ELSEVIER

Analytica Chimica Acta 391 (1999) 135–148

ANALYTICA CHIMICA ACTA

The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: from the Washington Conference to the laboratory

Ph. Hubert[1,a,*], P. Chiap[1,a], J. Crommen[a], B. Boulanger[1,b], E. Chapuzet[1,c], N. Mercier[1,c], S. Bervoas-Martin[1,d], P. Chevalier[1,e], D. Grandjean[1,f], P. Lagorce[1,f], M. Lallier[1,d], M.C. Laparra[1,g], M. Laurentie[1,h], J.C. Nivet[1,i]

[a]*Department of Analytical and Pharmaceutical Chemistry, Institute of Pharmacy, University of Liège, CHU, B36, B-4000 1 Liège, Belgium*
[b]*Lilly Development Centre, rue Granbompré, 11, B-1348 Mont-Saint-Guibert, Belgium*
[c]*QUALILAB, quai du Châtelet, 72, F-45000 Orléans, France*
[d]*INNOTHERA, avenue Paul Vaillant Couturier, 10, F-94110 Arcueil, France*
[e]*RHONE POULENC RORER, avenue Aron, 20, F-92165 Antony, France*
[f]*SANOFI, rue du Professeur Blayac, 371, F-34184 Montpellier cedex, France*
[g]*UPSA, B.P. 325, F-92506 Reuil Malmaison, France*
[h]*Agence du Médicament Vétérinaire, B.P. 203, F-35032 Fougères cedex, France*
[i]*CERB, chemin de Montifault, F-18800 Baugy, France*

Received 13 July 1998; received in revised form 12 November 1998; accepted 21 November 1998

Abstract

On the basis of the guidelines given in the Washington Conference report and the ICH (International Conference of Harmonisation) recommendations some suggestions about experimental design and data evaluation are proposed by an SFSTP Commission dedicated to the validation of chromatographic methods in bioanalysis. In a series of meetings, members of this Commission have tried to elaborate a rational, practical and statistically reliable strategy to assure the quality of the analytical results generated. This strategy has been formalised in a guide and the main suggestions made by the Commission are summarised in the present paper. The SFSTP guide has been produced to help analysts from the pharmaceutical industry to validate their bioanalytical methods. It is the result of a consensus between professionals having expertise in bioanalytical and/or statistical fields. The suggestions presented in this paper should therefore help the analyst to design and perform the minimum number of validation experiments needed to obtain all the required information to establish and demonstrate the reliability of its analytical procedure. The SFSTP guide suggests a validation strategy in two steps: a pre-validation and the validation itself. An experimental design is described for each of these steps and the main aspects discussed in the paper are related to the selection of the most appropriate calibration model to fit experimental data and the most suitable way to determine the limit(s) of quantitation and subsequently the calibration range as well as the optimum number of experiments to be performed in the validation phase. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Validation; Bioanalysis; Experimental design; Selectivity; Specificity; Absolute recovery; Response function; Accuracy; Linearity; Precision; Limit of detection; Limit of quantitation

*Corresponding author. Tel.: +32-4366-4346; fax: +32-4366-4347; e-mail: ph.hubert@ulg.ac.be

[1]Members of the SFSTP (Société Française Sciences et Techniques Pharmaceutiques) Commission, 5, rue Basse-des-Carmes, F-75005 Paris, France.

0003-2670/99/$ – see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0003-2670(99)00106-3

1. Introduction

The different European and American (Food and Drug Administration, FDA) authorities related to good laboratory practice (GLP) and good clinical practice (GCP) recommend that analytical data from studies intended to be included in new drug applications (NDAs) have to meet some acceptance criteria. As a consequence, assay methods must be validated and routinely performed analyses must be followed according to a quality control system.

Available documents defining validation criteria that must be assessed during the development of an analytical procedure are not precise enough [1–4]. They are usually restricted to general concepts and they do not provide any experimental approach. In order to help professionals from the pharmaceutical industry to validate their analytical procedures a specially dedicated commission from the Société Française des Sciences et Techniques Pharmaceutiques (SFSTP) elaborated a validation guide in 1992 [5].

With only a few exceptions [6–8], the situation is rather similar in the field of bioanalysis, i.e. the isolation and analysis of a drug and/or its metabolites from a biological matrix. Even if the currently available documents [9–24] have contributed to significant improvements in the validation of analytical methods in biological matrices, they are purposely general. The Washington Conference report [16] which is now being utilised as a basis for bioanalytical method validation gives the minimum requirements (e.g. three concentration levels covering the range of the expected concentrations, the minimum number of experiments to be performed at each concentration level, the acceptance criteria of 15% and 20% for precision and accuracy, etc.) but does not provide with any validation strategy. Under these conditions, the validation work performed in bioanalysis depends mainly on the analyst's experience, on his personal point of view and on the laboratory strategy [17]. Consequently, a new SFSTP Commission was created in 1995 to prepare a guide for the validation of bioanalytical methods.

In this paper a synopsis of the SFSTP guide is presented [25], which illustrates the consensus between both experimental and statistical points of view and describes the recommended strategy, allowing the analyst to perform the most profitable experimental work possible. This is achieved by drawing maximum information from the results and in routine use by reducing the re-analysis risk due to a lack of agreement with the validation criteria. It should be noted that the strategy presented in this paper is principally meant for chromatographic methods, but it could also be applied to other bioanalytical methods based on techniques such as capillary electrophoresis, fluorometry, atomic absorption spectrometry, etc. Moreover, even though the strategy was originally developed for pharmaceutical applications, it can reasonably be applied to other fields with similar specifications (environment, food products, etc.).

2. Validation goals

The aim of validation is to establish that the analytical method is suitable for its intended use [3,5]. In bioanalysis, the validation goal is to demonstrate reliable performance of the assay method and to prove the precision and accuracy of the results within well-defined limits. Methods used to quantitate low concentrations of the parent compound or of its metabolites have a great influence on the evaluation of different parameters in bioavailability, bioequivalence and pharmacokinetic/metabolism studies [3,16].

3. Validation criteria

The following generally accepted validation criteria [1–24,26,27] are listed in the SFSTP guide:

- specificity-selectivity,
- response function (calibration curve),
- linearity,
- assay range,
- accuracy,
- precision (repeatability and intermediate precision),
- limit of detection (LOD),
- limit of quantitation (LOQ),

In bioanalysis, additional criteria must be considered:

- absolute recovery (extraction efficiency),
- influence of the dilution [6,7,13,16,21],

- analyte stability in the biological matrix [6,7,13,16,20,21,23,24].

In the context of validation criteria, it must be borne in mind that these criteria have to be obtained from the same biological matrix as that of the samples to be analysed [16]. Every new analytical method must be validated for each animal species and for each type of matrix [3]. Moreover, each modification of a previously validated method automatically involves a re-validation, the extent of which depends on the modifications made and their possible influence on specific validation criteria [3,10,26].

The definitions of the validation criteria selected by the SFSTP Commission are those given in the ICH text on validation of analytical procedures: definitions and terminology [26] excepted for the two criteria described below.

3.1. Response function (calibration curve)

The response function [15] of an analytical method is, within the range, the existing relationship between the response (signal) and the concentration (quantity) of the analyte in the sample. The calibration curve is the simplest monotonous response function.

The response function can be linear (straight line), but non-linear models, sometimes related to by detection method [28] or to the particularly wide concentration range, can also be observed. The response function must however be monotonous, i.e. strictly increasing or decreasing. It should be noticed that the estimation of such a function using common fitting methods (e.g. least squares method) assumes that the response variance is constant irrespective of the analyte concentration (homoscedasticity). However, such conditions are not always met in bioanalysis. The function which meets these requirements and fits the response adequately is the calibration curve. The latter is then used to calculate the concentrations, i.e. the *results*. The stability of the calibration curve must be checked each time analyses are conducted.

3.2. Linearity (of the results)

The linearity of an analytical method is its ability within a definite range to obtain results directly proportional to the concentrations (quantities) of the analyte in the sample.

The linearity criteria must only be applied to the results [calculated concentration=f (introduced concentration)], not to the responses [signal=f (introduced concentration)]. A method must be accurate, but a linear relationship between calculated and introduced concentrations does not guarantee method accuracy (e.g. when a bias is present).

Another validation parameter is defined in the guide.

3.3. Absolute recovery [7,13,23,24]

The absolute recovery is the ratio of the response measured for a spiked sample treated according to the whole analytical procedure to that of a non-biological sample (e.g. aqueous solution) spiked with the same quantity of the targeted substance and directly injected into the chromatographic system [7,13].

It must be emphasised that absolute recovery can be calculated in one of these two ways: either by comparing extracted and unextracted standards as suggested in the given definition or by using radioactive analyte. The former method can be applied to virtually all bioanalytical procedures provided that a pure reference standard exists. The main advantage of a labelled analyte is that every fraction from the extraction procedure can be rapidly measured to control where any losses occur [13].

Whenever possible it is preferable to study the absolute recovery than the relative one for which the non-biological sample is treated following the same method as the biological sample. When a derivatisation procedure is used it is only possible to determine a relative recovery [7,13].

The absolute recovery is an important criteria when the procedure involves an extraction step [21] and because of its influence on the lowest quantifiable value, in case of a too low absolute recovery, it is recommended to undertake further method development. A rather low recovery can be accepted if a low limit of quantitation is not required or if the detection method is sensitive enough, but a repeatable recovery must be demonstrated throughout the range [20]. Most often a low recovery is related to a low precision [16].

4. Validation phases

The implementation of an assay procedure usually involves four steps: development, pre-validation, validation and routine analysis. Validation is a dynamic component of the development of a new analytical method. During the first three steps method performance improves and accuracy in the results increases. During routine use a quality control (QC) system is laid down in order to validate results from unknown samples and to follow-up the performance of the analytical system [10,12,13,15,17]. In this framework, the aim of the SFSTP Commission was to make an optimum use of the information obtained from the experiments usually performed during the pre-validation step in order to reduce the time-consuming validation step and introduce an integrated, efficient and cost-effective design for the validation of chromatographic methods taking into account the requirements drawn up in the Washington Conference report [16]. In order to be consistent with this objective, the experimental approach proposed by the SFSTP comprises two steps: a pre-validation phase and a validation phase (formal validation step).

4.1. Pre-validation phase

The aim of the pre-validation step is also to prepare the validation step. Consequently, before starting experiments the configuration of the equipment, the preparation of the stock solution and its dilutions, the validation standards (validation standards are QC samples used during the validation step) must be well defined in the validation protocol. As shown in Table 1, the pre-validation step allows the analyst to:

- identify the response function of the calibration curve (linear, non-linear, mathematical transformation, weighting),
- define the limit of detection (LOD),
- estimate the limit(s) of quantitation (LOQ),
- evaluate the range and the number of calibration levels,
- determine the (absolute) recovery,
- evaluate selectivity.

Table 1
Validation criteria to be studied in the pre-validation and validation steps

Criteria	Pre-validation step	Validation step
Absolute recovery	+	
Selectivity	(+)	+
Response function	(+)	+
Linearity		+
Accuracy		+
Precision		+
Limit of detection	+	
Limits of quantitation	(+)	+
Range	(+)	+

(+): estimation of the criteria; +: validation of the criteria.

4.2. Validation phase

The aim of the validation step (see Table 1) is to:

- demonstrate selectivity,
- validate the calibration model intended to be routinely used,
- evaluate accuracy,
- control linearity,
- evaluate precision (repeatability and intermediate precision),
- validate the limit(s) of quantitation.

5. Experimental

5.1. Pre-validation phase

5.1.1. Preparation of samples

As illustrated in Fig. 1, from a stock solution different diluted solutions are prepared to obtain at least six calibration levels. In order to be able to evaluate the limit of quantitation each calibration level must be treated at least as triplicates. To prepare the three minimum calibration curves (series) in the biological matrix, it is recommended to prepare three different stock solutions and to use three different sources of biological matrix. But if not possible (e.g. too little material available), according to the ICH text on validation of analytical procedure: methodology [27], a single stock solution can be prepared and three independent dilution series made from it. It should also be emphasised that the three series must be successively analysed, but not necessarily on three different days. To determine the variability, samples from the three series must not be pooled and must be

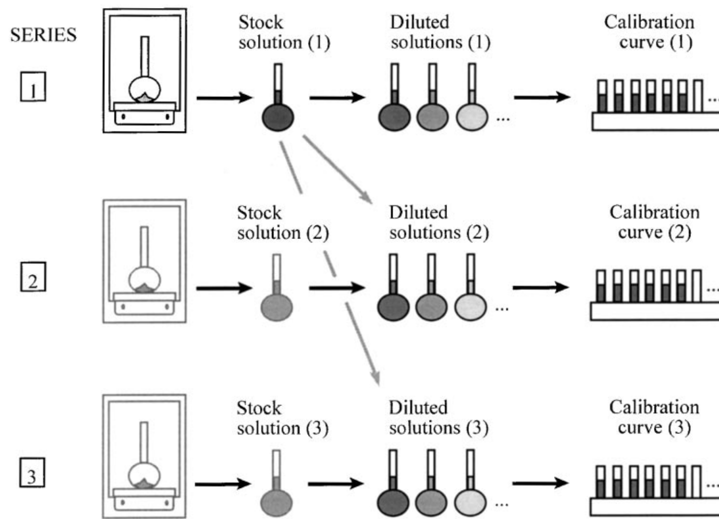

Fig. 1. Sample preparation for the pre-validation phase.

independently analysed (new series, buffers, mobile phases, etc.) (see Fig. 1).

The calibration points can be determined as follows:

- x_{i1k}, a low concentration level, close to the supposed limit of detection (LOD) and lower than the expected limit of quantitation (LOQ),
- x_{i2k}, a concentration level close to the lowest quantifiable value,
- x_{i3k}, a concentration level higher than the previous one (e.g. a twofold concentration level),
- x_{imk}, the maximum expected concentration level,
- intermediate concentration levels ($x_{i(m-1)k}$ to $x_{i(m-q)k}$ with q defined as the number of times that intermediate concentrations are calculated) can be determined as follows:

$$X_{i(j-1)k} = \frac{X_{i2k} + X_{ijk}}{2},$$

i.e. the medium concentration between the previous one and x_{i2k} (see Fig. 2), with $i \in [1, n_{jk}]$ the replication indices for concentration j in series k; $j \in [1, m]$ the concentration indices, and $k \in [1, p]$ are the series indices.

Finally, in order to evaluate the (absolute) recovery, a non-biological calibration curve (x^*_{ijk}) can be prepared using the stock solution (or one of the stock solutions) previously used to prepare the biological

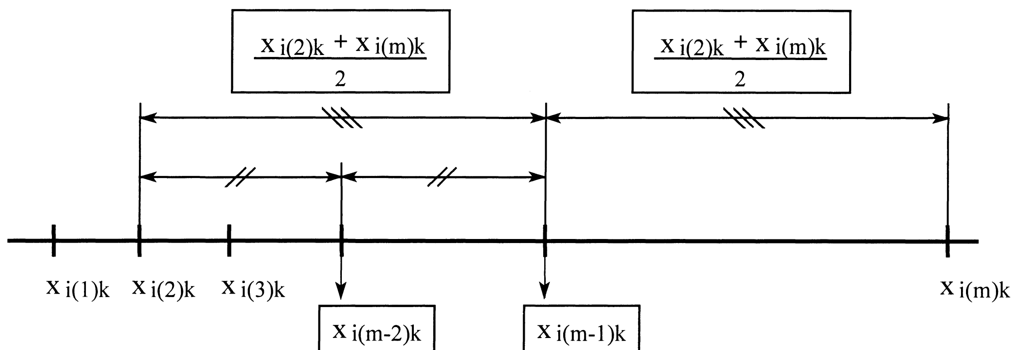

Fig. 2. Choice of intermediate concentrations.

Table 2
Experiments to be performed in the pre-validation step

	Biological samples									Non-biological samples		
	Series											
	1			2			3			3[a]		
	Replicates											
	1	2	3	1	2	3	1	2	3	1	2	3
x_{i1k}	x_{111}	x_{211}	x_{311}	x_{112}	x_{212}	x_{312}	x_{113}	x_{213}	x_{313}	x_{113}[a]	x_{213}[a]	x_{313}[a]
x_{i2k}	x_{121}	x_{221}	x_{321}	x_{122}	x_{222}	x_{322}	x_{123}	x_{223}	x_{323}	x_{123}[a]	x_{223}[a]	x_{323}[a]
x_{i3k}	x_{131}	x_{231}	x_{331}	x_{132}	x_{232}	x_{332}	x_{133}	x_{233}	x_{333}	x_{133}[a]	x_{233}[a]	x_{333}[a]
⋮												
x_{imk}	x_{1m1}	x_{2m1}	x_{3m1}	x_{1m2}	x_{2m2}	x_{3m2}	x_{1m3}	x_{2m3}	x_{3m3}	x_{1m3}[a]	x_{2m3}[a]	x_{3m3}[a]

[a] In the present table, the stock solution that was used to prepare the stock solution of the third series of biological samples was also used for determining the absolute recovery.

calibration curve. To minimise experimental errors it is recommended to use the same dilutions for the biological and the non-biological samples. The non-biological calibration points are directly introduced in the analytical system. All experiments performed during the pre-validation step are summarised in Table 2. According to this pre-validation experimental protocol, Y_{ijk} is the analytical response (peak area or peak height, ratios, etc.) for each biological sample and the subsequent analysis of the responses is shown in Fig. 3.

5.2. Validation phase

Formal validation experiments must be performed on different days (not necessarily consecutive ones but without non-documented rejection of intermediate results) under conditions as close as possible to those used in routine (apparatus, operator, etc.). Since the aim of the validation is also to evaluate intermediate precision, alternative experiments are advisable. Reproducibility, which is an inter-laboratory parameter, is not described in the SFSTP guide [25].

5.2.1. Preparation of the calibration samples

As shown in Table 3, for each of the p series (at least 3) made of m concentration levels (at least 5), the calibration points used in the validation phase are chosen as follows:

- x_{i1k}, the concentration level corresponding to the lowest quantifiable value (LLOQ) of the method as determined during the pre-validation step,
- x_{i2k}, a concentration level close to the lowest quantifiable value,
- x_{i3k}, a concentration level slightly higher than the previous one (e.g. a twofold concentration level),
- x_{imk}, the highest concentration level tested during the pre-validation step,
- x_{ijk}, a medium concentration between the highest (x_{imk}) and the limit of quantitation (x_{i1k}),
- $x_{i(j-1)k}$, an intermediate concentration between the medium and the limit of quantitation,

with

- $i \in [1, n_{jk}]$ indices of replication,
- $j \in [1, m]$ indices of concentration,
- $k \in [1, p]$ indices for series,

and n_{jk} is the number of replicates at the concentration level j of the calibration samples, in series k; n^{o}_{jk}, the number of replicates at the concentration level j of the validation standard samples, in series k.

It is highly recommended to take all n^{o}_{jk} to be equal.

Generally a standard curve constructed with six calibration points should adequately define the relationship between response and concentration. But, in case of a wide range, additional calibration points may be required. Under these conditions, the latter are iteratively determined using the medium value between the lowest quantifiable value and the previous value as described above (see Fig. 2). On the other hand, in order to validate the regression model defined during the pre-validation step, each calibration point

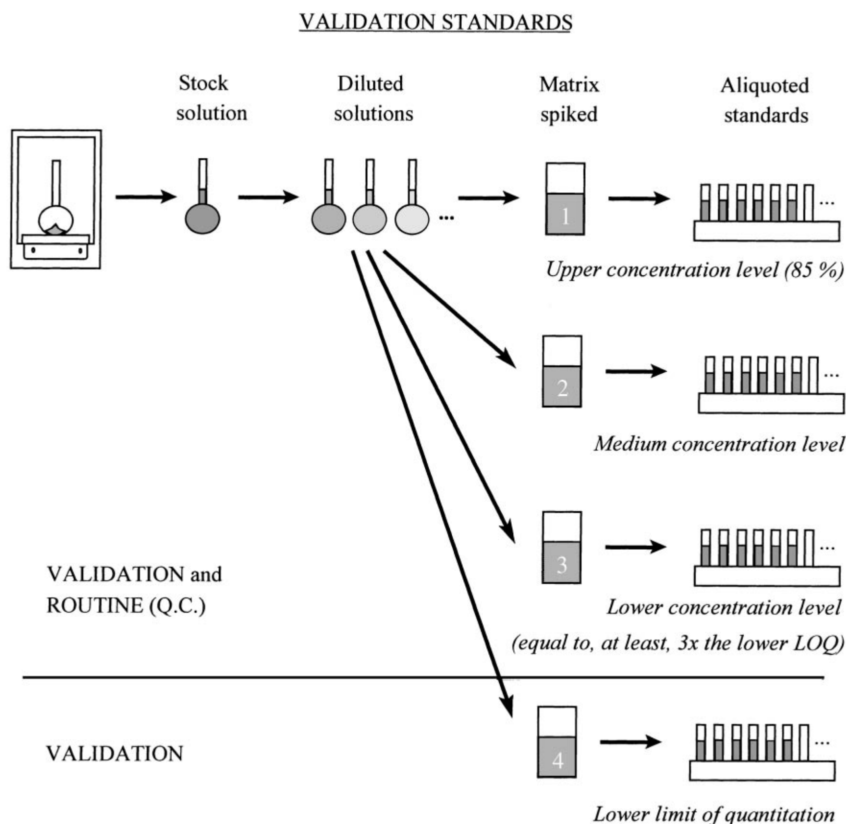

Fig. 3. Preparation of the validation standards.

must be analysed in replicate. However, if a lower number of calibration points is needed, only extreme calibration points must be analysed as replicates if a linear model is valid, whereas in the case of a quadratic model the medium point must also be replicated. Other calibration points are not necessarily replicated. It should also be noted that if only one stock solution can be prepared as mentioned above in the pre-validation phase, p independent series can be made from it but the resulting calibration points must be analysed on different days taking into account the stability of the stock solution.

5.2.2. Preparation of the validation standards

As shown in Fig. 3, except for the lowest quantifiable value (x^{o}_{i1k}), other validation standard levels are those of quality controls (QCs) intended to be used in routine:

- a concentration level equal to, at least, three times the lowest quantifiable value limit,
- $x_{i(50)k}$, a medium concentration level,
- $x_{i(85)k}$, a concentration level close to the highest concentration or the highest quantifiable value (around 85% of the highest concentration).

The basic experimental design for the validation step is summarised in Table 3. Validation standards must be prepared using a different source of matrix and each validation standard must be analysed at least four times. However, according to the results of the pre-validation step, it is possible to determine the optimum number of required replicates as mentioned in the complete version of the SFSTP guide [25,29]. In addition, in order to obtain a realistic evaluation of intermediate precision from this experimental design, the p series must be analysed under conditions as

Table 3
Experiments to be performed in the validation phase

Series	Calibration samples			Validation samples					
	Concentration	Replicates		Concentration	Replicates				
		1	2		1	2	3	...	n
1	x_{i1k}	x_{111}	x_{211}	x^{o}_{i1k}	x^{o}_{111}	x^{o}_{211}	x^{o}_{311}		x^{o}_{n11}
	x_{i2k}	x_{121}	x_{221}	$x^{o}_{i(3)k}$	$x^{o}_{1(3)1}$	$x^{o}_{2(3)1}$	$x^{o}_{3(3)1}$		$x^{o}_{n(3)1}$
	$x_{i(1/4)k}$	$x_{1(1/4)1}$	$x_{2(1/4)1}$	$x^{o}_{i(50)k}$	$x^{o}_{1(50)1}$	$x^{o}_{2(50)1}$	$x^{o}_{3(50)1}$		$x^{o}_{n(50)1}$
	$x_{i(1/2)k}$	$x_{1(1/2)1}$	$x_{2(1/2)1}$	$x^{o}_{i(85)k}$	$x^{o}_{1(85)1}$	$x^{o}_{2(85)1}$	$x^{o}_{3(85)1}$		$x^{o}_{n(85)1}$
	x_{imk}	x_{1m1}	x_{2m1}						
2	x_{i1k}	x_{112}	x_{212}	x^{o}_{i1k}	x^{o}_{112}	x^{o}_{212}	x^{o}_{312}		x^{o}_{n12}
	x_{i2k}	x_{122}	x_{222}	$x^{o}_{i(3)k}$	$x^{o}_{1(3)2}$	$x^{o}_{2(3)2}$	$x^{o}_{3(3)2}$		$x^{o}_{n(3)2}$
	$x_{i(1/4)k}$	$x_{1(1/4)2}$	$x_{2(1/4)2}$	$x^{o}_{i(50)k}$	$x^{o}_{1(50)2}$	$x^{o}_{2(50)2}$	$x^{o}_{3(50)2}$		$x^{o}_{n(50)2}$
	$x_{i(1/2)k}$	$x_{1(1/2)2}$	$x_{2(1/2)2}$	$x^{o}_{i(85)k}$	$x^{o}_{1(85)2}$	$x^{o}_{2(85)2}$	$x^{o}_{3(85)2}$		$x^{o}_{n(85)2}$
	x_{imk}	x_{1m2}	x_{2m2}						
⋮									
p	x_{i1k}	x_{11p}	x_{21p}	x^{o}_{i1k}	x^{o}_{11p}	x^{o}_{21p}	x^{o}_{31p}		x^{o}_{n1p}
	x_{i2k}	x_{12p}	x_{22p}	$x^{o}_{i(3)k}$	$x^{o}_{1(3)p}$	$x^{o}_{2(3)p}$	$x^{o}_{3(3)p}$		$x^{o}_{n(3)p}$
	$x_{i(1/4)k}$	$x_{1(1/4)p}$	$x_{2(1/4)p}$	$x^{o}_{i(50)k}$	$x^{o}_{1(50)p}$	$x^{o}_{2(50)p}$	$x^{o}_{3(50)p}$		$x^{o}_{n(50)p}$
	$x_{i(1/2)k}$	$x_{1(1/2)p}$	$x_{2(1/2)p}$	$x^{o}_{i(85)k}$	$x^{o}_{1(85)p}$	$x^{o}_{2(85)p}$	$x^{o}_{3(85)p}$		$x^{o}_{n(85)p}$
	x_{imk}	x_{1mp}	x_{2mp}						

representative as possible of the potential routine variability (day, operator, apparatus, etc.). For example, if two operators are involved during routine use, the simple design presented in Table 4 can be followed.

5.2.3. Selectivity

To demonstrate the selectivity of the analytical procedure with respect to endogenous substances additional experiments must be performed and at least six different biological matrix sources must be analysed [16,29,30]. Selectivity must be demonstrated towards metabolites, degradation products, etc.

Table 4
Example of experiment for intermediate precision

	Serie					
	1	2	3	4	5	6
Day	1	1	2	2	3	3
Operator	A	B	A	B	A	B

6. Analyses of results

6.1. Pre-validation phase

6.1.1. Determination of the response function

The first step of the pre-validation phase must be the thorough assessment of the relationship between the response and concentration in order to avoid serious difficulties in the estimation of other validation criteria. For instance, the selection of an inadequate calibration model can lead to an increase of the experimental error which could significantly affect the results and conclusions of the study. Once the analyses have been performed as proposed in the experimental protocol, one should evaluate the response function following the different steps presented in Fig. 4. According to the ICH methodology requirements [27], the analytical responses from the three series should be first plotted as a function of the concentrations (step A1) and a visual analysis of the graph allows an immediate detection of a lack of linearity. In such a case, a linearisation method or another type of regression model (polynomial, quadratic, etc.) can be used. Here are some useful trans-

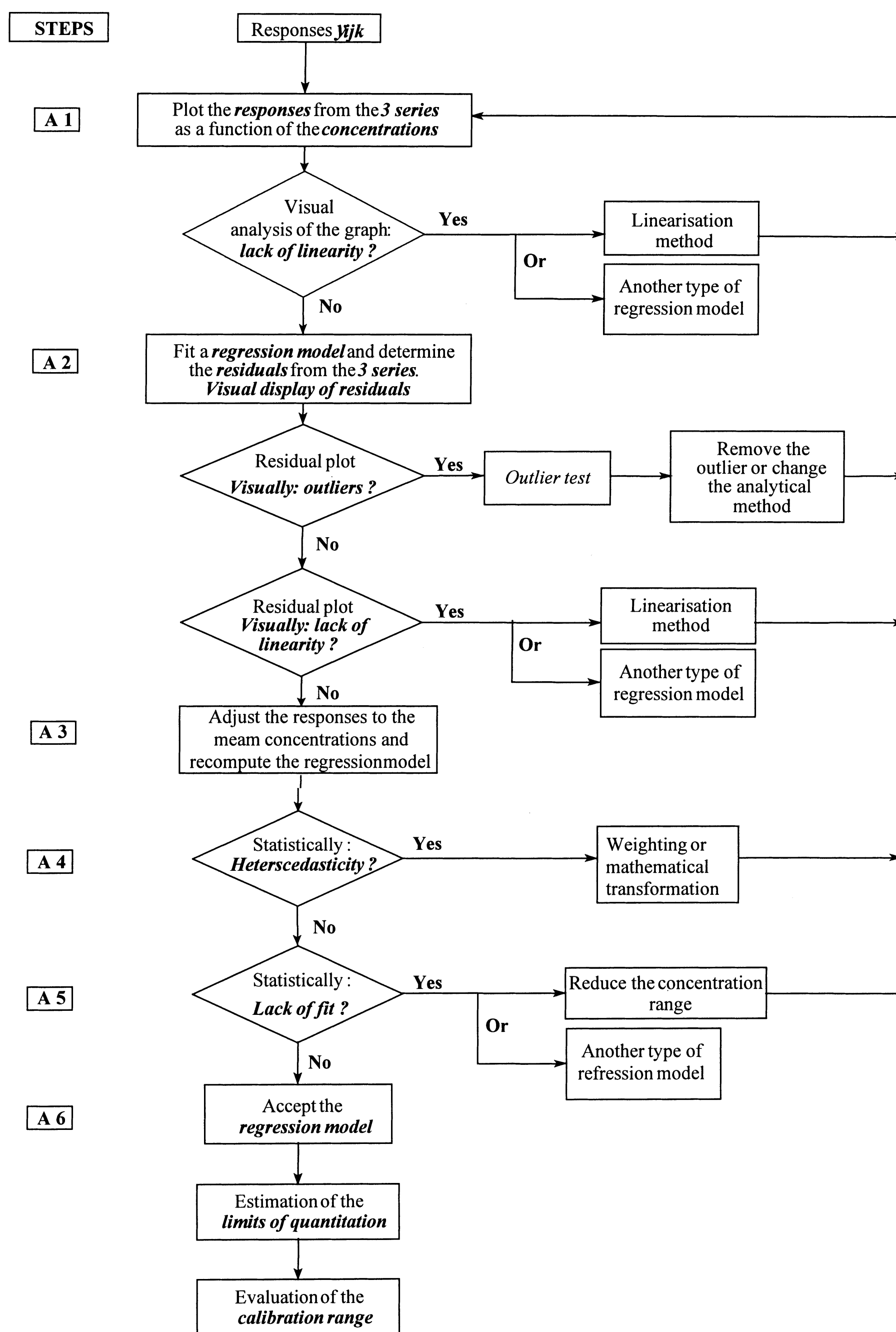

Fig. 4. Determination of the response function.

formations that can be considered for the concentrations (X) and responses (Y):

$$z' = \sqrt{z}, \quad z' = z^2, \quad z' = \frac{1}{z}, \quad z' = e^z, \quad z' = \ln(z).$$

Other transformations can also be envisaged but they must lead at least to strictly monotonous function, i.e. a strict order between the different values should be kept. The response graph can also help the analyst

to select an adequate transformation method. For instance, when a problem of variance homogeneity appears in the graph, a weighted regression model can be directly applied.

If no evident lack of linearity can be detected in this way, a default regression model is then fitted for each calibration curve (step A2). The well-known least squares regression model is first selected in accordance with the Washington Conference report statement that the simplest algorithm which fits the data should be used [16]. On the other hand, it must be emphasised that the use of a regression method with zero as the intercept can introduce a bias and therefore is not recommended. The residuals ($Y_{obs.}-Y_{est.}$) from the three calibration curves fitted by regression are then calculated and plotted as a function of the concentrations. Afterwards, the default model of the assumed response function can be evaluated by means of the plot of residuals around the fitted calibration curve. However, outliers must be removed before taking any decision about the model fit. That is why the detection of potential deviating values by visual analysis of the plot of residuals is first proposed. Such values related to any kind of analytical problem can be directly rejected. If no experimental problem or transcription error can be directly identified, the analyst can possibly use a statistical approach (e.g. Dixon's test [5,21], Grubbs' test [8,21]) because a sufficient number of replicates (n=12) is available at each concentration level. Indeed, more than four and six replicates are required by the single and paired Grubbs outlier test, respectively. In the case of the latter statistical approach, first one screens for values that are significant at the 5% significance level (α=5%). If an analytical reason can be found, the values may be discarded. These values should otherwise be retained for the calculations unless they are significant at α=1% [8]. It must however be emphasised that such a rejected value which is not apparently related to problems during sample preparation can be inherent to the method and may occur again. Therefore it is preferable to investigate its potential origin and try to solve the problem than to reject such a value on a statistical basis. When the assumed calibration model has been re-calculated without the identified outliers (see Fig. 4), the visual analysis of the plot of residuals allows the analyst to look for previously undetected linearity problems. In such a case a mathematical transformation can be applied.

Afterwards, the correctness of the applied regression model can also be investigated by means of the plot of residuals. To better evaluate this model when several stock solutions with weighing differences are performed (step A3), analytical responses have to be first adjusted to the mean concentration at each level, i.e. the responses obtained for each series must be transformed in order to have them aligned on the mean concentration.

To accept the regression model such as ordinary least square model, the pattern of residuals must be homogeneous (homoscedasticity), i.e. the variance of the residuals must be constant for all concentration levels, over the whole calibration range investigated (step A4). If a mathematical transformation has been previously envisaged for the response, the variance of the responses must be estimated from the transformed responses and not from the original ones. Indeed, the transformation of the responses modifies the corresponding variances. It should be noted that a transformation of the original response could in some cases eliminate the heterogeneity of the variances (heteroscedasticity) in the data. For example, when the variance of the responses increases with increasing concentration, the logarithmic transformation will make the variances homogeneous over the whole range. Even if the visual analysis of the plot of residuals allows inspection for the homogeneity of the response variances, a statistical test must be performed (Levene's test, Cochran's test) [25,29,31]. In case of heterogeneity of variances, a weighting method or a mathematical transformation can be applied. In order to select the most appropriate weighting factor, it is recommended to plot the response variance as a function of the concentration (see Fig. 5). The general shape of the curve allows the analyst to select an adequate weighting strategy ($1/X$ and $1/X^2$ being the most common). If the selection of a weighting factor remains ambiguous or doubtful, it is then proposed to plot the natural logarithm of the response variance as a function of the natural logarithm of the concentration. The weighting factor then becomes the inverse of the concentration at power λ ($1/X^{\lambda}$), where λ is the slope of the regression line fitted to the values at the logarithmic scale [25,29].

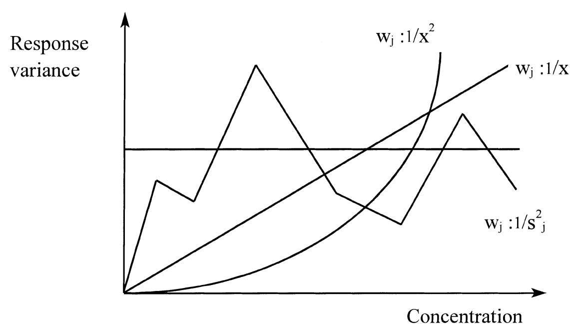

Fig. 5. Selection of the weighting factor.

The fitting of the regression model is then evaluated using an ANOVA lack-of-fit test (step A5) which determine the ratio of the residual error to the pure experimental error [25,28,29]. If the regression model is rejected, a new model is chosen on the basis of the previous graphs. Because extreme concentrations can introduce a lack of fit, it is sometimes necessary to reduce the range.

Finally, decisions are taken concerning the response function (step A6) which will be used during the validation step and then in routine.

6.1.2. Estimation of the limits of detection and quantitation [25,29,30]

Once the regression method to fit the relationship between the response (signal) and the concentration (quantity) has been selected, one can estimate the lower and upper quantitation limits of the analytical procedure in the way presented in Fig. 6 and explained below. For each of the three calibration curves (cf. Table 2) concentrations are back calculated from both analytical responses and regression equation. Afterwards, all back calculated concentrations resulting from the three calibration curves are pooled by concentration level in order to obtain not really an adequate but rather a realistic estimate of the intermediate precision variance for each selected concentration level. In the same way, mean accuracy ($R_j^{\%}$) is then determined at each concentration level

$$R_j^{\%} = \frac{1}{p}\sum_k^p \sum_i^{n_{jk}} \frac{R_{ijk}^{\%}}{n_{jk}},$$

where $R_{ijk}^{\%}$ is the recovery for each sample.

Then, for each concentration level, the unilateral confidence limits of a measure of accuracy at the 95% level are computed by introducing the estimation of standard deviation for intermediate precision as follows:

$$\mathrm{LCL}(R_j^{\%}) = \frac{\left(\bar{u}_{\cdot j\cdot} - t\left(0.1, \sum_k^p n_{jk} - p\right) \cdot s_{\mathrm{IP}}\right)}{X_J} \times 100,$$

$$\mathrm{UCL}(R_j^{\%}) = \frac{\left(\bar{u}_{\cdot j\cdot} + t\left(0.1, \sum_k^p n_{jk} - p\right) \cdot s_{\mathrm{IP}}\right)}{X_J} \times 100,$$

where for each concentration level, 0.1=10% significance level (α=10%) in the table; $\bar{u}_{\cdot j\cdot}$ is the mean estimated concentration; $R_j^{\%}$ the mean recovery (%); s_j(IP) the standard deviation for intermediate preci-

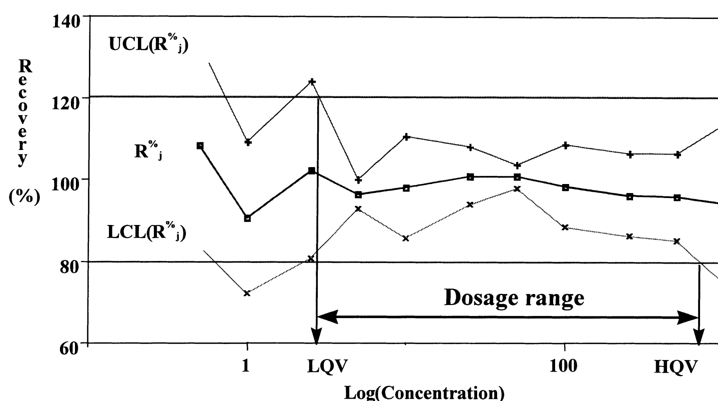

Fig. 6. Accuracy profile and estimation of the limit(s) of quantitation. LQV: lowest quantifiable value; HQV: highest quantifiable value; $R_j^{\%}$: mean recovery at each concentration level (%); LCL($R_j^{\%}$): lower confidence limit at the 95% level of the mean recovery (%); UCL($R_j^{\%}$): upper confidence limit at the 95% level of the mean recovery (%).

sion; X_J the introduced concentration; $\mathrm{LCL}(R_j^{\%})$ the lower confidence limit at the 95% level of the mean recovery (%); $\mathrm{UCL}(R_j^{\%})$ is the upper confidence limit at the 95% level of the mean recovery (%).

If for a specific concentration level the confidence limits of the mean percentage recovery are not included in the acceptance interval of 80–120%, this level of concentration must be excluded and the concentration range reduced. In contrast with what is reported in the literature [5], it is not examined here whether the 100% value is included within the confidence limits. The reason for this is that when poorly precise methods are used or when the sample size is too small, it is likely that unacceptable analytical methods may actually be accepted [19,32]. The present approach to select the LOQ is consistent with the recommendations of the Washington Conference [16]. Indeed, if the confidence limits are not greater than ±20%, then the point estimates for accuracy are obviously smaller than ±15%. In addition, this approach permits one to separate the total measurement error as defined in the guidelines (15% and 20% values) into its systematic and random elements and the minimisation of the risk of further problems when the method will be applied in routine, namely the rejection of many analytical runs.

It is also recommended in the SFSTP guide to represent graphically one of these measures of accuracy and its confidence limits as a function of the concentration as shown in Fig. 6. In this way, the lower and upper limits of quantitation can be easily defined. If the resulting values reduce the calibration range, the regression model must be determined again after rejection of the identified value(s).

6.1.3. Absolute recovery [23,24,29,30]

The recovery of the analyte from the biological matrix is an important factor in the validation of bioanalytical methods. That is why the determination of the absolute recovery of the analytical method on the whole calibration range is suggested at the end of the pre-validation. Absolute recovery is calculated for each selected concentration level using the ratio of each curve response from the biological calibration to the mean corresponding response from the non-biological calibration curve (see Table 2). Afterwards, coefficients of variation for the recovery at each concentration level are evaluated in order to demonstrate the reproducibility of the extraction procedure throughout the complete concentration range. Indeed, the reproducibility of the procedure recovery seems to be more essential than the percentage of recovery even if it is recognised that a poor recovery may result in poor analytical sensitivity. Once the reproducibility and the steadiness of the recovery have been established over the whole calibration range, the total recovery of the analytical procedure and its confidence interval can be determined in order to obtain a global estimate of this validation criterion.

6.1.4. Number of experiments to perform in the validation phase

At the end of the pre-validation phase, the opportunity is given to the analyst to determine the optimum number of experiments to be performed, i.e. the number of series and number of replicates by series needed in the validation phase. As can be seen from Table 5, these optimum numbers of experiments is

Table 5
Recommended number of series and replicates according to the pre-validation results

$CV_j(g)$	Number of series	$CV_j(r)$				
		4%	5%	6%	7%	8%
		Number of replicates				
4%	3	4	4	5	6	–
	4	4	4	4	5	9
	5	4	4	4	5	5
5%	3	4	4	4	5	–
	4	4	4	4	6	–
	5	4	4	4	5	8
	6	4	4	4	4	5
6%	3	4	4	6	10	–
	4	4	4	6	7	–
	5	4	4	5	7	–
	6	4	4	5	5	6
7%	3	6	8	–	–	–
	4	4	4	6	–	–
	5	4	4	5	7	–
	6	4	4	5	7	9
8%	4	9	–	–	–	–
	5	6	8	–	–	–
	6	4	5	8	–	–

$CV_j(r)$: repeatability coefficient of variation; $CV_j(g)$: between-series coefficient of variation.

based on the precision results obtained in the pre-validation phase. Indeed, performing too few experiments could lead to the rejection of an acceptable analytical method, giving results outside the acceptance limits of the Washington guidelines [16]. Conversely, a too high number of experiments, leading to an excessive power, will make the validation phase longer than necessary. Between these two extremes, there is an optimum number of experiments that can be calculated (see Table 5). This table was obtained by simulation at the 5% level of significance, assuming that the expected recovery will not be greater than 102% of the theoretical concentration. Optimum numbers for the series (p^*) and the replicates (n^*) are proposed as a function of the values of the precision estimates (repeatability coefficients of variation $CV_j(r)$ and between-series coefficients of variation $CV_j(g)$) obtained at each concentration level in the pre-validation phase. The higher set of series (p^*) and replicates (n^*), obtained for the different levels of concentration (j) will be used for the validation phase, i.e. $(p^*, n^*) = \max(p_j^*, n_j^*)$.

For coefficient of variation values smaller than those mentioned in Table 5, a minimum of three series and four replicates at each concentration level (using at least three concentration levels: a level close to the lowest quantifiable value, a medium level and a level close to the highest quantifiable value) must be envisaged for the validation phase. For the combination of values for which no number is proposed in the table, it is recommended to continue the development of the analytical method, otherwise the risk of never being able to validate the method according to the requirements of the Washington Conference guidelines [16] is high. If the expected percentage of recovery is very close to 100%, one replicate less than the number proposed in Table 5 can be envisaged.

6.2. Validation phase

At the end of the validation step, the results obtained are treated as follows:

- For each series, the calibration curve is adjusted using the model defined during the pre-validation step and the corresponding parameters are recorded.
- The fitting of the response function is confirmed and concentrations of the validation standards are calculated.
- The accuracy of the bioanalytical procedure is determined for each concentration level of validation standards. It is also possible to estimate global accuracy by controlling throughout the range the linearity of the relationship between calculated (measured) and spiked (known) concentrations [28].
- The precision is evaluated for each concentration level of validation standards.
- The precision and accuracy of the limit(s) of quantitation determined during the pre-validation step are recorded.
- The selectivity of the assay procedure is documented.

7. Conclusions

The lack of a clear experimental and statistical approach for the validation of bioanalytical methods has led scientists in charge of the development of these methods to propose a practical strategy to demonstrate and assess the reliability of chromatographic methods employed in bioanalysis. The aim of the SFSTP Commission was to provide simple to use approaches with a correct scientific background to improve the quality of the validation process. The Commission has also tried to take into account the practical constraints of the manipulations and experiments to be performed as well as the requirements from regulatory authorities. The strategy proposed in the SFSTP guide, the main aspects of which are summarised in this paper, divides the experiments to be performed after method development in two steps: the pre-validation and the validation phases. The so-called pre-validation allows the determination of the model for the standard curve, the limits of quantitation and the calibration range to be used in the validation phase. The actual validation phase is aimed at assessing the acceptance criteria usually described and recommended in the literature. In addition, the regression methods proposed in the guide include techniques which are increasingly used in bioanalytical methods (such as those involving MS or MS/MS) and which cannot be described by a simple linear regression.

Acknowledgements

The authors would like to thank Professor D.L. Massart (Vrije Universiteit Brussels, Belgium) for his constructive and useful comments.

References

[1] Current Concepts for the Validation of Compendial Assays, Pharmacopeial Forum, The United States Pharmacopeia Inc., Rockville, MD, 1986, p. 1241.
[2] Validation of Compendial Assays – Guidelines, Pharmacopeial Forum, The United States Pharmacopeia Inc., Rockville, MD, 1986, p. 4129.
[3] Explanatory Note EEC III/844/87-EN-final, August 1989.
[4] United States Pharmacopeia XXII, General Information <1225>, The United States Pharmacopeia Inc., Rockville, MD, 1990, p. 1710.
[5] J. Caporal-Gautier, J.M. Nivet, P. Algranti, M. Guilloteau, M. Histe, M. Lallier, J.J. N'Guyen-Huu, R. Russotto, STP Pharma Pratiques 2 (1992) 205.
[6] J.R. Lang, S. Bolton, J. Pharm. Biomed. Anal. 9 (1991) 357.
[7] Ph. Hubert, P. Chiap, J. Crommen, Validation of Bioanalytical Method, Validation Courses, Liège, 1993.
[8] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 17 (1998) 193.
[9] D. Dadgar, M.R. Smyth, Trends Anal. Chem. 5 (1986) 115.
[10] A.C. Metha, Talanta 34 (1987) 609.
[11] A.C. Mehta, J. Clin. Pharm. Ther. 14 (1989) 465.
[12] A.C. Causey, H.M. Hills, L.J. Phillfis, J. Pharm. Biomed. Anal. 8 (1990) 625.
[13] A.R. Buick, M.V. Doig, S.C. Jeal, G.S. Land, R.D. McDowall, J. Pharm. Biomed. Anal. 8 (1990) 629.
[14] H.T. Karnes, C. March, J. Pharm. Biomed. Anal. 9 (1991) 911.
[15] H.T. Karnes, G. Shiu, V.P. Shah, Pharm. Res. 8 (1991) 421.
[16] 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.
[17] P. Arnoux, R. Morrison, Xenobiotica 22 (1992) 757.
[18] H.T. Karnes, C. March, Pharm. Res. 10 (1993) 1420.
[19] C. Hartmann, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 12 (1994) 1337.
[20] D. Dadgar, P.E. Burnett, M.G. Choc, K. Gallicano, J.W. Hooper, J. Pharm. Biomed. Anal. 13 (1995) 89.
[21] H.H. Blume, K.K. Midha (Eds.), Bio-International 2, Post Conference Satellite Symposium, Update on Analytical Method Validation, Medpharm, Stuttgart, 1995, p. 319.
[22] K.A. Selinger, J. Pharm. Biomed. Anal. 13 (1995) 1427.
[23] D. Dadgar, P.E. Burnett, J. Pharm. Biomed. Anal. 14 (1996) 23.
[24] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, D.L. Massart, Anal. Chim. Acta 338 (1997) 19.
[25] E. Chapuzet, N. Mercier, S. Bervoas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, Ph. Hubert, Ph. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, STP Pharma Pratiques 7 (1997) 169.
[26] Text on validation of analytical procedures: definitions and terminology (Q2A), Tripartite International Conference on Harmonisation (ICH) Text, ICH Tech. Coordination, London, 26 October 1994.
[27] Validation of analytical procedure: methodology (Q2B), Tripartite International Conference on Harmonisation Text (ICH 3), ICH Tech. Coordination, London, 29 November 1995.
[28] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, in: B.G.M. Vandeginste, L. Kaufman (Eds.), Data Handling in Science and Technology, volume 2, Chemometrics: A Textbook, Elsevier, Amsterdam, 1988.
[29] E. Chapuzet, N. Mercier, S. Bervoas-Martin, B. Boulanger, P. Chevalier, P. Chiap, D. Grandjean, Ph. Hubert, Ph. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, STP Pharma Pratiques 8 (1998) 81.
[30] P. Chiap, Ph. Hubert, B. Boulanger, J. Crommen, Anal. Chem. Acta 391(2) (1999) 227–238.
[31] H. Levene, Robust test for equality of variances, in: I. Olkin (Ed.), Contributions to Probability and Statistics, Stanford University Press, 1960, p. 278.
[32] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Vankeerberghen, D.L. Massart, Anal. Chem. 67 (1995) 4491.