Journal of Chromatography B, 877 (2009) 2180–2190

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Review

Method validation across the disciplines—Critical investigation of major validation criteria and associated experimental protocols☆

Dietmar Stöckl, Heidi D'Hondt, Linda M. Thienpont*

Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Gent University, Harelbekestraat 72, B-9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 14 July 2008
Accepted 23 December 2008
Available online 31 December 2008

Keywords:
Analytical performance characteristics
Limit of detection
Limit of quantitation
Linearity
Precision
Trueness
Accuracy
Total error
Experimental protocols
Analytical acceptance criteria
Statistical significance testing

ABSTRACT

Analytical method development should aim at delivering reliable measurements within a given application. This implies that method validation is integrated in the development process, because it enables to establish a method's performance capabilities, and to demonstrate its fitness-for-purpose. Although analytical chemists mostly are familiar with the validation guidelines within the discipline of their responsibility, we believe that they may take advantage of a better acquaintance with recommendations among disciplines. Therefore, we review the guidance given in 4 disciplines (laboratory medicine, pharmacy, environment, and food), with emphasis on the proposed experimental protocols, acceptance criteria and interpretation strategies by statistical significance testing. Last but not least, we give incentive towards a modernized validation design.

Contents

☆ This paper is part of a special issue entitled "Method Validation, Comparison and Transfer", guest edited by Serge Rudaz and Philippe Hubert.

* Corresponding author. Tel.: +32 9 264 81 04; fax: +32 9 264 81 98.
E-mail address: linda.thienpont@ugent.be (L.M. Thienpont).

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.12.056

1. Introduction

According to the definition in the ISO 9000 standard series, validation is the "confirmation, through the provision of objective evidence, that requirements for a specific intended use or application have been fulfilled" [1]. Implicit in this definition is that the analytical validation process should (i) specify the intended use of a measurement procedure, (ii) define the analytical performance requirements, (iii) provide data from validation experiments (objective evidence), and (iv) interpret the validation data by use of a statistical test (confirmation that requirements have been fulfilled). While several reviews are available that discuss analytical method validation within a given discipline [2–10], the objective of this review is to compare guidelines across 4 disciplines, i.e., laboratory medicine, pharmacy, environment, and food, as well as some general guidelines. We restrict ourselves to the validation of the limit of detection (LoD), limit of quantitation (LoQ), linearity, precision and trueness (often wrongly termed accuracy). We shortly address the estimation of total error (accuracy) which, however, is typically not addressed in method validation guidelines. We investigate the experimental protocols recommended for estimating these performance characteristics, look into the defined acceptance criteria and verify whether statistical significance testing is described to assess whether the estimates pass the requirements. Additionally, we make proposals for modernizing the design of analytical method validation across the disciplines.

Note that the metrological terms/definitions adopted throughout this paper are from the "International Vocabulary of Metrology (VIM)" [11], unless differently stated [1,12]. However, being aware that sometimes analytical chemists are more familiar with a variety of discipline-specific technical terms, we refer the reader for a better understanding of the terminology in this review to Appendix A.

2. Investigation of validation guidelines in 4 disciplines

As mentioned in Section 1, this review applies to guidelines for validation of methods used in laboratory medicine-, pharmaceutical-, environmental-, and food analysis. From time to time, we also consider more general guidelines because of their broad applicability.

For each of the following performance characteristics, i.e., LoD, LoQ, linearity, precision, and trueness, we start with conceptual and/or theoretical considerations. Then we report on our scrutiny of the recommendations in the different validation guidelines regarding the experimental design, the acceptance criteria and statistical significance testing. We focus on the strong points in the concerned recommendations, but do not refrain to also reveal weak points and missing indispensable elements. Where applicable, we illustrate our statements in graphical form. For a compilation, the reader is referred to Appendix B, comprising 1 table per performance characteristic and the relevant literature citations. At the end of each section, we use our review in a constructive way by proposing a state-of-the-art validation design. Finally, we put special emphasis on certain aspects of the, in our opinion, key elements of a validation process, i.e., predefined acceptance criteria and significance testing. For what concerns significance testing, we give, at the end of each topic, always an example of possible statistical test. However, it is beyond the scope of the manuscript to detail the statistical aspects of the different tests.

2.1. Limit of detection (LoD)

Usually, the LoD is defined by k-times the standard deviation of blanks (SD_{Blank}) or low concentrated samples (SD_{Low}). While similar k-values are used in the different documents, the underlying statistical concepts may be different. This is illustrated in Fig. 1. When the LoD is estimated by blank measurements only (left population), the LoD may be defined by the one-sided z-value of that population (z = normal standard deviate), which corresponds to the α-error of detecting the analyte when it is not present. For example, a z-value of 1.645 (2.33) corresponds to an α-error of 5% (1%). In that case, however, one may better use the term "limit of the blank" (LoB) [13]. When the SD_{Blank} is estimated from few measurements (<20), only, then the LoD should be defined in terms of the t-value (t = Student's t-value). For example, when the SD_{Blank} is derived from 7 measurements, a one-sided t-value of 3.143 is used as statistical multiplier for an α-error of 1% [14,15]. On the other hand, when blanks and low concentrated samples are measured, the LoB is unsatisfactory as LoD, because of the β-error. This is the probability of not detecting the analyte when it is present. For the sample population at the LoB, the β-error would be 50%. It would be more suitable to position the LoD at a point that is twice as far as the LoB from the mean of the left distribution in Fig. 1, i.e., at $2 \times 1.645 = 3.3 \times SD_{Blank}$. Then the probability of each of the 2 kinds of error occurring (α-error of the blank and the β-error of a sample with an average concentration as shown by the right population) is 5%. For the calculation of the confidence intervals of the LoD defined by the α-error and/or α- and β-errors, we refer to Refs. [13,16].

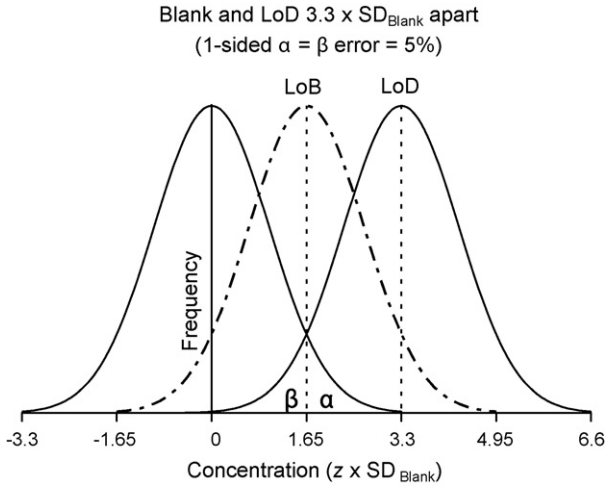

Fig. 1. The concept of the LoD based on the SD_{Blank} with consideration of the α- and β-error.

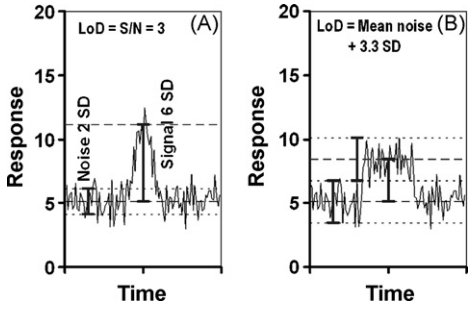

Fig. 2. Chromatographic LoD (S/N = 3) (A) compared with LoD from "blank" (=mean noise + 3.3 × SD) (B).

Several documents (for a compilation, see Table B.1 in Appendix B) do not address the LoD [17–21]. Most of the other documents define the LoD by $\sim 3 \times SD_{Blank}$ or $\sim 3 \times SD_{Low}$ [14,15,22–25]. When not explicitly stated, a SD multiplier of ~3 may suggest that the LoD was chosen on the basis of both α- and β-error considerations. However, some documents consider only the α-error [14,15,25]. Four documents define the LoD explicitly by α- and β-error considerations [13,26,27], but only 1 requires the measurement of blanks and samples with low concentrations [13]. Two documents also define the LoD of chromatographic procedures as S/N = 3 [22,27]. Note that in the S/N definition, the blank and LoD distributions are $6 \times SD_{Blank}$ apart from each other ($N = 2 \times SD_{Blank}$ and $S = 3N$), while only $3.3 \times SD_{Blank}$ in the α/β-error definition (see Fig. 2). For completeness, we add that, recently, a series of documents important for accredited food analytical laboratories has been published [29].

The number of measurements for estimating the LoD are grossly different: they range from no recommendation [22], over 6 [24], to 60 measurements [13]. Only 4 documents [13,15,26,28] address the measurement design, in terms of replicates (r) per day (d) (for example: $2r \times 5d = 10$ measurements in total).

With regard to acceptance criteria, only 2 documents address them [13,15], but, only 1 describes a non-parametric statistical test that allows verification whether the criteria are met [13].

We recommend that the analytical chemist should specify an acceptance criterion for every performance characteristic that is validated, therefore, also for the LoD. The definition should be based on both α- and β-error considerations (one-sided). For chromatographic methods, the LoD should be estimated from the S/N ratio. This is not only because the concept is so well known in that field, but in particular because it belongs to the major skills of an analytical chemist to be able to optimize the S/N, e.g., by improving the baseline and peak height, ensuring a sufficient number of detection cycles under the peak, etc. Naturally, like all tools, the concept has its limitations, e.g., when the procedural blank is contaminated with the analyte of interest. The estimation of the LoD should be done with sufficient measurements distributed over several days (for example, $1r \times 10d$ or $20d$) by measurement of suitable blanks and low samples. Usually, the intra-assay variation of the LoD is negligible compared to the inter-assay one, thus singlicate measurements may be performed per day. The estimated LoD should be validated versus the specified value by use of a statistical test (1-sample t-test, using the mean of the low-samples at LoD; non-parametric procedure [13]) or the respective confidence interval.

2.2. Limit of quantitation (LoQ)

Two different concepts are currently used to define the LoQ (for a compilation, see Table B.2 in Appendix B). The first defines LoQ in multiples of the LoD [14,22,23,25], the second defines LoQ in terms of values for R.S.D. and trueness (typically: 20% R.S.D. and trueness) [15,17–21,28] or total error [13]. Other documents specify 10% R.S.D. [30]. Three documents do not address or recommend the LoQ [24,26,27]. Moreover no direct metrological definition of LoQ is available in VIM [11] (note: the measuring interval indirectly defines the [lower and higher] limit of quantitation, see Appendix A).

The number of measurements for estimating the LoQ range from no recommendation [17,18,21,22], over 5 [19,20], 20 [23], to 40 measurements [13]. Only 2 documents [13,28] address the measurement design, in terms of replicates per day.

Regarding acceptance criteria for LoQ, only 3 documents address them. One defines it by a fraction of the maximum permissible concentration set by regulation [31], 1 by the lowest calibrator according to the standard procedure [15], and 1 uses analyte-specific acceptance criteria that are set by the user based on "peer" total error data [13]. Only 1 document describes a statistical test that allows the verification whether the acceptance criteria are met [13].

We recommend that the analytical chemist should specify an acceptance criterion for the LoQ. The definition of the LoQ should be based on values for precision and trueness or total error. The estimation of the LoQ should be done with sufficient measurements of suitable samples distributed over several days (for example, $2r \times 10d$ or $20d$). The estimated LoQ should be validated versus the specified value by use of a statistical test or confidence interval. Depending on the definition of the LoQ, one would use a 1-sample t-test (trueness criterion), a 1-sample F-test (precision criterion), or regression/accuracy profile based tests (total error criterion).

2.3. Linearity

All documents, with the exception of 1 [27], address linearity of the calibration function (for a compilation, see Table B.3 in Appendix B). Typically, it is recommended to establish the calibration itself from 5 or more calibration points. The experimental design for investigating linearity is described in 6 documents [24–26,32–34]. Usually, the measurement design recommends 2 or 3 replicates on 1 day (design: $2r$ or $3r \times 1d$). Only 1 document [26] recommends linearity investigation on several days ($3r \times 3d$).

Acceptance criteria are defined by visual assessment, statistical tests, trueness and precision and the correlation coefficient (r). Statistical tests for linearity are explicitly recommended in 5 documents [24–26,32,34], but only 1 discusses a test versus a user-defined limit [32]. While more and more documents deprecate the use of r for assessing linearity (e.g., [19,20,33]), it is still widely used in practice. The problem with r is that it is related to the range of the data (at constant standard deviation, r increases with the range). Because of the great diversity in the acceptance criteria for linearity, the problem shall be addressed by an example. Fig. 3 shows the simulation of a 5-points calibration curve that follows the theoretical function $y = -0.00018x^2 + x$; each point is determined as 4 replicates with a R.S.D. of 5% (note: the simulation in Fig. 3 does not perfectly match the theoretical function). The simulation is represented as scatter and residuals plot (mean residuals, in %) based on weighted linear (A1, A2) and 2nd order polynomial (B1, B2) regression. The figure shows that visual assessment of linearity is difficult in the scatter plot and, therefore, should be done with the residuals plot. The residual plot of the linear regression suggests that the data may be non-linear (curvature of the residuals), despite the fact that $r = 0.9996$. The residuals plot of the 2nd order polynomial regression shows a more random distribution of the residuals, indicating that the calibration function may be of 2nd order. A statistical test [32] gives a p-value of 0.065, which is borderline not significant (95% probability level). Using weighted linear regression for calibration would introduce ~4% error at the 2nd and the 5th calibration point. This may be acceptable from a validation point of view, however, as the calibration curve is the heart of a methods' trueness, one may wish to perform more experiments to verify whether the calibration function is of 2nd order.

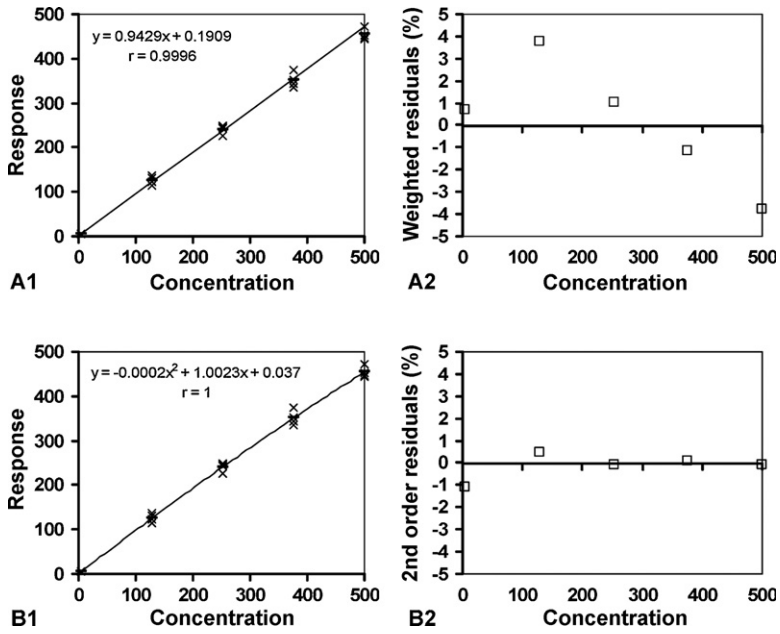

Fig. 3. Scatter (A1 and B1) and residuals (mean residuals in %, A2, B2) plot based on weighted linear (A1, A2) and 2nd order polynomial (B1, B2) regression analysis of simulated data for a 5-points calibration curve (each point determined as 4 replicates; R.S.D. 5%). The plots illustrate the difference between assessment of linearity on the basis of correlation coefficient *r* and by eye versus statistical testing as described in the text.

We recommend that the analytical chemist should specify an acceptance criterion for linearity (statistical or user-defined limit). The investigation of linearity should be done with sufficient measurements and repeated over several days (for example, $3r \times 5d$). First, visual assessment of linearity should be done with the residuals plot of linear regression. Validation of linearity should be done by use of a statistical test (null-hypothesis and/or user-defined limit). Note that the often used ANOVA lack-of-fit test is less suited than a *t*-test for significance of 2nd or 3rd order regression coefficients [34]. Testing of linearity in case of a presumed linear calibration function should not be mixed up with the situation of investigating the nature of a calibration curve, which may be non-linear, for example, S-shaped immunoassay curves (4 parameter logistic). For a detailed discussion about

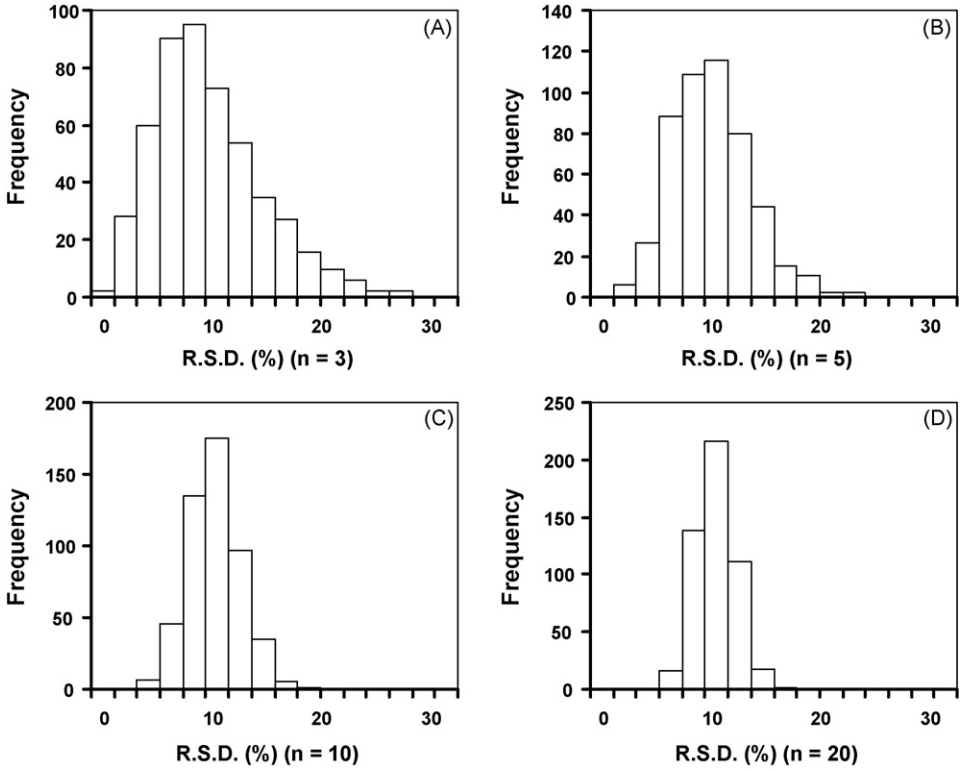

Fig. 4. Frequency plots of 500 precision estimates simulated with a "true" R.S.D. of 10% and different replicates ($n = 3$ (A), 5 (B), 10 (C) to 20 (D)). The plots illustrate the influence of the number of replicates on the shape of the distribution (skewed versus symmetric) and the agreement of the estimated with the true mean R.S.D.

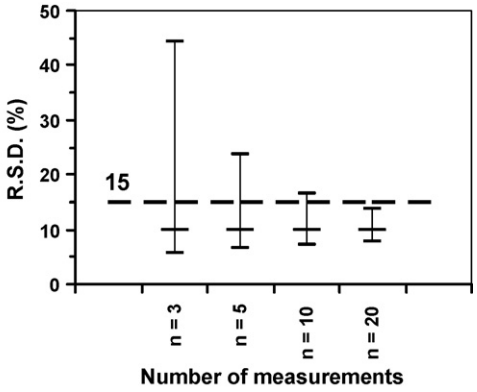

Fig. 5. Plot of the mean R.S.D. with confidence limits estimated from the simulations in Fig. 4. The plot shows the influence of the magnitude of the uncertainty on significance testing against a limit of 15%.

linearity and choice of calibration model the reader is referred to Ref. [2].

2.4. Precision

The estimation of precision is required by all guidelines (for a compilation, see Table B.4 in Appendix B). Typically, it is recommended to estimate precision at 3 different concentrations, however, according to varying experimental designs, i.e., with total number of replicates ranging from 3 [22] to 40 [35]. In this respect, it is important to know the influence of the number of replicates on the shape of the distribution of precision estimates and the agreement of the estimated with the "true" precision (=precision of the population). This is illustrated in Fig. 4, representing the frequency plot of 500 precision estimates, simulated with a "true" R.S.D. of 10%: in (A) the distribution of R.S.D.s from $n = 3$ is highly positively skewed, the mean R.S.D. is only 9.3%, and 311 estimates are <10%; only the distribution with $n = 20$ is fairly symmetric, has a mean R.S.D. of 9.9%, and 266 estimates <10% (D). Note also that the uncertainty of the estimate is high with only 3 replicates, in particular, for what concerns the upper confidence limit. This poses problems for significance testing, as shown in Fig. 5: with an estimated R.S.D. of 10% and using an acceptance criterion of 15%, only the R.S.D. estimate with $n = 20$ would pass the significance test (1-sample F-test, one-sided, 95%).

In the guidelines, often only the total number of replicates is given, without indication of the measurement design in terms of replicates per day [14,15,17–22,25]. Nevertheless, all guidelines require the estimation of the intra- and inter-assay precision, either explicitly or implicitly. Only 4, however, require the estimation of total precision [26,28,35,36]. The experimental design, however, has great influence on the reliability of the intra- and inter-assay component of precision (see Table 1). The table shows the degrees of freedom for the different precision estimates for 5 measurement designs with a total number of ~20 replicates. Note that the 1st design ($10r \times 1d$ and $1r \times 10d$) does not allow to calculate the inter-assay precision. Among the others, only the $2r \times 10d$ design gives balanced degrees of freedom for the intra- and inter-assay precision estimate. The importance of a balanced design has also been discussed elsewhere [2]. Another advantage of this design is that the intra-assay precision is the most representative one because it is estimated on 10 different days. Note that the degrees of freedom for the total precision must be calculated by the Satterthwaite approximation (Table 1 lists the maximum ones) [37]. Depending on the intra-/inter-assay precision ratio, lower values are calculated. The designs that were most commonly used in this Journal (we verified it for the period July to December 2007) were $5r \times 1d$ and $1r \times 5d$, $5r \times 5d$, and $3–6r \times 3d$.

Table 1
Degrees of freedom (df) of precision estimates for various experimental designs with a total of ~20 replicates.

Design (replicates × days)	Intra	Inter	Total (#maximum)
10 × 1 and 1 × 10	9 (1 day)	–	9
5 × 4	16	3	19[a]
7 × 3	18	2	20[a]
3 × 7	14	6	20[a]
2 × 10	10	9	19[a]

[a]The actual dfs should be calculated with the Satterthwaite approximation [37].

Acceptance criteria are found in all guidelines but 1 [22], however, significance testing is only applied in laboratory medicine and 1 ISO document [26,28,35,36]. Interestingly, analyte-specific criteria are used in laboratory medicine [26,35,36], while concentration-related criteria (according to the Horwitz function, (38)) are used in food analysis [27,33], and general criteria (in the order of 10–25%) in environmental and pharmaceutical analysis [14,15,17,18,31]. The practice of setting acceptance criteria will be discussed below in more detail.

We recommend that the analytical chemist should specify an acceptance criterion for precision that is commensurate with the intended use of the procedure (intra-/inter-assay –, and total precision). The experimental design should allow a sufficiently reliable estimation of precision and should have balanced degrees of freedom for the components (for example, $2r \times 10d$ or $2r \times 20d$). The estimated precision should be validated versus the specified value by use of a statistical test (1-sample F-test) or confidence interval of an estimated standard deviation.

2.5. Trueness

The estimation of trueness is required by all guidelines (for a compilation, see Table B.5 of Appendix B). Note, however, that many guidelines erroneously use the term accuracy instead of trueness. Accuracy refers to the distance between a (=1) measured quantity value and a true quantity value of the measurand, whereas trueness stands for the difference between the average of an infinite number of replicate measured quantity values and a reference quantity value (for the exact definitions, see Appendix A). Considering the definition of trueness, it is a fact that the estimates according to many guideline protocols are usually very weak (see below). Trueness assessment is typically combined with precision experiments. Therefore, the remarks about the experimental design, made before for the estimation of precision, also apply here. However, if trueness is estimated by a separate experiment, a reduced design (for example, $2r \times 5d$) may be used, since trueness estimates with few replicates are more reliable than the respective precision estimates.

Acceptance criteria are found in all guidelines but 1 [22], however, significance testing is only applied in laboratory medicine [26,35,36], 1 ISO document [28], and in the more general guidelines [24,25]. Similar to the precision case, analyte-specific criteria are used in laboratory medicine [26,36], concentration-related criteria (empiric) in food analysis [27,33], and general criteria (in the order of 10–25%) in environmental and pharmaceutical analysis [14,15,17,18,31]. Interestingly, in some guidelines similar or the same criteria are used for trueness and precision [14,15,17,18,21,31]. This poses problems when applying significance testing. For example, as shown in Fig. 6, if an acceptance criterion of 15% is used for an estimated trueness of 10% with a R.S.D. of 10%, 20 replicates are necessary before the confidence interval does not include the limit anymore (A), while only 5 replicates are necessary when the R.S.D. is 5% (B).

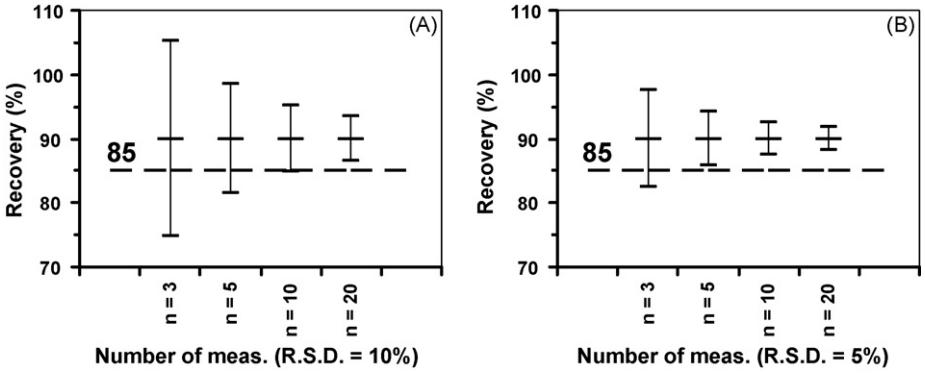

Fig. 6. Different cases of significance testing of a mean trueness of 10% (±confidence limits) (or [100 – 10]% recovery of the true quantity value) against an acceptance criterion of 15% (or [100 – 15] = 85% recovery). *I*n both plots the estimates for the mean trueness were obtained from respectively 3, 5, 10 and 20 replicates, but with a R.S.D. of respectively 10% (A) and 5% (B). They illustrate the relationship between the magnitude of the R.S.D. and n in significance testing.

We recommend that the analytical chemist should specify an acceptance criterion for trueness that is matched with the procedure's precision and is commensurate with the intended use of the procedure. The estimate may be obtained from the precision experiments when certified materials are used. Otherwise, a reduced design may be applied (for example, $2r \times 5d$). The estimated trueness should be validated versus the specified value by use of a statistical test (1-sample *t*-test) or confidence interval of the estimated average difference. In case that several reference materials are used for validation, one has to recognize that the probability of *t*-testing is inflated and that it may be necessary to adapt the significance level of the individual tests by the so-called Bonferroni correction [2,39].

2.6. Total error (accuracy)

The total error is a measure of (in)accuracy and, as already described before, it refers to the distance between a result and the true value of this result. It may be expressed as the sum of the observed bias and k^* imprecision (with k typically 1.96 or 2.58). The estimation of the total error is typically not addressed in validation guidelines, with the exception of those for laboratory medicine [40–43]. This discipline has a long tradition of estimating total error by dedicated method comparison studies using native samples assigned with values by a hierarchically higher reference measurement procedure (e.g., [26]). A typical experimental protocol (EP9-A2) makes use of 40 "real-world" samples measured in duplicate over 5 days (8 samples each day) [44]. Note that the method comparison approach is recommended over the use of matrix-based certified reference materials. It has namely been scientifically proven that the latter may be of limited utility for total error assessment of a hierarchically lower method [45]. This is partly due to the restricted number of concentration levels they cover (usually only 2–3), but mainly to the fact that the way they are prepared (by pooling, supplementation and/or processing) makes them prone to the so-called non-commutability. Therefore, it is generally accepted that before a certified reference material can be used with a method, its commutability has to be assessed. If this has not been done, it is impossible to decide whether any observed bias is genuine or an artifact due to the inadequacy of the material.

The importance of estimating total error is nowadays also recognized in other application fields [4,46–49), however, it has not yet entered into the respective guidelines. One of the reasons is that the statistical concepts for acceptable total error are complex and that, currently, no generally accepted concept for statistical treatment of total error exists.

2.7. Acceptance criteria

According to the definition of validation, acceptance criteria should be tailored to the "specific intended use or application" of a measurement procedure, or, as it is nowadays said, methods should be "fit-for-purpose". Principally, this would require that, for each application, dedicated performance specifications be established. In laboratory medicine, several approaches for setting performance specifications are discussed [50]. Among them, there is the principle of defining an analytical R.S.D. limit (R.S.D.a) for monitoring a patient as half the within-subject biological variation of the component of interest (R.S.D.w). Indeed, when the R.S.D.a = 0.5 × R.S.D.w, the analytical variation will increase the total variation of the result (R.S.D.t) by 12%, only ($\text{R.S.D.t} = \sqrt{[\text{R.S.D.a}^2 + \text{R.S.D.w}^2]} = 1.12 \times \text{R.S.D.w}$, if R.S.D.a = 0.5 × R.S.D.w). By way of example, for serum sodium analysis, a R.S.D.a as tight as 0.4% would be calculated, for serum estradiol analysis, a R.S.D.a of 9% would be sufficient [51]. This principle in fact can also be applied in other disciplines. Consider, for example, a pharmaceutical company where a drug with a narrow therapeutic range is under development. This would require that monitoring of the drug is performed in an early phase of development, e.g., as part of the clinical phase study. It would be logical that the analytical performance of the method to be used for that application is tailored on the therapeutic requirements. In that context, it has, for example, already been suggested that monitoring of lithium therapy requires a R.S.D.a <3%, while monitoring of primidone can be done with a R.S.D.a <11% [52]. Another field is that of environmental analysis. Suppose that a company has to "pay for pollution", e.g., on the basis of the nitrate content as determined in its wastewater effluent. For that application, the company would for sure be interested in keeping the bias of its analytical test under strict control (say <2%). On the other hand, for nitrate testing under field conditions, a much higher bias may be tolerable (say <10%). Also in food analytics, the principle applies. Suppose, for example, that the milk price is connected to the protein content. In that case, again, one would strive for small biases in the analytical method used for protein quantification.

In summary, we strongly advocate to introduce the establishment of analytical performance specifications from individual "fitness-for-purpose" criteria in all analytical disciplines. Ideally, laboratories should actively be involved in discussions to establish or at least to reassess regulatory limits. This is, for example, done in the discipline of food chemistry, more in particular with respect to maximum limits for undesirable contaminants in foodstuffs. These limits are not only derived from risk or human exposure assessment, but also account for the opinion of analytical chemists, e.g.,

Table 2
Analyte concentration and R.S.D. (%) according to the Horwitz equation [38].

Analyte ratio	x (g/g)	R.S.D. (%)
1.00	1 g	1.0
10^{-1}	100 mg	1.4
10^{-2}	10 mg	2.0
10^{-3}	1 mg	2.8
10^{-4}	100 μg	4.0
10^{-5}	10 μg	5.7
10^{-6}	1 μg	8.0
10^{-7}	100 ng	11.3
10^{-8}	10 ng	16.0
10^{-9}	1 ng	22.6
10^{-10}	100 pg	32.0
10^{-11}	10 pg	45.3
10^{-12}	1 pg	64.0

through the network of national reference laboratories in Europe. In disciplines where such consultation rounds are not yet common practice, and where the analytical field would consider that the legal performance specifications are too stringent, laboratory associations should try to influence regulatory authorities to expand them. In case they would be too loose, laboratories should be prudent to design their methods to a quality that is better than required by regulation. Naturally, it is extremely difficult to define what "is fit-for-purpose". It should be admitted that even in disciplines where the practice is generally established and where there is agreement about the models to establish acceptance criteria, e.g., in laboratory medicine [53], there is seldom a consensus about the actual numbers to be used. However, we consider it inevitable that all analytical fields engage in setting such specifications and move away from general "all purpose" criteria. In particular, we think that the precision criteria based on the Horwitz function [38] should be abandoned. The Horwitz function ($\%\text{R.S.D.} = 2^{(1-0.5\times\log C)}$) is regarded as a sort of "law of nature" to set general precision criteria based on analyte concentration. For a better understanding of the above statement, we will compare the acceptable values for the R.S.D. (%) calculated by the original Horwitz function for analyte ratios from 1 to 10^{-12} (see Table 2) with those achievable today. For example, for an analyte present in serum at a concentration of 10 pg/mL (mass fraction $\sim 10^{-11}$), the Horwitz function would predict an R.S.D. of 45%. When analyzing steroid hormones at that concentration by isotope dilution–gas chromatography/mass spectrometry, the R.S.D. values attainable with reasonable effort are typically of the order of 5% [54]. On the Horwitz scale, this R.S.D. would correspond to a concentration of $\sim$10 μg/mL (mass fraction $\sim 10^{-5}$). Because of these recognized deficiencies at mass fractions $<10^{-7}$, modifications of the Horwitz function were proposed [55]. They would predict a R.S.D. of 22% at a mass fraction of 10^{-11}, which, still, is deficient in predicting precision of analytical procedures tailored for "high performance". These facts, indeed, demonstrate that there is no "law of nature" that can predict the precision of a procedure from the analyte concentration. They rather show that it is perfectly possible to tailor the precision of a measurement procedure by careful choice of the measurement principle, the analytical equipment, and the sample size. Of course, the implication of pushing down the LoQ is that costs rise in a non-linear fashion. Similar considerations hold true for the empiric concentration-related criteria sometimes used for trueness [27,33].

2.8. Significance testing

As discussed before, with the exception of the discipline of laboratory medicine, significance testing is seldom part of validation guidelines. This fact has been criticized before (e.g., [48]) and we join that critique. However, being involved in method validation in laboratory medicine, we must say that even in this discipline, significance testing is seldom applied. By way of illustration, it was done in only 2% of papers that interpreted method comparison studies by use of the Bland and Altman approach [56].

On the other hand, over the years, a lot of attention has been paid to statistical issues connected to method validation. Terms associated with some of the newer concepts are interval hypothesis testing, equivalence testing, the "two-one-sided-*t* test approach" (TOST), and accuracy profiles. Interval hypothesis testing, which has been introduced in the 90ies in method validation [57], is based on the TOST approach developed earlier for the interpretation of bioequivalence studies [58,59]. This approach is also addressed in the literature under the general topic of "equivalence testing" [60–62]. Other approaches are based on the so-called accuracy profile constructed by use of the β-expectation tolerance interval [47,49,63,64]. This approach, however, is still under development and regularly refined [46,65,66]. Boiled down, the concepts recommend to consider α- and β-errors and tolerance intervals in method validation studies. A more detailed discussion of the statistical approaches for method validation is beyond the scope of this review. Beware that testing becomes, in particular, more complicated when uncertainties of reference values cannot be neglected. Thus, the "bad" news is that analytical chemists need to develop profound statistical skills [67]. Personally, we have found the NIST Special Publication 829 on the "Use of NIST Reference Materials for Decisions on Performance of Analytical Chemical Methods and Laboratories" one of the most useful documents in this regard [68].

3. Conclusion

In this paper it is documented that there is a variety of specific and more general guidance available for validation of methods used in laboratory medicine-, pharmaceutical-, environmental-, and food-analysis. A special aspect of the review is that it investigated the recommendations for validating the LoD, LoQ, linearity, precision, and trueness across the disciplines. Striking was the observation of quite some difference in experimental measurement designs (number of samples, concentration levels, replicates, spread over several days, balanced degrees of freedom, etc.), but even more striking was the difference in recommended acceptance criteria and significance testing, if any available. The review extracted from the investigated guidelines the most important elements to formulate recommendations that can lead to improvement/optimization of the currently available validation and interpretation protocols.

Acknowledgements

The authors acknowledge the information and expert advice received from Sarah De Saegher, Laboratory of Food Analysis, Faculty of Pharmaceutical Sciences, Ghent University (BE), Ingrid Temmerman, Vlaamse Milieumaatschappij (VMM), Gent (BE) and Joris Van Loco, Scientific Institute of Public Health, Brussels (BE).

Appendix A. Terminology/definitions

Terminology/definitions according to the VIM [11], except differently stated.

A.1. Validation

Confirmation, through the provision of objective evidence, that requirements for a specific intended use or application have been fulfilled [1].

A.2. Validation

Verification, where the specified requirements are adequate for a stated use.

A.3. Verification

Provision of objective evidence that a given item fulfils specified requirements, taking any measurement uncertainty into consideration.

A.4. Limit of detection

Measured quantity value, obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component in a material is β, given a probability α of falsely claiming its presence.

A.5. Limit of detection (in analysis)

The limit of detection, expressed as the concentration, *cL*, or the quantity, *qL*, is derived from the smallest measure, *xL*, that can be detected with reasonable certainty for a given analytical procedure. The value of *xL* is given by the equation $xL = xbi + k \bullet sbi$, where *xbi* is the mean of the blank measures, *sbi* is the standard deviation of the blank measures, and *k* is a numerical factor chosen according to the confidence level desired [12].

A.6. Measuring interval

Set of values of the quantities of the same kind that can be measured by a given measuring instrument or measuring system with specified instrumental uncertainty, under defined conditions.

Note: the measuring interval, indirectly defines the [lower and upper] limits of quantitation, LoQ.

A.7. Linear range

Concentration range over which the intensity of the signal obtained is directly proportional to the concentration of the species producing the signal [12].

A.8. Precision

Closeness of agreement between indications obtained by replicate measurements on the same or similar objects under specified conditions.

A.9. Accuracy

Closeness of agreement between a measured quantity value and a true quantity value of the measurand.

A.10. Trueness

Closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value.

A.11. Uncertainty

Parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used.

Appendix B. Protocols, acceptance criteria and statistical tests

Tables B.1–B.5.

Table B.1
Limit of detection (LoD)—protocols, acceptance criteria and statistical tests.

Guideline	Experimental protocol and definition	Acceptance criteria[a]	Test
Laboratory medicine			
[26]	Blank 2 replicates (r) × 5 days (d) k × SD (k depends on α- and β-error)	No	NA
[13]	Blanks and "several" samples near LoD Example: 4S and B × 3r × 5d (n = 60, each) LoD based on α- and β-error[b]	"Peer" LoD (analyte specific)	Yes
Pharma			
[22]	No n Blank: 3.3 × SD, or Low samples (chromatogr.): S/N = 2 or 3	No	NA
[17,18]	Not addressed	NA	NA
Environment			
[14]	Samples with low concentration 7r (no design) t (0.01, 6, 1-sided) (=3.143) × SD	No	NA
[15]	Low concentration sample 7r in 3 days t (0.01, 6, 1-sided) (=3.143) × SD	Analyte specific	No
[28]	Blank 2r × 10d 2 × √2 × t(0.05, 1-sided) × SD	No	NA
Food			
[23]	Blank 20r (no design) 3 × SD	No	NA
[19,20]	Not addressed	NA	NA
[27][c]	Low samples, 20r (no design) (a) Based on α- and β-error (b) Chromatographic: S/N = 3	No	NA
General			
[21]	Not addressed	NA	NA
[24]	Blank or low level sample 6r (no design) 3 × SD	No	NA
[25]	10 blank or low level samples 1r (no design) 3 × SD	No	NA

[a] Usually, LoD is used as descriptive parameter; nevertheless, the confidence interval should be calculated. NA: not applicable (not addressed or no acceptance criteria). [b] 5% α-error (blank)/5% β-error (LoD-sample). [c] Called decision limit, 1% α-error (blank)/5% β-error (LoQ-sample); other: by calibration function.

Table B.2
Limit of quantitation (LoQ)—protocols, acceptance criteria and statistical tests.

Guideline	Experimental protocol and definition	Acceptance criteria	Test
Laboratory medicine			
[26]	Not described	NA	NA
[13]	Reference materials near LoQ Example: 4RM × 2 replicates (r) × 5 days (d) (total n = 40)	"Peer" LoQ (analyte specific)	Yes
Pharma			

Table B.2 (*Continued*)

Guideline	Experimental protocol and definition	Acceptance criteria	Test
[22]	(a) Samples at 10 × SD, or S/N = 10 (chrom.); (b) samples with "acceptable" trueness and precision. No *n* given	No	NA
[17,18]	Samples with >5 × blank response No *n* given Defined by CV and trueness[$]	No	NA
Environment			
[14]	Samples with low concentration 7*r* (no design) 4 × LoD	No	NA
[15]	9*r* (no design) Defined by CV and trueness[$]	Lowest standard (given in standard method)	No
[28,31]	Samples at 1/10th/1/4th of limit (elements/organics)[§] $2r \times 10d$	10% (elements)[§] 25% (organics)	No
Food			
[23]	Blank 20*r* (no design) 10 × SD	No	NA
[19,20]	5*r* (no design) Defined by CV and trueness[$]	No	NA
[27]	Not addressed	NA	NA
General			
[21]	5 samples near lowest standard No *n* given Defined by CV and trueness[$]	No	NA
[24]	Not recommended	NA	NA
[25]	10 blank or low level samples 1*r* (no design) 5, 6, or 10 × SD (other: by uncertainty)	No	NA

NA: not applicable (not addressed or no acceptance criteria). [§][31] (no protocols given). [$]Typically: 20% CV and trueness.

Table B.3
Linearity—protocols, acceptance criteria and statistical tests.

Guideline	Experimental protocol	Acceptance criteria	Test
Laboratory medicine			
[26]	6-points calibration curve 3 replicates (*r*) × 3 days (*d*)	Statistical	Yes
[32]	5-points, equidistant (sample mixtures) $2r \times 1d$	Statistical or user-defined (analyte-specific)	Yes
Pharma			
[22]	5 points Design not given	(a) Visual (b) Investigate regression data	No
[17,18][$]	No *n* given	$r \geq 0.999$; give slope and intercept	No
Environment			
[14,69][§]	No *n* given	$r \geq 0.99$	No
[15]	Calibration curve (3–5 points) Design not given	Defined by precision and trueness criteria	No
[34]	10 calibration points 1*r* (note 1st and last point 10*r*, for testing homoscedasticity)	(a) Visual (b) Statistical	Yes
Food			
[33]	Calibration curve, 6 to 8 points $2r \times 1d$	Visual *r not* recommended	No
[19,20]	For calibration curves with ≥3 points (bracketing and 1-point calibration allowed) Design not given	Visual and residuals ≤20% from predicted value (10% at limit) *r* is *not* recommended	NA

Table B.3 (*Continued*)

Guideline	Experimental protocol	Acceptance criteria	Test
[27]	No (report calibration function and "goodness-of-fit-data")	NA	NA
General			
[21]	Determine response function by appropriate statistical tests (*n*: "single or replicates")	Statistical (simplest relationship)	No
[24]	At least 6 points 2 or $3r \times 1d$	(a) Statistical (b) Visual	Yes and visual
[25]	At least 10 points, 1*r* and 6 points, $3r \times 1d$	(a) Statistical (b) Visual	Yes and visual

[$]Only in [18]. [§]Only in [69]. NA: not applicable (not addressed or no acceptance criteria).

Table B.4
Precision—protocols, acceptance criteria and statistical tests.

Guideline	Experimental protocol	Acceptance criteria	Test
Laboratory medicine			
[26]	Low, mid, and high QC or patient samples 2 replicates (*r*) × 5 days (*d*) Inter, intra, and total (ANOVA)	Analyte-specific (tables given)	Yes
[36]	Low, and high QC or patient samples $3r \times 5d$ Intra, inter, and total (ANOVA)	"Peer" imprecision (analyte-specific)	Yes
[35]	Low, and high QC or patient samples $2r \times 20d$ Intra, inter, and total (ANOVA)	"Peer" imprecision (analyte-specific)	Yes
Pharma			
[22]	*Intra:* 3 concentrations, 3*r* (no design) 6*r* at 100% (no design) *Inter:* ANOVA (user decides on protocol)	No	NA
[17,18]	*Intra:* 3 concentrations, 5*r* (no design) *Inter:* Addressed, but no design given	CV ≤15% CV ≤20% at LoQ	No
Environment			
[14]	5 and 50 times LoQ in 3 matrices 7*r* (no design)	≤15% long-term ≤20% recovery duplicates	No
[15]	9*r* (no design)	≤20%: LoQ–10 × LoQ ≤10%: 10 × LoQ to highest calibrator	No
[28,31]	2 Concentrations $2r \times 10d$ Intra, inter, and total (ANOVA)	[§]10%/25% (elements/organics)	Yes
Food			
[33]	Reference materials *Intra* and *inter:* $2r \times 5d$	Concentration-related (Horwitz function)	No
[19,20]	2 Concentrations 5*r* (no design)	CV ≤20%	No
[27]	*Intra* and *inter:* 3 samples $6r \times 3d$ (Inter: different conditions every day)	Concentration-related (Horwitz function) Elements: 10–20%	No
General			
[21]	3 Concentrations 5*r* (no design)	≤15% ≤20% at LoQ	No
[24]	Intra and inter ANOVA (adapt design to application) Example: $2r \times 10d$	From respective regulation	No
[25]	Intra and inter 10*r* (no design)	From respective regulation	No

[§][31] (no protocols given). NA: not applicable (not addressed or no acceptance criteria).

Table B.5
Trueness—protocols, acceptance criteria and statistical tests.

Guideline	Experimental protocol	Acceptance criteria	Test
Laboratory medicine			
[26]	"Several" samples 2 replicates (r) × 5 days (d) (other: method comparison)	Analyte-specific (tables given)	Yes
[36,41,44]$	2 samples $2r \times 5d$	"Peer" trueness (analyte-specific)	Yes
Pharma			
[22]	3 concentrations $3r$ (no design)	No	NA
[17,18]	3 concentrations $5r$ (no design)	≤15% deviation (mean) ≤20% deviation at LoQ	No
Environment			
[14]	5 and 50 times LoQ in 3 matrices $7r$ (no design)	≤20%	No
[15]	$9r$ (no design)	≤25% at LoQ–10 × LoQ ≤15% at 10 × LoQ to highest calibrator	No
[28,31]	2 concentrations $2r \times 10d$	§10%/25% (elements/(organics)	Yes
Food			
[33]	Reference materials $2r \times 5d$	Concentration related (98–101%; 70–125%)	No
[19,20]	2 Concentrations $5r$ (no design)	Trueness 70–120%	No
[27]	Reference materials $6r$ (no design)	Concentration related (−20/+10%; −50/+20%) Elements: ±10%	No
General			
[21]	3 Concentrations $5r$ (no design)	≤15% deviation (mean) ≤20% deviation at LoQ	No
[24]	Adapt design to application Example: $2r \times 10d$	From respective regulation	Yes
[25]	$10r$ (no design)	From respective regulation	Yes

$Additional protocols: [44] (assessment of regression-predicted and average bias) and [41] (assessment of total error). § In Ref. [31] (no protocols given). NA: not applicable (not addressed or no acceptance criteria).

References

[1] ISO 9001, Quality Management Systems—Requirements, International Organization for Standards (ISO), Geneva, 2000.
[2] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 17 (1998) 193.
[3] K. Linnet, J.C. Boyd, in: C.A. Burtis, E.R. Ashwood, D.E. Bruns (Eds.), Tietz Textbook of Clinical Chemistry and Molecular Diagnostics, Elsevier Saunders, St Louis, MO, 2006, p. 353.
[4] J. Ermer, J.H.McB. Miller, Method Validation in Pharmaceutical Analysis, Wiley-VCH, Weinheim, 2005.
[5] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, J. Chromatogr. A 1158 (2007) 111.
[6] S. Chandran, R.S. Singh, Pharmazie 62 (2007) 4.
[7] S. Schmidt, Environ. Sci. Pollut. Res. Int. 10 (2003) 183.
[8] P. Hecq, A. Hulsmann, F.S. Hauchman, J.L. McLain, F. Schmitz, in: P. Quevauviller, K.C. Thompson (Eds.), Analytical Methods for Drinking Water, John Wiley & Sons, New York, NY, 2005, p. 1.
[9] J. Van Loco, in: S. Caroli (Ed.), The Determination of Chemical Elements in Food, John Wiley & Sons, New York, NY, 2007, p. 135.
[10] I. Taverniers, M. De Loose, E. Van, Bockstaele, Trends Anal. Chem. 23 (2004) 535.
[11] ISO/IEC Guide 99, International Vocabulary of Metrology, Basic and General Concepts and Associated Terms (VIM), International Organization for Standards (ISO), Geneva, 2007.
[12] IUPAC Compendium of Chemical Terminology (Gold book) <http://goldbook.iupac.org/> (accessed 11.11.08).
[13] EP17-A, Protocols for Determination of Limits of Detection. Clinical and Laboratory Standards Institute (CLSI), Wayne, PA, 2004.
[14] Guidance for Methods Development and Methods Validation for the Resource Conservation and Recovery Act (RCRA) Program (SW-846 methods), EPA, Washington, DC, 1995.
[15] EPA method 300.1, Determination of Inorganic Anions in Drinking Water by Ion Chromatography. Revision 1.0. Cincinatti, OH, 1997.
[16] L.A. Currie, Pure Appl. Chem. 67 (1995) 1699.
[17] Guidance for Industry: Bioanalytical Method Validation, US Department of Health and Human Services, Food and Drug Administration, Center for Biologics Evaluation and Research (CBER), Rockville, MD, 2001.
[18] Reviewer Guidance. Validation of Chromatographic Methods. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Rockville, MD, 1994.
[19] Residues: Guidance for Generating and Reporting Methods of Analysis in Support of Pre-registration Data Requirements for Annex II (part A, Section 4) and Annex III (part A, Section 5) of Directive 91/414, in Doc. SANCO/3029/99, 11 July 2000.
[20] Method Validation and Quality Control procedures for Pesticide Residues, in Doc. SANCO/2007/3131, 31 October 2007.
[21] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, Mark J. Rose, J. Sailstad,Vinod P. Shah, Jerome P. Skelly, Patrick G. Swann, R. Weiner. Workshop/Conference Report–Quantitative Bioanalytical Methods Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays The AAPS Journal 2007; 9 (1) Article 4 <http://www.aapsj.org>.
[22] ICH Topic Q2 (R1), Validation of Analytical Procedures: Text and Methodology, International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva, 2005.
[23] AOAC Peer-Verified Methods Program, Manual on Policies and Procedures, Arlington, VA, 1998.
[24] M. Thompson, S.L.R. Ellison, R. Wood, Pure Appl. Chem. 78 (2006) 145.
[25] The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics. EURACHEM, Middlesex, 1998.
[26] A. Vassault, D. Grafmeyer, C.I. Naudin, G. Dumont, M. Bailly, J. Henny, M.F. Gerhardt, P. Georges, Société Française de Biologie Clinique, Ann. Biol. Clin. 44 (1986) 720.
[27] 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. Off. J. Eur. Commun., L221, 2002, 8.
[28] ENV-ISO 13530, Water Quality–Guidance on Analytical Quality Control for Chemical and Physicochemical Water Analysis (under revision), International Organization for Standards (ISO), Geneva, 1998.
[29] ISO 11929, Determination of the Detection Limit and Decision Threshold for Ionizing Radiation Measurements—Part 8 in the Series, International Organization for Standards (ISO), Geneva, 2005.
[30] M. Thompson, S.L.R. Ellison, R. Wood, Pure Appl. Chem. 74 (2002) 835.
[31] Council Directive 98/83/EC; on the quality of water intended for human consumption, Off. J. Eur. Commun. L330, 1998, 32.
[32] EP6-A, Evaluation of the Linearity of Quantitative Measurement. Clinical and Laboratory Standards Institute (CLSI), Wayne, PA, 2003.
[33] Guidelines for Single Laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals, Association of Official Analytical Chemists (AOAC), 2002.
[34] ISO 8466-1, Water Quality—Calibration and Evaluation of Analytical Methods and Estimation of Performance Characteristics–Part 1: Statistical Evaluation of the Linear Calibration Function, International Organization for Standards (ISO), Geneva, 1990.
[35] EP5-A2, Evaluation of Precision Performance of Quantitative Measurement Methods. Clinical and Laboratory Standards Institute (CLSI), Wayne, PA, 2004.
[36] EP15-A2, User Verification of Performance for Precision and Trueness. Clinical and Laboratory Standards Institute, Wayne, PA, 2006.
[37] F. Satterthwaite, Biometr. Bull. 2 (1946) 110.
[38] W. Horwitz, Pure Appl. Chem. 67 (1995) 331.
[39] G.W. Snedecor, W.G. Cochran, Statistical Methods, 7th ed., Iowa State University Press, Ames, 1980.
[40] J.S. Krouwer, Setting performance goals and evaluating total analytical error for diagnostic assays, Clin. Chem. 48 (2002) 919.
[41] EP21-A, Estimation of Total Analytical Error for Clinical Laboratory Methods; Approved Guideline. Clinical and Laboratory Standards Institute, CLSI21-A, Wayne, PA, 2003.
[42] J.M. Bland, D.G. Altman, Lancet 1 (1986) 307.
[43] D. Stöckl, D. Rodríguez Cabaleiro, K. Van Uytfanghe, L.M. Thienpont, Clin. Chem. 50 (2004) 2216.
[44] EP9-A2, Method Comparison and Bias Estimation. Clinical and Laboratory Standards Institute, Wayne, PA, 2002.
[45] H.W. Vesper, W.G. Miller, G.L. Myers, Clin. Biochem. Rev. 28 (2007) 139.
[46] D. Hoffman, R. Kringle, Pharm. Res. 24 (2007) 1157.
[47] A.G. Gonzalez, M.A. Herrador, Trends Anal. Chem. 26 (2007) 227.
[48] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 82.
[49] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 48 (2008) 760 (part IV in a series of papers).
[50] D. Stöckl, H. Baadenhuijsen, C.G. Fraser, J.C. Libeer, P.H. Petersen, C. Ricós, Eur. J. Clin. Chem. Clin. Biochem. 33 (1995) 157.
[51] Desirable specifications for total error, imprecision, and bias, derived from biologic variation. <http://www.westgard.com/biodatabase1.htm> (accessed 11.11.08).
[52] C.G. Fraser, Clin. Chem. 33 (1987) 387.
[53] D. Kenny, C.G. Fraser, P. Hyltoft Petersen, A. Kallner, Scand. J. Clin. Lab. Invest. 59 (1999) 585.
[54] L.M. Thienpont, P.G. Verhaeghe, K.A. Van Brussel, A.P. De Leenheer, Clin. Chem. 34 (1988) 2066.

[55] M. Thompson, Analyst 125 (2000) 385.
[56] K. Dewitte, C. Fierens, D. Stöckl, L.M. Thienpont, Clin. Chem. 48 (2002) 799.
[57] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Vankeerberghen, D.L. Massart, Anal. Chem. 67 (1995) 4491.
[58] D.L. Schuirman, Biometrics 37 (1981) 617.
[59] W.J. Westlake, Biometrics 37 (1981) 589.
[60] S. Feng, Q. Liang, R.D. Kinser, K. Newland, R. Guilbaud, Anal. Bioanal. Chem. 385 (2006) 975.
[61] M. Meyners, Food Qual. Prefer. 18 (2007) 541.
[62] G.B. Limentani, M.C. Ringo, F. Ye, M.L. Bergquist, E.O. McSorley, Anal. Chem. 77 (2005) 221A.
[63] H.M. Feinberg, J. Chromatogr. A 1158 (2007) 174.
[64] R.M. Mee, Technometrics 26 (1984) 251.
[65] E. Rozet, V. Wascotte, N. Lecouturier, V. Préat, W. Dewé, B. Boulanger, P. Hubert, Anal. Chim. Acta 591 (2007) 239.
[66] T. Rebafka, S. Clemencon, M. Feinberg, Chemom, Intell. Lab. Systems 89 (2007) 69.
[67] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, 4th ed., Pearson Education Limited, Essex, 2000.
[68] D. Becker, R. Christensen, L. Currie, B. Diamondstone, K. Eberhardt, T. Gills, H. Hertz, G. Klouda, J. Moody, R. Parris, R. Schaffer, E. Steel, J. Taylor, R. Watters, R. Zeisler, NIST Special Publication 829. US Government Printing Office, Washington, DC, 1992.
[69] EPA Method 8000C, Determinative Chromatographic Separations (Revision 3). EPA, Washington, DC, 2003.