Available online at www.sciencedirect.com
ScienceDirect
Journal of Pharmaceutical and Biomedical Analysis 45 (2007) 82–96
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Harmonization of strategies for the validation of quantitative analytical procedures
A SFSTP proposal–Part III

Ph. Hubert [a,*], J.-J. Nguyen-Huu [b], B. Boulanger [c], E. Chapuzet [d], N. Cohen [e], P.-A. Compagnon [f], W. Dewé [g], M. Feinberg [h], M. Laurentie [i], N. Mercier [d], G. Muzard [j], L. Valat [k], E. Rozet [a]

[a] *Laboratory of Analytical Chemistry, Bioanalytical Chemistry Research Unit, Institute of Pharmacy, University of Liège, CHU, B36, B-4000 Liège, Belgium*
[b] *Sanofi-Aventis, quai Jules Guesde, B.P. 14, F-94403 Vitry sur Seine, France*
[c] *UCB Pharma SA, Chemin du Foriest, B-1420 Braine-L'alleud, Belgium*
[d] *Qualilab, rue de la Bergeresse, F-45160 Olivet (Orléans), France*
[e] *Expanscience, rue des quatre Filles, B.P. 25034, F-28231 Epernon, France*
[f] *French Agency for Health Products Safety (AFSSAPS), boulevard Anatole France, Les Portes de Pleyel, F-93285 St Denis, France*
[g] *GlaxoSmithKline Biologicals, 89 Rue de l'Institut, B-1330 Rixensart, Belgium*
[h] *Met@risk, National Institute for Agricultural Research (INRA), rue Claude Bernard, F-75231 Paris, France*
[i] *LERMDV, French Food Safety Agency (AFSSA), B.P. 90203, F-35032 Fougères, France*
[j] *Merck-Theramex, avenue Prince Héréditaire Albert, F-98007 Monaco, France*
[k] *Viatris-Manufacturing, avenue J.F. Kennedy, B.P. 100, F-33701 Mérignac, France*

Received 15 June 2007; accepted 23 June 2007
Available online 17 July 2007

Abstract

In the first two documents [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, J. Pharm. Biomed. Anal. 36 (2004) 579–586; 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, E. Rozet, J. Pharm. Biomed. Anal., in press], a recent SFSTP Commission on the validation of analytical procedure has introduced a harmonized approach for the validation of analytical procedures. In order to complete this guide, the statistical methodology allowing to correctly conclude about the validity of a procedure is proposed in this third part of the guide. Indeed all the steps to obtain the decision tool namely the accuracy profile are described and illustrated step by step by a numerical example. This tool, based on the concept of total error (bias + standard deviation) build with a β-expectation tolerance interval, allows to easily take the right decision and simultaneously minimizing the risk of the future use of the analytical procedure.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Analytical validation; Statistics; Total error; Risk; Accuracy profile

1. Introduction

All the statistical calculations necessary to implement the concepts presented in the parts 1 and 2 of the guide are developed in this third part [1,2]. These concepts are illustrated with an example, with the aim to make them easier to understand and put them into practice. To facilitate the understanding of the changes as well as the interest of this new approach, the authors decided to address in extenso the same set of data (see Tables 1 and 2) as the one selected in the previous guide issued in 1997 about bioanalytical methods validation [3–5]. The main regulatory guidelines followed by this guide are as follows:

- the ISO 5725 documents [6],
- the ISO 17025 documents [7],
- ICH documents [8],

* Corresponding author. Tel.: +32 4 366 43 16; fax: +32 4 366 43 17.
E-mail address: ph.Hubert@ulg.ac.be (Ph. Hubert).

0731-7085/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2007.06.032

Table 1
Calibration standards

Concentration (ng/ml)	Signal (ratio)		
	Serie 1	Serie 2	Serie 3
25.3533	0.0485	0.0358	0.0449
25.3533	0.0448	0.0402	0.0415
48.2417	0.0959	0.1025	0.0987
48.2417	0.087	0.0993	0.0892
96.4833	0.1974	0.2046	0.2036
96.4833	0.2057	0.1996	0.2082
223.8496	0.5589	0.5371	0.5095
223.8496	0.5667	0.5066	0.5756
437.8235	1.1041	0.9963	1.1726
437.8235	1.0961	1.0568	1.1772
964.8233	2.396	2.2877	2.4528
964.8233	2.3861	2.25	2.3147

- FDA documents [9,10],
- the articles referring to the Washington Conference [11,12].

It is important to remind that this approach aims at re-specifying the very objectives of an analytical method and its validation [13–16], re-focusing some validation criteria and proposing harmonized protocols by distinguishing, in particular, diagnosis rules and decision rules. These decision rules are based on the use of the accuracy profile, integrating in a statistically correct way in a single graph (or table), all the elements essential for the validation, i.e. the bias, the precision, the risk and the quantitation limits. This approach, not only simplifies the validation process of an analytical procedure, but, also allows to monitor risk related to its utilization. This method will facilitate the understanding by everyone involved in the process. These documents, as a whole, represent the consensus achieved and define what could be reasonable expectation from an analytical procedure validation by every member of the group. All along these works, the common focus has been to rationalize the decision making process, to improve the validity and the documentation of the choices made and thus, ultimately, the quality.

To make this process applicable at laboratory level, the practical aspect of the proposed experimental approach has also been taken into account. In this way, the protocols lead to a sufficient but realistic number of experiments. Thus, the gain in quality is obtained without detriment to the validation global cost. In a later part of this guide, this harmonization approach will also be illustrated with a number of examples covering different fields of application.

Table 2
Validation standards

Concentration (ng/ml)	Signal (ratio)		
	Serie 1	Serie 2	Serie 3
25.3533	0.0439	0.0371	0.0444
25.3533	0.0488	0.0422	0.0457
25.3533	0.048	0.0461	0.0502
25.3533	0.0484	0.0448	0.0475
48.2417	0.0949	0.0922	0.0956
48.2417	0.0927	0.0916	0.1023
48.2417	0.0887	0.0854	0.1007
48.2417	0.1015	0.0918	0.1092
437.8235	0.9873	0.9718	1.0392
437.8235	1.0136	1.0322	1.1132
437.8235	1.0288	1.0342	1.1419
437.8235	1.0173	1.0319	1.0751
838.6479	2.022	1.9252	2.1272
838.6479	1.9901	2.0284	2.2127
838.6479	2.0937	2.0127	2.2699
838.6479	2.0189	2.0273	2.2546

2. Example

The example addressed in the bioanalytical methods validation guide published in 1997 and 1999 [3–5] corresponds to the V5 protocol of Table 3 here after and also presented in the part 2 of the guide [2]. The validation phase experiments must be carried out on several series (not necessarily consecutive) and in conditions as close as possible to those that will be met during the routine analysis (apparatus, operators, etc.). Validation also aims at evaluating the intermediate precision, i.e. the precision in the same laboratory, under different conditions (days, different solvents, different apparatus, different operators). To that end, the use of an experimental design integrating these main sources of variation from a series to the other is recommended. Finally, reproducibility has not been considered in this document since it requires several inter-laboratories trials.

3. Statistics

3.1. Response function

Once the experiments have been carried out and the data collected, the relationship between the response (signal or response of the apparatus) Y and the quantity (concentration) X should, at first, be determined on the basis of the calibration standards (CSs). This relationship is characterized by a f function that must strictly be monotonic (strictly increasing or decreasing) over the considered determination interval [5,17]:

$$Y = f(X) + \varepsilon \quad (1)$$

where $\varepsilon \sim N(0, \sigma^2)$ is the error associated to the f response function, commonly called residual error.

Thus, the response function must be adjusted, i.e. the parameters of the model must be evaluated in such a way that the residual error is reduced to the minimum.

Two families of functions emerge from this set: the functions called "linear" in their parameters and the non-linear functions. A function is called "linear" if it is a linear combination of its parameters. It must be noted that the quadratic function, even though its graphic representation is not straight, is actually linear in its parameters. If this is not the case, like for logistic functions, the function is then called non-linear in its parameters. The way of fitting these response functions depends on this distinction. Different response functions can be considered during the method validation, as illustrated in Table 4. The choice depends

Table 3
Choice of number of calibration standards and validation standards depending on the selected protocol (see ref. [2])

Standards	Concentration levels	Protocol				
		V1	V2	V3	V4	V5
CSs without matrix	Low		*2*		*2*	
	Mid	*2*	*(2)*[a]	*2*	*(2)*[a]	
	High	*(2)*[b]	*2*	*(2)*[b]	*2*	
CSs within matrix	Low				*2*	**2**
	Mid			*2*	*(2)*[a]	**(2)**[a]
	High			*(2)*[b]	*2*	**2**
	Additional					**(2)**[c]
VSs within matrix	Low	*3*	*3*	*3*	*3*	**3**
	Mid	*3*	*3*	*3*	*3*	**3**
	High	*3*	*3*	*3*	*3*	**3**
Minimum number of series		*3*	*3*	*3*	*3*	**3**
Total number of experiments (minimum)		**33**	**45**	**39**	**63**	**45**

In bold is the protocol corresponding to the example illustrated throughout this paper.
[a] Considering the regression model selected (for example: simple regression line), the possible suppression of the mid-range concentration level depending on the regression model considered to express the response function (for example: model as the simple regression line). In this case, there are 39 experiments for the protocols V2 (without matrix) and V5 (within matrix). There are 51 experiments for the protocol V4.
[b] Selection of a concentration level higher than the target concentration in order to calibrate (for example: 120% of target concentration).
[c] Addition of a concentration level for a more complex response function (for example: four-parameter logistic regression).

on the type of method (physico-chemical, bioanalytical method, immunoassay, etc.).

Very likely, most of the physico-chemical methods will use the straight line (through zero point or not). As far as bioanalytical methods are concerned, the quadratic function could be considered in some cases, essentially because of the width of the dosing range. In the case of an immunoassay, the four- or five-parameter logistic functions should be preferred. Mathematical transformations could also be considered. For instance, the natural logarithm or the square root transformations could be applied to the X concentration and to the Y response. However, it is recommended to apply this kind of transformation only with the linear models of Table 4.

3.2. Response functions fitting

3.2.1. Linear functions

3.2.1.1. Quality of fitting and quality of results. The envisaged approach to fit these calibration models or response functions is the maximum likelihood method. The maximum likelihood method to assess parameters of a response function consists in finding the parameters values maximizing the function representing the likelihood to observe the generated data. Two working hypotheses are, in theory, required: the normality of the response at every concentration level and the homogeneity of the variances of the responses (homoscedasticity) in the selected concentrations interval. In theory only, because what matters really, is the quality of the inverse predictions (or back-calculated values) rather than the quality of fit. Preference must be given to a model giving good results rather than a model presenting good quality of fit, even if some required statistical hypothesis are to be infringed. Nevertheless, in practice, both aspects – quality of results and application of hypotheses – often go together, even if it is not possible to generalize.

Table 4
Examples of response functions

Type	Equation	Parameters	Linear
Straight line through 0	$Y=\beta X$	β	Yes
Straight line	$Y=\alpha+\beta X$	α, β	Yes
Quadratic function	$Y=\alpha+\beta X+\gamma X^2$	α, β, γ	Yes
Four parameters logistic	$y=\alpha+\frac{\delta-\alpha}{1+(x/\gamma)^{\beta}}$	$\alpha, \beta, \gamma, \delta$	No
Five parameters logistic	$y=\alpha+\frac{\delta-\alpha}{[1+(x/y)^{\beta}]^{\psi}}$	$\alpha, \beta, \gamma, \delta, \psi$	No

3.2.1.2. Weighting. The problem of normality of responses may possibly be addressed by the use of mathematical transformations as described above. In this case, only the differences in accuracy of the results, obtained between the different transformations, will allow deciding on the transformation that suits. The same applies for homoscedasticity: accuracy profiles achieved with and without weighting the observations, are to be compared in order to decide whether weighting is potentially useful. This approach consists in weighting every term of the likelihood function by a weight w that, usually, is an increasing function of the corresponding concentration level. The weights usually selected are $w = 1/X$ or $w = 1/X^2$. It must be noted that $w = 1$ is equivalent to no weighting at all. Modeling the variances as a function of the concentration level allows to find an efficient relationship of the $1/X^{\tau}$ type, where τ becomes a parameter to be assessed.

Practically, a regression line, between the natural logarithm of signals variances and the natural logarithm of the concentration is fitted and this straight line slope estimation (rounded or not),

is used as exponent τ of the inverse of the concentration to form the weighting factor. In this case:

$$w_{ijk} = \frac{1}{x_{ijk}^{\hat{\tau}}} \tag{2}$$

More simply, to avoid this variance modeling, it is possible to consider, a priori, several values for the exponent τ, for example 1 or 2, or stated differently, weight the observations by $1/X$ or $1/X^2$ and then compare the accuracy profiles obtained. This approach is practically still valid because, on the one hand, the results sensitivity to the different exponents τ is low and, on the other hand, it is admitted that, for most of the analytical methods, the variance generally increases according to the concentration (quantity) or, at most, according to the square of the concentration (quantity). This approach will be used to analyze the example presented here.

3.2.1.3. Parameters estimation. According to the considered linear model, the different parameters mentioned in Table 4 are estimated as follows:

For a line through 0, with weighting :

$$\hat{\beta}_i = \frac{\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}} w_{ijk}x_{ijk}y_{ijk}}{\sum_{j=1}^{m}\sum_{k=1}^{n_{ij}} w_{ijk}x_{ijk}^2} \tag{3}$$

with

- $i \in [1, p]$ series index,
- $j \in [1, m]$ concentration index,
- $k \in [1, n]$ repetition index.

It must be noted that, obviously, this formula is interesting and makes sense only if several levels of calibration are used to fit a straight line through 0. If, as it is frequent, only one level is used, then, weighting is of no interest. For the following models, only the solutions for the case where the experimental design is balanced are presented. In the case of an unbalanced design, the use of specialized statistical software is recommended:

For a regression line:

$$\hat{\beta}_i = \frac{g(x_{ijk}, y_{ijk})}{g(x_{ijk}, x_{ijk})}, \qquad \hat{\alpha}_i = \bar{y}_{i\cdot\cdot,w} - \hat{\beta}_i\bar{x}_{i\cdot\cdot,w} \tag{4}$$

For a quadratic model :

$$\hat{\gamma}_i = \frac{g(x_{ijk}, y_{ijk})g(x_{ijk}, x_{ijk}^2) - g(x_{ijk}^2, y_{ijk})g(x_{ijk}, x_{ijk})}{(g(x_{ijk}, x_{ijk}^2))^2 - g(x_{ijk}, x_{ijk})g(x_{ijk}^2, x_{ijk}^2)},$$

$$\hat{\beta}_i = \frac{g(x_{ijk}^2, y_{ijk}) - \hat{\gamma}_i g(x_{ijk}^2, x_{ijk}^2)}{g(x_{ijk}, x_{ijk}^2)},$$

$$\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \hat{\beta}_i\bar{x}_{i\cdot\cdot} - \hat{\gamma}_i\overline{x_{i\cdot\cdot}^2} \tag{5}$$

where

$$g(a_{ijk}, b_{ijk}) = \frac{1}{\sum_{j=1}^{m}\sum_{k=1}^{n} w_{ijk}} \left(\sum_{j=1}^{m}\sum_{k=1}^{n} w_{ijk}a_{ijk}b_{ijk} - \frac{1}{\sum_{j=1}^{m}\sum_{k=1}^{n} w_{ijk}} \left(\sum_{j=1}^{m}\sum_{k=1}^{n} w_{ijk}a_{ijk} \right) \times \left(\sum_{j=1}^{m}\sum_{k=1}^{n} w_{ijk}b_{ijk} \right) \right) \tag{6}$$

in which it is only necessary to replace a_{ijk} and b_{ijk} with the argument mentioned in the g functions, according to the model. The estimated regression parameters obtained with the example using the main response functions can be found in Table 5.

3.2.1.4. Residual error and determination coefficient. The residual variances at the different concentration levels are estimated, after alignment of the observations (see Section 3.2.3), as follows:

$$\hat{\sigma}_j^2 = \frac{1}{\sum_{k=1}^{p} n_{ij} - p}\sum_{i=1}^{p}\sum_{k=1}^{n_{ij}}(y_{ijk,c} - \bar{y}_{\cdot j\cdot,c})^2 \tag{7}$$

where $y_{ijk,c}$ is the aligned observation (response). If alignment is not required, $y_{ijk,c}$ is replaced by the observation y_{ijk}.

$\bar{y}_{j,c}$ is the average of the aligned observation of the j level. Similarly, $\bar{y}_{j,c}$ can be replaced by $\bar{y}_j$ if no alignment was required.

The degrees of freedom (d.f.) of residual error are obtained, by series, by withdrawing from the number of observations used in the regression, the number of parameters estimated in the model, i.e. 1, 2 or 3, in the case of the linear models presented in Table 4.

In order to satisfy most of the regulatory requirements (ICH in particular) it is also necessary to compute and report the determination coefficient for each i series. The determination coefficient r_i^2 expresses the part of responses total variance explained by the model. This coefficient is often incorrectly interpreted as an evaluation of the quality of fit of a model. This is wrong and to impose constraints such as $r^2 > 0.99$ is not a guarantee of quality for the results to be achieved [17]. It will be possible to refer to the following documents [3–5] to calculate the determination coefficient.

3.2.2. Functions non-linear in their parameters

For functions non-linear in their parameters, the fitting is more difficult because maximizing the likelihood function does not give always analytical solutions. In this case, initial values for the different parameters are required as well as using iterative methods to obtain estimates. However, there is no guarantee of finding a solution. To fit non-linear models, it is recommended to use reference books and specialized software as well [17].

3.2.3. Alignment of observations

If, for a given concentration level, the introduced quantities are not the same from one series to the other (often for weighing

Table 5
Regression parameters obtained for the different response functions

Model	Series	Intercept	Slope	Quad. term.	Weighting factor	r^2	Residual d.f.
Linear 0–max	Serie 1	–	2.478E−03	–	1	N/A	1
	Serie 2	–	2.352E−03	–	1	N/A	1
	Serie 3	–	2.471E−03	–	1	N/A	1
Linear 0–223	Serie 1	–	2.514E−03	–	1	N/A	1
	Serie 2	–	2.331E−03	–	1	N/A	1
	Serie 3	–	2.424E−03	–	1	N/A	1
Linear	Serie 1	−1.932E−02	2.510E−03	–	1	0.9996	10
	Serie 2	−1.758E−02	2.373E−03	–	1	0.9995	10
	Serie 3	−1.386E−02	2.520E−03	–	1	0.9963	10
Weighted linear, $1/X$	Serie 1	−2.305E−02	2.523E−03	–	$1/X$	0.9996	10
	Serie 2	−2.015E−02	2.382E−03	–	$1/X$	0.9995	10
	Serie 3	−2.538E−02	2.558E−03	–	$1/X$	0.9963	10
$\log(X) - \log(Y)$	Serie 1	−2.872	1.099	–	1	0.9983	10
	Serie 2	−2.905	1.104	–	1	0.9953	10
	Serie 3	−2.906	1.114	–	1	0.9975	10
$\text{sqrt}(X) - \text{sqrt}(Y)$	Serie 1	−4.585E−02	5.169E−02	–	1	0.9990	10
	Serie 2	−3.974E−02	5.005E−02	–	1	0.9991	10
	Serie 3	−4.858E−02	5.209E−02	–	1	0.9971	10
Quadratic	Serie 1	−3.389E−02	2.656E−03	−1.480E−07	1	0.9998	9
	Serie 2	−2.206E−02	2.418E−03	−4.554E−08	1	0.9995	9
	Serie 3	−5.271E−02	2.910E−03	−3.946E−07	1	0.9978	9
Weighted quadratic, $1/X$	Serie 1	−2.531E−02	2.570E−03	−6.084E−08	$1/X$	0.9997	9
	Serie 2	−2.171E−02	2.415E−03	−4.193E−08	$1/X$	0.9995	9
	Serie 3	−3.233E−02	2.705E−03	−1.875E−07	$1/X$	0.9974	9

r^2 = determination coefficient; d.f. = degrees of freedom.

reasons that must be independent), it is necessary to carry out an alignment on the mean concentration each time a variance must be calculated (repeatability and intermediate precision estimate). This consists in transforming observed instrumental responses ($y_{ijk} \rightarrow y_{ijk,c}$) in order to align them on this mean concentration. This alignment is carried out by interpolation, by adding to the observed response, the difference between the considered response function value, at the mean concentration, and this function value, at the introduced concentration.

In validation, the alignment applies to the responses obtained with the validation samples, by using the response equations or functions achieved with the calibration standards. Thus, the alignment of the n_{ij} repetitions of the j concentration level of the i series is carried out as follows:

$$y_{ijk,c} = y_{ijk} + f(\bar{x}_{ij}) - f(x_{ijk}) \quad (8)$$

Table 6
Alignment rules for different response functions

Response function	Alignment rule
Straight line through 0	$y_{ijk,c} = y_{ijk} + \hat{\beta}_i[\bar{x}_{ij\cdot} - x_{ijk}]$
Straight line	$y_{ijk,c} = y_{ijk} + \hat{\beta}_i[\bar{x}_{ij\cdot} - x_{ijk}]$
Quadratic function	$y_{ijk,c} = y_{ijk} + \hat{\beta}_i[\bar{x}_{ij\cdot} - x_{ijk}] + \hat{\gamma}_i[\bar{x}^2_{ij\cdot} - x^2_{ijk}]$
Four parameters logistic	$y_{ijk,c} = y_{ijk} + (\hat{\delta}_i - \hat{\alpha}_i)\left(\frac{1}{1+(\hat{\gamma}_i/\bar{x}_{ij\cdot})^{\hat{\beta}_i}} - \frac{1}{1+(\hat{\gamma}_i/x_{ijk})^{\hat{\beta}_i}}\right)$

To summarize, Table 6 give the equations to perform the alignment for the different response functions. It must be noted that, even if it is mentioned in Table 6, the alignment is rarely carried out for immunoassays and the different responses are considered to come from the same concentration.

3.2.4. Inverse prediction

Before carrying out the inverse predictions, i.e. computing the back-calculated concentrations with the response function, it is preferable to make sure that within a given concentration level, the concentrations are all identical. If it is not the case, then it is recommended to align them as described in the previous section. The inverse predictions for the different regression models are obtained as described in Table 7. If the observations have been aligned, y_{ijk} must be replaced with $y_{ijk,c}$ in Table 7.

Table 7
Computation of the inverse predictions for the different response functions

Response function	Calculated concentration
Straight line through 0	$x_{ijk,\text{calc}} = \frac{y_{ijk}}{\hat{\beta}_i}$
Straight line	$x_{ijk,\text{calc}} = \frac{y_{ijk} - \hat{\alpha}_i}{\hat{\beta}_i}$
Quadratic function	$x_{ijk,\text{calc}} = \frac{-\hat{\beta}_i + \sqrt{\hat{\beta}_i^2 - 4\hat{\gamma}_i(\hat{\alpha}_i - y_{ijk})}}{2\hat{\gamma}_i}$
Four parameters logistic	$x_{ijk,\text{calc}} = \hat{\gamma}_i\left(\frac{\hat{\delta}_i - \hat{\alpha}_i}{y_{ijk} - \hat{\alpha}_i} - 1\right)^{1/\hat{\beta}_i}$
Five parameters logistic	$x_{ijk,\text{calc}} = \hat{\gamma}_i\left(\left(\frac{\hat{\delta}_i - \hat{\alpha}_i}{y_{ijk} - \hat{\alpha}_i}\right)^{1/\psi_i} - 1\right)^{1/\hat{\beta}_i}$

Table 8
Computation of the inverse prediction for transformed data with logarithm or square root

Response function	Calculated concentration (logarithm)	Calculated concentration (square root)
Straight line	$x_{ijk,\text{calc}} = e^{\frac{\ln(y_{ijk})-\hat{a}_i}{\hat{b}_i}}$	$x_{ijk,\text{calc}} = \left(\frac{\sqrt{y_{ijk}}-\hat{a}_i}{\hat{b}_i}\right)^2$

By the same way, if a transformation has been used, inverse transformation after this back-calculation must not be forgotten. For instance, after a logarithm or square root transformation of the straight line, back-calculated concentrations are made as shown in Table 8. The inverse predictions obtained with the different response function used in the example are shown in Table 9.

3.3. Estimation of trueness and precision

3.3.1. Model

The estimate of the trueness and precision of the method is carried out with the back-calculated concentrations coming from the validation standards of the validation phase (or calibration standards themselves in pre-validation phase). This estimation is carried out at each of the considered j concentration levels, using the following statistical models:

$$X_{ijk} = \mu_j + \alpha_{ij} + \varepsilon_{ijk} \tag{9}$$

in which

- X_{ijk} is the kth back-calculated concentration of the i series j level.
- μ_j is the mean of the back-calculated concentrations of the j-concentration level.
- α_{ij} is, at j level, the difference between the ith series average and the μ_j; α_{ij} is considered as a normal random variable with 0 as average and $\sigma^2_{\text{B},j}$ as variance.
- ε_{ijk} is the experimental error considered as a normal random variable with an average of 0 and a variance of $\sigma^2_{\text{W},j}$.

The experimental error is supposed to be independent of the series. The $\sigma^2_{\text{B},j}$ and $\sigma^2_{\text{W},j}$ variances respectively represent the inter-series and intra-series variances. The restricted maximum likelihood method is used to estimate, at every j concentration level the parameters of the model μ_j, $\sigma^2_{\text{B},j}$ and $\sigma^2_{\text{W},j}$:

$$\hat{\mu}_j = \frac{1}{\sum_{i=1}^{p} n_{ij}} \sum_{i=1}^{p} \sum_{k=1}^{n_{ij}} x_{ijk,\text{calc}} \tag{10}$$

$$\text{MSM}_j = \frac{1}{p-1} \sum_{i=1}^{p} n_{ij} (\bar{x}_{ij\cdot,\text{calc}} - \bar{x}_{\cdot j\cdot,\text{calc}})^2 \tag{11}$$

$$\text{MSE}_j = \frac{1}{\sum_{i=1}^{p} n_{ij} - p} \sum_{i=1}^{p} \sum_{k=1}^{n_{ij}} (x_{ijk,\text{calc}} - \bar{x}_{ij\cdot,\text{calc}})^2 \tag{12}$$

In the case of a balanced design (the repetition number is identical in every series for each concentration level), the variance components are estimated as follows for each concentration level (n being the repetition number in each series):

If $\text{MSE}_j < \text{MSM}_j$, then:

$$\hat{\sigma}^2_{\text{W},j} = \text{MSE}_j \tag{13}$$

$$\hat{\sigma}^2_{\text{B},j} = \frac{\text{MSM}_j - \text{MSE}_j}{n} \tag{14}$$

otherwise

$$\hat{\sigma}^2_{\text{W},j} = \frac{1}{pn-1} \sum_{i=1}^{p} \sum_{j=1}^{k} (x_{ijk,\text{calc}} - \bar{x}_{\cdot j\cdot,\text{calc}})^2 \tag{15}$$

$$\hat{\sigma}^2_{\text{B},j} = 0 \tag{16}$$

The intra-series variance estimate provides the repeatability variance estimate while the intra- and inter-series variances estimates sum provides an estimation of the intermediate precision variance.

$$\text{repeatability} : \hat{\sigma}^2_{\text{Re},j} = \hat{\sigma}^2_{\text{W},j}$$

$$\text{intermediate precision} : \quad \hat{\sigma}^2_{\text{IP},j} = \hat{\sigma}^2_{\text{W},j} + \hat{\sigma}^2_{\text{B},j} \tag{17}$$

Table 10 gives the results of precision obtained from the example with the different response functions tested.

3.3.2. Maximum likelihood and least squares

For an unbalanced design, there is no analytical solution to obtain parameters estimators. An iterative method has to be used. This case is relatively rare as a validation is often planned in order to obtain the same repetition number in every series. However, some data may be missing and, as a result, the experimental design could become unbalanced. In this case, using statistical software is recommended and preferred to the classical, called least squares methods, given that, as it is known, the least square estimators are biased in the case of unbalanced designs. The maximum likelihood estimators, that are the standards and only the correct ones, are equal to the least squares estimators only in the case of a balanced experimental design. This condition has been ignored in the past, for practical calculation reasons and, too often, it has been implicitly understood that these two types of estimators were equivalent under any condition. This is not the case and the present availability of computers and appropriate software does not justify any longer this difference to still be ignored. This is the reason why we recommend the use of maximum likelihood estimators to validate analytical methods.

3.3.3. Trueness

The trueness of an analytical procedure (as opposed to accuracy of a result) expresses the closeness of agreement between the average (as opposed to one single observation) of the trials results with the method and the accepted reference value, also called conventional true value [6,17]. The trueness of a procedure (or bias), at each concentration level, is obtained by calculating the difference between the introduced concentrations (arithmetic) mean $\bar{x}_j$ and the calculated concentrations mean $\hat{\mu}_j$. The bias can be expressed in absolute or relative terms

Table 9
Inverse predictions (ng/ml) obtained with the main response functions tested

Sample number	Series	Concentration (ng/ml)	Response (ratio)	Linear 0–max	Linear 0–223	Linear	Weighted linear	log	sqrt	Quadratic	Weighted quadratic
VAL01	Serie 1	25.3533	0.0439	17.71	17.46	25.19	26.54	23.9	24.41	29.33	26.94
VAL02	Serie 1	25.3533	0.0488	19.69	19.41	27.14	28.48	26.32	26.63	31.18	28.85
VAL03	Serie 1	25.3533	0.048	19.37	19.09	26.82	28.17	25.92	26.27	30.88	28.54
VAL04	Serie 1	25.3533	0.0484	19.53	19.25	26.98	28.33	26.12	26.45	31.03	28.7
VAL05	Serie 1	48.2417	0.0949	38.29	37.75	45.5	46.76	48.21	46.88	48.62	46.82
VAL06	Serie 1	48.2417	0.0927	37.41	36.87	44.63	45.89	47.19	45.93	47.78	45.96
VAL07	Serie 1	48.2417	0.0887	35.79	35.28	43.03	44.3	45.33	44.2	46.27	44.4
VAL08	Serie 1	48.2417	0.1015	40.96	40.37	48.13	49.38	51.25	49.71	51.11	49.4
VAL09	Serie 1	437.8235	0.9873	398.4	392.7	401	400.5	406.3	404.4	393	397.7
VAL10	Serie 1	437.8235	1.0136	409	403.2	411.5	411	416.2	414.7	403.4	408.2
VAL11	Serie 1	437.8235	1.0288	415.1	409.2	417.6	417	421.8	420.6	409.4	414.2
VAL12	Serie 1	437.8235	1.0173	410.5	404.6	413	412.4	417.5	416.1	404.9	409.6
VAL13	Serie 1	838.6479	2.022	815.9	804.2	813.3	810.7	780.2	806.3	810.6	812.2
VAL14	Serie 1	838.6479	1.9901	803	791.5	800.5	798.1	769	794	797.4	799.3
VAL15	Serie 1	838.6479	2.0937	844.8	832.8	841.8	839.1	805.3	834	840.3	841.2
VAL16	Serie 1	838.6479	2.0189	814.7	803	812	809.5	779.1	805.1	809.3	810.9
VAL17	Serie 2	25.3533	0.0371	15.78	15.91	23.04	24.04	21.66	21.55	24.48	24.37
VAL18	Serie 2	25.3533	0.0422	17.95	18.1	25.19	26.18	24.34	23.99	26.59	26.48
VAL19	Serie 2	25.3533	0.0461	19.6	19.77	26.83	27.82	26.37	25.84	28.2	28.1
VAL20	Serie 2	25.3533	0.0448	19.05	19.22	26.28	27.27	25.7	25.23	27.66	27.56
VAL21	Serie 2	48.2417	0.0922	39.21	39.55	46.26	47.17	49.41	47.07	47.29	47.21
VAL22	Serie 2	48.2417	0.0916	38.95	39.29	46	46.92	49.12	46.8	47.04	46.96
VAL23	Serie 2	48.2417	0.0854	36.32	36.63	43.39	44.32	46.1	43.99	44.48	44.39
VAL24	Serie 2	48.2417	0.0918	39.04	39.38	46.09	47	49.22	46.89	47.13	47.05
VAL25	Serie 2	437.8235	0.9718	413.3	416.9	416.9	416.5	417.3	419.8	414.2	414.4
VAL26	Serie 2	437.8235	1.0322	438.9	442.8	442.3	441.8	440.7	444.9	439.6	439.8
VAL27	Serie 2	437.8235	1.0342	439.8	443.6	443.2	442.7	441.5	445.7	440.4	440.7
VAL28	Serie 2	437.8235	1.0319	438.8	442.6	442.2	441.7	440.6	444.8	439.5	439.7
VAL29	Serie 2	838.6479	1.9252	818.7	825.8	818.6	816.7	775.1	813.1	817.8	817.9
VAL30	Serie 2	838.6479	2.0284	862.6	870.1	862.1	860.1	812.6	855.5	861.9	861.9
VAL31	Serie 2	838.6479	2.0127	855.9	863.4	855.5	853.5	806.9	849.1	855.2	855.2
VAL32	Serie 2	838.6479	2.0273	862.1	869.6	861.6	859.6	812.2	855	861.4	861.4
VAL33	Serie 3	25.3533	0.0444	17.97	18.32	23.12	27.27	24.78	24.77	33.52	28.42
VAL34	Serie 3	25.3533	0.0457	18.5	18.86	23.64	27.78	25.43	25.36	33.98	28.9
VAL35	Serie 3	25.3533	0.0502	20.32	20.71	25.42	29.54	27.67	27.39	35.54	30.57
VAL36	Serie 3	25.3533	0.0475	19.23	19.6	24.35	28.48	26.33	26.18	34.6	29.57
VAL37	Serie 3	48.2417	0.0956	38.69	39.44	43.44	47.28	49.33	47.17	51.32	47.44
VAL38	Serie 3	48.2417	0.1023	41.41	42.21	46.1	49.9	52.42	50.02	53.66	49.94
VAL39	Serie 3	48.2417	0.1007	40.76	41.55	45.46	49.28	51.68	49.34	53.1	49.34
VAL40	Serie 3	48.2417	0.1092	44.2	45.05	48.83	52.6	55.58	52.94	56.07	52.5
VAL41	Serie 3	437.8235	1.0392	420.6	428.8	417.9	416.1	419.9	420.3	396.5	407.6
VAL42	Serie 3	437.8235	1.1132	450.6	459.3	447.2	445	446.6	448.8	425.2	436.6
VAL43	Serie 3	437.8235	1.1419	462.2	471.1	458.6	456.2	456.9	459.9	436.3	447.9
VAL44	Serie 3	437.8235	1.0751	435.1	443.6	432.1	430.1	432.9	434.2	410.4	421.7
VAL45	Serie 3	838.6479	2.1272	861	877.7	849.6	841.3	798.6	836.9	846.2	848.1
VAL46	Serie 3	838.6479	2.2127	895.6	912.9	883.5	874.8	827.4	869.5	884.6	884
VAL47	Serie 3	838.6479	2.2699	918.7	936.5	906.2	897.1	846.6	891.2	910.6	908.1
VAL48	Serie 3	838.6479	2.2546	912.5	930.2	900.2	891.1	841.4	885.4	903.6	901.7

Table 10
Trueness and precision estimators obtained with the main response functions studied

Model	Introduced concentration (ng/ml)	Mean calculated concentration (ng/ml)	Bias (ng/ml)	Relative bias (%)	Recovery (%)	Repeatability standard deviation (ng/ml)	Between series standard deviation (ng/ml)	Intermediate precision standard deviation (ng/ml)	CV repeatability (%)	CV intermediate precision (%)
Linear 0–max	25.35	18.72	−6.629	−26.15	73.85	1.229	0.00E+00	1.229	4.846	4.846
	48.24	39.25	−8.99	−18.64	81.36	1.978	1.442	2.447	4.1	5.073
	437.8	427.7	−10.13	−2.313	97.69	13.5	16.13	21.03	3.083	4.803
	838.6	855.5	16.82	2.005	102	21.81	37.43	43.32	2.601	5.166
Linear 0–223	25.35	18.81	−6.545	−25.81	74.19	1.24	0.00E+00	1.24	4.892	4.892
	48.24	39.45	−8.794	−18.23	81.77	1.986	2.115	2.902	4.118	6.015
	437.8	429.9	−7.964	−1.819	98.18	13.67	23.85	27.49	3.123	6.279
	838.6	859.8	21.17	2.524	102.5	22	52.12	56.58	2.624	6.746
Linear	25.35	25.33	−2.02E−02	−7.95E−02	99.92	1.242	1.026	1.61	4.897	6.352
	48.24	45.57	−2.669	−5.533	94.47	1.785	0.00E+00	1.785	3.701	3.701
	437.8	428.6	−9.192	−2.099	97.9	13.29	14.04	19.33	3.035	4.414
	838.6	850.4	11.77	1.404	101.4	21.49	32.26	38.76	2.563	4.622
Weighted linear, $1/X$	25.35	27.49	2.138	8.434	108.4	1.234	0.8223	1.483	4.866	5.848
	48.24	47.57	−0.6752	−1.4	98.6	1.93	1.646	2.537	4.001	5.259
	437.8	427.6	−10.24	−2.338	97.66	13.15	13.54	18.87	3.003	4.31
	838.6	846	7.325	0.8735	100.9	21.29	29.01	35.98	2.539	4.291
$\log(X) - \log(Y)$	25.35	25.38	2.49E−02	9.83E−02	100.1	1.545	0.1358	1.551	6.093	6.116
	48.24	49.57	1.327	2.75	102.8	2.257	2.044	3.045	4.679	6.312
	437.8	429.8	−7.985	−1.824	98.18	12.16	11.06	16.44	2.778	3.755
	838.6	804.5	−34.1	−4.066	95.93	18.47	20.71	27.75	2.202	3.309
$\mathrm{sqrt}(X) - \mathrm{sqrt}(Y)$	25.35	25.34	−1.31E−02	−5.15E−02	99.95	1.409	0.7473	1.595	5.556	6.29
	48.24	47.58	−0.6646	−1.378	98.62	2.092	1.701	2.696	4.337	5.589
	437.8	431.2	−6.638	−1.516	98.48	12.98	13.47	18.71	2.964	4.273
	838.6	841.3	2.626	0.3132	100.3	20.76	28.67	35.4	2.476	4.221
Quadratic	25.35	30.58	5.23	20.63	120.6	1.184	3.792	3.973	4.672	15.67
	48.24	49.49	1.247	2.586	102.6	1.802	3.527	3.961	3.736	8.211
	437.8	417.7	−20.09	−4.587	95.41	13.07	13.93	19.1	2.986	4.363
	838.6	849.9	11.25	1.341	101.3	23.17	34.02	41.16	2.763	4.908
Weighted quadratic, $1/X$	25.35	28.08	2.729	10.76	110.8	1.207	1.239	1.73	4.762	6.824
	48.24	47.62	−0.6233	−1.292	98.71	1.874	1.651	2.498	3.885	5.177
	437.8	423.2	−14.65	−3.345	96.65	13.19	12.22	17.98	3.013	4.107
	838.6	850.2	11.51	1.373	101.4	22.26	32.98	39.79	2.654	4.744

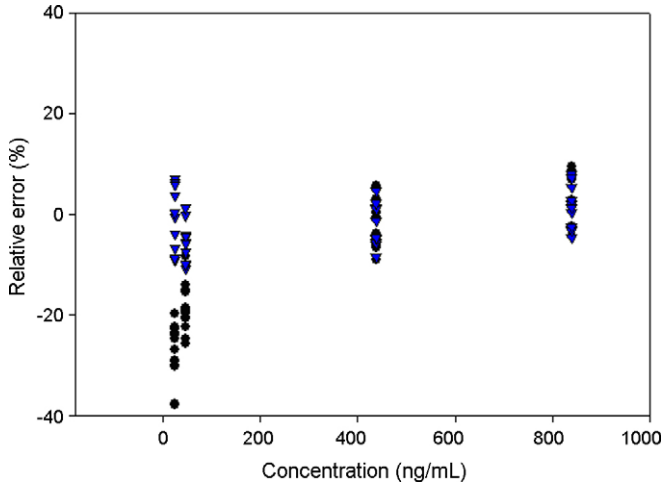

Fig. 1. Representation of individual relative errors as a function of theoretical concentration obtained by the same samples but with two different response functions: the simple linear model (triangles) and the linear through 0 and the highest level 964.8233 ng/ml (circles).

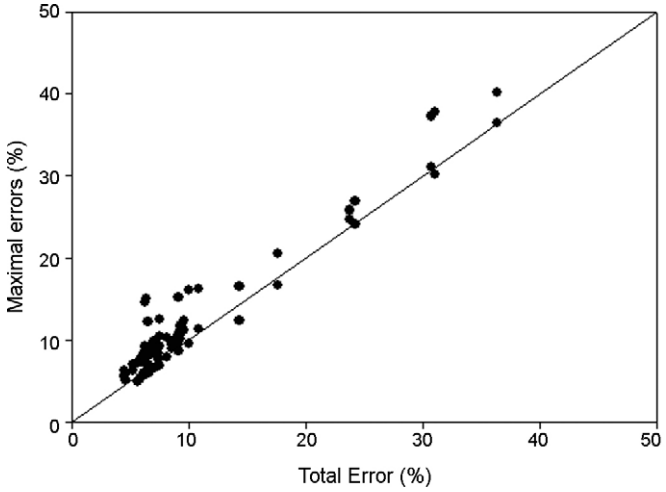

Fig. 2. Graphic of the highest and second highest maximal error in function of the relative total error for each concentration level and for each response function tested.

or in recovery terms, compared to the introduced quantities, as follows:

$$\text{bias}_j = \hat{\mu}_j - \bar{x}_{.j.} \tag{18}$$

$$\text{bias}_j\,(\%) = 100 \times \frac{\hat{\mu}_j - \bar{x}_{.j.}}{\bar{x}_{.j.}} \tag{19}$$

$$\text{recovery}_j\,(\%) = 100 \times \frac{\hat{\mu}_j}{\bar{x}_{.j.}} \tag{20}$$

Table 10 gives the results of trueness obtained from the example with the different response functions tested.

3.4. Estimation of accuracy

The accuracy [6,17] of a result x_i (as opposed to the analytical method) expresses the closeness of agreement between the trial result and the accepted reference value (μ), equally called conventionally true value, and this, for each measurement:

$$\text{accuracy} = x_i - \mu \tag{21}$$

Table 11 represents, for each model and each observation, the measurement accuracy, in relative value, i.e.:

$$\text{accuracy}\,(\%) = \frac{x_i - \mu}{\mu} 100 \tag{22}$$

Note that last line in Table 11 is the maximum relative unsigned error, observed for each model, on all the series. This maximum value already gives an idea of the relative efficiency of the different models. So, using the simple linear model gives a result whose maximal difference with the true value is 10.80%, among all the tested samples, while using the model through 0 and the highest concentration level, gives a maximal difference with the true value up to 37.76%, for the same data. This simple observation already shows the impact of the response function selection on the accuracy of the results (Fig. 1). By means of this simple criterion, it is already possible to see that the use of the simple linear model to calibrate data leads to more accurate results than the other model. This model will be preferred but one should then evaluate whether the results accuracy achieved this way is satisfactory: this will be seen later. The results accuracy like the maximum observed error, provide indications on the results obtained in the validation experiments.

3.4.1. Total error and total error profile

Each measurement obtained reflects the true value μ, the bias of the method and its precision, which is expressed as follows:

$$x_i = \mu + |\text{bias}|_{\text{procedure}} + \text{intermediate precision}_{\text{procedure}}$$

$$\Updownarrow$$

$$x_i - \mu = |\text{bias}|_{\text{procedure}} + \text{intermediate precision}_{\text{procedure}}$$

$$\Updownarrow$$

$$x_i - \mu = \text{total error}_{\text{procedure}}$$

The total error of an analytical procedure evaluates its ability to produce accurate results. Thus, the total error estimation of a procedure is fundamental to assess the validity of a method. This total error, as indicated above, is the sum of trueness (bias) and precision (Table 12). As shown in Fig. 2, the total error observed with each model and for each concentration level is closely linked to the corresponding observed maximum errors. It is normal that the maximum error observed on a large number of observations is noticeably bigger than the total error, given that these maximum errors represent rare occurrences, while the total error rather reflects the biggest errors that can be expected in most cases. Should the second biggest error be considered, then it appears that the points are clearly distributed around the bisecting line $y = x$, which clearly demonstrates that the total error estimates reflect the biggest errors produced by the method. Thus, an analytical procedure total error is clearly a good indicator of its produced results accuracy. It is for those reasons that we propose this criterion for a first simple assessment of an analytical procedure, as shown in Fig. 3.

Table 11
Values of relative accuracy (%) obtained with the principal response functions tested for each validation sample

Series	Concentration (ng/ml)	Linear 0–high	Linear 0–223	Linear	Weighted linear	log(X) − log(Y)	sqrt(X) − sqrt(Y)	Quadratic	Weighted quadratic
Serie 1	25.3533	−30.15	−31.13	−0.64	4.68	−5.73	−3.72	15.69	6.26
Serie 1	25.3533	−22.34	−23.44	7.05	12.33	3.81	5.04	22.98	13.79
Serie 1	25.3533	−23.60	−24.70	5.79	11.11	2.24	3.62	21.80	12.57
Serie 1	25.3533	−22.97	−24.07	6.42	11.74	3.02	4.33	22.39	13.20
Serie 1	48.2417	−20.63	−21.75	−5.68	−3.07	−0.07	−2.82	0.78	−2.95
Serie 1	48.2417	−22.45	−23.57	−7.49	−4.87	−2.18	−4.79	−0.96	−4.73
Serie 1	48.2417	−25.81	−26.87	−10.80	−8.17	−6.04	−8.38	−−4.09	−7.96
Serie 1	48.2417	−15.09	−16.32	−0.23	2.36	6.24	3.04	5.95	2.40
Serie 1	437.8235	−9.00	−10.31	−8.41	−8.52	−7.20	−7.63	−10.24	−9.16
Serie 1	437.8235	−6.58	−7.91	−6.01	−6.13	−4.94	−5.28	−7.86	−6.77
Serie 1	437.8235	−5.19	−6.54	−4.62	−4.76	−3.66	−3.93	−6.49	−5.40
Serie 1	437.8235	−6.24	−7.59	−5.67	−5.81	−4.64	−4.96	−7.52	−6.45
Serie 1	838.6479	−2.71	−4.11	−3.02	−3.33	−6.97	−3.86	−3.34	−3.15
Serie 1	838.6479	−4.25	−5.62	−4.55	−4.83	−8.30	−5.32	−4.92	−4.69
Serie 1	838.6479	0.73	−0.70	0.38	0.05	−3.98	−0.55	0.20	0.30
Serie 1	838.6479	−2.86	−4.25	−3.18	−3.48	−7.10	−4.00	−3.50	−3.31
Serie 2	25.3533	−37.76	−37.25	−9.12	−5.18	−14.57	−15.00	−3.44	−3.88
Serie 2	25.3533	−29.20	−28.61	−0.64	3.26	−4.00	−5.38	4.88	4.44
Serie 2	25.3533	−22.69	−22.02	5.82	9.73	4.01	1.92	11.23	10.83
Serie 2	25.3533	−24.86	−24.19	3.66	7.56	1.37	−0.49	9.10	8.70
Serie 2	48.2417	−18.72	−18.02	−4.11	−2.22	2.42	−2.43	−1.97	−2.14
Serie 2	48.2417	−19.26	−18.56	−4.65	−2.74	1.82	−2.99	−2.49	−2.66
Serie 2	48.2417	−24.71	−24.07	−10.06	−8.13	−4.44	−8.81	−7.80	−7.98
Serie 2	48.2417	−19.07	−18.37	−4.46	−2.57	2.03	−2.80	−2.30	−2.47
Serie 2	437.8235	−5.60	−4.78	−4.78	−4.87	−4.69	−4.12	−5.40	−5.35
Serie 2	437.8235	0.25	1.14	1.02	0.91	0.66	1.62	0.41	0.45
Serie 2	437.8235	0.45	1.32	1.23	1.11	0.84	1.80	0.59	0.66
Serie 2	437.8235	0.22	1.09	1.00	0.89	0.63	1.59	0.38	0.43
Serie 2	838.6479	−2.38	−1.53	−2.39	−2.62	−7.58	−3.05	−2.49	−2.47
Serie 2	838.6479	2.86	3.75	2.80	2.56	−3.11	2.01	2.77	2.77
Serie 2	838.6479	2.06	2.95	2.01	1.77	−3.79	1.25	1.97	1.97
Serie 2	838.6479	2.80	3.69	2.74	2.50	−3.15	1.95	2.71	2.71
Serie 3	25.3533	−29.12	−27.74	−8.81	7.56	−2.26	−2.30	32.21	12.10
Serie 3	25.3533	−27.03	−25.61	−6.76	9.57	0.30	0.03	34.03	13.99
Serie 3	25.3533	−19.85	−18.31	0.26	16.51	9.14	8.03	40.18	20.58
Serie 3	25.3533	−24.15	−22.69	−3.96	12.33	3.85	3.26	36.47	16.63
Serie 3	48.2417	−19.80	−18.25	−9.95	−1.99	2.26	−2.22	6.38	−1.66
Serie 3	48.2417	−14.16	−12.50	−4.44	3.44	8.66	3.69	11.23	3.52
Serie 3	48.2417	−15.51	−13.87	−5.77	2.15	7.13	2.28	10.07	2.28
Serie 3	48.2417	−8.38	−6.62	1.22	9.03	15.21	9.74	16.23	8.83
Serie 3	437.8235	−3.93	−2.06	−4.55	−4.96	−4.09	−4.00	−9.44	−6.90
Serie 3	437.8235	2.92	4.91	2.14	1.64	2.00	2.51	−2.88	−0.28
Serie 3	437.8235	5.57	7.60	4.75	4.20	4.36	5.04	−0.35	2.30
Serie 3	437.8235	−0.62	1.32	−1.31	−1.76	−1.12	−0.83	−6.26	−3.68
Serie 3	838.6479	2.67	4.66	1.31	0.32	−4.78	−0.21	0.90	1.13
Serie 3	838.6479	6.79	8.85	5.35	4.31	−1.34	3.68	5.48	5.41
Serie 3	838.6479	9.55	11.67	8.05	6.97	0.95	6.27	8.58	8.28
Serie 3	838.6479	8.81	10.92	7.34	6.25	0.33	5.57	7.74	7.52
Maximal error		37.76	37.25	10.80	16.51	15.21	15.00	40.18	20.58

As illustrated in Fig. 3, if it is desired, for instance, that the procedure maximum error does not exceed 20% of the true value, then using the linear model as response function provides us with the desired guarantees, given that this model presents a 10% maximum total error. This obviously is not the case with the linear model through 0.

3.4.2. Tolerance interval calculation

However, what matters in validation, is not the validity of the results obtained by means of the calculated total error, but, rather, the guarantee or a representation of what results will be produced by the same analytical procedure in the future, i.e. in routine analysis. This is the tolerance interval role.

The parameters estimate μ_j, $\sigma^2_{B,j}$ and $\sigma^2_{W,j}$, at every j concentration level are used to estimate the expected proportion of observations that will fall within the predefined acceptance acceptance $[-\lambda, +\lambda]$, i.e.:

$$E_{\hat{\mu},\hat{\sigma}}\{P[|X - \mu_T| < \lambda]\hat{\mu}_M, \hat{\sigma}_M\} \geq \beta \quad (23)$$

Table 12
Values of precision, trueness, total error and maximal error observed, obtained with the main response functions tested

Model	Concentration (ng/ml)	Intermediate precision (ng/ml)	Trueness (ng/ml)	Absolute total error (ng/ml)	Relative total error (%)	Maximal observed error (%)
Linear 0–max	25.35	1.229	−6.629	7.858	31.0	37.76
	48.24	2.447	−8.99	11.437	23.7	25.81
	437.8	21.03	−10.13	31.16	7.1	9
	838.6	43.32	16.82	60.14	7.2	9.55
Linear 0–223	25.35	1.24	−6.545	7.785	30.7	37.25
	48.24	2.902	−8.794	11.696	24.2	26.87
	437.8	27.49	−7.964	35.454	8.1	10.31
	838.6	56.58	21.17	77.75	9.3	11.67
Linear	25.35	1.61	−2.02E−02	1.63016	6.4	9.12
	48.24	1.785	−2.669	4.454	9.2	10.8
	437.8	19.33	−9.192	28.522	6.5	8.41
	838.6	38.76	11.77	50.53	6.0	8.05
Weighted linear, 1/*X*	25.35	1.483	2.138	3.621	14.3	16.51
	48.24	2.537	−0.6752	3.2122	6.7	9.03
	437.8	18.87	−10.24	29.11	6.6	8.52
	838.6	35.98	7.325	43.305	5.2	6.97
log(*X*) − log(*Y*)	25.35	1.551	2.49E−02	1.57591	6.2	14.57
	48.24	3.045	1.327	4.372	9.1	15.21
	437.8	16.44	−7.985	24.425	5.6	7.2
	838.6	27.75	−34.1	61.85	7.4	8.3
sqrt(*X*) − sqrt(*Y*)	25.35	1.595	−1.31E−02	1.60806	6.3	15
	48.24	2.696	−0.6646	3.3606	7.0	9.74
	437.8	18.71	−6.638	25.348	5.8	7.63
	838.6	35.4	2.626	38.026	4.5	6.27
Quadratic	25.35	3.973	5.23	9.203	36.3	40.18
	48.24	3.961	1.247	5.208	10.8	16.23
	437.8	19.1	−20.09	39.19	9.0	10.24
	838.6	41.16	11.25	52.41	6.2	8.58
Weighted quadratic, 1/*X*	25.35	1.73	2.729	4.459	17.6	20.58
	48.24	2.498	−0.6233	3.1213	6.5	8.83
	437.8	17.98	−14.65	32.63	7.5	9.16
	838.6	39.79	11.51	51.3	6.1	8.28

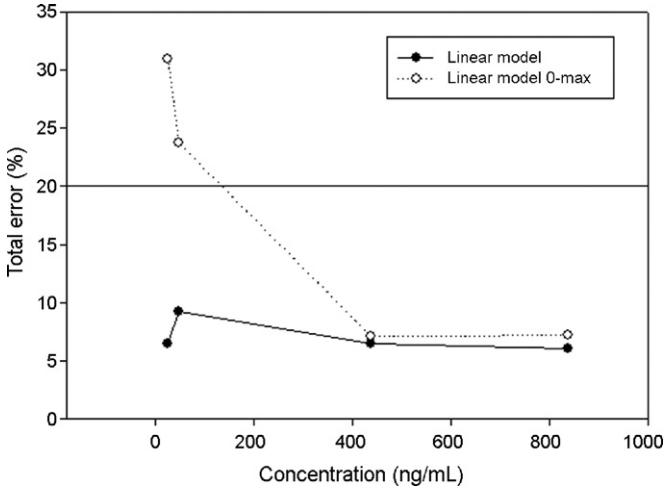

Fig. 3. Graphic of the relative total error by concentration level when the response function is the simple linear model or the linear model through 0 and the highest level.

However, there is no exact solution to estimate this expected proportion. The solution already proposed by several authors [1,14,16,18,19] consists in calculating the tolerance interval (β-expectation tolerance interval) as proposed by Mee [20]:

$$E_{\hat{\mu}_M,\hat{\sigma}_M}\{P_X[\hat{\mu}_M - k\hat{\sigma}_M < X < \hat{\mu}_M + k\hat{\sigma}_M|\hat{\mu}_M, \hat{\sigma}_M|]\} = \beta \quad (24)$$

where k is chosen in order that the proportion of future results expected to fall into the interval limits, equals β.When this tolerance interval is totally included within the limits $[-\lambda, +\lambda]$, i.e. if ($\hat{\mu}_M - k\hat{\sigma}_M > -\lambda$ and $\hat{\mu}_M + k\hat{\sigma}_M < +\lambda$), then the expected proportion of results within the acceptance limits will be higher than β. The estimation of trueness and precision parameters μ_j, $\sigma^2_{B,j}$ and $\sigma^2_{W,j}$ at each concentration level, is not an end in itself, but a necessary step to calculate the expected proportion of results located within the acceptance limits. The tolerance interval will be computed, at each considered concentration level of the validation standards. In practice, the tolerance interval is

computed as follows [19,20], in absolute value:

$$\left[\hat{\mu}_j - Q_t\left(\nu; \frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{pnB_j^2}}\hat{\sigma}_{IP,j}; \hat{\mu}_j + Q_t\left(\nu; \frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{pnB_j^2}}\hat{\sigma}_{IP,j}\right] \quad (25)$$

where

- $\hat{\sigma}^2_{FI,j} = \hat{\sigma}^2_{W,j} + \hat{\sigma}^2_{B,j}$;
- $R_j = \frac{\hat{\sigma}^2_{B,j}}{\hat{\sigma}^2_{W,j}}$;
- $B_j = \sqrt{\frac{R_j+1}{nR_j+1}}$;
- $\nu = \frac{(R+1)^2}{(R+(1/n))^2/(p-1)+(1-(1/n))/pn}$ [21];
- $Q_t\left(\nu; \frac{1+\beta}{2}\right)$ is the β quantile of the Student t distribution with ν degrees of freedom.

The same interval as Eq. (25), in relative scale, becomes:

$$\left[\text{bias}_j(\%) - Q_t\left(\nu; \frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{pnB_j^2}}\text{CV}_{IP,j}; \text{bias}_j(\%) + Q_t\left(\nu; \frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{pnB_j^2}}\text{CV}_{IP,j}\right] \quad (26)$$

Two terms are contained in the tolerance interval: one is the trueness and the other one, up to a factor, is the intermediate precision coefficient variation. For this reason the tolerance interval may be thus considered as expressing the results accuracy. But the tolerance interval integrates an additional dimension, the chance (or risk), for future results, conditionally to past results, to fall within (outside) the acceptance limits. Then the method can be considered accurate, at β chance level, for the concentration level in question, if the tolerance interval is included within the limits $[-\lambda, +\lambda]$ defined a priori, according to the method objectives. Table 13 contains the values of the tolerance intervals of each concentration level for all the response functions tested in the example.

3.4.3. Accuracy profile and decision

3.4.3.1. Calculation. According to Eq. (26), these intervals boundaries are

$$L_j = \text{bias}_j(\%) - Q_t\left(\nu; \frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{pnB_j^2}}\text{CV}_{IP,j} \quad (27)$$

$$U_j = \text{bias}_j(\%) + Q_t\left(\nu; \frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{pnB_j^2}}\text{CV}_{IP,j} \quad (28)$$

The analytical method's accuracy profile is achieved by joining, on the one hand, the lower limits L_j between themselves ($L_1 \to L_2 \to \cdots \to L_m$), and, on the other hand, the upper limits U_j between themselves ($U_1 \to U_2 \to \cdots \to U_m$) as represented in Figs. 4 and 5.

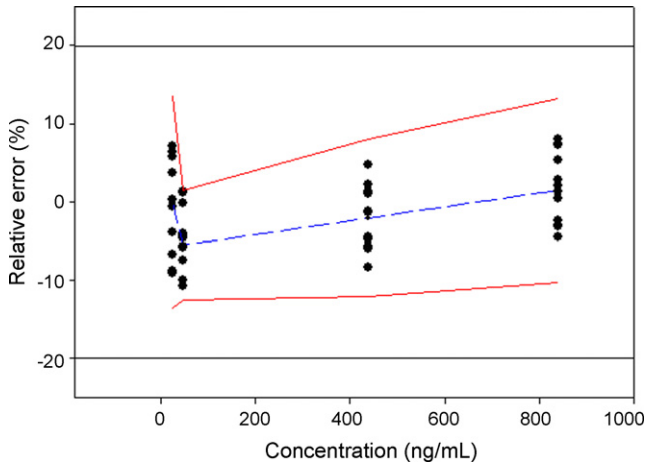

Fig. 4. Ninety-five percent accuracy profile of the analytical procedure results when the simple linear model is chosen as response function. Acceptance limits are set at ±20%. The dashed line is the relative bias.

3.4.3.2. Selection of the response function. When examining Fig. 6, it can be seen that, using some response functions does not allow the analytical procedure to achieve its objectives, given that, for some concentrations, the tolerance limits go beyond the acceptance limits [−20%, +20%] for this example. In addition, among the acceptable response functions, it could be noted that some of them provide better results than others and thus will have to be selected. We also insist on the fact that for all these models, the determination coefficient r^2 is always higher than 0.99 and here, unrelated to the quality of the results. Once again, we would like to stress, that this coefficient is not a relevant indication of the quality of the results that the procedure will produce. As indicated above, the accuracy profile, directly reflecting the analytical procedure potential, makes possible to appreciate the adequacy of different practices and allows to make decisions. For instance, one may wonder whether linear regression has to be weighted or not, as suggested in [5] and how to weight it.

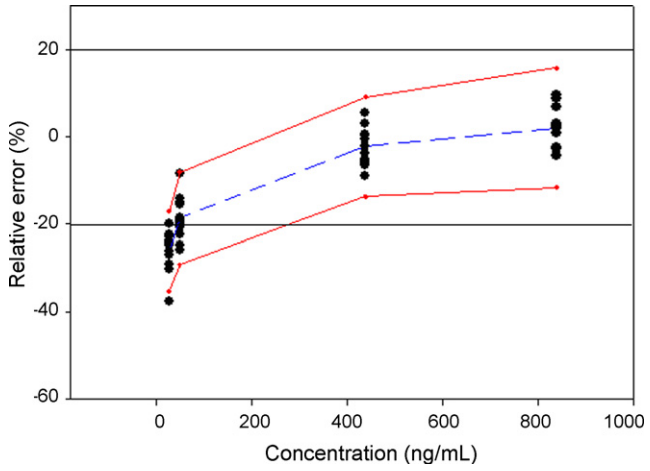

Fig. 5. Accuracy profile, at 95%, of the analytical procedure results, when the model selected is the linear model through 0 and the highest (964.8233 ng/ml) concentration level of the calibration range. The acceptance limits have been set at ±20% and the intersections between the accuracy profile and the acceptance limits define the lower quantitation limit (LQL = 278 ng/ml) and Upper limits of quantitation (UQL = 838 ng/ml).

Table 13
Tolerance intervals obtained with each response function

Model	Concentration level (ng/ml)	Mean introduced concentration (ng/ml)	Absolute tolerance limits (ng/ml)	Relative tolerance limits (%)
Linear 0–max	25.4	25.35	[16.42, 21.03]	[−35.23, −17.06]
	48.2	48.24	[34.16, 44.34]	[−29.19, −8.084]
	437.8	437.8	[378.0, 477.4]	[−13.66, 9.037]
	838.6	838.6	[740.8, 970.2]	[−11.67, 15.68]
Linear 0–223	25.4	25.35	[16.48, 21.13]	[−34.98, −16.64]
	48.2	48.24	[32.83, 46.07]	[−31.96, −4.501]
	437.8	437.8	[356.7, 503.0]	[−18.52, 14.89]
	838.6	838.6	[696.2, 1023]	[−16.98, 22.03]
Linear	25.4	25.35	[21.89, 28.77]	[−13.64, 13.49]
	48.2	48.24	[42.23, 48.92]	[−12.47, 1.405]
	437.8	437.8	[384.6, 472.6]	[−12.15, 7.953]
	838.6	838.6	[752.0, 948.9]	[−10.33, 13.14]
Weighted linear, $1/X$	25.4	25.35	[24.46, 30.53]	[−3.534, 20.40]
	48.2	48.24	[42.11, 53.03]	[−12.72, 9.916]
	437.8	437.8	[384.9, 470.2]	[−12.08, 7.404]
	838.6	838.6	[757.4, 934.6]	[−9.692, 11.44]
$\log(X) - \log(Y)$	25.4	25.35	[22.47, 28.29]	[−11.38, 11.58]
	48.2	48.24	[42.92, 56.22]	[−11.03, 16.53]
	437.8	437.8	[393.9, 465.8]	[−10.03, 6.384]
	838.6	838.6	[740.2, 868.9]	[−11.73, 3.603]
$\mathrm{sqrt}(X) - \mathrm{sqrt}(Y)$	25.4	25.35	[22.18, 28.50]	[−12.52, 12.41]
	48.2	48.24	[41.84, 53.32]	[−13.27, 10.52]
	437.8	437.8	[388.8, 473.6]	[−11.20, 8.165]
	838.6	838.6	[753.7, 928.8]	[−10.13, 10.75]
Quadratic	25.4	25.35	[18.39, 42.78]	[−27.47, 68.72]
	48.2	48.24	[38.59, 60.39]	[−20.00, 25.18]
	437.8	437.8	[374.1, 461.3]	[−14.55, 5.373]
	838.6	838.6	[746.1, 953.7]	[−11.04, 13.72]
Weighted quadratic, $1/X$	25.4	25.35	[24.17, 31.99]	[−4.648, 26.18]
	48.2	48.24	[42.20, 53.03]	[−12.52, 9.935]
	437.8	437.8	[383.7, 462.7]	[−12.37, 5.675]
	838.6	838.6	[749.5, 950.8]	[−10.63, 13.37]

With the accuracy profiles, an answer could simply be obtained by comparing what could be obtained in the different scenarios, as presented in Fig. 7. As it appears, (1) weighting by $1/X$ or $1/X^2$ has very little influence, (2) even if the responses variances are very heterogeneous from a concentration to the other, the model without weighting provides very good results and, in addition, has the advantage to be simple to use. It is this model that will be preferred, among these three options. The question of the transformation to be applied to the data, or the matrix potential role on the quality of the results, can be examined in the same way, i.e. comparing the profiles, as in Figs. 6 and 7.

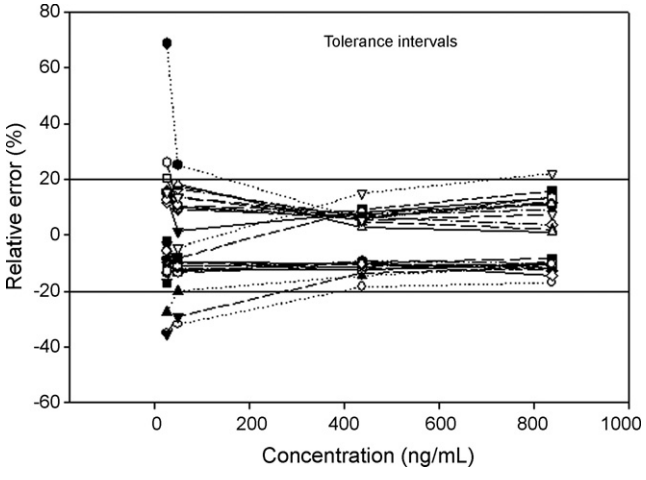

Fig. 6. Superposition of the different accuracy profiles (tolerance limits) for the different response function tested. Acceptance limits are set at ±20%.

3.5. Linearity

The linearity of an analytical procedure is its ability, within a given measurement interval, to obtain results directly proportional to the quantity (e.g. concentration) in analyte within the sample [17]. It must be reminded that the linearity requirement applies to the results (computed concentration = *f*(introduced concentration)), and not to the responses (signal = *f*(introduced concentration)). This is a prerequisite to the trueness assessment. Conversely, the existence of a linear relationship between the estimated concentration and the introduced concentration does not mean the method has adequate trueness, e.g. is not biased. In the case of the example using the simple linear model as a

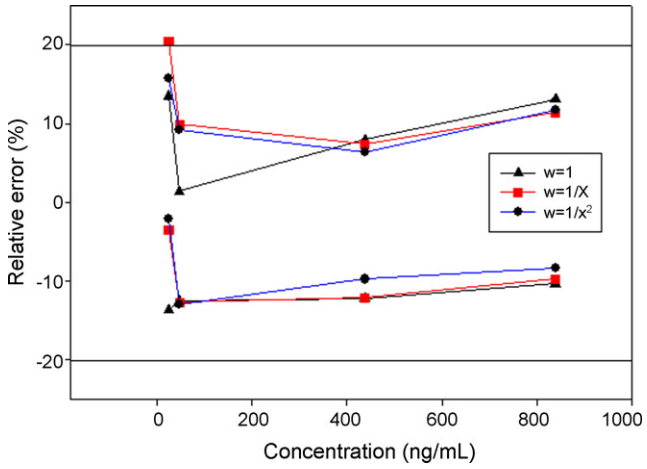

Fig. 7. Accuracy profiles obtained with the simple linear regression ($w = 1$), the weighted simple linear regression with $w = 1/X$ and with $w = 1/X^2$. Acceptance limits are set at ±20%.

response function, the linearity graphic is represented in Fig. 8, with the intercept, the slope (close to 1) and r^2 estimations.

3.6. The quantitation limits

The accuracy profile, built up from the expected measurements tolerance intervals allows, as illustrated in Fig. 5, to decide for which concentration levels a procedure is able to provide results within the acceptance limits. So, by definition, when it occurs, the intersection between the accuracy profile and the acceptance limits defines the lower limit of quantitation of the procedure (LQL) as well as the upper limit of quantitation of the procedure (UQL). Between these two limits, there is, obviously, the measurement interval or dosing range. Consequently, the quantitation limits are clearly the extreme values that can be quantified with a defined accuracy. It is to be noted that should the simple linear model be selected, as in the case in Fig. 4, the quantification limits then become the extreme concentration values investigated during the validation experiments.

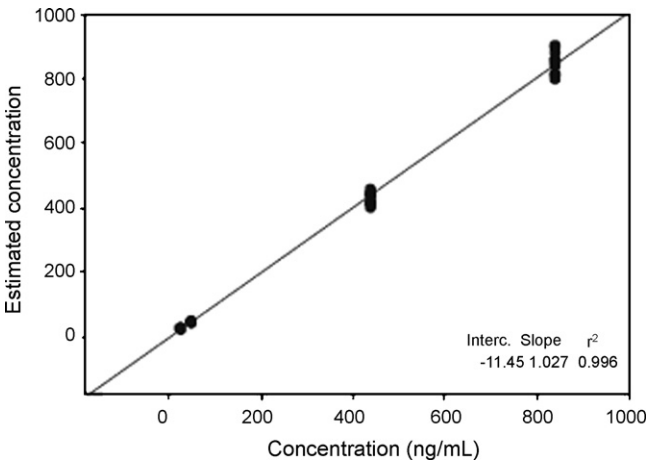

Fig. 8. Graphical display of the linearity of the procedure, with the estimations of the intercept, the slope and r^2.

4. Conclusion

4.1. One decision, one statistics

As already underlined in the parts 1 and 2 of this guide [1,2], the intention was not only to simplify the approach for analyzing the data of the validation of an analytical procedure, but, even more, to make statistics appropriate and consistent with the objective of validation. At the end of the validation, only one decision has to be made – whether the analytical procedure is considered as valid or not – and thus, only one statistics has to support this decision. This is the role played by the accuracy profile. This approach contrasts with the complicated strategies that require the computation of a large number of statistics, often badly interpreted, sometimes contradictory, when not erroneous. Facing this mosaic of numbers, the analyst founds himself alone to make a decision difficult to defend later. With the accuracy profile, the analytical interpretation is easy and all the useful required statistics, such as trueness, precision, quantitation limits, risk, linearity, are integrated. In addition, the accuracy profile makes possible a visual representation of the future performances of the procedure.

4.2. Quality of results rather than quality of statistics

Bringing back the results and their quality in the centre of our preoccupations also constitutes something new, compared to the approaches recommended so far. A drift, that started a long time ago, induced the computation of numerous statistics that could be obtained from the observations, such as trueness, precision, lack of fit test, variances homogeneity, aberrant values, the r^2, etc., so many statistics and properties that have been thought to confusedly be related to the quality of the results, even though such a relationship has never been clearly demonstrated. In the approach presented here, only the quality of the future results matters, because the very objective of an analytical procedure is to provide results, not statistics. And the relationship described in Eq. (23) is the very relationship that correctly links statistics with results.

4.3. Customer and laboratory risks controlled

New dimensions are also introduced in this guide: the notions of risks and in particular the customer risk—i.e. the user of the results. Actually, guides such as the ICH Q2R1 one [8], are focusing their attention on the quality of the analytical procedure, by calculating analytical performance criteria such as procedure trueness and precision. But almost nothing is proposed, in clear terms, to assess the quality of the results produced by a procedure: and yet, this is the only objective of an analytical procedure. It is too often implicitly understood that, if the procedure is “good”, then the results it will produce will be “good”. Formally, it is not true and no guarantee can be given. Conversely, if it is possible to demonstrate that the results are “good”, then they can only be obtained by a “good” procedure, it is exactly the approach adopted in this guide. The fundamental and practical consequence is, if it is possible to demonstrate by

using accuracy profiles, that the results are good, then automatically, all the performance criteria that must be calculated and reported to meet the regulatory requirements will also be met. Thus, the advantage is double: understanding the results quality and meeting the regulatory requirements.

4.4. Generalization to any types of methods

Finally, as it will be seen in part 4 of this guide, this approach applies to any type of analytical procedures. Indeed, there is no reason why the way of assessing quantitative procedure results is different according to the very nature of the procedure, physicochemical or biological. Some technical aspects can vary, the response functions for instance, but, finally, the ending point is identical: the accuracy profile of the results.

Acknowledgment

The authors wish to thank the members of the SFSTP office for the organization of the meetings.

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, J. Pharm. Biomed. Anal. 36 (2004) 579–586.
[2] 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, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 70–81.
[3] E. Chapuzet, N. Mercier, S. Bervioas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, Ph. Hubert, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, STP Pharma Prat. 7 (1997) 169–194.
[4] E. Chapuzet, N. Mercier, S. Bervioas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, Ph. Hubert, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, STP Pharma Prat. 8 (1997) 81–107.
[5] 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) 135–148.
[6] ISO 5725, Accuracy (trueness and precision) of measurement methods and results—Parts1–4, Part 6, ISO, Geneva, Switzerland, 1994.
[7] ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories, ISO, Geneva, 2000.
[8] International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use, Topic Q2 (R1): Validation of Analytical Procedures: Text and Methodology, Geneva, 2005.
[9] Guidance for Industry: Bioanalytical, Methods Validation for Human Studies, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), 2001.
[10] Guidance for Industry: Analytical, Procedures and Methods Validation (Draft guidance), U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Août, 2000.
[11] V.P. Shah, K.K. Midha, S. Dighe, I. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman, J. Pharm. Sci. 81 (1992) 309–312.
[12] V.P. Shah, K.K. Midha, J.W.A. Findlay, H.M. Hill, J.D. Hulse, I.J. McGilveray, G. McKay, K.J. Miller, R.N. Patnaik, M.L. Powell, A. Tonelli, C.T. Viswanathan, A. Yacobi, Pharm. Res. 17 (2000) 1551.
[13] B. Boulanger, W. Dewé, P. Chiap, J. Crommen, Ph. Hubert, J. Pharm. Biomed. Anal. 32 (2003) 753–765.
[14] M. Feinberg, B. Boulanger, W. Dewé, Ph. Hubert, Anal. Bioanal. Chem. 380 (2004) 502–514.
[15] E. Rozet, C. Hubert, A. Ceccato, W. Dewé, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 126–137.
[16] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 111–125.
[17] S. Huet, E. Jolivet, A. Méssean, La régression non linéaire, Éditions INRA, Paris, 1992.
[18] 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 Prat. 13 (2003) 101–138.
[19] D. Hoffman, R. Kringle, J. Biopharm. Stat. 15 (2005) 283–293.
[20] R. Mee, Technometrics 26 (1984) 251–254.
[21] F. Satterthwaite, Psychometrika 6 (1941) 309–316.