Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance

Isabel Taverniers, Marc De Loose, Erik Van Bockstaele[1]

It is internationally recognized that validation is necessary in analytical laboratories. The use of validated methods is important for an analytical laboratory to show its qualification and competency. In this update on analytical quality, we place validation of analytical methodologies in the broader context of quality assurance (QA). We discuss different approaches to validation, giving attention to the different characteristics of method performance. We deal with the concepts of single-laboratory or in-house validation, inter-laboratory or collaborative study, standardization, internal quality control (IQC), proficiency testing (PT), accreditation and, finally, analytical QA (AQA).

This article provides a good, complete, up-to-date collation of relevant information in the fields of analytical method validation and QA. It describes the different aspects of method validation in the framework of QA. It offers insight and direct help to anyone involved in any analytical methodologies, whether they are an academic researcher or in the industrial sector.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Accreditation; Analytical method validation; Performance parameter; Quality assurance; Quality control

**Isabel Taverniers*,
Marc De Loose,
Erik Van Bockstaele**
Department for Plant Genetics
and Breeding (DvP),
Centre for Agricultural
Research (CLO),
Ministry of the Flemish
Community,
Caritasstraat 21, B-9090 Melle,
Belgium

*Corresponding author:
Tel.: +32-9-272-2876;
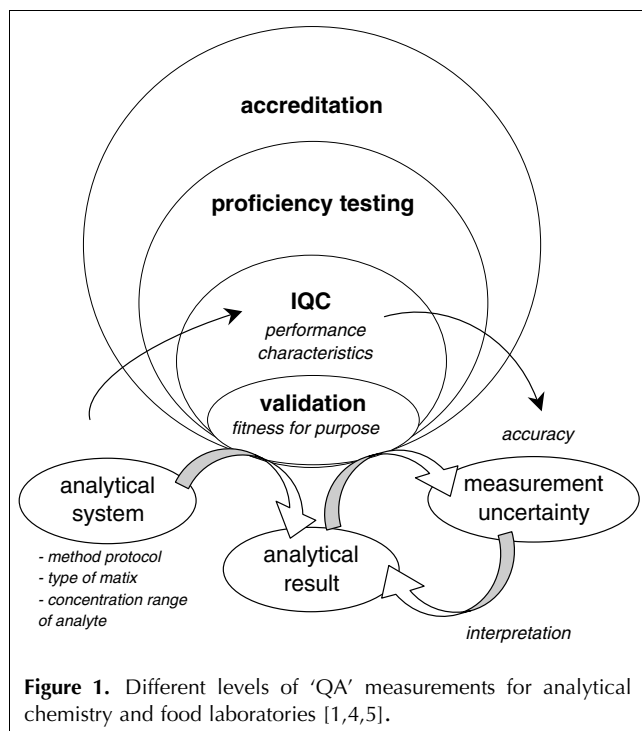Fax: +32-9-272-2901;
E-mail: i.taverniers@clo.fgov.be
[1] Department for Plant Production, Ghent University, Coupure Links 653, B-9000 Gent, Belgium.

Abbreviations: AOAC, Association of Official Analytical Chemists; AQA, Analytical quality assurance; ASTM, American Society for Testing and Material; c, Concentration of measurand; CCMAS, Codex Committee on Methods of Analysis and Sampling; CC_α, Decision limit; CC_β, Detection capability; CEN, European Committee for Normalization; CITAC, Cooperation on International Traceability in Analytical Chemistry; CRM, Certified reference material; CV, Coefficient of variation (= % RSD); EA, European Cooperation for Accreditation; EC, European Community; EN, European Norm; EPA, Environmental Protection Agency; EQC, External quality control; EU, European Union; FAO, Food and Agricultural Organization; FDA, Food and Drug Administration; FSA, Food Standards Association; GLP, Good Laboratory Practices; GMP, Good Manufacturing Processes; HACCP, Hazard Analysis Critical Control Points; ICH, International Conference on Harmonization; IEC, International Electrotechnical Commission; ILAC, International Laboratory Accreditation Cooperation; IQC, Internal quality control; ISO, International Standardization Organization; IUPAC, International Union of Pure and Applied Chemistry; k, Numerical factor used in formulae for LOD and LOQ; LIMS, Laboratory information management system; LOD, Limit of detection; LOQ, Limit of quantification; MRL, Maximum residue level (= PL); MRPL, Minimum required performance limit; MU, Measurement uncertainty; PL, Permitted limit (= MRL); PT, PTS, Proficiency Testing (Scheme); q, Quantity of measurand; QA, Quality assurance; QC, Quality control; r, Repeatability value or limit; R, Reproducibility value or limit; RM, Reference material; RSD, Relative standard deviation; SD (s), Standard deviation(s); SOP, Standard operating procedure; USP, United States Pharmacopeia; WHO, World Health Organization; x, Measured response or signal

1. The role of method validation in AQA

The terms validation and quality assurance (QA) are widely used. However, a lot of analysts and laboratories do not know the exact meaning neither the difference nor the relationship between the two terms. Validating a method is investigating whether the analytical purpose of the method is achieved, which is obtaining analytical results with an acceptable uncertainty level [1]. Analytical method validation forms the first level of QA in the laboratory (Fig. 1). AQA is the complete set of measures a laboratory must undertake to ensure that it can always achieve high-quality data. Besides the use of validation and/or standardized methods, these measures are: effective IQC procedures (use of reference materials (RMs), control charts, etc.); participation in PT schemes; and, accreditation to an international standard, normally ISO/IEC 17025 [1–3].

Figure 1. Different levels of 'QA' measurements for analytical chemistry and food laboratories [1,4,5].

The different levels in Fig. 1 represent the different measures a laboratory must undertake to ensure that it is qualified and competent to perform analytical measurements that satisfy their agreed requirements. A laboratory must be capable of providing analytical data of the required quality. The 'agreed requirement' of an analytical method and the 'required quality' of an analytical result refer to the 'fitness for purpose' of the method [1,4,5].

The ISO definition of validation is 'confirmation by examination and provision of objective evidence that the particular requirements of a specified intended use are fulfilled' [4,6]. Method validation is needed to 'confirm the fitness for purpose of a particular analytical method', i.e. to demonstrate that 'a defined method protocol, applicable to a specified type of test material and to a defined concentration rate of the analyte' – the whole is called the "analytical system" – 'is fit for a particular analytical purpose' [1]. This analytical purpose reflects the achievement of analytical results with an acceptable standard of accuracy. An analytical result must always be accompanied by an uncertainty statement, which determines the interpretation of the result (Fig. 1). In other words, the interpretation and the use of any measurement fully depend on the uncertainty (at a stated level of confidence) associated with it [5]. Validation is thus the tool used to demonstrate that a specific analytical method actually measures what it is intended to measure, and thus is suitable for its intended purpose [6–8].

In the first place, validation is required for any new method. As the definition says, validation always concerns a particular 'analytical system'. This means that, for a particular type of material and a particular operating range of concentrations, the method must be able to solve a particular analytical problem [1]. As a consequence, 'revalidation' is needed whenever any component of the analytical system is changed or if there are indications that the established method does not perform adequately any more [4,8,9].

Method validation is closely related to method development. When a new method is being developed, some parameters are already being evaluated during the 'development stage' while in fact this forms part of the 'validation stage' [4]. However, a validation study may indicate that a change in the method protocol is necessary, and that may then require revalidation [10].

Before any method validation is started, the scope of validation must be fixed, comprising both the "analytical system" and the "analytical requirement". A description of the analytical system includes the purpose and the type of method, the type and the concentration range of analyte(s) being measured, the types of material or matrices for which the method is applied, and a method protocol. The basis of a good analysis rests on a clear specification of the analytical requirement. The latter reflects the minimum fitness-for-purpose criteria or the different performance criteria that the method must meet in order to solve the particular problem. For example, a minimum precision (RSD, see later) of 5% may be required, or a limit of detection (LOD) of 0.1% (w/w) [1,2,4,10]. The established criteria for performance characteristics form the basis of the final acceptability of analytical data and of the validated method [10].

Validation of a new analytical method is typically done at two levels. The first is pre-validation, aimed at fixing the scope of the validation. The second is an extensive, "full" validation performed through a collaborative trial or inter-laboratory study. The objective of full validation, involving a minimum number of laboratories, is to demonstrate that the method performs as was stated after the pre-validation.

2. Guidelines and guidance on AQA

As shown in Fig. 1, using validated methods is the first level of QA, required within a system of IQC. The latter is needed for participation in PT schemes, which, in turn, form a prerequisite for accreditation [1].

For the different levels of QA presented in Fig. 1, guidelines and requirements are well described in detail by several regulatory bodies, standardization agencies

Table 1. Overview of European and international regulatory bodies and their guidelines and standards on different aspects of AQA

Body	Full name	Guidance on	References
Eurachem CITAC EA	A Focus for Analytical Chemistry in Europe Cooperation of International Traceability in Analytical Chemistry European Cooperation for Accreditation	• Method validation • Proficiency testing • Quality Assurance • Accreditation	[2,4,11,12]
CEN	European Committee for Normalization	• Standardization	[13]
IUPAC ISO AOAC International	International Union of Pure and Applied Chemistry International Standardization Organisation Association of Official Analytical Chemists	• Method validation • Standardization • Internal quality control • Proficiency testing • Accreditation	[1,5,14–23]
FDA USP ICH	United States Food and Drug Administration United States Pharmacopeia International Conference on Harmonization	• Method validation	[7,24–26]
FAO/WHO: Codex/CCMAS	Food and Agricultural Organization/World Health Organisation: Codex Committee on Methods of Analysis and Sampling	• Method validation	[27–31]
ILAC	International Laboratory Accreditation Cooperation	• Proficiency testing • Accreditation	[32–34]

and working groups or committees. Just as for traceability and measurement uncertainty (MU) (Part I of this review, Trends Anal. Chem. 23 (2004)), relevant guidelines are given in Table 1. Eurachem guides are published on quality in the laboratory in general [2], method validation [4] and PT [11]. A guideline on PT from the joint Eurachem-Eurolab-EA group is also available [12]. On the European level, there is also the CEN, that is working through different technical committees and working groups on standardization of analytical methods for all sectors [13].

On the international level, we distinguish IUPAC, ISO and AOAC International. All three bodies develop validation and standardization frameworks for analytical chemistry. AOAC International introduced the 'AOAC Peer Verified Methods Program' [14]. IUPAC, ISO and the AOAC International together developed different harmonized guidelines and protocols [1,5,15–19], in addition to a number of ISO standards [20–23]. The US FDA, USP and ICH developed guidelines specifically for pharmaceutical and biotechnological methods [7,24–26].

The international Codex Alimentarius Commission within the United Nations FAO/WHO Food Standards Program has a Codex Committee on Methods of Analysis and Sampling (CCMAS). CCMAS works out criteria for evaluating the acceptability of methods of analysis as well as guidelines on single-laboratory and inter-laboratory validation of methods [27–31]. For single-laboratory validation, CCMAS defends the harmonized IUPAC guidelines [1]. On the international level, also ILAC provides guidelines on PT [2] and on accreditation [33,34] (Table 1).

3. Approaches for evaluating acceptable methods of analysis

The purpose of an analytical method is the delivery of a qualitative and/or quantitative result with an acceptable uncertainty level, so, theoretically speaking, 'validation' boils down to 'measuring uncertainty'. In practice, method validation is done by evaluating a series of method-performance characteristics, such as precision, trueness, selectivity/specificity, linearity, operating range, recovery, limit of detection (LOD), limit of quantification (LOQ), sensitivity, ruggedness/robustness and applicability. Calibration and traceability have been mentioned also as performance characteristics of a method [1,2]. To these performance parameters, MU can be added, although MU is a key indicator for both fitness-for-purpose of a method and constant reliability of analytical results achieved in a laboratory (IQC). MU is a comprehensive parameter covering all sources of error and thus more than method validation alone. In practice, data from method validation and collaborative studies form the basis for, but are only a part of, MU estimation. MU is thus more than just a "method-performance parameter", as described extensively in Part I of this review. Over the years, the concept of MU has won attention in all analytical areas and this has led to two different approaches currently accepted and used for analytical method validation. The traditional 'criteria approach' is to identify specific performance parameters and to assign numeric values to these. These numeric values represent cut-off or threshold values that the method parameters must meet, in order for the method to be acceptable. The alternative approach is

focused on fitness-for-purpose and MU. In this 'fitness-for-purpose approach', the overall MU is estimated as a function of the analyte concentration (see Part I of this review).

Generally, the criteria approach is used for rational methods, i.e. methods where the measurement result can be obtained independently of the method used. Opposite to rational methods, there are empirical methods, in which the measured value depends on the method used. For empirical methods, the criteria approach cannot simply be applied. Instead, precision data from collaborative studies are normally used as basis for MU estimation and validation [27,29].

Validation is needed to demonstrate that the analytical method complies with established criteria for different performance characteristics [35]. When these different characteristics are being evaluated individually, this is generally done for the analytical method as such – where the input is the purified or isolated analyte and the output is the analytical result. However, MU covers the whole analytical procedure, starting from the original sample lot. The assessment of MU (see Part I) is in line with the so-called 'modular validation approach'. Modular validation refers to the ''modularity'' of an analytical procedure, divided up into several sequential steps needed to analyze the material. These may be sample preparation, analyte extraction and analyte determination (Fig. 2). Each step in the procedure can be seen as an ''analytical system'' and can thus be validated separately and combined later on with other ''modules'' in a flexible way. Modular validation is thus a stepwise validation of a whole procedure, taking into consideration all possible difficulties or uncertainty factors at each level in the procedure. The concept of modular validation originates from the domain of predictive microbiology and is now being proposed for methods of analysis for genetically modified organisms (GMOs) [36]. We show the relationship between the three approaches to validation described above in Fig. 2.

4. Method-performance characteristics and the 'criteria approach'

4.1. The extent of validation depends on the type of method

On the one hand, the extent of validation and the choice of performance parameters to be evaluated depend on the status and experience of the analytical method. On the other hand, the validation plan is determined by the analytical requirement(s), as defined on the basis of customer needs or as laid down in regulations.

When the method has been fully validated previously according to an international protocol [15,20], the laboratory does not need to conduct extensive in-house validation studies. It must verify only that it can achieve the same performance characteristics as outlined in the collaborative study.

As a minimum, precision, bias, linearity and ruggedness studies should be undertaken. Similarly, limited validation is required in cases where it concerns a fully validated method applied to a new matrix, a well-established but non-collaboratively studied method and a scientifically published method with characteristics given. More profound validation is needed for methods published in the literature as validation methods, without any characteristic given, and for methods developed in-house [37].

Which performance criteria have to be evaluated depends also on the purpose of the method. Different ICH/USP guidelines are set up for: (1) identification tests;

Figure 2. Schematic representation of the 'analytical method' within the 'analytical procedure', and of different approaches for validation. MU = measurement uncertainty, f = function (of), conc = concentration, LOD = limit of detection, and LOQ = limit of quantification.
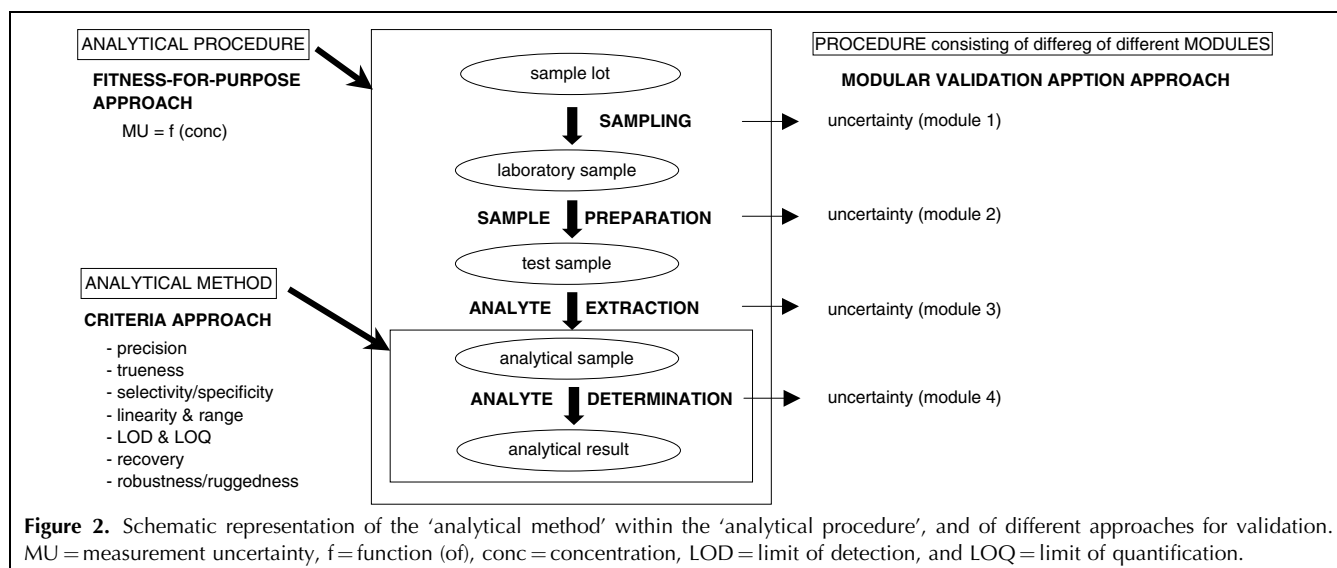
Table 2. Criteria to establish for different categories of methods of analysis [8]

Method-performance parameter	Identification test	Impurity test		Assay test
		Limit impurity test	Quantitative impurity test	
Precision	$-^a$	$-$	$+$	$+$
Trueness	$-$	$-^a$	$+$	$+$
Specificity	$+$	$+$	$+$	$+$
LOD	$-^a$	$+$	$-$	$-$
LOQ	$-^a$	$-$	$+$	$-$
Linearity	$-^a$	$-$	$+$	$+$
Range	$-^a$	$-^a$	$+$	$+$
Ruggedness	$+$	$+$	$+$	$+$

^{a} May be performed.

(2) impurity tests; and, (3) assay tests. An identification test ensures the identity of an analyte in a sample, by comparing it to a known RM. An impurity test is intended to confirm the identity of (limit impurity test) or to accurately quantify (quantitative impurity test) an impurity, defined as an entity 'which may normally not be present'. An assay test finally applies to the major component or active ingredient in a sample and quantifies the drug substance as such, as a whole, or the drug substance in a drug product.

For an assay test, where the major component or active ingredient is supposed to be present at high levels, other criteria than for an impurity test should be investigated. The same is valid for quantitative tests versus identification and limit impurity tests (Table 2) [7,8,24,38].

The literature gives a wide range of practical guidelines for the evaluation of method-performance characteristics [10]. Besides the diversity of approaches, the terminology and the way of reporting results vary widely. Differences may occur depending on the purpose and the application field of the method, and validation studies may become more difficult as the complexity of the analysis increases [39]. In what follows, terms and formulae are taken from the accepted IUPAC 'Nomenclature for the presentation of results of chemical analysis' [18]. For each validation parameter, we set out in Table 3 definitions, ways of expression, determination guidelines and acceptance criteria.

4.2. Accuracy

4.2.1. Precision and bias studies. Precision and bias studies, which form a part of the MU estimate, are the most important validation criteria.

Precision measures are divided into:

1. repeatability precision measures s or SD (s_r or SD_r) and RSD (RSD_r);
2. intra-laboratory reproducibility precision or 'intermediate precision' measures, SD and RSD; and,
3. inter-laboratory reproducibility precision s or SD (s_R or SD_R) and RSD (RSD_R) [18].

Besides standard deviations and coefficients of variation, repeatability/reproducibility values or limits (r,R) are additional parameters of high value in the assessment of precision (for formulae, see Table 3. These criteria mean that the absolute variation between two independent results – obtained within the same laboratory respectively between different laboratories – may exceed the value of r respectively R in a maximum of 5% of the cases [2]. Another measure of precision is the confidence interval, in which all measurements fall with a certain probability or confidence level 1-α (α is often 0.05, giving a probability here of 95%) [18].

Calculated repeatability, intermediate precision and reproducibility values can be compared with those of existing methods. If there are no methods with which to compare the precision parameters, theoretical relative reproducibility and repeatability standard deviations can be calculated from the Horwitz equation and the Horrat value (Table 3). Horwitz RSD values are reported in Table 4. Higher variability is expected as the analyte levels approach the LOD (see below). Next to the Horwitz equation, the AOAC's Peer Verified Program proposes its own levels of acceptability of %RSD, as function of analyte concentration level [8,24].

Precision data can be documented in bar charts or control charts, such as Shewhart control charts (see also Section 5.3 on IQC). Bar charts plot %RSD values with their corresponding confidence interval. Control charts plot the individual measurement results respectively the means of sets of measurements with their confidence level (or with horizontal lines representing 'limits', see further), as a function of the measurement number respectively the run number [4,7,8,10,24,38].

Precision relates to the random error of a measurement system (see Fig. 3) and is a component of MU (see Part I) [2].

4.2.2. Trueness. Trueness is expressed in terms of bias or percentages of error. Bias is the difference between the mean value determined for the analyte of interest and the accepted true value or known level actually present [40]. It represents the systematic deviation of the

Table 3. Summary of method-performance parameters: definitions, ways of expression, requirements or acceptance criteria and guidelines for practical assessment (For more details, see text) [1,4,7,8,24,27–29]. (*) $t_{p,v}$ is the Student factor corresponding to the confidence level $1 - \alpha$ and v degrees of freedom. The symbol p represents the percentile or percentage point of the t-distribution. For 1-sided intervals, $p = 1 - \alpha$; for 2-sided intervals, $p = 1 - \alpha/2$. Values of t can be found in the IUPAC Nomenclature ($t = 2.776$ for $n = 5$ and $t = 3.182$ for $n = 4$ at $p = 0.95$) [19]. (**) X is the mean determined value and n is the number of measurements for which the SD was calculated. If standard deviation data of the certified RMs are not available, 95% confidence limits may be used as an estimate of CRM standard deviation (see second form of formula for the z-score) [28]. (***) x_{bl} is the mean of the blank measurements, s_{bl} is the standard deviation (SD) on the blank measurements and S is the sensitivity of the method or the slope of the calibration function. The calibration function is the relationship between the measured response x_L and the concentration c_L or amount q_L [8,24,49,50]

PARAMETER	DEFINITION	EXPRESSION	REQUIREMENTS	PRACTICAL ASSESSMENT
Accuracy	The closeness of agreement between a test result and the accepted reference value			
1. Precision	The closeness of agreement between independent test results obtained under stipulated conditions	- standard deviation s or **SD** - relative standard deviation s_{rel} or **RSD** - % coefficient of variation **%CV** or **%RSD** = 100 * $RSD = 100 * SD / \bar{x}$ - repeatability limit **r** = 2.83 x SD_r and reproducibility limit **R** = 2.83 x SD_R - confidence interval **CI** = $\bar{x} \pm C$ (*) $$C = \frac{s \times t_{p,v}}{\sqrt{n}}$$	RSD_r = 0.5–0.6 times theoretical values determined by the **Horwitz function** = 2 exp (1 - 0.5 log C) with C= analyte concentration in decimal fraction RSD_R = 0.5–2 times theoretical values determined by the Horwitz function **Horrat** value = RSD_R(trial)/RSD_R(Horwitz) $\gtrless$ 2 RSD_r and RSD_R values according to AOAC Peer Verified Program	- minimum of 3 concentration levels over full range of analytical method (~at LOD, in middle, high) - minimum of 3 repeats per concentration level - calculate *repeatability precision* SD_r, RSD_r, r= 2.8 x SD_r, C, CI - calculate *intermediate precision* SD_{int}, RSD_{int}, r= 2.8 x SD_{int}, C, CI - calculate *reproducibility precision* SD_R, RSD_R, r= 2.8 x SD_R, C, CI - document in **bar chart** or in **control chart**
1.1. Repeatability precision (intra-run precision)	*Precision under conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time*			
1.2. Intermediate precision (inter-run precision)	*Precision under conditions where independent test results are obtained with the same method on identical test items in the same laboratory but by different analysts, using different equipment over an extended period of time*			
1.3. Reproducibility precision (inter-laboratory precision)	*Precision under conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment*			

2. Trueness	The closeness of agreement between the expectation of the test result (expected mean value) and an accepted reference value (true value)	if CRMs are used:
		bias = % error: difference between measured value and true value
		Z-score: difference between measured value and certified reference value % (**)
		
		$$z = \frac{X_{found} - X_{certified}}{\sqrt{\left(\frac{SD_{found}}{n_{found}} + \frac{SD_{certified}}{n_{certified}}\right)^2}}$$
		
		$$z = \frac{X_{found} - X_{certified}}{\sqrt{\frac{SD_{found}}{n_{found}} + \left(\frac{CI}{2}\right)^2}}$$
		
		if no CRMs are available:
		% recovery of known, spiked amount of analyte
			general: compare results with second, validated reference method
			
			if CRMs are used:
			Z-score $\leqslant$ |2|
			
			if spiking method is used:
			% recovery according to USP/ICH guidelines (depending on analyte concentration level)
				- minimum of 3 concentration levels over full range of analytical method (~at LOD, in middle, high)
				- minimum of 3–10 repeats per concentration level
				- use of CRMs: calculate bias (% error) and/or Z-score
				- no CRMs: *spike* a typical matrix and a blank sample with known amount of analyte; calculate %recovery= 100x [spike$_{matrix\ sample}$/spike$_{blank}$]

Recovery	The fraction of analyte added to the test sample (fortified or spiked) prior to analysis, which is measured by the method	**% recovery** of known, spiked amount of analyte	**% recovery** according to USP/ICH guidelines (depending on analyte concentration level)	- minimum of 6 repeats of matrix blanks or samples unfortified and fortified with the analyte at different concentrations - calculate %recovery= 100x [spike$_{matrix\ sample}$/spike$_{blank}$]
Specificity	The ability of the method to determine accurately and specifically the analyte of interest in the presence of other components in a sample matrix (that may be expected to be present in the sample matrix) under the stated conditions of the test (specificity= 100% selectivity)	NOTE! cannot be expressed, must be demonstrated; depends on type and purpose of method for identification tests: **% correct classification** of ▪ non-analyte containing samples as negatives ▪ analyte containing samples as positives the **ability to discriminate** with compounds of closely related structures (negative results) for quantitative tests: **% recovery** of samples, spiked with possible interferants	for identification tests: **% correct classification** ~100 for quantitative tests: **% recovery** according to USP/ICH guidelines (depending on analyte concentration level)	identification tests: - determine % of false positives for a minimum number of blank samples - determine % of false negatives for a minimum number of positive samples (RMs) - test a minimum number of structurally similar or closely related samples which must be negative impurity and assay tests: - spike different samples with possible interferents and calculate % recovery - if no interfering compounds are available, compare with second method

Table 3 (continued)

Term	Definition	Notes / expression	Range	Procedure
Limit of detection LOD	The lowest concentration or amount of analyte - that can be reliably distinguished from zero or - that can be detected/measured with reasonable statistical certainty	NOTE! LOD is expressed as a concentration or amount of analyte, which is derived from the measured signal (***) **lowest signal**: $x_L = x_{bl} + 3 s_{bl}$ **lowest concentration/amount** $LOD = q_L(c_L) = x_L/S = 3 s_{bl}/S$ NOTE: a signal of 3 times the SD_{blank} corresponds to a % RSD of 33%		- measure, each once, a minimum of either 10 independent sample blanks ($x_L = x_{bl} + 3 s_{bl}$) or 10 independent sample blanks fortified at lowest acceptable concentration ($x_L = 3 s_r$) - choose a number of low concentration levels near the *expected LOD* - set up the calibration curve: signal $x = A$.concentration + B and thus $x_L = A$.LOD determine the slope A and calculate the LOD= $x_L/A = [x_{bl} + 3 s_{x,bl}]/A$
Limit of quantitation LOQ	The lowest concentration or amount of analyte that can be determined quantitatively with an acceptable level of repeatability precision and trueness	NOTE! LOQ is expressed as a concentration or amount of analyte, which is derived from the measured signal **lowest signal**: $x_L = x_{bl} + 10 s_{bl}$ **lowest concentration/amount** $LOQ = q_L(c_L) = x_L/S = 10 s_{bl}/S$ NOTE: a signal of 10 times the SD_{blank} corresponds to a % RSD of 10%		- measure a minimum of 10 independent sample blanks each once ($x_L = x_{bl} + 10 s_{bl}$) - choose a number of low concentration levels near the calculated LOD - set up the calibration curve: signal $x = A$.concentration + B and thus $x_L = A$.LOQ determine the slope A and calculate the LOQ= $x_L/A = [x_{bl} + 10 s_{bl}]/A$
Linearity	The ability of the method to obtain test results proportional to the concentration of analyte (within a given range)	NOTE! linearity cannot be expressed, must be demonstrated		first rough estimation of *linear range*: - measure blanc + min. of 6 independent conc. levels, set up calibration curve and estimate linear range visually - choose CRMs or spiked samples of min. 6 different concentrations within *estimated linear range* - calculate **residual Y values** from calibration curve and plot them as function of concentration *if randomly distributed -> linearity; if systematic trends -> non-linearity* - calculate **relative signals**= signal/conc. and plot them as function of concentration *if horizontal line -> linearity; linearity limits correspond to 95% and 105% relative signal-values -> range*
Linear range = Working range = Linearity limits	The range of concentrations/amounts of analyte over which - the method gives test results proportional to the concentration of analyte OR - a linear calibration model can be applied with a known confidence level	range: a **range** of concentrations between lower and upper **linearity limits**	for assay tests: **range**= 80–120% of analyte level for impurity tests: **range**= LOQ(or 50%)–150%, 50–150% or 0–150% according to USP/ICH and IUPAC guidelines **lower linearity limit**= LOQ	
Ruggedness	The (intra-laboratory tested) behavior of an analytical process when small changes in the environmental and/or operating conditions are made (generally used term)	ruggedness: measure for variability (reproducibility of results obtained under a variety of conditions); expressed as **% RSD** (inter-laboratory)		- evaluate the effect of different small changes in parameters (in days, instruments, analysts, reagents, material, amount of sample material used,....) individually - calculate precision data
Robustness	A measure of the capacity of the analytical procedure to remain unaffected by small, but deliberate variations in method-performance parameters, which provides an indication of its reliability during normal usage (term used by USP/ICH only)			
Sensitivity	The change in the response of a measuring instrument divided by the corresponding change in the stimulus	the **slope** of the calibration curve (arbitrary)		

Table 4. Horwitz function as an empirical relationship between the precision of an analytical method and the concentration of the analyte regardless of the nature of the analyte, matrix and the method used. Acceptable RSD_R and RSD_r values according to [27] and to AOAC International [8,14] (PVM = Peer Verified Methods (Program))

Analyte %	Analyte ratio	Unit	Horwitz %RSD	AOAC PVM %RSD
100	1	100%	**2**	1.3
10	1.00E − 01	10%	**2.8**	2.8
1	1.00E − 02	1%	**4**	2.7
0.1	1.00E − 03	0.10%	**5.7**	3.7
0.01	1.00E − 04	100 ppm	**8**	5.3
0.001	1.00E − 05	10 ppm	**11.3**	7.3
0.0001	1.00E − 06	1 ppm	**16**	11
0.00001	1.00E − 07	100 ppb	**22.6**	15
0.000001	1.00E − 08	10 ppb	**32**	21
0.0000001	1.00E − 09	1 ppb	**45.3**	30

measured result from the true result. Method trueness is also an indicator of utility and applicability of that method with real samples [41].

Different sources of systematic errors contribute to the overall bias (Fig. 3). Thompson and Wood [5] describe 'persistent bias' as the bias affecting all data of the analytical system, over longer periods of time and being relatively small but continuously present. Different components contribute to the persistent bias, such as laboratory bias, method bias and the matrix-variation effect. Next to persistent bias, the larger 'run effect' is the bias of the analytical system during a particular run [1,4,5].

One or more of these bias components are encountered when analyzing RMs. In general, RMs are divided into certified RMs (CRMs, either pure substances/solutions or matrix CRMs) and (non-certified) laboratory RMs (LRMs), also called quality control (QC) samples [42]. CRMs can address all aspects of bias (method, laboratory and run bias); they are defined with a statement of uncertainty and traceable to international standards. CRMs are therefore considered as useful tools to achieve traceability in analytical measurements, to calibrate equipment and methods (in certain cases), to monitor laboratory performance, to validate methods and to allow comparison of methods [1,4,43]. However, the use of CRMs does not necessarily guarantee trueness of the results. The best way to assess bias practically is indeed by replicate analysis of samples with known concentrations, such as RMs (see also Part I of this article). The ideal RM is a matrix CRM, as this is very similar to the samples of interest (the latter is called 'matrix matching'). However, a correct result obtained with a matrix CRM does not guarantee that the results of unknown samples with other matrix compositions will be correct [1,42].

The usefulness of CRMs for validation (in particular for trueness assessment) and traceability purposes has been debated for years. This is illustrated by the enormous number of papers published on this topic. We mention here only some interesting references

[42,44,45]. Examples of the use of pure substance RMs, matrix CRMs or LRMs can be found in Special Issues of *Accreditation and Quality Assurance* (Volume 9, 2004) and *Analytical and Bioanalytical Chemistry* (Volume 278, 2004) on 'Biological and Environmental Reference Materials' and the Special Issue of *TrAC* (volume 23, 2004) on 'Challenges for achieving traceability of environmental measurements'.

If no such (certified) RMs are available, a blank sample matrix of interest can be 'spiked' with a known amount of a pure and stable in-house material, called the 'spike' or 'surrogate'. Recovery is then calculated as the percentage of the measured spike of the matrix sample relative to the measured spike of the blank control or the amount of spike added to the sample. The smaller the recovery %, the larger the bias that is affecting the method and thus the lower the trueness [1,8,24,46,47].

An indication of trueness can also be obtained by comparing the method with a second, well-characterized reference method, under condition that the precision of the established reference method is known. Results from the two methods, performed on the same sample or set of samples, are compared. The samples may be CRMs, in-house standards or just typical samples [4]. A comparison between the three ways of establishing bias, is also given in Part I of this article. It should be clear that the use of recovery estimates and comparing methods, are alternative ways which encompass serious limitations. They can give an idea about data comparability; however, trueness cannot be assured [42].

Trueness or exactness of an analytical method can be documented in a control chart. Either the difference between the mean and the true value of an analyzed (C)RM together with confidence limits, or the percentage recovery of the known, added amount can be plotted [8,14]. Here again, special caution should be taken concerning the reference used. Control charts may be useful to achieve trueness only if a CRM, which is in principle traceable to SI units, is used. All other types of references only allow traceability to a 'consensus' value,
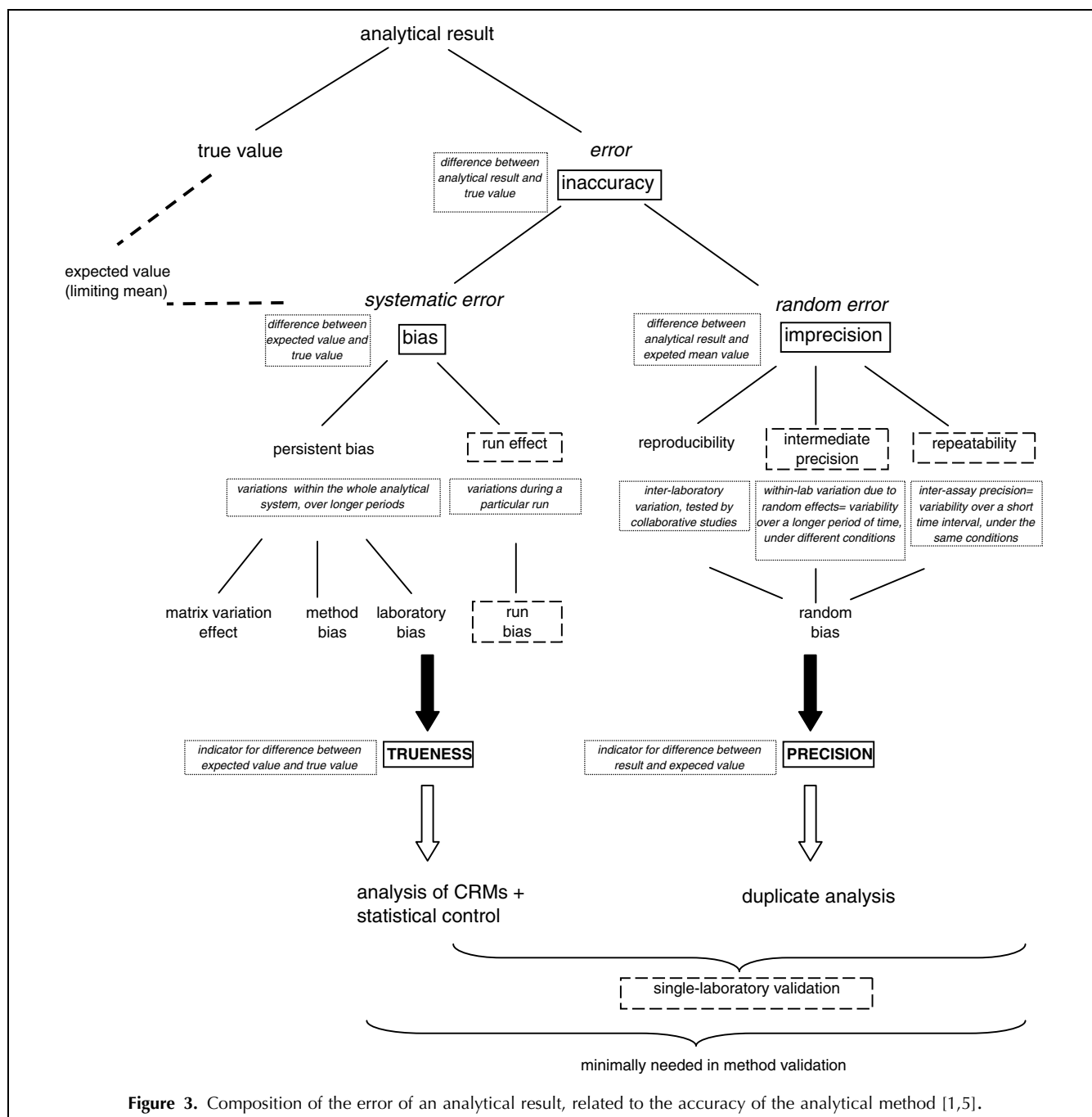
Figure 3. Composition of the error of an analytical result, related to the accuracy of the analytical method [1,5].

which is assumed to be equal to the 'true' value, although this is not necessarily the case [42]. The expected trueness or recovery % values depend on the analyte concentration. Trueness should therefore be estimated for at least three different concentrations. If recovery is measured, values should be compared to acceptable recovery rates, as outlined by the AOAC Peer Verified Methods program (Table 5) [8,14]. Besides bias and % recovery, another measure for the trueness is the z-score (Table 3). It is important to note that a considerable component of the overall MU will be attributed to MU on the bias of a system, including uncertainties on RMs (Fig. 3) [2].

4.2.3. Recovery. Recovery is often treated as a separate validation parameter (Table 3). Analytical methods aim to estimate the true value of the analyte concentration with an uncertainty that is fit for purpose. However, in such analytical methods, the analyte is transferred from the complex matrix to a simpler solution, whereby there is a loss of analyte. As a consequence, the measured value will be lower than the true concentration present in the original matrix. Therefore, assessing the efficiency of the method in detecting all of the analyte present is a part of the validation process. Eurachem, IUPAC, ISO and AOAC International state that recovery values

Table 5. Acceptable recovery percentages as a function of the analyte concentration [8]

Analyte%	Analyte ratio	Unit	Mean recovery (%)
100	1	100%	**98–102**
10	1.00E – 01	10%	**98–102**
1	1.00E – 02	1%	**97–103**
0.1	1.00E – 03	0.10%	**95–105**
0.01	1.00E – 04	100 ppm	**90–107**
0.001	1.00E – 05	10 ppm	**80–110**
0.0001	1.00E – 06	1 ppm	**80–110**
0.00001	1.00E – 07	100 ppb	**80–110**
0.000001	1.00E – 08	10 ppb	**60–115**
0.0000001	1.00E – 09	1 ppb	**40–120**

should always be established as a part of method validation. Recovery or spiking studies should be performed for different types of matrices, several examples of each matrix type and for each matrix type at different levels of analyte concentration [1,2,4].

4.3. Specificity and selectivity

Specificity and selectivity both give an idea of the reliability of the analytical method (for definitions, see Table 3). Some authors give different definitions for both terms while, for others, they are identical. The term 'specific' generally refers to a method that produces a response for a single analyte only, while the term 'selective' is used for a method producing responses for different chemical entities or analytes which can be distinguished from each other. A method is called 'selective' if the response is distinguished from all other responses. In this case, the method is perfectly able to measure accurately an analyte in the presence of interferences [8,48]. According to Eurachem, specificity and selectivity essentially reflect the same characteristic and are related very closely to each other in such a way that specificity means 100% selectivity. In other words, a method can only be specific if it is for 100% selective. Another related term is 'confirmation of identity', which is the proof that 'the measurement signal, which has been attributed to the analyte, is only due to the analyte and not to the presence of something chemically or physically similar or arising as coincidence' [4]. A method must first show high specificity before true quantification can be performed [40].

There is no single expression for specificity. It is rather something that must be demonstrated. The way that this is done depends on the objective and the type of analytical method (see also below). For identification tests, the goal is to ensure the identity of an analyte. Specificity is here the ability to discriminate between compounds of closely related structures that can be present.

For impurity tests (limit impurity test; quantitative impurity test) and assay tests, the accent is on the ability to determine or to discriminate for the analyte in the presence of other interferants. Selectivity can be assessed by spiking samples with possible interferants (degradation products,···) [7,8,24].

4.4. LOD

There is no analytical term or parameter for which there is a greater variety of terminology and formulations than for LOD and quantification. The limits of detection, or detection limit, is the terminology most widely used, as accepted by Eurachem. ISO uses 'minimum detectable net concentration', while IUPAC prefers 'minimum detectable (true) value' [4]. However, all official organizations refer to the same definition: 'the lowest amount of an analyte in a sample which can be detected but not necessarily quantified as an exact value'. In general, the LOD is expressed as a concentration c_L or a quantity q_L, derived from the smallest signal x_L which can be detected with reasonable certainty for a given analytical procedure. The lowest signal x_L is the signal that lies k times SD_{blank} above the mean blank value, whereby k is a numerical factor chosen according to the level of confidence required [8,24,49–51].

The larger the value of k, the larger the confidence level. Eurachem and IUPAC recommend a value of 3 for k, meaning that the chance that a signal more than $3s$ above the sample blank value is originating from the blank is less than 1%. The LOD is thus the concentration or amount corresponding to a measurement level (response, signal) three s_{bl} units above the value for zero analyte (Table 3). At the concentration or amount three times the s_{bl}, the relative standard deviation or coefficient of variation on the measured signal is 33% (measure for uncertainty) [1,2,4,27,49,52]. According to USP/ICH, the LOD corresponds to that signal where the 'signal-to-noise ratio' is 2:1 or 3:1 [24,38].

It is not true – as is often thought – that detection or quantification is impossible below the determination limit; but, at these lower levels, the uncertainty of the detection/quantification measurement is higher than the actual value itself [28]. In this context, Huber [8] also defines the LOD as the point at which a measured value is larger than the uncertainty associated with it. According to Krull and Swartz [41], the LOD is a concentration point where only the qualitative identification is possible but not accurate and precise quantification.

For qualitative methods, the LOD is defined as the 'threshold concentration at which the test becomes

unreliable'. Each of a series of blank samples, spiked with different concentrations of the analyte, is analyzed at least 10 times. The threshold or 'cut-off' concentration is determined visually based on a response curve, plotting the % positive results versus the concentration. In this respect, the LOD is also defined as the concentration at which 95% of the experiments give a clearly positive signal [4].

The LOD may not be confused with the sensitivity of the method. The latter is 'the capability of the method to discriminate small differences in concentration or mass of the test analyte' and is equal to the slope of the calibration curve (see below) [8].

4.5. LOQ

For the LOQ or 'limit of determination', definitions and formulas are very similar to those of LOD, except that for LOQ, k is taken to be 5, 6 or even 10 [1,2,4,8,24,50]. A value of 10 for k means that the relative standard deviation (%RSD) at the LOQ is 10%. The LOQ thus corresponds to that concentration or amount of analyte, quantifiable with a variation coefficient not higher than 10% [52]. The LOQ is always higher than the LOD and is often taken as a fixed multiple (typically 2) of the LOD [1]. Also, the determination limit is referred to as the signal 10 times above the noise or background signal, corresponding to a 'signal-to-noise ratio' of 10:1 [24,38].

In practice, the LOQ can be calculated analogously to the LOD, as indicated in Table 3. An alternative way of practically assessing the LOD and LOQ is the following. In a first step, 10 independent sample blanks are each measured once, the blank standard deviation s_{bl} is calculated and the lowest signals corresponding to both the LOD and the LOQ are calculated as $x_{LOD} = x_{bl} + 3s_{bl}$ respectively as $x_{LOQ} = x_{bl} + 10 s_{bl}$. In a second step, sample blanks are spiked with various analyte concentrations (e.g., 6) close to the LOD. Per concentration, 10 independent replicates are measured and the standard deviation of the measured signals calculated. These standard deviations s (or the relative standard deviations %RSD) are then plotted against the concentration. LOD and LOQ values are those concentrations of analyte corresponding to %RSD values of 33% and 10%, respectively [4,28].

As was said for LOD, it is not true that at and below the LOQ, quantification becomes impossible. Quantification is possible, but it becomes unreliable as the uncertainty associated with it at these lower levels is higher than the measurement value itself. Quantification becomes reliable as soon as the MU is lower than the value measured [28].

4.6. Decision limit and detection capability: for specific sectors only

In the context of analytical method validation, the terms decision limit (CC_α) and detection capability (CC_β) as well as minimum required performance limits (MRPLs) are often used and need some clarification. These terms are applicable for the measurement of organic residues, contaminants and chemical elements in live animals and animal products, as regulated within the EU by the Council Directives 96/23/EC [53], 2002/657/EC [35] and 2003/181/EC [54]. The Commission distinguishes 'Group A substances', for which no permitted limit (PL) (maximum residue level, MRL) has been established, and 'Group B substances' having a fixed PL.

CC_α is the limit at and above which it can be concluded with an error probability of α that a sample is non-compliant. If a PL has been established for a substance (Group B or the regulated compounds), the result of a sample is non-compliant if the decision limit is exceeded ($CC_\alpha = x_{PL} + 1.64\ s_{MRL}$). If no PL has been established (Group A), the decision limit is the lowest concentration level at which the method can discriminate with a statistical certainty of $1-\alpha$ that the particular analyte is present ($CC_\alpha = x_{bl} + 2.33\ s_{sample}$). CC_β is the smallest content of the substance that may be detected, identified and/or quantified in a sample with an error probability of β ($CC_\beta = CC_\alpha + 1.65\ s_{sample}$).

MRPLs have been established for substances for which no PL has been fixed and in particular for those substances the use of which is not authorized or even prohibited within the EU (Group A). A MRPL is the minimum content of an analyte in a sample, which at least has to be detected and confirmed. A few MRPLs for residues of certain veterinary drugs have been published so far in Directive 2003/181/EC.

For Group A substances (no PL established), CC_α and CC_β are comparable with LOD and LOQ, respectively, as their concentrations correspond to measured signals laying y times above the blank signal. For substances having a PL (Group B), CC_α and CC_β are not related to LOD and LOQ but are expressed in relation to this PL. It is important to note that these terms apply specifically to inspection of animals and fresh meat for the presence of residues of veterinary drugs and specific contaminants and are therefore different from LOD and LOQ [35,53–56].

4.7. Linearity and range

For assessment of the linearity of an analytical method, linear regression calculations are not sufficient. In addition, residual values should be calculated (Table 3). The latter represent the differences between the actual y value and the y value predicted from the regression curve, for each x value. If residual values, calculated by simple linear regression, are randomly distributed about the regression line, linearity is confirmed, while systematic trends indicate non-linearity. If such a trend or pattern is observed, this suggests that the data are best treated by weighted regression. For either simple or weighted linear regression, linearity supposes that the intercept is not significantly different from zero [1,4,27,28].

An alternative approach to establishing linearity is to divide the response by the respective concentrations and to plot these 'relative responses' as a function of the concentration, on a log scale. The line obtained should be horizontal over the full linear range, with a positive deviation at low concentrations and a negative deviation at high concentrations. By drawing parallel horizontal lines, corresponding to, e.g., 95% and 105% of the horizontal relative response line, the intersection points can be derived where the method becomes non-linear [8].

It is important that a linear curve is repeatable from day to day. However, linear ranges may be different for different matrices. The reason for this is a possible effect of interferences inherent to the matrix. A test for general matrix effect can be performed by means of 'standard additions' or the method of analyte additions. For a set of samples, obtained by adding different concentrations of analyte to a certain matrix, the slope of the calibration curve is compared with the slope of the usual calibration function. A lack of significance (curves are parallel) means that there is no matrix effect [27,28].

4.8. Ruggedness and robustness

Although the terms ruggedness and robustness are often treated as the same and used interchangeably, separate definitions exist for each, as indicated in Table 3.

To have an idea about the ruggedness, Eurachem recommends introducing deliberate variations to the method, such as different days, analysts, instruments, reagents, variations in sample preparation or sample material used. Changes should be made separately and the effect evaluated of each set of experimental condi-
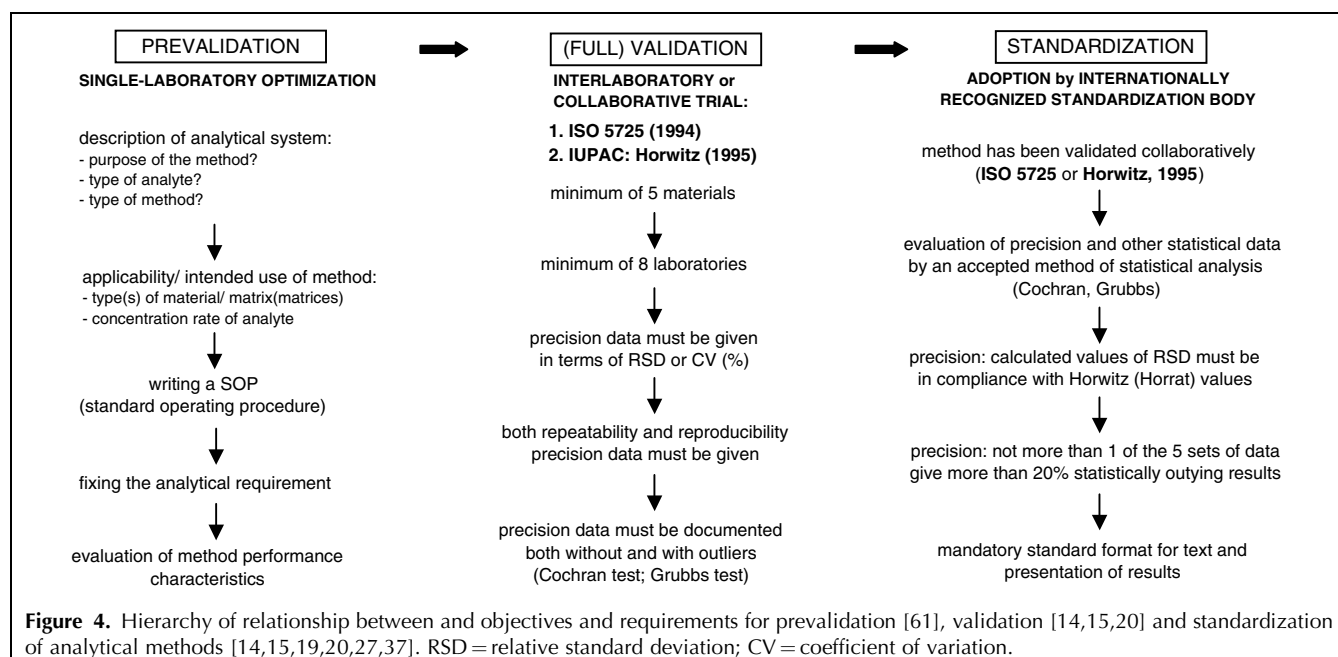
tions on the precision and trueness [1,4,38]. To examine the effects of different factors, a "factorial design" methodology can be applied, as described by von Holst et al. [57]. By combining changes in conditions and performing a set of experiments, one can determine which factors have a significant or even critical influence on the analytical results.

In ICH/USP guidelines, ruggedness is not defined separately but treated under the same denominator as reproducibility precision: it is 'the degree of reproducibility of the results obtained under a variety of conditions, expressed as %RSD' [8,38].

Robustness is a term introduced by USP/ICH [41]. Although Eurachem has included the term robustness in its official list of definitions, the term is not used by official organizations other than USP/ICH. According to Eurachem, both parameters do present the same and are thus synonyms [4,24,38].

4.9. Sensitivity

The sensitivity of a method is the gradient of the response curve. In practical terms, sensitivity refers to the slope of the calibration curve. Sensitivity is often used together with LODs and LOQs. Indeed, the slope of the calibration curve is used for the calculation of LODs and LOQs. A method is called sensitive if a small change in concentration or amount of analyte causes a large change in the measured signal [1,4,28]. Sensitivity is not always mentioned as a validation parameter in official guidelines. According to Thompson et al. [1], it is not useful in validation because it is usually arbitrary, depending on instrument settings. USP/ICH does not mention sensitivity at all.

Figure 4. Hierarchy of relationship between and objectives and requirements for prevalidation [61], validation [14,15,20] and standardization of analytical methods [14,15,19,20,27,37]. RSD = relative standard deviation; CV = coefficient of variation.

5. AQA

QA is the complete organizational infrastructure that forms the basis for all reliable analytical measurements [5]. It stands for all the planned and systematic activities and measures implemented within the quality system [2,58]. A quality system has a quality plan, which emphasizes the implementation of Good Laboratory Practice (GLP). GLP is comparable to the Good Manufacturing Process (GMP) and the larger HACCP (Hazard Analysis Critical Control Point) quality systems of food-production factories. Attention goes to all aspects of quality management in the laboratory organization, including staff training, the maintenance and calibration of all equipment used, the laboratory environment, safety measures, the system of sample identification, record keeping and storage – the latter may be simplified by the use of laboratory information management systems (LIMS), the use of validated and standardized methods and the documentation of these methods and of all information concerning the followed procedures (standard operating procedures, SOP).

QA embraces both QC and 'quality assessment'. QC is defined as the mechanism or the practical activities undertaken to control errors, while quality assessment is 'the mechanism to verify that the system is operating within acceptable limits'. Quality assessment and QC measures are in place to ensure that the measurement process is stable and under control [2,5].

Within QC, we distinguish between internal and external QC. In general, QA comprises the following topics, as also schematized in Fig. 1.

5.1. The use of validated methods: in-house versus inter-laboratory validation

Wherever possible or practically achievable, a laboratory should use methods that have been ''fully validated'' through a collaborative trial, also called inter-laboratory study or method-performance study. Validation in collaborative studies is required for any new analytical method before it can be published as a standard method (see below). However, single-laboratory validation is a valuable source of data usable to demonstrate the fitness-for-purpose of an analytical method. In-house validation is of particular interest in cases where it is inconvenient or impossible for a laboratory to enter into or to organize itself a collaborative study [1,59].

On the one hand, even if an in-house validated method shows good performance and reliable accuracy, such a method cannot be adopted as a standard method. In-house validated methods need to be compared between at least eight laboratories in a collaborative trial.

On the other hand, a collaborative study should not be conducted with an un-optimized method [10]. Inter-laboratory studies are restricted to precision and trueness while other important performance characteristics, such as specificity and LOD, are not addressed [60]. For these reasons, single-laboratory validation and inter-laboratory validation studies do not exclude each other but must be seen as two necessary and complementary stages in a process, as presented in Fig. 4. The added value of single-laboratory validation is that it simplifies the next step – inter-laboratory validation – and thereby minimizes the gap between internally (validated or not) developed methods and the status of inter-laboratory validation. By optimizing the method first within the laboratory, as a kind of preliminary work, an enormous amount of collaborators time and money is saved [10].

The importance of conducting such a single-laboratory preliminary validation step is increasingly highlighted by international standardization agencies. IUPAC and AOAC International include a 'Preliminary Work' paragraph in their guidelines for collaborative studies,

Table 6.	Mandatory text format for standardized methods according to ISO Guide 78/2 [2,10]	
1	Scope	States briefly what the method determines
2	Definitions	Precise definition of the analyte or parameter determined by the method
3	Fields of application	Type of materal(s)/matrix(ces) to which the method is applicable
4	Principle	Basic steps involved in the procedure
5	Apparatus	Specific apparatuses required for the determination are listed
6	Reagents	Analytical reagent-grade reagents needed for the determination are listed
7	Sampling	Description of the sampling procedure
8	Procedure	Divided into numbered paragraphs or sub-clauses; includes a 'preparation of test sample' step and a reference to 'quality assurance' procedures
9	Calculation and expression of results	Indication of how the final results are calculated and of units in which the results are to be expressed
10	Notes	Additional information as to the procedure; may be in the form of notes, placed here or in the body of the text
	Annex	Includes all information on analytical quality control, such as precision clauses (repeatability and reproducibility data), table of statistical data outlining the accuracy (trueness and precision) of the method
	References	References to the report published on the collaborative study carried out prior to standardization of the method

stating that within-laboratory testing data are required on precision, bias, recovery, applicability. Additionally, a clear description of the method including statements on the purpose of the method, the type of method and the probable use of the method is required within this preliminary work [14,15]. However, it is not only in the harmonized guidelines for collaborative trials (see below) that the link between a single-laboratory pre-validation step and the collaborative trial is emphasized. Separate guidelines for single-laboratory validation of methods of analysis have recently been published by IUPAC, ISO and AOAC International [1]. The IUPAC guidelines have also been considered and accepted by the Codex Committee on Methods of Analysis and Sampling (CCMAS) [30]. In addition, specific, individual working groups or scientists are presenting their own framework for 'pre-, single-laboratory validation' of methods of analysis. The latter do not concern official, published guidelines but can be found on the internet (e.g., [10,38]) or are distributed through national or international specific working groups. The objectives of a single-laboratory or in-house validation process are depicted in Fig. 4. Depending on the type of method (Table 2), data can be obtained for all criteria except for the reproducibility (inter-laboratory) precision. However, it is this 'among-laboratories variability' or reproducibility which is the dominating error component in analytical measurement and which underlies the need for inter-laboratory validation [61].

Inter-laboratory or collaborative validation studies can be organized by any laboratory, institute or organization, but should preferably be conducted according to one of the following recognized protocols:

(1) ISO 5725 on Accuracy (Trueness and Precision) of Measurement Methods and Results [20]; or,
(2) IUPAC Protocol for the Design, Conduct and Interpretation of Method-Performance Studies [15].

The latter revised, harmonized guidelines have been adopted also by AOAC International as the guidelines for the AOAC Official Methods Program [14]. The main requirements for collaborative studies outlined in these guidelines are shown in Fig. 4.

Precision plays a central role in collaborative studies. Wood [37] defines a collaborative trial as 'a procedure whereby the precision of a method of analysis may be assessed and quantified'. Precision is the objective of inter-laboratory validation studies, and not trueness or any other method-performance parameter. Evaluation of the acceptability of precision data is important for the standardization of methods (see hereafter).

5.2. The use of standardized methods
The first level of AQA is the use of validated or standardized methods. The terms validated and standardized here refer to the fact that the method-performance characteristics have been evaluated and have proven to meet certain requirements. At least, precision data are documented, giving an idea of the uncertainty and thus of the error of the analytical result. In both validated and standardized methods, the performance of the method is known.

Validated methods can be developed by the laboratory itself, or by a standardization organization after inter-laboratory studies. Standardized methods are developed by organizations such as the AOAC International, ISO,

Table 7. Differences between method-performance studies and proficiency testing (PT) schemes [5,31,62]

Characteristic	Collaborative/interlaboratory studies (method performance studies)	Proficiency testing schemes
Main objective	**Validation** of new methods	**Competency check** of analytical laboratories
Application	• new methods	• routinely used (validated and/or standardized methods)
	• required for full validation and standardization	• recommended within IQC and QA system
	• first prerequisite for IQC and QA	
Results aimed at	**Precision**: multiple results, both repeatability and reproducibility → % RSD is compared to theoretical Horwitz and Horrat values	**Trueness**: 1 single result per test material → calculation of Z-score as measure for the bias
Method & protocol used	1 single, prescribed method for which SOP must strictly be followed	Multiplicity of methods; participants have free choice of (validated and/or standardized) method
Test materials	• minimum of five different materials	• no minimum; often less than 5 test samples per round
	• no stipulations for homogeneity and stability of test samples	• homogeneity and stability of materials must be assured
Participating laboratories	• minimum of eight	• no minimum; variety in participants is possible throughout 1 scheme (different rounds)
	• are assumed to be equally competent	• not assumed to have equal competency (will be tested)

USP (see Table 1), US Environmental Protection Agency (EPA), American Society for Testing and Material (ASTM) or Food Standards Association (FSA) [8]. This is exactly where the difference lies between a validated and a standardized method: an analytical method can only be standardized after it has been validated through inter-laboratory comparisons. The main prerequisite for a standards organization is that a method has been adequately studied and its precision shown to meet a required standard, as summarized in Fig. 4. The format of a standard method, as outlined in ISO Guide 78/2 [10] [10], as shown in Table 6 [2]. A specific IUPAC Protocol [19] describes in detail how to present AQA data, such as the performance characteristics.

5.3. Effective IQC

In the IUPAC Harmonized Guidelines for IQC, Thompson and Wood [5] define IQC as a 'set of procedures undertaken by the laboratory staff for the continuous monitoring of operation and the results of measurements in order to decide whether results are reliable enough to be released'. IQC guarantees that methods of analysis are fit for their intended purpose, meaning the continuous achievement of analytical results with the required standard of accuracy. The objective of IQC is the elongation of method validation: continuously checking the accuracy of analytical data obtained from day to day in the laboratory. In this respect, both systematic errors, leading to bias, as well as random errors, leading to imprecision, are monitored. In order to be able to monitor these errors, they should remain constant. Within the laboratory, such constant conditions are typically achieved in one analytical run. The word 'internal' in IQC implicates that repeatability conditions are achieved. Thus, monitoring the precision as an objective of IQC does not concern reproducibility or inter-laboratory precision, but only repeatability or intra-laboratory precision. The monitoring of accuracy of an analytical method in IQC can be translated into the monitoring of the analytical system [5].

Two aspects are important for IQC:

(1) the analysis of 'control materials', such as RMs or spiked samples, to monitor trueness; and,
(2) replication of analysis to monitor precision.

Also of high value in IQC are blank samples and blind samples. Both IQC aspects form a part of statistical control, a tool for monitoring the accuracy of an analytical system. In a control chart, such as a 'Shewhart control chart', measured values of repeated analyses of an RM are plotted against the run number. Based on the data in a control chart, a method is defined either as an 'analytical system under control' or as an 'analytical system out of control'. This interpretation is possible by drawing horizontal lines on the chart: $\bar{x}$(mean value),

$\bar{x} + s$(SD) and $\bar{x} - s$, $\bar{x} + 2s$ (upper warning limit) and $\bar{x} - 2s$ (lower warning limit), $\bar{x} + 3s$ (upper action or control limit) and $\bar{x} - 3s$ (lower action or control limit). An analytical system is under control if no more than 5% of the measured values exceeds the warning limits [2,3,38].

5.4. Participation in proficiency testing (PT) schemes

PT is the periodic assessment of the competency or the analytical performance of individual participating laboratories [11]. An independent coordinator distributes individual test portions of a typical uniform test material. The participating laboratories analyze the materials by their method of choice and return the results to the coordinator. Test results obtained by different laboratories are subsequently compared with each other and the performance of each participant evaluated based on a single competency score [16,62]. International harmonized protocols exist for the organization of PT schemes [11,12,16,21,32].

Participation in PTs is not a prerequisite or an absolute substitute for IQC measures, or vice versa. However, participation in PT is meaningless without a well developed IQC system. IQC underlies participation in PT schemes, while IQC and participation in PT schemes are both important substitutes of AQA (Fig. 1). It is shown that laboratories with the strongest QC procedures score significantly better in PT schemes [5,62]. Participation in PT can to a certain extent improve laboratory performance. However, unsatisfactory performance in schemes (up to 30% of all participants) has been reported. This means that there is no correlation between good analytical performance and participation in PT [63]. However, PT has a significant educational function, as it helps a laboratory to demonstrate competency to an accreditation body or another third party [12].

The terms 'PT schemes' and 'collaborative trials' are often confused with each other, as, in both QA measures, a number of different laboratories is involved. However, there is a clear distinction between both. The main differences with respect to objective and application, results, used method, test materials and participating laboratories are summarized in Table 7. It is important to note also that the results obtained from PT schemes, as well as those from collaborative performance studies, can be used for assessing the MU (see Part I).

5.5. External QC (EQC) and accreditation

Participation in PT schemes is an objective means of evaluating the reliability of the data produced by a laboratory. Another form of external assessment of laboratory performance is the physical inspection of the laboratory to ensure that it complies with externally imposed standards. Accreditation of the laboratory indicates that it is applying the required QA principles. The 'golden standard' ISO/IEC 17025 [23], which is the

revised version of ISO Guide 25 [22], describes the general requirements for the competence of calibration and testing laboratories. In Europe, the accreditation criteria have been formalized in European Standard EN45001 [64]. Participation in PT schemes forms the basis for accreditation, because PT is a powerful tool for a laboratory to demonstrate its competency. Accreditation guides use the information obtained by PT schemes [2,12,16,23].

Accreditation is formal recognition that a laboratory is competent to carry out specific (types of) calibrations or tests [2]. After the use of validated and standardized methods, the introduction and use of appropriate IQC procedures and the participation in PT schemes, accreditation to ISO/IEC 17025 is the fourth basic principle related to laboratory QA in general [1]. Guidelines on the implementation of ISO/IEC 17025, including the estimation of MU (see also Part I), are published in the literature and by official accreditation bodies, such as Eurachem, CITAC, EA, Eurolab and ILAC (see Table 1) [2,12,33,34,65]. It is worthwhile to mention that accreditation, just like participation in PT schemes, does not necessarily indicate that the laboratory has a good performance [63].

6. Summary

Together with the fast development of analytical methodologies, nowadays great importance is attached to the quality of the measurement data. Besides the necessary reporting of any result with its MU and traceability of the results to stated standards or references (Part I of this review), a third crucial aspect of analytical methods of any type is their validation status. It is internationally recognized that validation is necessary in analytical laboratories. However, less is known about what validation is and what should be validated, why validation is important, when and by whom validation is performed and, finally, how it is carried out practically. This article tries to answer these questions.

We define method validation in detail and describe different approaches to evaluating the acceptability of analytical methods. We attach great importance to the different method-performance parameters, their definitions, ways of expression and approaches for practical assessment.

Validation of analytical methodologies is placed in the broader context of QA. The topics of standardization, internal and external QC and accreditation are discussed, as well as the links between these different aspects. Because validation and QA apply to a specific analytical method, it is important to approach each method on a case-by-case basis. An analytical method is a complex, multi-step process, starting with sampling and ending with the generation of a result. Although every method has its specific scope, application and analytical requirement, the basic principles of validation and QA are the same, regardless the type of method or sector of application. The information in this work is mainly taken from the analytical chemistry, although it applies to other sectors as well.

This second part on quality in the analytical laboratory provides a good, complete, up-to-date collation of relevant information in the fields of analytical method validation and QA. It is useful for the completely inexperienced scientist as well as for those involved in this topic for a long time, but having somewhere lost their way in the labyrinth, looking for more explanation on a particular aspect, or longing for deeper insight and knowledge.

Acknowledgements

References

[1] M. Thompson, S. Ellison, R. Wood, Pure Appl. Chem. 74 (2002) 835.

[2] CITAC/Eurachem Guide: Guide to Quality in Analytical Chemistry – An Aid to Accreditation, 2002 (http://www.Eurachem.bam.de).

[3] R.J. Mesley, W.D. Pocklington, R.F. Walker, Analyst (Cambridge, UK) 116 (1991) 975.

[4] Eurachem Guide: The fitness for purpose of analytical methods. A laboratory guide to method validation and related topics, LGC, Teddington, Middlesex, UK, 1998 (http://www.Eurachem.bam.de).

[5] M. Thompson, R. Wood, Pure Appl. Chem. 67 (1995) 649.

[6] J. Fleming, H. Albus, B. Neidhart, W. Wegschieder, Accred. Qual. Assur. 1 (1996) 87.

[7] ICH-Q2A, Guideline for Industry: Text on Validation of Analytical Procedures, 1995 (http://www.fda.gov/cder/guidance/index.htm).

[8] L. Huber (Ed.), Validation and Qualification in Analytical Laboratories, Interpharm Press, East Englewood, CO, USA, 1998.

[9] R.J. Wells, Accred. Qual. Assur. 3 (1998) 189.

[10] M. Green, Anal. Chem. 68 (1996) 305A.

[11] Eurachem Guide: Selection, use and interpretation of PT schemes by laboratories, 2000 (http://www.Eurachem.bam.de).

[12] Eurachem/Eurolab/EA Guide EA-03/04: Use of proficiency testing as a tool for accreditation in testing, August 2001 – rev.01, 18pp.

[13] CEN, European Committee for Normalisation, 2004 (http://www.cenorm.be/cenorm/index.htm).

[14] AOAC International, Method Validation Programs (OMA/PVM Department), including Appendix D: Guidelines for collaborative

study procedures to validate characteristics of a method of analysis, 2000 (http://www.aoac.org/vmeth/devmethno.htm).

[15] W. Horwitz, Pure Appl. Chem. 67 (1995) 331.

[16] M. Thompson, R. Wood, Pure Appl. Chem. 65 (1993) 2123 (Also published in J. AOAC Int. 76 (1993) 926).

[17] L.A. Currie, Pure Appl. Chem. 67 (1995) 1699.

[18] L.A. Currie, G. Svehla, Pure Appl. Chem. 66 (1994) 595.

[19] W.D. Pocklington, Pure Appl. Chem. 62 (1990) 149.

[20] ISO Guide 5725, 'Accuracy (trueness and precision) of measurement methods and results', ISO, Geneva, Switzerland, 1994.

[21] ISO Guide 43, 'Development and operation of laboratory proficiency testing', ISO, Geneva, Switzerland, 1984.

[22] ISO Guide 25, 'General requirements for the competence of calibration and testing laboratories', ISO, Geneva, Switzerland, 1990.

[23] ISO/IEC 17025, 'General requirements for the competence of calibration and testing laboratories', ISO, Geneva, Switzerland, 1999.

[24] ICH-Q2B, Guidance for Industry: Validation of Analytical Procedures: Methodology, 1996 (http://www.fda.gov/cder/guidance/index.htm).

[25] ICH-Q6B, Harmonised Tripartite Guideline: Specifications: Test procedures and acceptance criteria for biotechnological/biological products, 1996 (http://www.fda.gov/cder/guidance/index.htm).

[26] FDA/CDER/CVM, Guidance for Industry – Bioanalytical Method Validation, 2001, 22pp (http://www.fda.gov/cder/guidance/index.htm).

[27] CX/MAS 01/4, Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling, Criteria for evaluating acceptable methods of analysis for Codex purposes, Agenda Item 4a of the 23rd Session, Budapest, Hungary, 26 February–2 March, 2001.

[28] CX/MAS 02/4, Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Proposed draft guidelines for evaluating acceptable methods of analysis, Agenda Item 4a of the 24th Session, Budapest, Hungary, 18–22 November 2002 + CX/MAS 02/4-Add.2 Dispute situations.

[29] CX/MAS 02/5, Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Criteria for evaluating acceptable methods of analysis for Codex purposes, Agenda Item 4b on the 24th Session, Budapest, Hungary, 18–22 November 2002 + CX/MAS 02/5-Add.1 Proposed amendment – Government comments.

[30] CX/MAS 02/11, Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Requirements for single-laboratory validation for Codex purposes. Agenda Item 8b of the 24th Session, Budapest, Hungary, 18–22 November 2002.

[31] CX/MAS 02/12, Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling (FAO/WHO), Validation of methods through the use of results from proficiency testing schemes, Agenda Item 8c on the 24th Session, Budapest, Hungary, 18–22 November 2002.

[32] ILAC-G13: Guidelines for the requirements for the competence of providers of proficiency testing schemes, ILAC Technical Accreditation Issues Committee, 2000, 23pp (http://www.ilac.org/).

[33] ILAC-G15: Guidance for accreditation to ISO/IEC 17025, ILAC Technical Accreditation Issues Committee, 2001, 16pp (http://www.ilac.org/).

[34] ILAC-G18: The scope of accreditation and consideration of methods and criteria for the assessment of the scope in testing, ILAC Technical Accreditation Issues Committee, 2002, (http://www.ilac.org/www).

[35] European Commission, Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, Off. J. Eur. Commun. L 221/8, 17.8.2002.

[36] A. Holst-Jensen, K.G. Berdal, J. AOAC Int. 87 (2004) 109.

[37] R. Wood, Trends Anal. Chem. 18 (1999) 624.

[38] Waters Corporation. Validation Guidelines: Terminology and Definitions (http://www.waters.com/WatersDivision/).

[39] D.B. Hibbert, Accred. Qual. Assur. 4 (1999) 352.

[40] J. Fleming, B. Neidhart, H. Albus, W. Wegscheider, Accred. Qual. Assur. 1 (1996) 135.

[41] I.S. Krull, M. Swartz, Anal. Lett. 32 (1999) 1067.

[42] Ph. Quevauviller, Trends Anal. Chem. 23 (2004) 171.

[43] M. Segura, C. Camara, Y. Madrid, C. Rebollo, J. Azcarate, G.N. Kramer, B.M. Gawlik, A. Lamberty, Ph. Quevauviller, Trends Anal. Chem. 23 (2004) 194.

[44] G. Holcombe, R. Lawn, M. Sargent, Accred. Qual. Assur. 9 (2004) 198.

[45] M. Lauwaars, E. Anklam, Accred. Qual. Assur. 9 (2004) 253.

[46] J. Fleming, H. Albus, B. Neidhart, W. Wegschieder, Accred. Qual. Assur. 1 (1996) 191.

[47] M. Thompson, S.L.R. Ellison, A. Fajgelj, P. Willets, R. Wood, Pure Appl. Chem. 71 (1999) 337.

[48] J. Vessman, J. Pharmac. Biomed. Anal. 14 (1996) 867.

[49] Analytical Methods Committee, Analyst (Cambridge, UK) 112 (1987) 199.

[50] J. Fleming, H. Albus, B. Neidhart, W. Wegschieder, Accred. Qual. Assur. 2 (1997) 51.

[51] W. Huber, Accred. Qual. Assur. 8 (2003) 213.

[52] I. Kuselman, F. Sherman, Accred. Qual. Assur. 4 (1999) 124.

[53] European Commission, Council Directive 96/23/EC of 29 April 1996 on measures to monitor certain substances and residues thereof in live animals and animal products and repealing Directives 85/258/EEC and 86/469/EEC and Decisions 89/187/EEC and 91/664/EEC, Off. J. Eur. Commun. L 125 0010-0032, 23.05.1996.

[54] European Commission, Commission Decision 2003/181/EC amending Decision 2002/657/EC as regards the setting of minimum required performance limits (MRPLs) for certain residues in food of animal origin, Off. J. Eur. Commun. L 71/17 0017-0018, 15.3.2003.

[55] J.P. Antignac, B. Le Bizec, F. Monteau, F. Andre, Anal. Chim. Acta 483 (2003) 325.

[56] K. De Wasch, H.F. De Brabander, D. Courtheyn, N. Van Hoof, S. Poelmans, H. Noppe, in: Proc. Euro Food Chem. XII: Strategies on Safe Food, vol. 1, Bruges, Belgium, 24–26 September, 2003, p. 45.

[57] C. von Holst, A. Müller, E. Björklund, E. Anklam, Eur. Food Res. Technol. 213 (2001) 154.

[58] E. Prichard, H. Albus, B. Neidhart, W. Wegscheider, Accred. Qual. Assur. 2 (1997) 348.

[59] R. Battaglia, Accred. Qual. Assur. 1 (1996) 256.

[60] H. van der Voet, J.A. van Rhijn, H.J. van de Wiel, Anal. Chim. Acta 391 (1999) 159.

[61] W. Horwitz, R. Albert, J . AOAC Intern. 79 (1996) 589.

[62] M. Thompson, P.J. Lowthian, Analyst (Cambridge, UK) 118 (1993) 1495.

[63] B. King, N. Boley, G. Kannan, Accred. Qual. Assur. 4 (1999) 280.

[64] EN45001, General criteria for the operation of testing laboratories, CEN/CENELEC, The Joint European Standard Institution, Brussels, Belgium, 1989.

[65] J. Pritzkow, Accred. Qual. Assur. 8 (2003) 25.

Isabel Taverniers graduated in Agricultural and Applied Biological Sciences from the University of Gent, Belgium, in 1999. Until April

2001, she worked at AgriFing, a joint spin-off laboratory of Gent University, Hogeschool Gent and the Department for Plant Genetics and Breeding (DvP); she specialized in DNA fingerprinting technologies. She is now preparing a Ph.D. thesis in the Laboratory of Applied Plant Biotechnology of the Department for Plant Genetics and Breeding (CLO, Flemish Community).

Erik Van Bockstaele is Head of the Department for Plant Genetics and Breeding and Professor at the Faculty of Agricultural and Applied Biological Sciences of the University of Gent.

Marc De Loose is Head of the Section of Applied Plant Biotechnology at the Department for Plant Genetics and Breeding.