory sample;
- known interferences which prevent or limit the use of the method;
- the instrumental technique used in the method;
- the minimum sample size.

The food sector [35] uses the concept 'applicability' as a synonym for scope and defines it as "the analytes, matrices, and concentrations for which a method of analysis may be used satisfactorily".

A.6 (Normative) references

This clause shall give a list of those documents which are necessary for the application of the method. Documents which have merely served as references in the preparation of the method shall be indicated in a bibliography at the end of the document.

A.7 Definitions

Give any definitions of terms used in the text that may be necessary for its complete understanding. Use ISO definitions wherever possible. Quote sources. Analytical structures can be included here if relevant.

A.8 Principle

Outline the essential steps of the method, the principle by which the analytical technique operates. A flow chart or cause-and-effect diagram may help. This section should be written so as to allow an at-a-glance summary of how the method works. Include an explanation on the principle of the calculation. Where appropriate to clarify the working of the method or calculations, include details of any relevant chemical reactions (for example, this may be relevant where derivatisation is involved, or in titrimetry).

E.g. "The concentration is derived from a 6 point calibration curve by reading off the concentration, corresponding to the sample absorbance, corrected for the blank value, and multiplying it by the concentration factor."

A.9 Reactions

This clause shall indicate the essential reactions, if they are considered necessary for the comprehension of the text or the calculations. They justify the calculations made from the data obtained in the determinations and may lead to a better understanding of the method, especially if several successive changes occur in the state of oxidation of the element being determined. When titrations are involved, they are particularly useful in indicating the number of equivalents in each mole of reactant.

A.10 Reagents and materials

List all reagents and materials required for the analytical process, together with their essential characteristics (concentration, density, etc.) and numbered for later reference. List:

- Chemical Abstract Service (CAS) Registry numbers (if available);
- details of any associated hazards including instructions for disposal;
- analytical grade or purity;
- need for calibration and QC materials to come from independent batches;

- details of preparation, including need to prepare in advance;
- containment and storage requirements;
- shelf life of raw material and prepared reagent;
- required composition with notes of type of concentration or other quantity;
- labelling requirements.

A.11 Apparatus

Describe individual equipment and how they are connected in sufficient detail to enable unambiguous set-up. Number the items for later reference. Diagrams and flowcharts may assist clarity. Any checking of the functioning of the assembled apparatus shall be described in the "Procedure" clause in a subclause headed "Preliminary test" or "Check test" (see A.13).

List minimum performance requirements and verification requirements, cross-referenced to the calibration section (A.13) and any relevant instrument manuals. If appropriate, refer to International Standards or other internationally acceptable documents concerning laboratory glassware and related apparatus. Include environmental requirements (fume cupboards etc.).

A.12 Sampling

The sampling in this protocol includes both the sampling to obtain the laboratory sample and the subsampling in the laboratory to obtain the test sample from which the test portion will be drawn.

If sampling for the preparation of the laboratory sample is independent of the chemical analysis as such, it is generally sufficient to refer informatively to the relevant procedure dealing specifically with this question. If no such relevant procedure exists, the sampling clause may include a sampling plan and sampling procedure, giving guidance on how to avoid alteration of the product and taking into account requirements concerning the application of statistical methods.

The sampling clause should give all the information necessary for the preparation of the test sample from the laboratory sample. Include storage, conditioning/pretreatment and disposal details. If this stage is particularly complicated, a separate document describing individual steps may be justified.

A.13 Procedure

Describe each sequence of operations. If the method to be described is already given in another standard, the phrase "use the method specified in ISO 12345" or "use one of the methods specified in ISO 12345" shall be used, with an indication of any modification, if necessary. Mention operations for which special safety precautions are necessary. The 'Procedure' clause shall normally include subclauses on the following.

- test portion (its preparation from the test sample or laboratory sample and the required mass or volume);
- blank tests (conditions and limitations);
- preliminary test or check test (e.g. to verify the performance of a measuring instrument);
- determination(s) or test(s). This includes mentioning the number of measurements or tests (e.g. duplicate) and detailed description of all steps;
- calibration. Identify the critical parts of the analytical process. These will have to be controlled by careful operation and calibration. Cross-reference to the relevant sections above. Include calibration of equipment – what needs to be calibrated, how, with what, and how often? Consider appropriate metrological traceability of calibrants.

A.14 Calculation

Describe how the result(s) are calculated. Include information about the units in which the result and other quantities are to be expressed; the equation used for the calculation; the meanings of the algebraic symbols used in the equation; the number of decimal places or significant figures to which the result is to be given. The symbols of quantities shall be in accordance with ISO 80000 [14].

A.15 Precision

For methods that have been subjected to an interlaboratory comparison, the precision data (i.e. the repeatability and reproducibility) shall be indicated. The precision data shall be calculated, and should preferably also be published, in accordance with the relevant part of ISO 5725 or in accordance with

another suitable International Standard (which shall be referenced). Clearly state whether the precision values are expressed in absolute or relative terms, or as precision limits.

A.16 Quality assurance and quality control

One outcome from the validation exercise should be a description of the internal and external (proficiency testing) quality control procedures to follow. Explain what form the quality control takes, frequency of quality control checks during batch analysis, pass/fail criteria, action to take in the event of a failure. Cross-reference to the relevant sections above.

A.17 Special cases

Include any modifications to the procedure necessitated by the presence or absence of specific components in the product to be analysed. The modifications shall already have been referred to in the "Scope" clause. Each special case shall be given a different title.

A.18 Test report

This clause should specify the information to be given in the test report. The following aspects of the test should normally be included.

- a reference to the method used;
- the result(s) and an indication of the associated quality (precision, specified uncertainty; confidence interval) if applicable, including a reference to the "Calculation" clause;
- any deviations from the procedure;
- any unusual features observed;
- the date of the test.

A.19 Annexes

To improve readability, some information is more conveniently presented in an annex. It shall be clearly stated whether the annex is normative or informative. Examples of information which can be annexed are data from the method validation work, risk analysis and uncertainty calculations. For the latter, the major sources of uncertainty relating to the method should be identified and the assigned values listed. Insignificant contributions not used in the final calculation should be mentioned. The combined standard uncertainty and/or the expanded uncertainty should be listed together with an explanation of how it was derived. A more detailed treatment may be in a cross-referenced file.

A.20 Bibliography

If informative references are considered necessary, these may be given at the point in the text at which they are referred to or, if there are several, in a bibliography at the end of the document.

Annex B – Statistical basis of limit of detection calculations*

Quick Reference 2 in Section 6.2.3 indicated that the limit of detection (LOD) can be calculated by multiplying a suitable standard deviation by a factor of 3. This Annex describes the statistical basis for this factor.

The aim when determining the LOD is typically to establish the lowest concentration of the analyte present in a sample that can be detected, using a given measurement procedure, with a specified level of confidence. Defining the LOD is a two-step process. First a 'critical value' is established. This value is set so that the probability of obtaining a measurement result that exceeds the critical value is no greater than α, if a sample actually contains *none* of the analyte. The critical value sets a criterion for declaring a sample to be 'positive'. A false positive probability of $\alpha = 0.05$ is generally used; this leads to a critical value of approximately $1.65s$ (where s is the standard deviation of a large number of results for a blank sample or a sample containing a low concentration of the analyte, and 1.65 is the one-tailed Student t-value for infinite degrees of freedom at a significance level, $\alpha = 0.05$). The critical value is most conveniently expressed in terms of concentration, though in principle it may be any observation, such as peak area. Any result exceeding the critical value should be declared positive.

However, if the true value for the concentration in a sample were exactly equal to the critical value (expressed in terms of concentration), approximately half of the measurement results would be expected to fall below the critical value, giving a false negative rate of 50 %. A false negative rate of 50 % is obviously too high to be of practical use; the method does not reliably give results above the critical value if the concentration is equal to the critical value. The LOD is intended to represent the true concentration for which the false negative rate is acceptable given the critical value. The false negative error, β, is usually set equal to the false positive error, this is largely for historical reasons (IUPAC recommends default values of $\alpha = \beta = 0.05$ [49]). Using $\alpha = \beta = 0.05$, the LOD needs to be $1.65s$ above the value specified for the critical value. The factor for calculating the LOD with $\alpha = \beta = 0.05$ is thus $1.65+1.65 = 3.30$. This is frequently rounded to give the '$3s$' calculation shown in Quick Reference 2. This approach is based on several approximations which are described in the literature [49].

The multiplier of 3, as calculated in the previous paragraph, arises from the one-tailed Student t-value for infinite degrees of freedom, rounded down to one significant figure. For a statistically rigorous estimate of the LOD, the multiplying factor used should take into account the number of degrees of freedom associated with the estimate of s. For example, if s is obtained from 10 replicate measurements, the Student t-value at $\alpha = 0.05$ is 1.83 (9 degrees of freedom). This leads to an LOD calculated as $3.7s$.

* The text is based on the Eurachem Guide on Terminology in Analytical Measurement [8].

Annex C – Analysis of variance (ANOVA)

The central idea behind 'analysis of variance' (ANOVA) is that where a set of replicate data can be grouped in some way, e.g. by analyst, instrument, day, laboratory, method etc., the total variation in the whole set can be represented as the combination of the variances (s^2) between and within the groups. ANOVA can be used to evaluate results from the type of experimental study shown in Figure C 1. In this 'nested design', replicate measurements (typically obtained under repeatability conditions) are repeated in different measurement runs to provide p groups of data. To estimate intermediate precision from such a study there should be maximum variation in conditions between the runs (different days, analysts, etc.).

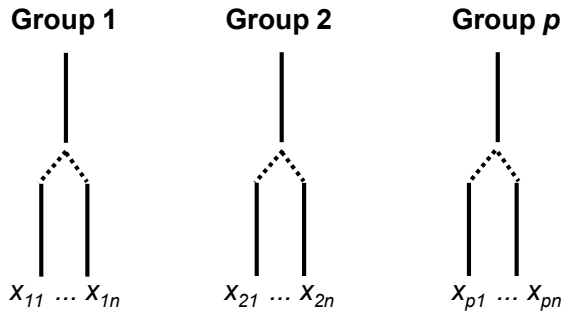


Figure C 1 – Example of a 'nested design' for an experiment from which different precision measures can be evaluated using ANOVA

The general form of a table for one-way ANOVA, for a total of N results in p groups of n observations, and with ν degrees of freedom, is shown in Figure C2. Each line of the table relates to a different source of variation. The first row relates to variation between the means of the groups; the second describes the variation within the groups and the third describes the variation of the data set as a whole. Spreadsheet programmes and statistical software also provide the F and F critical value, and corresponding P (probability) value.

Source of variation	**Sum of squares (SS)**	**ν**	**Mean square (MS)**	***F***	***P***	F_{crit}
Between groups	SS_b	p-1	$MS_b = SS_b/(p\text{-}1)$	MS_b/MS_w		
Within group (residuals)	SS_w	N-p	$MS_w = SS_w/(N\text{-}p)$			
Total	$SS_{tot} = SS_b+SS_w$	N-1				

Figure C2 – Anatomy of a table for a one-way ANOVA

The values related to the between-group variation are almost always either referred to as 'between-group' terms or are identified by the grouping factor (e.g. analyst, day or laboratory). Several different terms are used in software, textbooks etc. to describe the within-group variation – 'within-group', 'residual', 'error' or 'measurement' being the most common.

Assuming that the nested design shown in Figure C 1 is executed by a single laboratory, that the replicates within each group were obtained under repeatability conditions, and that the analytical conditions were varied between the groups, the repeatability and intermediate precision can be calculated as follows.

1. The repeatability standard deviation s_r, is obtained by taking the square root of the within-group mean square term which represents the within-group variance:

$$s_r = \sqrt{MS_w} \qquad \text{(Eq. C1)}$$

2. The contribution to the total variation from the grouping factor (s_{between}) is also obtained from the ANOVA table:

$$s_{\text{between}} = \sqrt{\frac{MS_{\text{b}} - MS_{\text{w}}}{\text{n}}} \qquad \text{(Eq. C2)}$$

3. The intermediate precision s_I can now be calculated by combining the within- and between-group variance components above:

$$s_I = \sqrt{s_r^2 + s_{\text{between}}^2} \qquad \text{(Eq. C3)}$$

The experiment referred to in Section 6.6.4 can be illustrated as follows. As part of a method validation exercise in a single laboratory, duplicate measurements were carried out during each of eight days (Table C1). The measurements on each day were performed under repeatability conditions but with different analysts, different equipment etc. on the different days, in order to mimic the conditions under which the method will be used routinely.

Table C1 – Example of experimental set-up that enables repeatability and intermediate precision to be evaluated using one-way ANOVA with acceptable degrees of freedom

Day:	1		2		3		4		5		6		7		8	
Result:	$x_{1,1}$	$x_{1,2}$	$x_{2,1}$	$x_{2,2}$	$x_{3,1}$	$x_{3,2}$	$x_{4,1}$	$x_{4,2}$	$x_{5,1}$	$x_{5,2}$	$x_{6,1}$	$x_{6,2}$	$x_{7,1}$	$x_{7,2}$	$x_{8,1}$	$x_{8,2}$

A one-way ANOVA can be used to separate the variation inherent within the method (repeatability) and the variation due to differences in the measurement conditions, i.e. different analysts, equipment, extended timescale (intermediate precision). Note that with this approach, it is not possible to draw conclusions about which of the parameters – analyst, equipment, time – contributes most to the intermediate precision but this is normally not needed at the validation stage.

Applying a one-way ANOVA to the results in Table C1 will provide a results table similar to that in Figure C2. The *F,* critical *F* and *P* values allow direct conclusions to be drawn on whether the variation between results obtained on different days is significantly greater than the variation in results obtained on the same day. The values for the two precision measures (s_r and s_I) are then readily calculated from Eq. C1 – Eq. C3 above. The associated number of degrees of freedom (ν) will be N-p = 16-8 = 8 for s_r. The value of ν for the intermediated precision is more complex but will not be smaller than p-1, i.e. 7 in this example (see Figure C2). This results in a reasonable compromise between workload and the uncertainty of the precision estimates.

Annex D – Notes on qualitative analysis

Qualitative analysis follows the basic principles of quantitative analysis but unique concepts need to be applied when describing the properties of the method and in the interpretation of the results. This appendix introduces qualitative analysis briefly and points to relevant guidance.

Qualitative analysis is defined by IUPAC as: *analysis in which substances are identified or classified on the basis of their chemical or physical properties, such as chemical reactivity, solubility, molecular weight, melting point, radiative properties (emission, absorption), mass spectra, nuclear half-life, etc.* [17]. This means that results are expressed on a nominal scale, which is inferior to expressing results on a ratio scale. Qualitative analysis, instead of quantitative analysis, is therefore recommended primarily for screening purposes using low-cost methods or at analyte concentrations near to the limit of detection (LOD).

A 'qualitative method' gives effectively a 'Yes'/'No' answer at a given cut-off concentration of an analyte [55]. Validation involves identification of the cut-off concentration in order to **classify/diagnose a condition**, e.g. the presence or absence of a polluting agent in water where there is a directive, law etc. defining which cut-off concentration applies.

In order to characterise the properties of a qualitative method, a quantitative method with superior metrological properties (confirmatory method), e.g. lower LOD, is optimal in order to determine the true state of with- or without a condition. Properties of the qualitative method should be determined at a number of concentrations, below, at and above the cut-off concentration. The use of a confirmatory quantitative method is preferable to the use of spiked and non-spiked blank samples.

For qualitative methods, precision cannot be expressed as a standard deviation or relative standard deviation, but may be expressed as true and false positive rates, and true and false negative rates [55, 85, 86, 87]. This is illustrated in Figure D1.

	Samples above cut-off	**Samples below cut-off**	
Positive test	**True positive tests**	**False positive tests (type I error)**	**Total number of positive tests**
Negative test	**False negative tests (type II error)**	**True negative tests**	**Total number of negative tests**
	Total number of samples above cut-off	**Total number of samples below cut-off**	

Figure D1 – A 2 × 2 table serving as the basis for calculating false positive and false negative rates

The 'diagnostic sensitivity' is the proportion of samples with a condition, e.g. concentration above cut-off, which have positive qualitative test results. The diagnostic sensitivity is a fundamental feature of a qualitative method, which expresses its ability to detect small amounts of the analyte in a sample to produce the binary Yes/No response at a predefined level of probability.

$$Diagnostic\ sensitivity = \frac{number\ of\ true\ positive\ samples}{total\ number\ of\ samples\ with\ condition} \qquad \text{(Eq. D1)}$$

The 'diagnostic specificity' is the proportion of samples without a condition, e.g. concentration below cut-off, which have negative qualitative test results

$$Diagnostic\ specificity = \frac{number\ of\ true\ negative\ samples}{total\ number\ of\ samples\ without\ condition} \qquad \text{(Eq. D2)}$$

Data from a confirmatory method comparison should be used if available. Otherwise, spiked and non-spiked blank samples can be measured.

The important parameters for the measurement quality in qualitative analysis are the LOD and the cut-off limit (Figure D2). The LOD is similarly defined as in quantitative analysis; the concentration of an analyte which provides a signal that can be statistically distinguished from the mean signal of relevant blank samples. The cut-off limit, if correctly determined, is where false negative rates for concentrations above the limit are low – with a stated probability. In the validation the proposed cut-off limit given in the documented procedure is assessed.

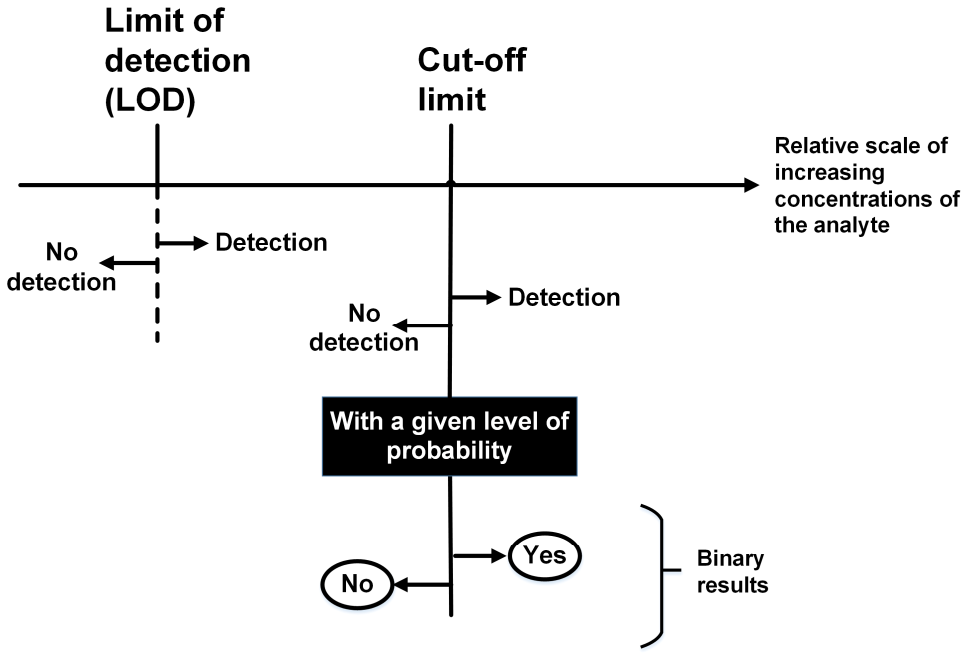


Figure D2 – There are two quantitative references that produce a binary response in the sample qualification/classification type of qualitative analysis: 1. The limit of detection (LOD) which is inherent to the method, 2. The cut-off limit given in the documented procedure. They are placed in an imaginary increasing concentration scale. In the detection zone, above the detection limit, the cut-off limit allows one to distinguish concentration zones of the component in which the correct binary response is produced: i.e. No below the limits and Yes above them.

Several additional concepts are used in qualitative analysis (Table D1). The **predictive values** of the results can be increased by increasing the prevalence of concentration above cut-off in the samples tested by the qualitative method, e.g. by other sources of information than the qualitative chemical method. This will substantially improve the practical value of the qualitative measurement method.

The **selectivity** of a qualitative method is an ordinal concept: the extent to which analytes other than the one included in the specification interferes with the analysis. This fundamental feature of the method can also be defined as its ability to produce results which are not influenced by matrix effects. The better the selectivity, the better the certainty of identity and sample classification.

Table D1 – Definition and calculation of concepts describing the diagnostic properties of measurement methods, including qualitative measurement methods

Concept (symbol)	Description	Formula
Positive likelihood ratio (*LR*+)	The ratio of the true positive rate to the false positive rate.	$LR+= \frac{dignostic\ sensitivity}{1 - diagnostic\ specificity}$
Negative likelihood ratio (*LR*-)	The ratio of the false negative rate to the true negative rate.	$LR-= \frac{1 - diagnostic\ sensitivity}{diagnostic\ specificity}$
Diagnostic odd ratio (*DOR*)	This combines the concepts of diagnostic sensitivity, diagnostic specificity and likelihood ratios into a single number.	$DOR = \frac{LR+}{LR-}$
Positive predictive value (*PPV*)	The proportion of the samples with a positive qualitative test result which have the condition. This takes into account the prevalence of the condition in the target population of samples.	$PPV = \frac{Number\ of\ true\ positives}{Total\ number\ of\ positives}$
Negative predictive value (*NPV*)	The proportion of the samples with negative qualitative test results which do not have the condition. This takes into account the prevalence of the condition in the target population of samples.	$PPV = \frac{Number\ of\ true\ negatives}{Total\ number\ of\ negatives}$

Bibliography

(For update of current most important references please refer to the Eurachem Reading List placed under Publications at the Eurachem website, www.eurachem.org.)

1. ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories, ISO Geneva.
2. ISO 15189:2012 Medical laboratories – Requirements for quality and competence, ISO Geneva.
3. ISO 15195:2003 Laboratory medicine – Requirements for reference measurement laboratories, ISO Geneva.
4. J. N. Miller, Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. A review, Analyst, 1991, 116, 3.
5. J. C. Miller, J. N. Miller, Statistics and chemometrics for analytical chemistry, 6th ed., Pearson, Harlow, 2010, ISBN 978-0-273730422.
6. S. L. R. Ellison, V. J. Barwick, T. J. Duguid Farrant, Practical statistics for the analytical scientist. A bench guide, 2nd ed., RSC Publishing, Cambridge, 2009, ISBN 978-0-85404-131-2.
7. International vocabulary of metrology – Basic and general concepts and associated terms (VIM), JCGM 200:2012, www.bipm.org. A previous version is published as ISO/IEC Guide 99:2007, ISO Geneva.
8. V. J. Barwick, E. Prichard (eds.), Eurachem Guide: Terminology in analytical measurement – Introduction to VIM 3, Eurachem, 2011, ISBN 978-0-948926-29-7, www.eurachem.org.
9. ISO 9000:2005 Quality management systems – Fundamentals and vocabulary, ISO Geneva.
10. ISO 9001:2008 Quality management systems – Requirements, ISO Geneva.
11. ISO online browsing platform (OBP), https://www.iso.org/obp/ui/.
12. M. Thompson, S. L. R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC technical report), Pure Appl. Chem., 2002, **74**(5), 835.
13. Validation of analytical procedures: Text and methodology Q2(R1), ICH harmonised tripartite guideline, 2005, www.ich.org.
14. ISO 80000-1:2009 Quantities and units – Part 1: General, ISO Geneva.
15. M. H. Ramsey and S. L. R. Ellison (eds.), Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches, Eurachem, 2007, ISBN 978-0-948926-26-6, www.eurachem.org.
16. AMC technical brief No. 19, March 2005, M. Thompson (ed.), Terminology – the key to understanding analytical science. Part 2: Sampling and sample preparation, www.rsc.org.
17. Compendium of chemical terminology (IUPAC Gold Book), www.iupac.org.
18. Compendium of analytical nomenclature (IUPAC orange book), www.iupac.org.
19. Method validation of U.S. Environmental Protection Agency microbiological methods of analysis. Prepared for The EPA forum on environmental measurements (FEM). The FEM Microbiology Action Team, FEM Document Number 2009-01, 7 Oct., 2009.
20. ISO 10012:2003 Measurement management systems - Requirements for measurement processes and measuring equipment, ISO Geneva.
21. Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM), JCGM 100:2008 (corrected version 2010), www.bipm.org. Printed as ISO/IEC Guide 98-3:2008, ISO Geneva.
22. S. L. R. Ellison, A. Williams (eds.), Eurachem/CITAC Guide CG4: Eurachem/CITAC, Quantifying uncertainty in analytical measurement, 3rd ed., Eurachem, 2012, www.eurachem.org.
23. Guide to method validation for quantitative analysis in chemical testing laboratories, INAB Guide PS15, 3 April 2012, www.inab.ie.

24. CLSI, User verification of performance for precision and trueness; Approved guideline – 2nd ed. CLSI document EP15-A2. Wayne PA, Clinical and Laboratory Standards Institute 2005, www.clsi.org.
25. AOAC Guidelines for collaborative study procedures to validate characteristics of a method of analysis, 2002, www.aoac.org.
26. Protocol for the design, conduct and interpretation of method-performance studies, (IUPAC technical report), Pure Appl. Chem., 1995, **67**(2), 331.
27. ASTM E1601-12 Standard practice for conducting an interlaboratory study to evaluate the performance of an analytical method, 2012, www.astm.org.
28. CEN/TR 10345:2013 Guideline for statistical data treatment of inter laboratory tests for validation of analytical methods, CEN Brussels.
29. ISO 5725 Accuracy (trueness and precision) of measurement methods and results – Parts 1-6, ISO Geneva.
30. ISO Guide 30:1992/Amd 1:2008 Terms and definitions used in conjunction with reference materials, ISO Geneva.
31. M. Thompson, P. J. Lowthian, Notes on statistics and data quality for analytical chemists, Imperial College Press, 2011, ISBN 978-1848166172.
32. E. Mullins, Statistics for the quality control chemistry laboratory, RSC, Cambridge, 2003, ISBN 978-0-854074-671-3.
33. W. Funk, V. Dammann, G. Donnevert, Quality assurance in analytical chemistry: Applications in environmental, food, and materials analysis, biotechnology, and medical engineering, 2nd ed., Wiley-VCH, Weinheim, 2006, ISBN 978-3-527-31114-9.
34. A. Kallner, Laboratory statistics. Handbook of formulas and terms (1st ed.), Elsevier, 2013, ISBN 978-0-12-416971-5.
35. Codex Alimentarius Commission, Procedural manual 21st ed., 2013.
36. Council Directive 98/83/EC (3 November 1998) on the quality of water intended for human consumption.
37. Commission Directive 2009/90/EC (31 July 2009) laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status.
38. Commission Decision 2002/657/EC (12 August 2002) implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results.
39. SANCO/12571/2013 (19 Nov. 2013) Guidance document on analytical quality control and validation procedures for pesticide residues analysis in food and feed.
40. AMC technical brief No. 17, July 2004, M. Thompson (ed.), The amazing Horwitz function, www.rsc.org.
41. Selectivity in analytical chemistry (IUPAC recommendations 2001), Pure Appl. Chem., 2001, **73**(8), 1381.
42. NATA – Technical report #17 – Guidelines for the validation and verification of quantitative and qualitative methods, 2012.
43. E. Theodorsson, Validation and verification of measurement methods in clinical chemistry, Bioanalysis, 2012, **4**(3), 305.
44. AMC technical brief No. 37, March 2009, M. Thompson (ed.), Standard additions: myth and reality, www.rsc.org.
45. ISO 11843-1:1997/Cor 1:2003 Capability of detection – Part 1: Terms and definitions, ISO Geneva.
46. ISO 11843-2:2007 Capability of detection – Part 2: Methodology in the linear calibration case, ISO Geneva.

47. ISO 11843-3:2002 Capability of detection – Part 3: Methodology for determination of the critical value for the response variable when no calibration data are used, ISO Geneva.
48. ISO 3534 Statistics – Vocabulary and symbols – Parts 1-3, ISO Geneva.
49. Nomenclature in evaluation of analytical methods, including detection and quantification capabilities (IUPAC Recommendations 1995), Pure Appl. Chem., 1995, **67**, 1699.
50. L. A. Currie, Detection in analytical chemistry – Importance, theory, and practice, ACS Symposium Series 361, American Chemical Society, Washington, DC 1988.
51. Analytical Methods Committee, Recommendations for the definition, estimation and use of the detection limit, Analyst, 1987, **112**, 199.
52. A. Shrivastava, V. B. Gupta, Methods for the determination of limit of detection and limit of quantitation of the analytical methods, Chronicles of Young Scientists, 2011, **2**(1), 21.
53. United States Pharmacopeia, Validation of compendial methods, 26th revision, National Formulary, 21st ed. Rockville, MD: The United States Pharmacopeial Convention Inc., 2003.
54. Commission Regulation (EC) No 333/2007 (28 March 2007) laying down the methods of sampling and analysis for the official control of the levels of lead, cadmium, mercury, inorganic tin, 3-MCPD and benzo(a)pyrene in foodstuffs, Off. J. EU, L 88/29, 29 March 2007.
55. M. Valcárcel, S. Cárdenas, D. Barceló et al., Metrology of qualitative chemical analysis, report EUR 20605 EN, European Commission, 2002, ISBN 92-894-5194-7.
56. H. Sahai, R. P. Singh, The use of R2 as a measure of goodness of fit: An overview, Virginia Journal of Science, 1989, **40**(1), 5.
57. Analytical Methods Committee, Uses (proper and improper) of correlation coefficients, Analyst, 1988, **113**, 1469.
58. ISO 11732:2005 Water quality – Determination of ammonium nitrogen – Method by flow analysis (CFA and FIA) and spectrometric detection, ISO Geneva.
59. A. Menditto, M. Patriarca, B. Magnusson, Understanding the meaning of accuracy, trueness and precision, Accred. Qual. Assur., 2007, **12**, 45.
60. D. T. Burns, K. Danzer, A. Townshend, Use of the terms "recovery" and "apparent recovery" in analytical procedures (IUPAC Recommendations 2002), Pure Appl. Chem., 2002, **74**(11), 2201.
61. S. L. R. Ellison, B. King, M. Rösslein, M. Salit, A. Williams (eds.), Eurachem/CITAC Guide Traceability in chemical measurement. A guide to achieving comparable results in chemical measurement, 1st ed, Eurachem, 2003, www.eurachem.org.
62. P. De Bièvre, R. Dybkaer, A. Fajgelj, D. Brynn Hibbert, Metrological traceability of measurement results in chemistry: Concepts and implementation (IUPAC Technical Report), Pure Appl. Chem., 2011, **83**(10), 1873.
63. AMC technical brief No. 21, Sept. 2008, M. Thompson (ed.), The estimation and use of recovery factors, www.rsc.org.
64. ISO Guide 33:2000 Uses of certified reference materials, ISO Geneva.
65. T. Linsinger, Application note 1, Rev. 3 2010. Comparison of a measurement result with the certified value, www.erm-crm.org.
66. B. Magnusson, T. Näykki, H. Hovind, M. Krysell, Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report TR 537 (ed. 3.1) 2012, www.nordtest.info.
67. ISO 21748:2010 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, ISO Geneva.
68. Eurolab, Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation, Technical report No. 1/2007, www.eurolab.org.
69. S. L. R. Ellison, A. Williams, Measurement uncertainty: the key to the use of recovery factors? From "The use of recovery factors in trace analysis", M. Parkany (ed.), RSC, Cambridge, 1996, ISBN 0-85404-736-0.

70. V. J. Barwick, S. L. R. Ellison, Measurement uncertainty: approaches to the evaluation of uncertainties associated with recovery, Analyst, 1999, **124**, 981.
71. S. L. R. Ellison, V. J. Barwick, Estimating measurement uncertainty: Reconciliation using a cause and effect approach, Accred. Qual. Assur., 1998, **3**, 101-105.
72. G. E. O'Donnell, D. B. Hibbert, Treatment of bias in estimating measurement uncertainty, Analyst, 2005, **130**, 721.
73. B. Magnusson, S. L. R. Ellison, Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements, Anal. Bioanal. Chem., 2008, **390**, 201.
74. W. J. Youden, E. H. Steiner, Statistical Manual of the AOAC, AOAC International, 1975, ISBN 0-935584-15-3.
75. Harmonised guidelines for internal quality control in analytical chemistry laboratories, (IUPAC technical report), Pure Appl. Chem., 1995, **67**(4), 649.
76. H. Hovind, B. Magnusson, M. Krysell, U. Lund, and I. Mäkinen, Internal quality control – Handbook for chemical laboratories, Nordtest technical report 569, 4th ed., 2011, www.nordtest.info.
77. ISO 7870 Control charts – Parts 1-5, ISO Geneva.
78. AMC technical brief No. 9, Feb. 2002, M. Thompson (ed.), A simple fitness-for-purpose control chart based on duplicate results obtained from routine test materials, www.rsc.org.
79. ISO/IEC 17011:2004 Conformity assessment – General requirements for accreditation bodies accrediting conformity assessment bodies, ISO Geneva.
80. ISO/IEC 17043:2010 Conformity assessment – General requirements for proficiency testing, ISO Geneva.
81. I. Mann, B. Brookman (eds.), Eurachem Guide: Selection, use and interpretation of proficiency testing (PT) schemes by laboratories, 2nd ed., Eurachem, 2011, www.eurachem.org.
82. EA-4/18 TA, Guidance on the level and frequency of proficiency testing participation, European co-operation for Accreditation, 2010, www.european-accreditation.org.
83. ISO 78-2:1999 Chemistry – Layouts for standards - Part 2: Methods of chemical analysis, ISO Geneva.
84. S.L.R. Ellison, A. Williams (eds.), Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment, Eurachem, 2007, www.eurachem.org.
85. R. R. Galen, S. R. Gambino, Beyond normality: The predictive value and efficiency of medical diagnoses, John Wiley and Sons, 1975, ISBN 978-0471290476.
86. M. S. Pepe, The statistical evaluation of medical tests for classification and prediction, Oxford University Press, Oxford, 2003, ISBN 978-0-19-850984-4.
87. X-H. Zhou, N.A. Obuchowski, D.K. Mcclish, Statistical methods in diagnostic medicine, 2nd ed., Wiley-Interscience, New York, 2011, ISBN 978-0-470-18314-4.

ISBN: 978-91-87461-59-0