A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles

A. Gustavo González, M. Ángeles Herrador

The objective of analytical method validation is to ensure that every future measurement in routine analysis will be close enough to the unknown true value for the content of the analyte in the sample. Classical approaches to validation only check performance against reference values, but this does not reflect the needs of consumers. A holistic approach to validation also takes into account the expected proportion of acceptable results lying inside predefined acceptability intervals.

In this article, we give a detailed step-by-step guide to analytical method validation, considering the most relevant procedures for checking the quality parameters of analytical methods. Using a holistic approach, we also explain the estimation of measurement uncertainty and accuracy profiles, which we discuss in terms of accreditation requirements and predefined acceptability limits.

Keywords: Accuracy profile; Analytical method validation; Measurement uncertainty; Performance characteristic

Abbreviations: ANOVA, Analysis of variance; βETI, β-expectation tolerance interval; CS, Calibration standard; DTS, Draft technical specification; ELISA, Enzyme-linked immunosorbent assay; FDA, Food and Drug Administration; ICH, International Conference on Harmonization; ICP, Inductively coupled plasma; ISO, International Organization for Standardization; IUPAC, International Union of Pure and Applied Chemistry; LGC, Laboratory of Government Chemist; LOD, Limit of detection; LOQ, Limit of quantitation; RSD, Relative standard deviation; SAM, Standard-addition method; SB, System blank; SFSTP, Societé Française des Sciences et Techniques Pharmaceutiques; TR, Tolerated ratio; TYB, Total Youden blank; USP, United States Pharmacopoeia; VAM, Valid analytical measurement; VS, Validation standard; YB, Youden blank.

A. Gustavo González*, M. Ángeles Herrador
Department of Analytical Chemistry, University of Seville, E-41012 Seville, Spain

*Corresponding author.
Tel.: +34 954557173;
Fax: +34 954557168;
E-mail: agonzale@us.es

1. Introduction

The final goal of the validation of an analytical method is to ensure that every future measurement in routine analysis will be close enough to the unknown true value for the content of the analyte in the sample. [1]. Accordingly, the objectives of validation are not simply to obtain estimates of trueness or bias and precision but also to evaluate those risks that can be expressed by the measurement uncertainty associated with the result [2]. Accuracy, according to the ISO 5725 definition [3], comprises two components – trueness and precision – but, instead of assessing these independently, it is possible to assess accuracy in a global way according to the concept of acceptability limits and accuracy profiles [4–8]. Accuracy profiles and measurement uncertainty are related topics, so either can be evaluated using the other. In a holistic sense, as Feinberg and Laurentie pointed out [9], method validation, together with uncertainty measurement or accuracy-profile estimation, can provide a way to check whether an analytical method is correctly fit for the purpose of meeting legal requirements. Fitness for purpose is the extent to which the performance of a method matches the criteria that have been agreed between the analyst and the end-user of the data or the consumer and that describe their needs [10]. Classical approaches to validation consisted of checking the conformity of a performance measure to a reference value, but this does not reflect the consumer's needs, mentioned above. By contrast, the holistic approach to validation establishes the expected proportion of acceptable results lying between predefined acceptability limits. Many excellent papers and guides have been written about the validation of analytical methods but no attention has been paid to the holistic paradigm. We aim to provide to the analyst with a practical

0165-9936/$ - see front matter © 2007 Elsevier Ltd. All rights reserved. doi:10.1016/j.trac.2007.01.009

guide to performing the validation of analytical methods using this holistic approach.

2. Practical approach to global method validation

For the sake of clarity, we have divided the content of the guide into four sections that we will outline and explain, as follows and as shown in Fig. 1:

(1) applicability, fitness for purpose and acceptability limits;
(2) specificity and selectivity;
(3) calibration study, involving the goodness of the fit of the calibration function and dynamic concentration range, sensitivity and detection and determination limits, as well as assessment for matrix effects; and,
(4) accuracy study, involving trueness, precision and robustness as well as the estimation of measurement uncertainty and accuracy profiles

2.1. Applicability, fitness for purpose and acceptability limits

The method applicability is a set of features that cover, apart from the performance specifications, information about the identity of analyte (e.g., nature and speciation), concentration range covered, kind of matrix of the material considered for validation, the corresponding protocol (describing equipment, reagents, analytical procedure, including calibration, as well as quality procedures and safety precautions) and the intended application with its critical requirements [10]. The method applicability must be consistent with the ''golden rules'' for method validation proposed by Massart et al. [11], namely:

(1) the analytical procedure has to be validated as a whole, including sample treatments prior to analysis;
(2) the analytical procedure has to be validated covering the full range of analyte concentrations specified in the method scope; and,
(3) the analytical procedure has to be validated for each kind of matrix where it will be applied.

Fitness for purpose [10,12,13] is the extent to which the method performance matches the agreed criteria or requirements. A laboratory must be capable of providing results of the required quality. The agreed requirements of an analytical method and the required quality of the analytical result (i.e. its accuracy) refer to the fitness for purpose of the analytical method. The accuracy can be assessed in a global way, as indicated above, by using the concept of acceptability limit λ [6–8]. Thus, analytical result Z may differ from unknown ''true value'' T to an extent less than the acceptability limit:

$$|Z - T| < \lambda \quad (1)$$

Limit λ depends on the goals of the analytical procedure: 1% for bulk materials; 5% for determination of active ingredients in dosage forms; and, 15% in bio/environmental analysis [8]. A procedure can be validated if it is very likely that the requirement given by (1) is fulfilled, i.e.:

$$P(|Z - \mu| < \lambda) \geqslant \beta \quad (2)$$

β being the probability that a future determination falls inside the acceptability limits. It is possible to compute the so-called ''β-expectation tolerance interval'' (βETI) (i.e. the interval of future results that meet Equation (2)) by using the accuracy profiles that we will describe later (in Section 2.4., devoted to accuracy study and measurement uncertainty). The use of acceptability limits together with accuracy profiles is an excellent way to check the fitness for purpose of the validated method.

2.2. Specificity and selectivity

Selectivity is the degree to which a method can quantify the analyte accurately in the presence of interferences under the stated conditions of the assay for the sample matrix being studied. As it is impracticable to consider every potential interference, it is advisable to study only the worst cases that are likely [10]. The absolute absence of interference effects can be taken as ''specificity'', so specificity = 100% selectivity [13]. The selectivity of a method can be quantitatively expressed by using the maximum tolerated ratio (TR_{max}) [14] (i.e. the concentration ratio of interference (C_{int}) to analyte (C_a) leading to a disturbance (systematic error) on the analytical response that yields a biased estimated analyte concen-

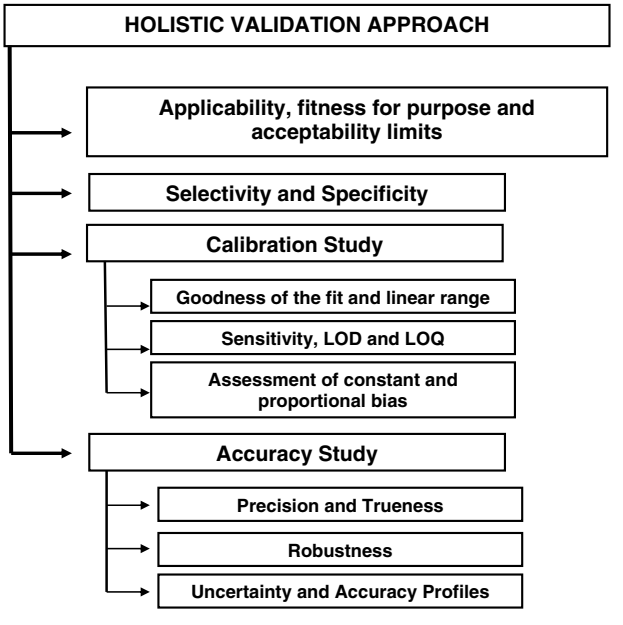

Figure 1. Scheme for the holistic approach to validation.

tration, $\widehat{C}_a$, falling outside the confidence interval derived from the expanded uncertainty U at a given probability confidence level):

$$TR_{\max} = \frac{C_{\text{int}}}{C_a} \qquad \widehat{C}_a \notin [C_a - U, C_a + U] \tag{3}$$

In this way, selectivity is supported by uncertainty, and, at least, a crude estimation of the uncertainty is needed for the nominal concentration of analyte (C_a) used in the selectivity study.

The selectivity, when applying separative analytical methods (e.g., chromatography or capillary electrophoresis), is envisaged as the ability of the method to measure accurately the analyte in presence of all the potential sample components (e.g., placebo formulation, synthesis intermediates, excipients, degradation products, and process impurities) [15], leading to pure, symmetric peaks with suitable resolution [16]. For example, in chromatographic methods, the analyte peak in the mixture should be symmetric with a baseline resolution of at least 1.5 from the nearest eluting peaks.

Non-separative analytical methods (e.g., spectrophotometric methods) are susceptible to selectivity problems (with the exception of highly selective fluorescence-based techniques), but, in some cases, first- and second-derivative techniques may overcome these handicaps [17]. In electroanalytical methods, selective electrodes or amperometric sensors are practical examples of specific devices. For voltammetric methods, pulsed differential votammetry and square-wave techniques are suitable to enhance selectivity [18].

However, sometimes the sensitivity (slope of the calibration function) of the analyte is affected by the sample matrix, leading to another kind of interference for which the matrix acts as a whole. These effects may be circumvented by using in situ calibration following the method of standard additions and we will consider them in Sections 2.3 and 2.4., on calibration study and the accuracy study, respectively, because matrix effects may lead to systematic additive and proportional errors.

2.3. Calibration study

The response function or calibration curve of an analytical method is, within the range, a monotonic relationship between the analytical signal (response) and the concentration of analyte [4]. Response function can be linear, but non-linear models, such as in enzyme-linked immunosorbent assay (ELISA) or inductively coupled plasma (ICP) techniques are also observed.

The response function is obtained using calibration standards (CSs) prepared in absence of matrix sample and relating the response and the concentration. According to the harmonized procedure published by the Societé Française des Sciences et Techniques Pharmaceutiques (SFSTP) [1], several experimental designs are available for CSs and validation standards (VSs). For example, a typical experimental design for calibration consists of preparing duplicate solutions at N concentration levels and replicating over three days or three conditions: $3 \times N \times 2$ [8]. VSs are prepared in the matrix with maybe a different experimental design $p \times m \times n$ (p conditions, m levels and n repetitions), as we will consider in Section 2.4. on accuracy study.

To obtain the best adapted calibration function, several mathematical models can be tested, involving mathematical transformations as well as weighted regression techniques in which the response variance varies as a function of analyte concentration.

2.3.1. Goodness of fit. Suitable regression analysis of the analytical signal (Y) on the analyte concentrations (Z) established in the calibration set yields the calibration curve for the predicted responses ($\widehat{Y}$). The simplest model is the linear one, very often found in analytical methodology, leading to predicted responses according to Equation (4):

$$\widehat{Y} = a + bZ \tag{4}$$

where a is the intercept and b the slope, with standard deviations s_a and s_b, respectively. However, as we stated above, non-linear models can be also applied in a number of analytical techniques, as given in Equation (5):

$$\widehat{Y} = f(\mathrm{Z}) \tag{5}$$

f being the non-linear function to be tested.

Equation (4), or (5), must be checked for goodness of fit. The correlation coefficient, although commonly used, especially in linear models, is not appropriate [19], so some more suitable criteria should be considered. A simple way to diagnose the regression model is to use residual plots [20–22]. For an adequate model, the residuals are expected to be normally distributed, so a plot of them on a normal probability graph may be useful. Any curvature suggests a lack of fit due to a non-linear effect. A segmented pattern indicates heteroscedasticity in data, so weighted regression should be used to find the straight line for calibration [23]. In the latter case, the non-homogeneity of variances can also be checked by using the Cochran criterion [24], when the number of observations is the same for all concentration levels. Another advantage of using calibration designs with repetition is the possibility of estimating a pooled sum of squares due to pure errors, SS_{PE}. The best way to test the goodness of fit is by comparing the variance of the lack of fit against the pure-error variance [22].

The residual sum of squares of the model

$$SS_R = \sum_{i=1}^{N} (\widehat{Y}_i - Y_i)^2 \tag{6}$$

can be decomposed into the sum of squares corresponding to pure error (SS_{PE}) and the sum of squares corresponding to the lack of fit (SS_{LOF}), hence:

$$SS_{LOF} = SS_R - SS_{PE}$$
$$v_{LOF} = v_R - v_{PE} \quad (7)$$

v_{PE} and v_R being the degrees of freedom for estimating the sum of squares of pure error and residuals, respectively.

The pure-error variance is SS_{PE}/v_{PE}, and the variance of the lack of fit is SS_{LOF}/v_{LOF}. In order to estimate the adequacy of the model, the Fisher F-test is applied:

$$F = \frac{(SS_R - SS_{PE})/(v_R - v_{PE})}{SS_{PE}/v_{PE}} \quad (8)$$

The calibration model is considered suitable if F is less than the one-tailed tabulated value $F_{tab}(v_R - v_{PE}, v_{PE}, P)$ at a P selected confidence level.

It is possible, from this proof, to devise a connection with correlation coefficient r by remembering that $1 - r^2$ accounts for the ratio of the residual sum of squares to the total sum of squares of the deviation about the mean, as González et al. pointed out [25].

Always within the realm of linear models, and taking into account the term ''linearity'' in the context of linearity of the response function, some authors consider two features: in-line and on-line linearity [26]. In-line linearity refers to the linearity of the model assessed by the goodness of fit (absence of curvature), and on-line linearity refers to the dispersion of the data around the calibration line and is based on the relative standard deviation of the slope, $RSD_b = s_b/b$. This value is taken as another characteristic parameter of performance and depends on the maximum value accepted for RSD_b, so a typical threshold may be $RSD_b \leqslant 5\%$. Once both ''in-line'' and ''on-line'' linearity are assessed, a test for an intercept significantly different from zero is generally performed by applying the Student t-test [27].

Once the calibration curve is obtained, an inverse prediction equation is then built to predict the actual concentrations of the VSs by considering corrections terms, if needed, owing to the presence of constant and proportional bias.

2.3.2. Linear range. The procedure described by Huber [28] is quite efficient at giving the proper linear dynamic range of analyte from the calibration data. It consists of evaluating so-called response factors RF_i obtained by dividing the signal responses by their respective analyte concentrations. A graph is plotted with the response factors on the y-axis and the corresponding concentrations on the x-axis. The line obtained should be of near-zero slope (horizontal) over the concentration range. This behavior is supported assuming that the model is linear and without intercept, $\widehat{Y}_i = bZ_i$, so $RF_i = \frac{Y_i}{Z_i} \simeq b$. At high concentrations, negative deviation from linearity is expected. Two parallel horizontal lines are drawn in the graph at 0.95 and 1.05 times the average value of the response factors (very close to b, as indicated above) in a fashion similar to the action limits of control charts. The linear range of responses corresponds to the analyte concentrations from the point intersecting the line $y = 1.05b$ up to the point that intersects the line $y = 0.95b$. It is possible that no intersections are found, and, in this case, the linear range applies to the full range being studied, if the minimum concentration level is higher than the limit of detection (LOD) in case of trace analysis.

In routine analysis, linear ranges are established in assay methods as 80–120% of the analyte level. For impurity tests, the lower linearity limit is the limit of quantitation (LOQ), so the range of linearity begins with the LOQ and finishes about the 150% of the target level for the analyte, according to USP/ICH and IUPAC guidelines [13].

2.3.3. Sensitivity, detection limit and quantitation limit. Sensitivity is the change in the analytical response divided by the corresponding change in analyte concentration; i.e. at a given value of analyte concentration Z_0:

$$Sensitivity = \left(\frac{dY}{dZ}\right)_{Z_0} \quad (9)$$

If the calibration is linear, the sensitivity is just calibration slope b at every value of analyte concentration. In addition to sensitivity, there are two parameters reciprocally derived from sensitivity, much more often used for performance characteristics: the limit of detection (LOD) and the limit of determination or quantitation (LOQ).

LOD is the lowest concentration of analyte that can be detected and reliably distinguished from zero (or the noise level of the system), but not necessarily quantified; the concentration at which a measured value is larger than the uncertainty associated with it. LOD can be expressed in response units (Y_{LOD}) and is taken typically as three times the noise level for techniques with continuous recording (e.g., chromatography). Otherwise, it is commonly estimated by using the expression [29]:

$$Y_{LOD} = Y_{blank} + 3s_{blank} \quad (10)$$

where Y_{blank} and s_{blank} are the average value of the blank signal and its corresponding standard deviation, respectively, obtained by measuring at least a minimum of 10 independent sample blanks. Alternatively, when sample blank cannot produce any response (i.e. voltammetry), 10 independent sample blanks fortified at the lowest acceptable concentration of the analyte are

measured and then, $Y_{LOD} = 3s$, s being the standard deviation of the set of measurements.

Nevertheless, LODs expressed in signal units are awful to handle. It is more advisable to use LODs in analyte-concentration units. Thus Y_{LOD} values are converted to Z_{LOD} by using the calibration function:

$$Z_{LOD} = \frac{Y_{LOD} - a}{b} \tag{11}$$

Then, the final LOD value, considering zero intercept, gives:

$$Z_{LOD} = \frac{3s_{blank}}{b} \tag{12}$$

If the errors associated with the calibration line are taken into account [30], another expression can be applied [31]:

$$Z_{LOD} = \frac{2t(v,P)[s_{blank}^2 + s_a^2 + (a/b)^2 s_b^2]^{1/2}}{b} \tag{13}$$

LODs have to be determined only for impurity methods but not for assay methods.

LOQ is the lowest concentration of analyte that can be determined quantitatively with an acceptable level of precision [31]. The procedure for evaluating LOQs is equivalent to that of LODs, by measuring at least 10 independent sample blanks and using the factor 10 instead of 3 for calculations:

$$Y_{LOQ} = Y_{blank} + 10s_{blank} \tag{14}$$

To express LOQ in concentration units, relationships equivalent to Equations (12) and (13) can be applied by changing LOD into LOQ. The reason for the factor 10 comes from IUPAC considerations [31], assuming a relative precision of about 10% in the signal. However, in order to obtain an LOQ more consistent with the definition, it is advisable to make a prior estimation of the RSD of the response against the analyte concentration (near to the unknown LOQ). Thus, a series of blanks are spiked at several analyte concentrations and measured in triplicate. For every addition, the %RSD is calculated. From the plot of %RSD versus the spiked analyte concentration, the amount that corresponds to a previously defined precision RSD is interpolated and taken as the Z_{LOQ} [32]. As indicated above, LOQs are immaterial for assay methods, but, for impurity tests and trace analysis, the lower linearity limit is always the LOQ [13].

2.3.4. Checking proportional and constant bias derived from matrix effects. As stated above, the use of external calibration enormously simplifies the protocol because calibration standards are prepared as simple solutions of the analyte. However, the effects of possible matrix effects coming from the sample material must be checked. A very useful tool for testing constant and proportional bias due to matrix effects is the standard-addition method (SAM) and the Youden plot [2,33–35].

Consider the external standard calibration curve, obtained by plotting the signal or analytical response of different standard solutions of the analyte. Let us assume that the standard calibration relationship is linear within a given concentration range of analyte, so the analytical response follows Equation (4).

Consider now the application of the analytical procedure to a dissolved test portion of a unknown sample within the linear working range. Assuming that the sample matrix does not contribute to the signal as an interfering agent [36] and that there is no interaction between the analyte and the matrix, the analytical response can be now modeled as:

$$\widehat{Y} = A + BZ \tag{15}$$

where A and B are sample constants. A is a constant that does not change when the concentration of the analyte and/or the sample change [37]. It is called the ''true sample blank'' [38] and can be evaluated from the Youden's sample plot [39–41]. B is the fundamental term that justifies the analytical procedure, and it is directly related to the analytical sensitivity [42]. If both constant and proportional bias are absent, then $A = a$ and $B = b$. In order to assess the absence of proportional bias, a homogeneous bulk spiked sample from a matrix that contains the analyte is used. The analyte (here, surrogate) has to be spiked at several concentration levels in order to cover the concentration range of the method scope. Here the SAM can be suitably used to estimate the recovery of spiked samples [2,33–35], so, for a spiked sample, Equation (15) may be rewritten as:

$$\begin{aligned}\widehat{Y} &= A + B(C_{native} + C_{spike}) = A + BC_{native} + BC_{spike} \\ &= a_{SAM} + b_{SAM}C_{spike}\end{aligned} \tag{16}$$

where C_{native} is the concentration of the analyte in the unspiked sample, C_{spike} the concentration of the spiked analyte, and a_{SAM} and b_{SAM} are the intercept and the slope of the SAM calibration straight line.

From Equation (16), we get:

$$\begin{aligned} b_{SAM} &= B \\ a_{SAM} &= A + b_{SAM}C_{native}\end{aligned} \tag{17}$$

If we try to estimate the analyte concentration of a spiked sample by using the external calibration line (Equation (4)), we obtain an estimation of the total observed analyte concentration:

$$\widehat{C}_{obs} = \frac{\widehat{Y} - a}{b} = \frac{(a_{SAM} - a) + b_{SAM}C_{spike}}{b} \tag{18}$$

For the unspiked sample ($C_{spike} = 0$), an estimation of the native analyte concentration is obtained:

$$\widehat{C}_{native} = \frac{a_{SAM} - a}{b} \tag{19}$$

According to Equations (18) and (19), the spiked concentration of analyte is estimated from the external calibration as:

$$\widehat{C}_{spike} = \widehat{C}_{obs} - \widehat{C}_{native} = \frac{b_{SAM}}{b} C_{spike} \quad (20)$$

The relationship established by Equation (20) is of the utmost importance because it leads to a measure of the overall consensus recovery:

$$R = \frac{\widehat{C}_{spike}}{C_{spike}} = \frac{b_{SAM}}{b} \quad (21)$$

The absence of proportional bias corresponds to $b_{SAM} = b$, or, in terms of recovery, $R = 1$. This must be checked for statistical significance [43]:

$$t = \frac{|R-1|}{u(R)} \quad (22)$$

with the uncertainty given by:

$$u(R) = \sqrt{\frac{u^2(b_{SAM})}{b^2} + \frac{b_{SAM}^2 u^2(b)}{b^4}} \quad (23)$$

According to the LGC/VAM protocol [44], if the degrees of freedom associated with the uncertainty of consensus recovery are known, t is compared with the two-tailed tabulated value, $t_{tab}(v,P)$ for the appropriate number of degrees of freedom at P% confidence. If $t \leqslant t_{tab}$, the consensus recovery is not significantly different from 1. Alternatively, instead of t_{tab}, coverage factor k may be used for the comparison. Typical values are $k = 2$ or $k = 3$ for 95% or 99% confidence, respectively [45], so

- if $\frac{|R-1|}{u(R)} \leqslant k$, the recovery is not significantly different from 1; and,
- if $\frac{|R-1|}{u(R)} > k$, the recovery is significantly different from 1 and the analytical result must be corrected by R.

Recovery is sometimes considered a separate validation parameter, but, in any case, it should always be established as a part of method validation [13]. Apart from the statistical significance given above, there are published acceptable recovery percentages as a function of the analyte concentration [28], as shown in Table 1.

In any case, the relative uncertainty for proportional bias due to matrix effects, according to the SAM, is taken as $\frac{u(R)}{R}$.

As mentioned above, in the presence of sample matrix, the relationship between the analytical response and the analyte concentration is given by Equation (15). The independent term ''A'' is the true sample blank because it is determined when both the native analyte and the matrix are present. The SAM calibration, indicated by Equation (22), includes this term within the intercept ($a_{SAM} = A + b_{SAM}C_{native}$). The Youden's plot [39–41] consists of plotting the instrumental response (Y) against the amount of sample (the weight or volume of sample test portion to be dissolved up to the assay volume):

Table 1. Acceptable recovery percentages depending on the analyte level

Analyte (%)	Analyte fraction	Unit	Recovery range (%)
100	1	100%	98–102
10	10^{-1}	10%	98–102
1	10^{-2}	1%	97–103
0.1	10^{-3}	0.1%	95–105
0.01	10^{-4}	100 ppm	90–107
0.001	10^{-5}	10 ppm	80–110
0.0001	10^{-6}	1 ppm	80–110
0.00001	10^{-7}	100 ppb	80–110
0.000001	10^{-8}	10 ppb	60–115
0.0000001	10^{-9}	1 ppb	40–120

$$Y = A + b_{Youden} w_{sample} \quad (24)$$

The intercept of the plot is an estimation of the total Youden blank (TYB), which is the sum of the system blank (SB) corresponding to the intercept of the standard calibration (a) and the Youden blank (YB) associated with the constant bias in the method [34,46]. Thus, we can equate $TYB = A$, $SB = a$ and $YB = A - a$. The constant bias in the method is defined as [34]:

$$\theta_c = \frac{YB}{b} = \frac{A-a}{b} \quad (25)$$

The uncertainty of the constant bias can be obtained by the law of variance propagation:

$$u(\theta_c) = \sqrt{\frac{u^2(A)}{b^2} + \frac{u^2(a)}{b^2} + \frac{(A-a)^2 u^2(b)}{b^4} + \frac{2(A-a)}{b^3} cov(a,b)} \quad (26)$$

The uncertainties, $u^2(A)$, $u^2(a)$ and $u^2(b)$, are obtained from the statistical parameters of the straight line fits: $s^2(A)$ from the Youden's plot; and, $s^2(a)$ and $s^2(b)$ from the external calibration plot. Also $cov(a,b)$ is computed from the external calibration straight line.

Once uncertainty $u(\theta_c)$ is evaluated, the constant bias in the method is tested for significance in a way very similar to recovery:

- if $\frac{|\theta_c|}{u(\theta_c)} \leqslant k$, the constant bias is not significantly different from 0; and,
- if $\frac{|\theta_c|}{u(\theta_c)} > k$, the constant bias is significantly different from 0 and the analytical result should be corrected by θ_c

As Maroto et al. [47] pointed out, even if the analytical procedure is free from constant bias, its uncertainty must be included in the overall uncertainty budget for future determinations. The same applies to the absence of proportional bias. Thus, consider Z_{found}, the analyte concentration obtained by applying the analytical procedure to a sample by using the external calibration function. If, in the matrix-effect study, there are both

proportional bias (recovery R significantly different from 1) and constant bias (offset θ_c significantly different from 0), then the corrected estimated concentration value for the analyte will be:

$$Z = \frac{Z_{found} - \theta_c}{R} \quad (27)$$

Another way to obtain the corrected concentration directly is to use the corrected calibration equation for the measured response Y:

$$Z = \frac{Y - A}{b_{SAM}} \quad (28)$$

The values of A and b_{SAM} are previously established for every kind of matrix subjected to validation and established in the method scope and applicability.

Equation (28) is the inverse prediction equation to be used for estimating the actual concentrations of VSs. Note that once it has been demonstrated that both constant and proportional bias are absent for the matrices considered for validation, CSs could be taken as VSs without problem.

2.4. Accuracy study

As stated above, method validation scrutinizes the accuracy of results by considering both systematic and random errors. Accuracy is therefore studied as an entity with two components – trueness and precision – but considered as a global entity, the uncertainty [48], from which the βETI will be estimated as well as the accuracy profiles once the acceptability limits have been established.

We are interested in estimating the accuracy profile from the uncertainty measurement of the analytical assay from validation data according the LGC/VAM protocol [44] and the ISO/DTS 21748 guide [49]. The basic model for the uncertainty of measurand Z is given by three terms for intra-laboratory measurements:

$$u^2(Z) = S_R^2 + u^2(\delta) + u_{rob}^2(Z) \quad (29)$$

where S_R is the intra-laboratory-reproducibility standard deviation (intermediate precision), $u(\delta)$ is the uncertainty associated with the bias or trueness of the procedure, and u_{rob} is the uncertainty coming from a robustness exercise.

As both precision and trueness are assessed within a single laboratory, the uncertainty due to laboratory transfer should be taken into account. This can be obtained from the robustness study, which considers changes in the variables of the analytical procedure (called factors) expected in a transfer between laboratories. According to the International Conference on Harmonization (ICH Q2A document) [50], the robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate, variations in method parameters, and provides an indication of its reliability during normal usage. Robustness tests can be considered to be intra-laboratory simulations of inter-laboratory studies, if the alterations introduced are suitably selected. Usually, robustness tests that yield significant effects for the measurand lead to further optimization of the method, so uncertainty evaluation should be performed only after the method has been shown to be robust.

In the case of inter-laboratory validation, the expression for uncertainty is:

$$u^2(Z) = S_R^2 + u^2(\delta) \quad (30)$$

Now S_R is the inter-laboratory reproducibility, and hence it is not needed any robustness study.

Because inter-laboratory measurements can be done only when inter-laboratory exercises are available, we will focus on the intra-laboratory estimation of uncertainty.

The estimation of bias and reproducibility is performed using suitable VSs, prepared in the same matrix as that expected for future samples. Certified or internal reference materials represent the best way to obtain a VS, but spiked samples can be considered as a suitable alternative [8]. In the case of pharmaceutical formulations (or other manufactured products) where a ''placebo'' is available, the bias or precision study should be carried out using spiked placebos. But, when a placebo is not available, selected stable samples fortified to a suitable level of the analyte may be prepared. VSs must be stable, homogeneous and as similar as possible to the future samples to be analyzed, and they represent, in the validation phase, the future samples that the analytical procedure will have to quantify. Each VS must be prepared and treated independently as a future sample. This independence is essential for a good estimation of the between-series variance. Indeed, the analytical procedure is not developed to quantify routinely with the same operator and on the same equipment a single sample unknown on one day but a very large number of samples through time, thus often implying several operators and several equipments.

2.4.1. Intermediate precision and trueness studies. Both intermediate precision and trueness studies can be performed using the prediction of actual concentrations from the VSs selected for the analytical assay. Following the golden rules of validation, the analytical procedure should be validated separately for each kind of matrix considered, covering the full range of analyte concentrations.

Accordingly, a suitable way to perform the intermediate precision study is to consider a single sample matrix and a range of analyte concentrations. It is advisable for there to be at least three concentration levels m (low, medium and high) covering the dynamic working range, with a number of n replicates at each

concentration. The ICH Q2B document recommends three replicates [51] and the FDA document on bioanalytical validation considers five replications [52], so 3–5 replications are advisable. Calculations of intermediate precision must be carried out on results instead of responses.

Considering the different conditions p (here, the days) chosen as the main source of variation, an analysis of variance (ANOVA) is then performed for each concentration. Accordingly, for each concentration level, m, we consider the results of the analysis, according to inverse-prediction Equation (28), z_{ij} with two indices: i (from 1 to p) corresponding to the different days and j (from 1 to n) accounting for the repetitions. From the ANOVA, we can easily obtain [53,54] estimations of within-condition variance (S_W^2) and between-condition variance (S_B^2). The within-condition, also known as repeatability, variance (S_r^2) is given by:

$$S_W^2 = S_r^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n}(z_{ij} - \bar{z}_i)^2}{p(n-1)} \quad (31)$$

with

$$\bar{z}_i = \frac{\sum_{j=1}^{n} z_{ij}}{n} \quad (32)$$

The between-condition variance is estimated according to:

$$S_B^2 = \frac{\sum_{i=1}^{p}(\bar{z}_i - \bar{\bar{z}})^2}{p-1} - \frac{S_r^2}{n} \quad (33)$$

with

$$\bar{\bar{z}} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n} z_{ij}}{pn} \quad (34)$$

The intra laboratory reproducibility or intermediate precision can be taken as:

$$S_R^2 = S_r^2 + S_B^2 \quad (35)$$

From these data, the corresponding relative standard deviations, RSD_r and RSD_R, are calculated. These values can be compared with the expected values issued from the Horwitz equation and the ''Horrat'' [55,56]. Horwitz [57] devised an expression to predict the expected value of the relative standard deviation for inter-laboratory reproducibility ($PRSD_R$) according to:

$$PRSD_R = 2^{(1-0.5\log C)} \quad (36)$$

where C is the analyte concentration in decimal fraction units. The Horwitz value is now widely used as a benchmark [58] for the performance of analytical methods via a measure called the ''Horrat'', which is defined as the ratio of the actual relative standard deviation, RSD (either for repeatability or reproducibility), calculated from the analytical data to the Horwitz value:

$$Horrat = \frac{RSD}{PRSD_R} \quad (37)$$

Apart from Horwitz's parameters, expected values of RSD according the AOAC Peer Verified Methods program are also considered [59]. These two approaches, as a function of the analyte concentration, are presented in Table 2. Some practical requirements concerned with inter-laboratory studies are [13]:

- RSD_r = 0.5–0.6 times $PRSD_R$
- RSD_R = 0.5–2 times $PRSD_R$

For intra-laboratory validations, a quick rule is to divide the interval by 2 [60], leading to:

- RSD_r = 0.2–0.3 times $PRSD_R$
- RSD_R = 0.2–1 times $PRSD_R$

In the bias calculation, the $\bar{\bar{z}}$ value is taken as the final result for Z corresponding to the VS of estimated ''true'' concentration T, so the average bias for the VS is obtained from the elemental bias $\delta_{ij} = z_{ij} - T$ according to:

$$\delta = \frac{1}{pn}\sum_{i=1}^{p}\sum_{j=1}^{n}\delta_{ij} = \frac{1}{pn}\sum_{i=1}^{p}\sum_{j=1}^{n} z_{ij} - T = \bar{\bar{z}} - T = Z - T \quad (38)$$

The bias uncertainty can be estimated from the same ANOVA design, according to ISO/DTD 21748 guide [49], as:

$$u^2(\delta) = \frac{S_R^2(1 - \gamma + \frac{\gamma}{n})}{p} \quad (39)$$

with

$$\gamma = \frac{S_r^2}{S_R^2} \quad (40)$$

Accordingly, the only term we need to estimate to evaluate $u(Z)$ from Equation (29) is the robustness uncertainty.

Table 2. Acceptable RSD percentages obtained from the Horwitz function and from the AOAC Peer Verified Methods (PVM) program on the analyte level

Analyte (%)	Analyte fraction	Unit	Horwitz %RSD	AOAC PVM %RSD
100	1	100%	2	1.3
10	10^{-1}	10%	2.8	1.8
1	10^{-2}	1%	4	2.7
0.1	10^{-3}	0.1%	5.7	3.7
0.01	10^{-4}	100 ppm	8	5.3
0.001	10^{-5}	10 ppm	11.3	7.3
0.0001	10^{-6}	1 ppm	16	11
0.00001	10^{-7}	100 ppb	22.6	15
0.000001	10^{-8}	10 ppb	32	21
0.0000001	10^{-9}	1 ppb	45.3	30

2.4.2. Robustness study. Robustness [61,62], considered in the sense of internal validation, deals with the effect of experimental variables, called factors, inherent in the analytical procedure (e.g., temperature, mobile-phase composition, detection wavelength, and pH), on the analytical result. A robustness study examines the alteration of these factors, as expected in a transfer between laboratories, so is of the utmost importance in the uncertainty budget. In experiments to study the main effects of factors, screening designs are used. Screening designs are two-level saturated fractional factorial designs centered on the nominal analytical conditions [63]. Plackett and Burmann [64] developed such designs for studying f factors in $N = f + 1$ experiments, where N is any multiple of 4 less than 100 (except for 92) [65]. Plackett-Burmann designs are very useful tools for a robustness study of analytical procedures. However, these designs cannot deal with factor interactions, so they are suitable only when the interactions are negligible or when considering a key set of dominant factors [25].

The strategy for carrying out a robustness study is based on a landmark procedure suggested by Youden [66,67]:

(1) identify the influential factors;
(2) for each factor, define the nominal and the extreme values expected in routine work and encode them as follows: nominal value = 0, high value = +1 and low value = − 1;
(3) arrange the experimental design by using a two-level 2^{7-4} fractional Plackett-Burmann matrix; and,
(4) perform the experiments in random order on a control sample with analyte concentration halfway in the concentration range of the method scope.

According to the definition of robustness, the interval under investigation is very short (−1, +1; e.g., pH 3.8–4.2). Under these conditions, it must be stressed that no quadratic effect is generally observed, so a linear model can be used. It is one of fundamental differences between the robustness study and the optimization study, in which the interval under investigation is wider.

Youden selected a 2^{7-4} Plackett-Burmann design because it enabled the study of up to seven factors in eight experiments. The corresponding matrix design is illustrated in Table 3. The eight runs are split into two groups of four runs on the basis of levels +1 or −1. The effect of every factor x_k is estimated as the difference of the mean result obtained at the level +1 from that obtained at the level −1.

$$D(x_k) = \frac{1}{4}\left(\left(\sum_{i=1}^{N} Z_i\right)_{(x_k=+1)} - \left(\sum_{i=1}^{N} Z_i\right)_{(x_k=-1)}\right) \qquad (41)$$

Once effects $D(x_k)$ have been estimated, to determine whether variations have a significant effect on the result, a significance t-test is used [68]:

$$t(x_k) = \frac{\sqrt{2}|D(x_k)|}{S_R} \qquad (42)$$

The t-value is compared with the 95%-confidence level two-tailed tabulated value with the degrees of freedom coming from the precision study for each concentration: $v = pn - 1$. If $t(x_k) \leqslant t_{tab}$, then the procedure is robust against the factor x_k. In this case, the uncertainty on measurand Z coming from factor x_k, $u(Z(x_k))$ is evaluated from [44]:

$$u(Z(x_k)) = \frac{t_{crit} S_R}{1.96\sqrt{2}} \frac{\delta_{real}(x_k)}{\delta_{test}(x_k)} \qquad (43)$$

Here δ_{real} is the change in the factor level that would be expected when the method is operating routinely, and δ_{test} is the change in the factor level specified in the robustness study. Once the contributions of the influential factors have been estimated, the relative uncertainty of the robustness study is calculated as:

$$RSD_{rob} = \sqrt{\frac{\sum_k u^2(Z(x_k))}{\overline{Z}^2}} \qquad (44)$$

$\overline{Z}$ is the average value of the eight results obtained in the robustness study. By taking this value for any future concentration, Z, the uncertainty of robustness is $u_{rob}(Z) = Z\ RSD_{rob}$. This leads to the final uncertainty budget according to Equation (29). The expanded uncertainty is obtained as the product of the standard uncertainty $u(Z)$ and the coverage factor k selected as the β quantile of the

Table 3. Arrangement of factor levels for a 2^{7-4} Plackett-Burmann design

Runs	Factors x_k (k = 1 to f)							Response
N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Z_i (i = 1 to N)
1	+1	+1	+1	+1	+1	+1	+1	Z_1
2	+1	+1	−1	+1	−1	−1	−1	Z_2
3	+1	−1	+1	−1	+1	−1	−1	Z_3
4	+1	−1	−1	−1	−1	+1	+1	Z_4
5	−1	+1	+1	−1	−1	+1	−1	Z_5
6	−1	+1	−1	−1	+1	−1	+1	Z_6
7	−1	−1	+1	+1	−1	−1	+1	Z_7
8	−1	−1	−1	+1	+1	+1	−1	Z_8

Student t distribution, so the final expression for the confidence interval of the analytical result is

$$Z \pm ku(Z) = Z \pm U(Z) \quad (45)$$

2.4.3. Study of accuracy profiles. Considering that the concentrations of VSs are taken as reference values T, after analyzing them, the interval for the bias of the result is given, according to Equation (38), as:

$$(Z \pm U(Z)) - T = (Z - T) \pm U(Z) = \delta \pm U(Z) \quad (46)$$

Remembering Equation (39) and condition (2), the corresponding βETI is constructed as:

$$|\delta \pm U(Z)| < \lambda \quad (47)$$

When the βETI is evaluated for a number of concentration levels (VSs) covering the whole range of analyte concentrations specified in the method scope, the accuracy profile can be constructed, according to González and Herrador [69]: for each concentration level, the βETI is calculated from Equation (47). Upper and lower tolerance-interval limits are then connected by straight lines to interpolate the behavior of the limits between the levels studied. The quantification limits are located at the intersections between the interpolating lines and the acceptance limits. If we deal with three levels – low (T_l), medium (T_m) and high (T_h) – three βETIs are computed and expressed as percentage values – $100\frac{\delta_l \pm U(Z_l)}{Z_l}$, $100\frac{\delta_m \pm U(Z_m)}{Z_m}$ and $100\frac{\delta_h \pm U(Z_h)}{Z_h}$. The points corresponding to the upper limits have coordinates $(T_l, 100\frac{\delta_l + U(Z_l)}{Z_l})$, $(T_m, 100\frac{\delta_m + U(Z_m)}{Z_m})$ and $(T_h, 100\frac{\delta_h + U(Z_h)}{Z_h})$, and they are connected by linear segments. The same procedure is done for the lower limits $(T_l, 100\frac{\delta_l - U(Z_l)}{Z_l})$, $(T_m, 100\frac{\delta_m - U(Z_m)}{Z_m})$ and $(T_h, 100\frac{\delta_h - U(Z_h)}{Z_h})$. The intersections between these two segmented lines with the

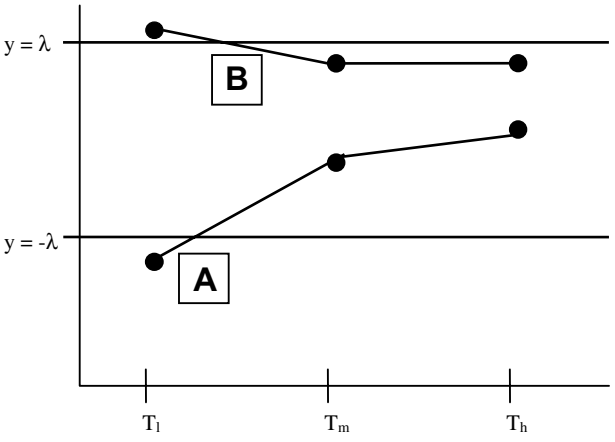

Figure 2. Accuracy profile when intersection occurs between medium and low level. LQL = B (B > A) and UQL = T_h (no intersections in this zone).

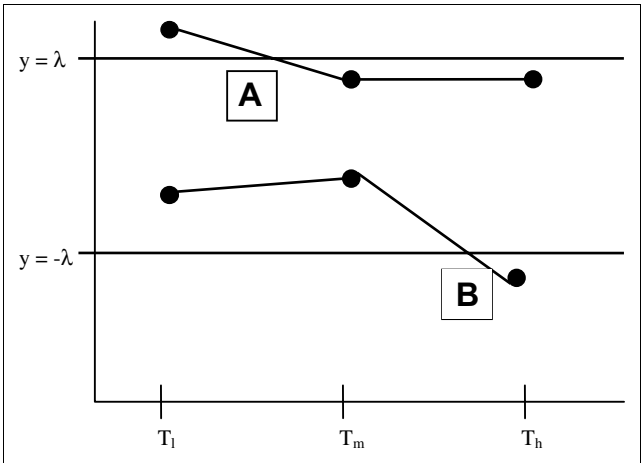

Figure 3. Accuracy profile when two intersections occur, one between medium and low level and another between medium and high level. LQL = A and UQL = B.

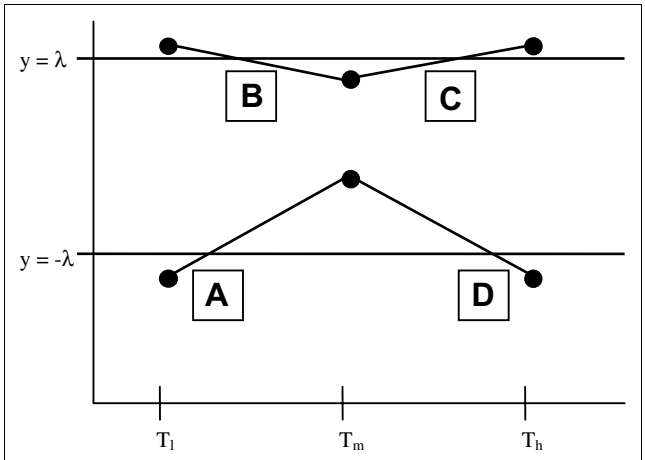

Figure 4. Accuracy profile when several intersections take place. LQL = B (B > A) and UQL = C (C < D).

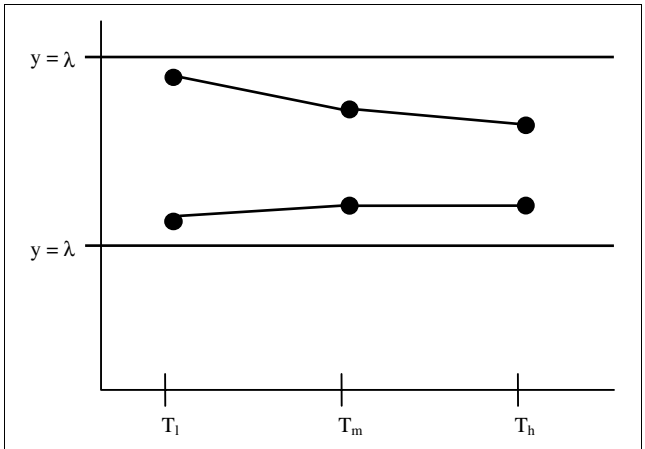

Figure 5. Accuracy profile when no intersections occur in the working range of concentration.

acceptance-limit straight lines $y = \lambda$ and $y = -\lambda$ (in %) give the aforementioned quantification limits. Some typical patterns are shown in Figs. 2–5. The domain of valid concentrations for future assays corresponds to the quantification limits LQL (lower) and UQL (upper) that can be extracted from accuracy profiles.

3. Summary

In this article, we have presented in detail a holistic approach to validate analytical methods including uncertainty measurement and accuracy profiles.

In a first step, we outlined the scope of an analytical method, considering its applicability, fitness for purpose and the given acceptability limits, paying special attention to the concept of acceptability limit that enables us to estimate the βETI interval and the accuracy profiles.

We briefly considered features of specificity and selectivity because modern analytical methods used for routine analysis are selected just because of selectivity issues.

We considered calibration study as a separate stage because of its importance in method validation. At this stage, we also considered goodness of the fit of the calibration function, together with the linearity assessment, linear range, LOD and LOQ. Moreover, checking for matrix effects taking into account possible constant and proportional bias are carried out at this stage in order to correct the calibration function suitably for direct application to routine analysis of samples.

We outlined accuracy study from a contemporary, holistic perspective [70], covering the study of precision, trueness and robustness and the estimation of measurement uncertainty as well as the derivation of βETI intervals and accuracy profiles.

References

[1] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, STP Pharma Pratiques 13 (2003) 101.

[2] A.G. González, M.A. Herrador, A.G. Asuero, Talanta 65 (2005) 1022.

[3] International Organization for Standardization (ISO), ISO 5725-1. Accuracy (trueness and precision) of measurement method and results. Part 1. General principles and definitions, ISO, Geneva, Switzerland, 1994.

[4] Ph. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, Anal. Chim. Acta 391 (1999) 135.

[5] B. Boulanger, P. Chiap, W. Dewé, J. Crommen, Ph. Hubert, J. Pharm. Biomed. Anal. 32 (2003) 753.

[6] B. Boulanger, W. Dewe, P. Hubert, Objectives of pre-study validation and decision rules, AAPS Conf., APQ Open Forum, Indianapolis, USA, 2000.

[7] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (2004) 579.

[8] M. Feinberg, B. Boulanger, W. Dewe, P. Hubert, Anal. Bioanal. Chem. 380 (2004) 502.

[9] M. Feinberg, M. Laurentie, Accred. Qual. Assur. 11 (2006) 3.

[10] M. Thompson, S.L.R. Ellison, R. Wood, Pure Appl. Chem. 47 (2002) 835.

[11] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, The Netherlands, 1997.

[12] Eurachem, Guide: The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics, 1998 (http://www.eurachem.ul.pt/guides/valid.pdf).

[13] I. Taverniers, M. De Loose, E. Van Bockstaele, Trends Anal. Chem. 23 (2004) 535.

[14] R. Kellner, J.-M. Mermet, M. Otto, M. Valcárcel, H.M. Widmer (Editors), Analytical Chemistry: A Modern Approach to Analytical Science, 2nd Ed., Wiley-VCH, Weinheim, Germany, 2004, p. 34.

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

[16] M. Martin-Smith, D.R. Rudd, Acta Pharm. Jugoslav. 40 (1990) 7.

[17] J.A. Murillo, J.M. Lemus, L.F. García, Anal. Lett. 24 (1991) 683.

[18] R.A. Couch, C.L. Olson, Chapter 6, Electroanalytical methods of pharmaceutical analysis, in: J.W. Munson (Editor), Pharmaceutical Analysis. Modern Methods, Part B, Marcel Dekker, New York, USA, 1984, pp. 322–329.

[19] A.G. Asuero, A. Sayago, A.G. González, CRC Anal. Chem. 36 (2006) 41.

[20] Analytical Methods Committee, Analyst (Cambridge, U.K.) 113 (1988) 1469.

[21] J.R.J. Belloto, T.D. Sokolovski, Am. J. Pharm. Educ. 49 (1985) 295.

[22] M. Meloun, J. Militky, M. ForinaChemometrics for Analytical Chemistry, Vol. 2, Ellis Horwood, Chichester, West Sussex, UK, 1994 pp. 64–69.

[23] A.G. Asuero, A.G. González, Microchem. J. 40 (1989) 216.

[24] G.W. Snedecor, W.G. Cochran, Statistical Methods, 8th Ed., Iowa State University Press, USA, 1989.

[25] A.G. González, M.A. Herrador, A. Sayago, A.G. Asuero, Accred. Qual. Assur. 11 (2006) 256.

[26] L. Cuadros Rodríguez, A.M. García Campaña, J.M. Bosque Sendra, Anal. Lett. 29 (1996) 1231.

[27] J.C. Miller, J.N. Miller, Statistics for Analytical Chemistry, 3rd Ed., Ellis Horwood, Chichester, West Sussex, UK, 1993 p. 222.

[28] L. Huber, Validation and Qualification in Analytical Laboratories, Interpharm Press, East Englewood, CO, USA, 1998.

[29] IUPAC, Compendium of Analytical Nomenclature, Definitive Rules 1987, Blackwell Scientific Publications, Oxford, UK, 1997.

[30] L.G. Long, J.D. Winefordner, Anal. Chem. 55 (1983) 712A.

[31] IUPAC, Compendium of Analytical Nomenclature, Definitive Rules 1997, 3rd Edition, Blackwell Science, Oxford, UK, 1998.

[32] Eurachem, Guidance Document No. 1, WELAC Guidance Document No WGD 2: Accreditation for Chemical Laboratories: Guidance on the Interpretation of the EN 45000 Series of Standards and ISO/IEC Guide 25, 1993 (https://citeseer.ist.psu.edu/cache/papers/cs/18190/http:zSzzSzwww.european-accreditation.orgzSzpdfzSzEA-4-05.pdf/eurachem-guidance-document-no.pdf).

[33] A. Maroto, R. Boqué, J. Riu, F.X. Rius, Anal. Chim. Acta 446 (2001) 133.

[34] L. Cuadros-Rodríguez, A.M. García-Campaña, F. Alés Barrero, C. Jiménez Linares, M. Román Ceba, J. AOAC Int. 78 (1995) 471.

[35] R.C. Castells, M.A. Castillo, Anal. Chim. Acta 423 (2000) 179.

[36] Miscellània Enric-Casassas, Bellaterra, Universitat Autónoma de Barcelona, 1991, pp. 147–150.
[37] A.G. González, M.A. Herrador, Talanta 48 (1999) 729.
[38] M.J. Cardone, Anal. Chem. 58 (1986) 438.
[39] W.J. Youden, Anal. Chem. 19 (1947) 946.
[40] W.J. Youden, Biometrics 3 (1947) 61.
[41] W.J. Youden, Mater. Res. Stand. 1 (1961) 268.
[42] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A.
[43] T.J. Farrant, Statistics for the Analytical Scientist. A Bench Guide, Royal Society of Chemistry, Cambridge, UK, 1997.
[44] V.J. Barwick, L.R. Ellison, VAM Project 3.2.1, Development and Harmonisation of Measurement Uncertainty Principles. Part d: Protocol for Uncertainty Evaluation from Validation Data, Report No: LGC/VAM/1998/088, January 2000.
[45] A. Maroto, R. Boqué, J. Riu, F.X. Rius, Analyst (Cambridge, U.K.) 128 (2003) 373.
[46] R.C. Castells, M.A. Castillo, Anal. Chim. Acta 423 (2000) 179.
[47] A. Maroto, J. Riu, R. Boqué, F.X. Rius, Anal. Chim. Acta 391 (1999) 173.
[48] I. Taverniers, E. Van Bockstaele, M. De Loose, Trends Anal. Chem. 23 (2004) 480.
[49] International Organization for Standardization (ISO), ISO/DTS 21748, Guide to the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation, ISO, Geneva, Switzerland, 2003.
[50] http://www.fda.gov/cder/guidance/ichq2a.pdf.
[51] http://www.fda.gov/cder/guidance/1320fnl.pdf.
[52] http://www.fda.gov/cder/guidance/4252fnl.htm.
[53] International Organization for Standardization (ISO), ISO-5725-2, Accuracy (Trueness and Precision) of Measurement Methods and Results - Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method, ISO, Geneva, Switzerland, 1994.
[54] T. Luping, B. Schouenborg, Methodology of Inter-comparison Tests and Statistical Analysis of Test Results. Nordtest Project No. 1483-99, SP Swedish National Testing and Research Institute, SP report 2000:35, Borås, Sweden, 2000.
[55] M. Thompson, The Amazing Horwitz Function, AMC Technical Brief No. 17 Royal Society of Chemistry, Cambridge, UK, July 2004.
[56] R. Wood, Trends Anal. Chem. 18 (1999) 624.
[57] W. Horwitz, Anal. Chem. 54 (1982) 67A.
[58] Codex Alimentarius Commission, Codex Committee on Methods of Analysis and Sampling, CX/MAS 01/4, Criteria for evaluating acceptable methods of analysis for Codex purposes, Agenda Item 4a, 23rd Session, Budapest, Hungary, 26 February-2 March 2001.
[59] 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).
[60] E. Pritchard, Quality in the Analytical Chemistry Laboratory, ACOL series, Wiley, Chichester, West Sussex, UK, 1995.
[61] M.B. Sanz, L.A. Sarabia, A. Herrero, M.C. Ortiz, Talanta 56 (2002) 1039.
[62] L.C. Rodríguez, R. Blanco, A.M. García, J.M. Bosque, Chemometrics. Intell. Lab. Syst. 41 (1998) 57.
[63] J.M. Bosque, M. Nechar, L. Cuadros, Fresenius' J. Anal. Chem. 365 (1999) 480.
[64] R.L. Plackett, J.P. Burmann, Biometrica 33 (1946) 305.
[65] D.C. Montgomery, Design and Analysis of Experiments, 3rd ed., Wiley, New York, USA, 1991.
[66] W.Y. Youden, Statistical Techniques for Collaborative Tests, Association of Official Analytical Chemists (AOAC), Washington D.C., USA, 1967.
[67] I. García, M.C. Ortiz, L. Sarabia, C. Vilches, E. Gredilla, J. Chromatogr., A 992 (2003) 11.
[68] Y. Vander Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, Anal. Chim. Acta 312 (1995) 245.
[69] A.G. González, M.A. Herrador, Talanta 70 (2006) 896.
[70] A. Menditto, M. Patriarca, B. Magnusson, Accred. Qual. Assur. 12 (2007) 45.

A. Gustavo González received his Ph.D. in Chemistry from the University of Seville, Spain, in 1986. His research has focused on several lines: solvent effects on dissociation and drug solubility; chemometric pattern recognition covering methodological developments (including software) and its application to the authentication of foods and beverages; experimental design and optimization of the analytical process; and, quality assurance, validation of analytical methods and estimation of the measurement uncertainty. He has published over 120 scientific papers on these subjects.

M. Ángeles Herrador obtained her Ph.D. in Pharmacy from the University of Seville in 1985. Her research interests comprise the development of methodologies for pharmaceutical and biomedical analysis, and quality assurance for drug assays. In 2002, she obtained the title of Pharmacist Specialist in Analysis and Control of Drugs from the Ministerio de Educación, Cultura y Deporte in Spain. She has published about 70 scientific papers on these topics.