Journal of Pharmaceutical and Biomedical Analysis 55 (2011) 848–858

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Review

Advances in validation, risk and uncertainty assessment of bioanalytical methods

E. Rozet[a,*,1], R.D. Marini[a], E. Ziemons[a], B. Boulanger[b], Ph. Hubert[a]

[a] Analytical Chemistry Laboratory, CIRM, Institute of Pharmacy, University of Liège, Liège, Belgium
[b] Arlenda s.a., Liège, Belgium

ARTICLE INFO

Article history:
Received 20 October 2010
Received in revised form 9 December 2010
Accepted 10 December 2010
Available online 21 December 2010

Keywords:
Bioanalytical method validation
Results accuracy
Standard curve
Incurred sample reproducibility
Measurement uncertainty

ABSTRACT

Bioanalytical method validation is a mandatory step to evaluate the ability of developed methods to provide accurate results for their routine application in order to trust the critical decisions that will be made with them. Even if several guidelines exist to help perform bioanalytical method validations, there is still the need to clarify the meaning and interpretation of bioanalytical method validation criteria and methodology. Yet, different interpretations can be made of the validation guidelines as well as for the definitions of the validation criteria. This will lead to diverse experimental designs implemented to try fulfilling these criteria. Finally, different decision methodologies can also be interpreted from these guidelines. Therefore, the risk that a validated bioanalytical method may be unfit for its future purpose will depend on analysts personal interpretation of these guidelines. The objective of this review is thus to discuss and highlight several essential aspects of methods validation, not only restricted to chromatographic ones but also to ligand binding assays owing to their increasing role in biopharmaceutical industries. The points that will be reviewed are the common validation criteria, which are selectivity, standard curve, trueness, precision, accuracy, limits of quantification and range, dilutional integrity and analyte stability. Definitions, methodology, experimental design and decision criteria are reviewed. Two other points closely connected to method validation are also examined: incurred sample reproducibility testing and measurement uncertainty as they are highly linked to bioanalytical results reliability. Their additional implementation is foreseen to strongly reduce the risk of having validated a bioanalytical method unfit for its purpose.

Contents

* Corresponding author at: Laboratory of Analytical Chemistry, Institute of Pharmacy, Université de Liège, CHU, B 36, B-4000 Liège, Belgium. Tel.: +32 4 3664320; fax: +32 4 3664317.
E-mail address: eric.rozet@ulg.ac.be (E. Rozet).
[1] F.R.S.-FNRS Postdoctoral Researcher (Belgium).

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.12.018

1. Introduction

Analytical method validation is a mandatory step to evaluate the ability of developed methods to provide accurate results for their routine application. Indeed, without results of adequate quality or reliability, the critical decisions that will be made during routine application of the method will be untrustworthy leading to either over or under estimations of effect of new drugs, to inadequate monitoring of patient statues, to erroneous conclusions of clinical studies and so on. For the last 15–20 years, method validation has been the subject of many discussions. A general guidance on bioanalytical method validation was given at the 1990 AAPS/FDA workshop [1]. In 2000, new workshops addressed the validation of bioanalytical methods [2] as well as the special case of assays dedicated to macromolecules [3]. In 2001, the United States Food and Drug Administration (FDA) published its well known guidance document on bioanalytical method validation [4]. Recently in 2006, a new AAPS/FDA workshop was held in order to renew the topic of bioanalytical method validation for both small molecules and macromolecules and lead to a consensus white paper [5]. For interested readers, a definitive history of bioanalytical method validation and regulations can be found in [6].

Acknowledging the lack of guidance in Europe for the validation of bioanalytical methods, the European Medicines Agency (EMA) published a concept paper in December 2008 on this topic [7]. This concept paper relies evidently on the FDA guidance [4] which is now almost generally accepted by the biopharmaceutical industries as the gold standard method validation approach [8]. This initiative has led to an international motivation to globally discuss and harmonise bioanalytical method validation guidelines [9].

Nonetheless, even if a tentative to standardise globally bioanalytical method validation guidelines is in-progress, there is still the need to clarify the meaning and interpretation of bioanalytical method validation criteria and methodology.

Indeed, there are several risks inherent to bioanalytical methods validation. The first one resides in the different interpretations that can be made of the validation guidelines such as the FDA guidance [4] or the recent FDA/AAPS white paper [5]. Various interpretations can be made for the definitions of the validation criteria [10–12], diverse experimental designs may be implemented to try fulfilling these criteria [11,13,14], and different decision methodologies can also be interpreted from these guidelines [14–16]. Therefore, the risk that a validated bioanalytical method may be unfit for its future use will depend on subjective interpretation of these guidelines. Additionally, the general use of spiked biological matrix samples in method validation studies may also increase the risk of having validated an inadequate bioanalytical method for the analyses of real in-study samples by neglecting at least the important inter-subject variability [17,18]. Finally, even with the best validation approach applied, results obtained from bioanalytical methods are prone to uncertainties. In this context, knowing the measurement uncertainty [19] related to each analytical result may also reduce the risks of inadequate interpretations of these results when comparing them to regulatory compliance limits or product specification limits.

The objective of this review is thus to discuss and highlight several essential aspects of methods validation, not only restricted to chromatographic ones but also to ligand binding assays (LBAs) owing to their increasing role in biopharmaceutical industries. The points that will be reviewed are the common validation criteria, which are selectivity, standard curve, trueness, precision, accuracy, limits of quantification and range, dilutional integrity and analyte stability. Definitions, methodology, experimental design and decision criteria are reviewed. Two other points closely connected to method validation are also examined: incurred sample reproducibility (ISR) testing and measurement uncertainty as they are highly linked to bioanalytical results reliability. Their additional implementation is foreseen to strongly reduce the risk of having validated a bioanalytical method unfit for its purpose.

2. Validation criteria

2.1. Selectivity

The terms “selectivity” and “specificity” are often used interchangeably while their significances are different [20–26]. Selectivity is something that can be graded while specificity is an absolute characteristic. Specificity can be considered as the ultimate selectivity [20,23,27]. For this reasons, selectivity should be preferred and is the recommended terminology.

Any bioanalytical method may be subject to interferences. It is thus crucial to first document the selectivity of the analytical method. Selectivity is the documented demonstration of the ability of the bioanalytical procedure to discriminate the analyte from interfering components. It is usually defined as “*the ability of the bioanalytical method to measure unequivocally and to differentiate the analyte(s) in the presence of components, which may be expected to be present*” [4]. Typically, these might include metabolites, impurities, degradants, matrix components, etc. [4]. These interferences may arise from the constituent of the biological matrix under study. They may depend on characteristics of the individual under study, be it an animal (age, sex, race, ethnicity, etc.) or a plant (development stage, variety, nature of the soil, etc.), or they could also depend on environmental exposure (climatic conditions such as UV-light, temperature and relative humidity). In clinical studies, co-medications can also be the causes of potential interferences. For biotechnological products, the variability of the sources of raw biomaterials is also the basis of interfering compounds. Types of containers and preservatives may also be the sources of matrix effect [28].

Demonstrating and documenting method selectivity can be made through several approaches. The first one is by demonstrating the absence of detector signal in a blank matrix [1,2,14,29–35]. The actual FDA guidance for bioanalytical method validation requires the use of at least six independent sources of matrix to demonstrate methods selectivity [4]. It is however evident that all potential sources of interferences arising from the diversity of the matrix under study will not be present in the six sources studied. Therefore different research groups have recommended to analyse from 10 to 20 different sources of blank matrices [14,30]. These blank plasmas, if not naturally containing the potential interfering compounds, should be spiked with these at their maximum likely concentrations, e.g., drugs co administered in clinical studies, key metabolites, etc. Comparison of the blank plasma signal with these plasma spiked with the analyte of interest at the expected lower

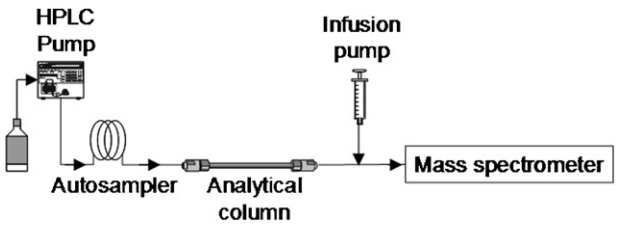

Fig. 1. Schematic for the post-column infusion analytical set-up.

limit of quantification (LLOQ) will allow to document methods selectivity.

For separative techniques, orthogonal modes of separation can help in documenting the selectivity of a procedure. Some examples of orthogonal mode of separation are reversed phase HPLC and normal phase HPLC, chromatography and electrophoresis. If, when using two orthogonal modes of separation, it is shown that the same number of peaks is present, it is a good start in documenting method selectivity. For methods implying UV detection, diode array detectors will allow to check the peak purity of key peaks in the chromatogram obtained. Mass detectors will also allow to evaluate the peak purity, usual by analysing the mass spectra at the beginning, apex and end of the chromatographic peak of interest.

Another approach to assess method selectivity is not to try demonstrating the complete absence of interferences but to allow a small amount of interfering compounds. This approach requires also to analyse up to 20 different sources of matrices, spiked with supplemental potential interfering compounds as well as the analyte under study at its lower limit of quantification (LLOQ). Selectivity is then demonstrating that, at the LLOQ, the trueness, precision and thus accuracy of the results is still deemed acceptable [14,30]. Some authors also proposed that for chromatographic methods, the peak response of interfering compounds in blank matrix at the retention time of the analyte should be at most 20% of the response of the lower limit of quantification sample [36].

For hyphenated MS(-MS) techniques it is well known that matrix effects is far from being negligible [18,37–40]. This effect of the matrix on the MS signal is usually studied during the method development by using continuous post column infusion (see Fig. 1) and monitoring decreased or increased detector signal when interfering compounds that elute from the column suppress or enhance ionisation, respectively [41–48]. This will for example orientate the choice of a sample preparation and clean-up procedure.

For method validation, a confirmation of this matrix effect could be valuably added. Matuszweski et al. [18] proposed a methodology to assess this matrix effect by measuring the signal of analyte in three different types of samples:

- Type 1: matrix free standards.
- Type 2: blank matrices of different sources extracted and spiked with the analyte after extraction.
- Type 3: same sources of blank matrices spiked before extraction.

Then they define matrix effect ME as the ratio of the response of Type 2 samples over Type 1, recovery as the ratio of the response of Type 3 over Type 2 and process efficiency as the ratio of the detector signal of Type 3 samples over those of Type 1 signal [18]. The methodology proposed by Matuszweski et al. [18] was recently improved by adding a supplementary solution by Marchi et al. [49] that allows to evaluate the extraction yield (EY) which is particularly useful in case of multianalytes determination. In this improvement, a fourth type of sample is analysed:

- Type 4: matrix free standards submitted to extraction.

EY is then defined as the ratio of the response of Type 4 samples over Type 1 [49].

Interestingly, Matuszweski et al. [18] proposed to investigate this matrix effect on at least five different sources of matrices. Finally, they insist that the combined effect of the sample matrix and of the extraction recovery variability should be evaluated on the quality of the assay results [18,50], that is results accuracy. An alternative approach in assessing and measuring matrix effect may be to compare the variability of slopes of standard curves prepared in different sources of matrices [50] or to compare these external standard curves slopes to standard addition performed on the samples [39].

Similarly, for ligand binding assays (LBAs), lack of selectivity can also induce enhancement or suppression of the signal (from binding proteins, endogenous analogs, concomitant drugs, similar class of immunoglobulin, etc.). Assessment of matrix effect in such situations could also be performed by comparing assay results of matrix samples with the drug analyte without the potential interfering compounds and with their presence at various concentration levels [51–53].

As can be seen, it is more and more essential to investigate the selectivity of the assay under validation through the use of several sources of matrices. This selectivity assessment should not be limited to qualitative aspects (adequate resolution, absence of interferences, absence of signal enhancement or suppression, etc.) but also over the quantitative performance of the method and mainly on the quality of the results generated by the assay.

2.2. Standard curve

The standard curve for a bioanalytical procedure is the existing relationship, within a specified range, between the response (signal, e.g., area under the curve, peak height, absorption) and the concentration (quantity) of the analyte in the sample. This standard or calibration curve should be described preferably by a simple monotonic (i.e. strictly increasing or decreasing) response function that gives reliable measurements, i.e. accurate results as discussed thereafter. The calibration standards used should generally be matrix-matched and analyte-matched, i.e. prepared by spiking the same blank matrix as the one that will be encountered during routine analyses with the analyte under study. The standard curve is widely and frequently confounded with the linearity criterion [10]. The linearity criterion refers to the relationship between the quantity introduced and the quantity back-calculated from the standard curve. This is different from a standard curve which refers to the relationship between the instrumental response and the concentration.

2.2.1. Type of standard curve

To systematically force a linear function is not required, often irrelevant and may lead to large errors in measured results (e.g., for bioanalytical methods using LC–MS/MS or ligand binding assays) where the linear range can be different from the working or dosing range [54,55]. A significant source of bias and imprecision in analytical measurements can be caused by the inadequate choice of the statistical model for the calibration curve [56]. The FDA guidance on Bioanalytical Method Validation issued in May 2001 [4] requires only that "*The simplest model that adequately describes the concentration–response relationship should be used*." Depending on the analytical techniques being used several functions can be used to obtain the standard curve.

The most common one is the ordinary least square (OLS) linear regression and the simplest one is the linear regression forced through the origin using a single concentration level as shown in Fig. 2. Applicability of this last one is not uncommon for bioanalyses and performs relatively well by comparison to multiple-point calibration in many instances [57]. OLS linear regression is commonly used with HPLC methods with UV, fluorescence and MS(-MS)

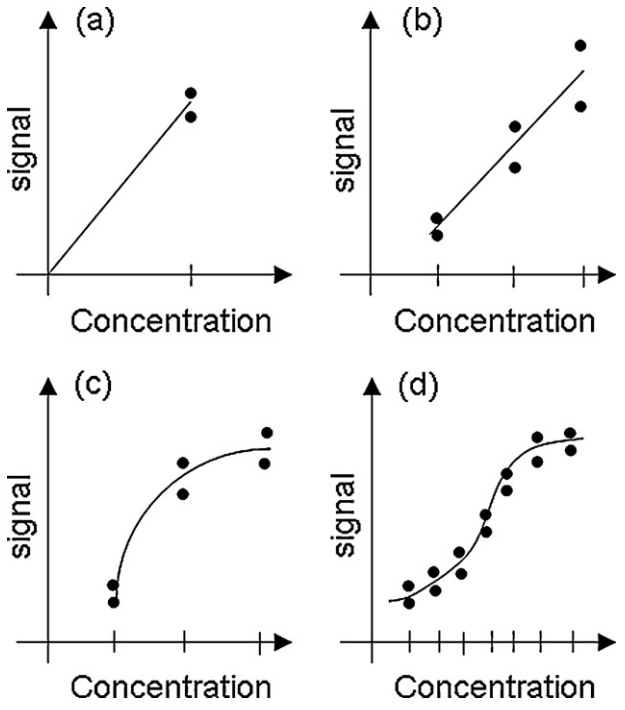

Fig. 2. Representation of different possible standard curves that may be used by bioanalytical methods. (a) A simple linear regression using a single calibration concentration level forced through the origin; (b) a simple linear regression showing heteroscedasticity; (c) a quadratic regression model; (d) a 4 parameter logistic regression model.

detectors. It is usually the first guess and the starting point for the selection of the adequate standard curve. It works well in relatively narrow range of concentration. Indeed, when the concentration range investigated by the bioanalytical method is increased, heteroscedasticity is observed leading to inadequate fit of the OLS linear regression and thus potential inaccurate analytical results [56,58,59]. An alternative is thus the more general weighted least square (WLS) linear regression that can be found useful [56,58,59]. The weight used describes the speed of the increase in signal variance when the concentration increases. Several possible common weights are:

$$\frac{1}{x^{0.5}}, \frac{1}{x}, \frac{1}{x^2}, \frac{1}{y^{0.5}}, \frac{1}{y}, \frac{1}{y^2},$$

where x is the concentration and y is the detector response.

Alternatively, transformations of the signal and/or of the concentration may be found useful to obtain an adequate standard curve [56]. The idea behind transformation is multiple. It can be in order to linearise the relationship between the signal and concentration, to obtain residues with a distribution closer to a normal one as well as reducing heteroscedasticity. It is first aimed at improving the fit of the mathematical function to the observed data. Most common transformations are logarithmic or square root ones [60–64]. Transformations of the Box–Cox family are also potential candidates [65–67]. Depending on the detector used or for large concentration range, quadratic standard curves $y = ax^2 + bx + c$ (see Fig. 2) that can be weighted or not, with or without transformations, may be used [56,64]. While combining transformations and weighting is a solution, care should be made when such combination is performed as transformations and weighting may have opposite effect on variances. Therefore, residuals should be monitored in order to diagnostic any pervert effect of these combinations.

For LBAs, standard curves are generally non linear in their parameters, and usually sigmoidal. The most used calibration model is the 4 parameters logistic (4-PL) model as illustrated in Fig. 2 [52,68,69]. A 5-PL model may also be useful when supplemental asymmetry in the relationship between concentration and detector signal is observed [69,70]. For these models, weighting may also be appropriate and are generally more complex than those used for physico-chemical assays such as chromatographic based assays. The variance relationship is usually a power of the mean response: $\sigma_y \propto \mu^\theta$ where σ_y is the standard deviation of the responses y, μ is the mean response and θ a power to be estimated from the data [68,71].

2.2.2. Design for the standard curve

The smallest and highest calibration standards of the standard curve define the maximum concentration range over which the bioanalytical method may be able to be validated as extrapolation is rarely accepted [2,14,16,72,73]. They must thus be chosen according to the specific aim of the method. Nonetheless, the validation range can also be included inside the calibration range and it is the validation standard concentration levels used that will define the definitive concentration range over which the method may be validated. For each model of standard curve there is an optimal design to obtain the most accurate back-calculated results [74]. However, an efficient rule of thumb is to place the standards at equidistance from each others. It has been shown to provide close to optimal design even for complex standard curves such as 4 parameters logistic functions [74]. At least two replicates should be analysed per concentration level, although it is essential that the same procedure should be used for method validation as it will be done during routine analyses [14,16,72]. It should be avoided to use in method validation more repetitions and more concentration levels for the standard curve than that will be used during routine application of the method. Indeed, it has to be kept in mind that method validation should mimic as much as possible the behaviour of the analytical procedure and the accuracy of the results that will be generated during its daily use. The idea of validation is also to confirm that the adequate standard curve has been selected together with the design used to generate it. Generally, it is advisable to use fewer concentration levels with more replicates than vice versa [14,74].

D-optimal designs can be found when the mathematical model used as standard curve is sufficiently known. For linear ones, the design that will provide analytical results with the best precision are two extreme calibration points [15,74]. For quadratic models, the D-optimal design is a calibration design with standards at the extreme concentration and one standard at the middle of the calibration concentration range [15,74]. The use of replicates for calibration standards allows to improve the quality of the estimates of the regression parameters. For LBAs, if it is assumed that 4-PL is the correct model for the standard curves then 5–8 calibration concentration levels should be used, with two or three replicates per levels to increase the precision of the estimators of the parameters [57,68]. Anchor calibration points may also be useful to increase the fit. For LBAs, the position of the calibration standards over the plate is also crucial [68]. While complete randomisation of all types of samples should be the ideal scenario [75], it is not practically feasible and compromise designs could be performed [68].

Finally, practical aspects may also orientate the choice of the number of calibration points to use: cost and availability of material used to prepare the calibration curve (reference substances, biomatrices, etc.), minimising the number of runs or space allocated to calibration standards to increase the throughput for incurred samples and so on. However these criteria should not impair the accuracy of the results generated by the assay.

2.2.3. Selecting the adequate standard curve

With all these possible standard curves at hand, how to select the good one? It is essential that a standard curve must be evaluated on its ability to provide accurate measurements. A significant source

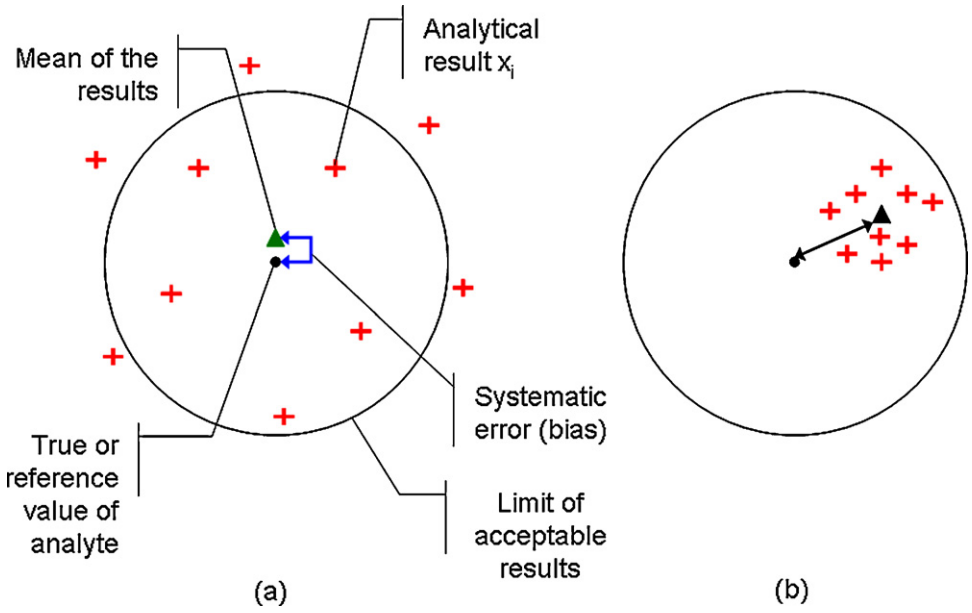

Fig. 3. Schematic illustrating that a small systematic error or bias as shown in panel (a) does not guarantee that all results are included in a specified target. Whereas panel (b) illustrates a situation with greater systematic error but nonetheless each result is included within the target.

of bias and imprecision in analytical measurements can be caused by the inadequate choice of the statistical model for the standard curve. The statistical criteria such as coefficient of determination "R^2", lack-of-fit test or any other statistical test to demonstrated quality of fit of a model are only informative and barely relevant for the objective of the assay [10,11,14,54,59,68,76–78]. The general idea to select the adequate standard curve is to demonstrate that this curve (and also the whole bioanalytical procedure) is able to provide results of adequate quality or accuracy. This is achieved by analysing independent validation standards prepared at several concentration levels spanning all or part of the calibration curve range. Then several methodologies exist. A first one is to compare the bias and relative standard deviation at each validation standard concentration level to a priori acceptance limits with or without their respective confidence intervals [4,14,15,52,53]. Finally, several authors [10,79–81] have introduced the use of the accuracy profile based on statistical tolerance intervals to decide if a calibration model will give results of sufficient quality. The models should be retained or rejected based on the accuracy of the back-calculated results, which is the final purpose of any quantitative bioanalytical method.

2.3. Accuracy: trueness (bias) + precision

2.3.1. Trueness (bias)

Trueness is related to systematic errors [10,79,82–84]. Indeed, it is expressed as the distance from the average value of a series of measurements ($\bar{x}_i$), i.e. the average of the spiked QC samples (or validation standards) at a defined concentration level and a reference value μ_T, i.e. the concentration of the spike. This concept is measured by a bias, relative bias or recovery. Spiking blank matrices is one of the most popular method to assess trueness of a method. However other approaches are available:

- Using a certified reference material.
- Comparing results with a reference method.

When possible, these alternative approaches should also be used to assess the trueness of the assay [11,12].

For the FDA Bioanalytical Method Validation document [4] trueness is mixed up with the concept of accuracy but this should be avoided [10,13,82]. It is essential to distinguish the difference between a result and an average value. The results of an analytical procedure are its very objective [10,79]. This average value only gives the central location of the distribution of results of the same true content, not the position of each individual result as shown in Fig. 3. By extension, the bias, relative bias or recovery will locate the center of the results distribution produced by the analytical procedure relative to the accepted true value.

2.3.2. Precision

Precision (or sometimes called imprecision) is expressed as standard deviation (s), variance (s^2), relative standard deviation (RSD) or coefficient of variation (CV). It measures the random error linked to the analytical procedure, i.e. the dispersion of the results around their average value [10,79,82,83].

The document of the FDA [4] distinguishes "within-run, intra-batch precision or repeatability, which assesses precision during a single analytical run", and "between-run, inter-batch precision or repeatability, which measures precision with time, and may involve different analysts, equipment, reagents, and laboratories". As can be seen in this document the same word, namely repeatability, is used twice for both component of variability which is certainly not free of confusion for the analyst. Furthermore this document considers at the same level the variability in a single laboratory or in different laboratory.

In order to evaluate correctly the two components of variability of an analytical procedure during the validation phase, the ANalysis Of VAriance (ANOVA) by concentration level investigated is recommended. As long as the design is balanced, i.e. the same numbers of replicates per series for a concentration level, the least square estimations of the variance components can be used. However when this condition is not met the maximum likelihood estimates of those components should be preferred [64].

It is important to note that the misapplications of known variance formula are still widely used and can lead to dramatic under or overestimation of the variance components [53,85,86]. For example, using the data of Table 1, when computing intermediate precision using the common variance formula neglecting the

Table 1
Data used to illustrate the misuse of variance formulas for the estimation of intermediate precision relative standard deviation (RSD).

Day 1	Day 2	Day 3
34.10	32.34	30.49
34.75	31.35	29.38
33.21	32.33	30.39
33.18	33.50	29.97
34.25	32.45	31.12

fact that the experiments were realised over several days give a RSD value of 5.2% while using the ANOVA model thus considering the series effect provides a RSD value of 6.0%. Here the misuse of variance formulas underestimates the intermediate precision thus giving a too optimistic vision of the method quantitative performances that is nonetheless influenced by external sources of variations included into the "days" effects. Another misuse would be to choose from one day the repeatability variance, rather than using all the data through an ANOVA model to estimate this repeatability variance.

As can be seen in the regulatory documents what makes the difference between within run repeatability and between run repeatability is the concept of series or runs. These series or runs are composed at least of different days with eventually different operators and/or different equipments. A run or series is a period during which analyses are executed under repeatability conditions that remain constant. In this context, one can consider that repeatability is a natural or intrinsic variation of the analytical procedure while maintaining constant all other modifiable analytical conditions herein cited. The rational to select the different factors which will compose the runs/series is to mimic conditions that will be encountered during the routine use of the analytical procedure. It is evident that the analytical procedure will not be used only one day, may be with more than one operator and over different equipments. Thus, different factors representing the normal variability of the procedure that will be used during the routinely performed analysis will be introduced in the validation protocol, leading to a representative estimation of the variability of the analytical procedure. Statistical experimental design can be defined in order to optimise the number of runs or series to account for the main effects of these factors with a cost effective analysis time.

2.3.3. Accuracy

In the document FDA Bioanalytical Method Validation [4], accuracy is defined as ". . .the closeness of mean test results obtained by the method to the true value (concentration) of the analyte. (. . .) The mean value should be within 15% of the actual value except at LLOQ, where it should not deviate by more than 20%. The deviation of the mean from the true value serves as the measure of accuracy." As already mentioned in the previous sections, this definition corresponds to the analytical method trueness. For bioanalytical methods, earlier reviews have already stressed the problem of the crucial difference between the definition of accuracy and trueness [10–12,14,79,87].

For most uses it does not matter whether a deviation from the true value is due to random error (lack of precision) or to systematic error (lack of trueness), as long as the total quantity of error remains acceptable. Thus, the concept of total analytical error or accuracy which is the combination of random and systematic errors is essential. Furthermore, every analyst wants to ensure that the total amount of error of the method will not affect the interpretation of the test result and compromise the subsequent decision [29,53,54,79,16,88]. Decision based on the separate evaluation of the trueness and precision criteria cannot achieve this [89]. Only evaluation of the accuracy of the results which takes into account the total error concept, gives guarantees to both laboratories and regulatory agencies on the ability of the method to achieve its final purpose. The computational details of total error is out of the scope of the present paper, however the interested reader is referred to the following documents that detail all the computational steps to obtain total error: [64,80,81].

2.3.4. Experimental design for accuracy: trueness + precision

Accuracy and thus trueness and precision are estimated from the analysis of Validation Standards (VS; sometimes these samples are also called Quality Control – QC-samples, see e.g., [4]). These VS samples should match the matrix of the real samples that will be studied in routine application of the bioanalytical method. Indeed, they have to mimic as close as possible the incurred samples. Ideally, the blank matrix used to prepare the VS samples should be different than the one that has been used to prepare the standard curve, as this will be the case during the daily life of the procedure.

The extreme smallest and maximum VS levels should be selected within or equal to the smallest and highest calibration standard as it is generally not acceptable to perform extrapolation [4,5]. Also at least a third VS level must be used, near the middle of the validation range [4,5]. Nonetheless, more VS levels can valuably be added near the smallest VS levels if the validity of the bioanalytical procedure at the expected LLOQ is felt unsure. For instance the FDA guide [4] proposes to use 4 VS levels to validate analytical methods: one at the expected LLOQ, one at three times this expected LLOQ, a middle one and a maximum one. For LBA assays at least five concentrations levels are recommended [52,53].

The FDA guide as well as the recent white paper requests that at least five replicates should be analysed at each VS level [4,5]. However, no recommendation is proposed for the number of runs to perform to assess inter-run precision. A minimum of three runs should be used in order to compute an almost meaningful between-run standard deviation. However, many authors suggested increasing the number of different runs. Some proposed 8 runs [14] others at least 5 runs [32,52] or 6 runs [53]. The NCCLS proposed a design involving 2 repetitions for each concentration level during 20 runs performed during 20 days [90]. Two replicates is also the minimum proposed for LBA [53]. It has to be reminded that each run can be or should be performed by different analyst, equipments or reagents when relevant, with different sources of blank matrices and during different days to address the normal sources of variability expected during the routine application of the assay. Failing to do so will artificially provide an over optimistic estimation of precision and hence of results accuracy.

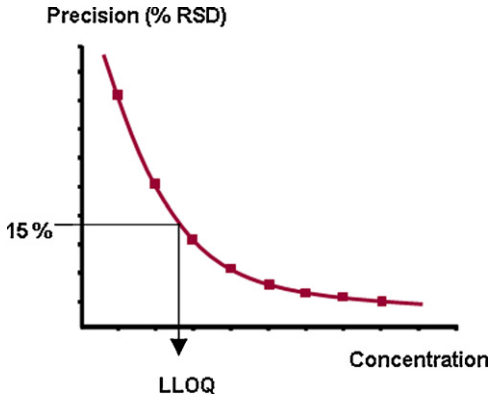

Fig. 4. Example of precision profile, showing the evolution of the bioanalytical method precision with respect to analyte concentration. The lower limit of quantification (LLOQ) is for example the smallest concentration with a RSD not exceeding 15%.

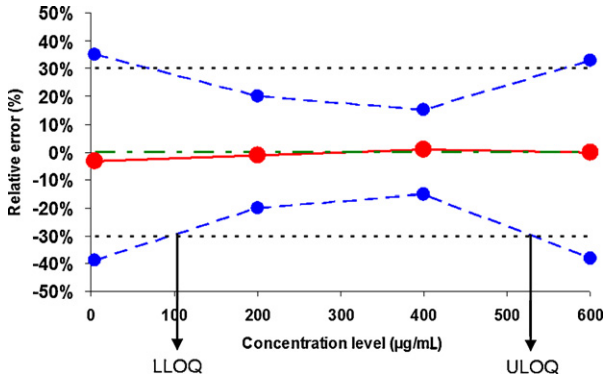

Fig. 5. Schematic accuracy profile obtained from a method validation. The plain line represents the relative bias or trueness. The dashed lines represent total error or accuracy. The dotted curves are the acceptance limit set at ±30%. The lower and upper limit of quantification (LLOQ and ULOQ) are the smallest and highest concentration levels where total error and thus results accuracy does not exceed 30% error, respectively.

2.4. LOQ and range

The FDA Bioanalytical Method Validation document defines the lower limit of quantification and the upper limit of quantification which are *the lowest* (*highest*) *amount of an analyte in a sample that can be quantitatively determined with precision and accuracy* [4]. Several approaches exist in order to estimate the lower limit of quantification (LLOQ).

A first approach is based on the well known signal-to-noise (S/N) ratio approach. A 10:1 S/N is considered to be sufficient to discriminate the analyte from the background noise [91]. The other approaches are based on the "*Standard Deviation of the Response and the Slope*". The computation for LLOQ is:

$$\text{LLOQ} = \frac{10\sigma}{S}$$

where σ is the standard deviation of the response and S = the slope of the calibration curve.

Another approach to estimate the LLOQ is to plot the RSD versus concentrations close to the expected LLOQ. As shown in Fig. 4, the LLOQ is then the concentration for which the RSD corresponds to a maximum target RSD value [92].

These methods of estimating the LLOQ are well known for their important pitfalls [10,12,93,94]. None of these approaches fulfill the definition of the LLOQ. Indeed, they do not estimate the LLOQ by assessing the acceptability of trueness and precision or accuracy at this level. Therefore, if using such approaches to define the LLOQ, further experiments should be performed to demonstrate that accuracy of the results at this level is acceptable, or at least showing that precision and trueness of the method at this level are adequate.

A last approach which allows an adequate estimation of the LLOQ as well as of the upper limit of quantification (ULOQ) is the use of the accuracy profile approach as illustrated in Fig. 5 [10,74,79,88,95,96]. Indeed this methodology demonstrates that the total error of the results is known and acceptable at these concentration levels, i.e. both an acceptable level of systematic and random errors. The quantification range is "*the range of concentration, including ULOQ and LLOQ, that can be reliably and reproducibly quantified with accuracy and precision through the use of a concentration-response relationship*" [4]. Thus, the quantification range is deducted from all the experiments realised during method validation.

2.5. Dilutional integrity

When the analyte is present in the sample at concentrations above the ULOQ, the samples should be diluted in order to bring back the sample concentration within the valid concentration range. Several dilution procedures and factors should therefore be validated [4,5,53]. It should be demonstrated that this additional sample manipulation does not impair the trueness and precision of the method and thus the accuracy of the results obtained in such a way. To demonstrate this, blank matrix samples are spiked at concentration above the ULOQ and then submitted to the dilutional procedure. Then, the same methodology to evaluate methods precision and trueness or results accuracy should be used to assess the dilutional integrity.

2.6. Stability

Assessment of analyte stability is a prerequisite to obtain a reliable bioanalytical method. It is thus essential to evaluate analyte stability during the method validation step at least. The FDA guide on bioanalytical method validation [4] as well as the recent AAPS/FDA white paper [5] require evaluating analyte stability at different stages. They include the evaluation of the analyte stability in the biological matrix through several freeze–thaw cycles, bench-top stability (i.e. under the conditions of sample preparation), long term stability at for example −20 °C or −70 °C (i.e. during storage conditions of the samples) and stability of samples on the auto-sampler.

Generally, stability should be evaluated at least at two concentration levels, using blank biological matrix matched samples spiked at a low and high concentration level [4,5,53]. It should be assessed in each matrix and species in which the analyte will be quantified. Bench-top stability should be representative of the conditions over which study samples will be prepared, for instance at room temperature and/or refrigerator temperature when relevant. Freeze–thaw stability is generally evaluated during three freeze–thaw cycles and should always mimic the way samples will be handled during daily application of the bioanalytical method.

Stability is evaluated by using freshly prepared calibration standards to compute recovery either from the nominal concentration of the samples assessed at the beginning of the stability study or by using quality control samples prepared at each time point of analysis at the same nominal concentration level as the stability samples [97]. At each time point, analysis of at least six replicates for both the stability samples and the VS (or QC) ones are generally recommended [97]. Common rule to accept the stability of analytes is to monitor the ratios between stability samples and VS (or QC) samples and to verify that two thirds of individual stability samples ratio falls inside specified acceptance limits such as 80–120% or 85–115% [4,5,53]. Alternatively the mean of the recoveries together with their confidence interval should be included within the acceptance limits. Finally, another approach modelling the analyte instability through a simple linear regression of the recoveries versus storage times together with its confidence interval has been proposed that is then compared to the pre-specified acceptance limits as illustrated in Fig. 6 [97]. These last two approaches are statistical equivalence tests that provide an adequate framework for stability assessment by efficiently controlling the risk of falsely concluding stability [97].

3. Incurred sample reproducibility

Although the analysis of incurred samples (or real study samples) is not formally required in the FDA guidance of bioanalytical method validation [4], FDA requires more and more to assess

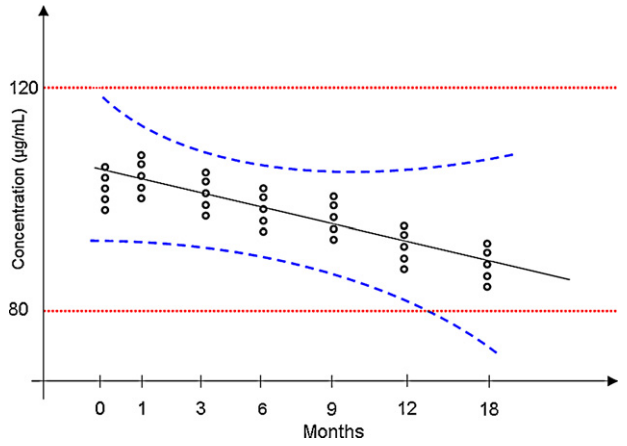

Fig. 6. Schematic illustrating a stability profile. The continuous line is the linear regression model showing the evolution of the stability of the analyte. The two dashed lines are the two-sided 95% confidence interval and the two dotted lines are the acceptance limits given at 80 and 120 µg/ml. The open circles represent the calculated concentrations obtained from the stability samples. On this example, stability is acceptable up to around 12 months, where the lower confidence interval crosses the lower acceptance limit.

the variability of bioanalytical methods in real situations due to high discrepancies observed during audits or reviews [98]. This paramount variability observed when re-analysing incurred samples during different runs impairs the validity status of the bioanalytical methods. Therefore, recent discussions and white papers promote the inclusion of incurred sample reproducibility (ISR) during the bioanalytical method validation [5,69,97]. Indeed one of the limitations of bioanalytical methods validation step is that they are usually performed using spiked samples that only more or less mimic the real in-study samples. Therefore, these spiked samples may not effectively provide adequate estimates of the bioanalytical method quantitative performances such as precision and trueness obtained with study or incurred samples. The aim of ISR testing is to demonstrate that the bioanalytical method will produce consistent results from study samples when re-analysed on a separate occasion. The actual common acceptance criteria adopted by biopharmaceutical laboratories is the 4–6–λ rule, with $\lambda = \pm 20\%$ acceptance limits for small molecules and $\lambda = \pm 30\%$ for large ones [99,100]. This means that two-thirds of the re-analysed incurred samples must lie within ±20% (or ±30%) of the original result.

To test ISR, it is recommended to select re-tested samples randomly in order to cover the valid range of concentrations and from sufficient different subjects to include the most inter-subject matrix variability [99–101]. Rather than defining a fixed percentage of samples of a study to re-analyse, the sample size should be defined in order to reduce the risk to abusively reject ISR testing and to increase the probability to accept adequately ISR testing. Simulations, such as those proposed by Hoffman [100] could be used to achieve this. For instance they showed that a sample size of 40 incurred samples allowed to reject 90% of times truly non reproducible methods. For such a sample size the risk to reject truly reproducible method with true RSD of 10% is about 1% [100]. Incurred samples should be analysed over as many analytical runs as practicable within a laboratory in order to reduce the risk of erroneously failing ISR testing [100].

While the 4–6–λ rule is a non statistical decision rule with very poor performances [102], other rigorous statistical methodologies are available. These approaches strictly control the risk of incorrectly accepting truly non-reproducible bioanalytical methods. They are the tolerance intervals and the probability approaches [81,100,103–105].

The tolerance interval approach will define a region where the differences "repeated sample result minus original result" will fall with a probability β. This probability could be set at $\beta = 66.7\%$ in the case of ISR testing [100]. This interval is then compared to acceptance limits. Hoffman [100] has proposed an acceptance limits of 21.2% (see [100] for rational). Details of statistical computations for tolerance interval may be found in: [64,81,100,104]. The probability approach estimates directly the probability to obtain the differences "repeated sample result minus original result" within pre-defined acceptance limits and compare it to a minimum quality level, for instance 66.7%. Here also the interested reader is referred to the following references for computational details: [100,104,105].

4. Measurement uncertainty

Uncertainty defined by the ISO Guide on Uncertainty of Measurement (GUM) [19,106] is "a parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand". This parameter is usually a standard deviation, a given multiple of it, or the width of a confidence interval. This uncertainty expanded by a factor, 2 for e.g., is interpreted as an interval in which the true value of the result of a measurement resides with a defined probability. For instance, when the coverage factor is 2, there is about 95% probability that the true measurement result is within this interval, assuming a normal distribution of the results. More detailed explanations about measurement uncertainty can be found in various guides or articles [107–110]. In few lines, the initial idea is to perform a bottom-up approach by determining all the individual sources of uncertainty linked to the final results, realising an uncertainty budget and combining all these uncertainties through the law of error propagation. Each uncertainty contribution is expressed as a standard deviation either based on experimental data (type A evaluation) or based on scientific judgment using a priori selected distributions (type B evaluation). This bottom-up procedure has the advantage to define and measure all sources of uncertainty and allows the improvements of the analytical procedure by reducing the most important sources of uncertainties. It also allows a detailed understanding of the analytical process involved. However, direct application of the GUM in bioanalytical laboratories is found tedious and laborious and opposed to the competitiveness required for bioanalytical laboratories to survive. Therefore other approaches, called top-down approaches have been proposed [107,111–113]. They use experiments coming from various studies such as trueness, precision, validation, robustness or inter-laboratories ones, or even combinations of them. These top-down approaches allow providing measurement uncertainty estimates to bioanalytical laboratories in a cost-effective manner. If the bioanalytical method validation is well designed, including several operators, performed on different equipments and on different days of analyses, then the estimation of measurement uncertainty should be adequate. Some authors have further demonstrated the mathematical link between statistical tolerance intervals used in the accuracy profile validation methodology and measurement uncertainty [95]. It should also be noted that a prerequisite from the GUM is that the analytical method should be totally free of systematic error, and if it is not the case results should be corrected for it [19]. This can thus orientate the selection of the calibration curve by choosing the one that provides no bias or the least bias. Then, having eliminated all sources of systematic influences, the uncertainty can be expressed as a standard deviation. While suppressing all sources of systematic errors is not always possible or pragmatic, then method bias should be included into the uncertainty budget [107,114–116].

Measurement uncertainty defines thus a region around a routine result obtained from an incurred sample where it is highly probable to observe the real unknown true result. Measurement uncertainty may have a good place in the daily application of bioanalytical methods by assessing whether results obtained from incurred samples are compliant with specification limits, or legal thresholds [109,117–120]. However measurement uncertainty is rarely used in practice although for laboratories aiming at achieving ISO 17025 [121] or ISO 15189 [122] accreditations, it is a mandatory step.

5. Conclusion

Bioanalytical methods must be validated to objectively demonstrate their fitness for their intended use. This review aimed at reporting and clarifying the interpretations and methodologies employed to assess the main validation criteria for bioanalytical methods that are selectivity, standard curve, trueness, precision, accuracy, limits of quantification and range, dilutional integrity and analyte stability. Two other equally important elements to reduce the risks to validate an unfit bioanalytical method were also discussed. The first one, incurred sample reproducibility testing, is recommended to assess the consistency of assays over real routine samples and therefore increase the reliability of the analytical results. The second one, measurement uncertainty, reminds analysts that an analytical result is only an estimation of the real concentration of a sample. Therefore measurement uncertainty quantifies the doubt about a result in order to help the analyst to make reliable decisions, knowing the risks of false compliance and non-compliance.

Acknowledgements

A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is gratefully acknowledged as well as the Walloon Project PPP (Convention OPTIMAL DS No. 917007) to R.D. Marini.

References

[1] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman, S. Spector, Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies, Pharm. Res. 9 (1992) 588–592.
[2] 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, Bioanalytical method validation—a revisit with a decade of progress, Pharm. Res. 17 (2000) 1551–1557.
[3] K.J. Miller, R.R. Bowsher, A. Celniker, J. Gibbons, S. Gupta, J.W. lee, S.J. Swanson, W.C. Smith, R.S. Weiner, Workshop on bioanalytical methods validation for macromolecules: summary report, Pharm. Res. 18 (2001) 1373–1383.
[4] Guidance for Industry: Bioanalytical Method Validation, US Department of Health and Human Services, US Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville, May 2001.
[5] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, M.J. Rose, J. Sailstad, V.P. Shah, J.P. Skelly, P.G. Swann, R. Weiner, Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays, AAPS J. 9 (2007) E30–E42.
[6] V.P. Shah, The history of bioanalytical method validation and regulation: evolution of a guidance document on bioanalytical methods validation, AAPS J. 9 (2007) E43–E47.
[7] European Medicines Agencey (EMA) Concept paper/recommendations on the need for a (CHMP) guideline on the validation of bioanalytical methods, EMA/CHMP/EWP/531305/2008 (2008), www.ema.europa.eu/pdfs/human/ewp/53130508en.pdf.
[8] G. Smith, Review of the 2008 European Medicines Agency concept paper on bioanalytical method validation, Bioanalysis 1 (2009) 877–881.
[9] P. van Amsterdam, B. Lausecker, S. Luedtke, Ph. Timmerman, M. Brudny-Kloeppel, Towards harmonized regulations for bioanalysis: moving forward!, Bioanalysis 4 (2010) 689–691.
[10] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, Ph. Hubert, Analysis of recent pharmaceutical regulatory documents on analytical method validation, J. Chromatogr. A 1158 (2007) 111–125.
[11] D. Stöckl, H. D'Hondt, L.M. Thienpont, Method validation across the disciplines—critical investigation of major validation criteria and associated experimental protocols, J. Chromatogr. B 877 (2009) 2180–2190.
[12] P. Araujo, Key aspects of analytical method validation and linearity evaluation, J. Chromatogr. B 877 (2009) 2224–2234.
[13] F.T. Peters, O.H. Drummer, F. Musshoff, Validation of new methods, Forensic Sci. Int. 165 (2007) 216–224.
[14] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, Validation of bioanalytical chromatographic methods, J. Pharm. Biomed. Anal. 17 (1998) 193–218.
[15] B. Boulanger, P. Chiap, W. Dewé, J. Crommen, Ph. Hubert, An analysis of the SFSTP guide on validation of chromatographic bioanalytical methods: progresses and limitations, J. Pharm. Biomed. Anal. 32 (2003) 753–765.
[16] B. Boulanger, E. Rozet, F. Moonen, S. Rudaz, Ph. Hubert, A risk-based analysis of the AAPS conference report on quantitative bioanalytical methods validation and implementation, J. Chromatogr. B 877 (2009) 2235–2243.
[17] C.A. James, M. Breda, E. Frigerio, Bioanalytical method validation: a risk-based approach? J. Pharm. Biomed. Anal. 35 (2004) 887–893.
[18] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC–MS/MS, Anal. Chem. 75 (2003) 3019–3030.
[19] Guide to the Expression of Uncertainty in Measurement, ISO, Geneva, 1995.
[20] J. Vessman, Selectivity or specificity? Validation of analytical methods from the perspective of an analytical chemist in the pharmaceutical industry, J. Pharm. Biomed. Anal. 14 (1996) 867–869.
[21] J. Vessman, Selectivity—the hallmark of an analytical chemist: the current situation in the analytical sciences, Accred. Qual. Assur. 6 (2001) 522–527.
[22] B.-A. Persson, J. Vessman, Generating selectivity in analytical chemistry to reach the ultimate-specificity, Trends Anal. Chem. 17 (1998) 117–119.
[23] B.-A. Persson, J. Vessman, The use of selectivity in analytical chemistry—some considerations, Trends Anal. Chem. 20 (2001) 526–532.
[24] J. Vessman, R.I. Stefan, J.F. Van Staden, K. Danzer, W. Lindner, D.T. Burns, A. Fajgelj, H. Müller, Selectivity in analytical chemistry, Pure Appl. Chem. 73 (2001) 1381–1386.
[25] D. Alan, McNaught and Andrew Wilkinson. IUPAC Compendium of Chemical Terminology, 2nd ed., Blackwell, Oxford, 1997.
[26] H.Y. Aboul-Enein, Selectivity versus specificity in chromatographic analytical methods, Accred. Qual. Assur. 5 (2000) 180–181.
[27] WELAC Guidance Documents WG D2, Eurachem/Western European Laboratory Accreditation Cooperation (WELAC) Chemistry, 1st ed., Teddington, UK, 1993.
[28] H. Mei, Y. Hseih, C. Nardo, et al., Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: application to drug discovery, Rapid Commun. Mass Spectrom. 17 (2003) 97–103.
[29] H.T. Karnes, G. Shiu, V.P. Shah, Validation of bioanalytical methods, Pharm. Res. 8 (1991) 421–426.
[30] D. Dadgar, P.E. Burnett, Issues in evaluation of bioanalytical method selectivity and drug stability, J. Pharm. Biomed. Anal. 14 (1995) 23–31.
[31] D. Dadgar, P.E. Burnett, M.G. Choc, K. Gallicano, J.W. Hooper, Application issues in bioanalytical method validation, sample analysis and data reporting, J. Pharm. Biomed. Anal. 13 (1995) 89–97.
[32] J. Wieling, G. Hendriks, W.J. Tamminga, J. Hempenius, C.K. Mensink, B. Oosterhuis, J.H. Jonkman, Rational experimental design for bioanalytical methods validation. Illustration using an assay method for total captopril in plasma, J. Chromatogr. A 730 (1996) 381–394.
[33] F. Bressolle, P.M. Bromet, M. Audran, Validation of liquid chromatographic and gas chromatographic methods. Applications to pharmacokinetics, J. Chromatogr. B 686 (1996) 3–10.
[34] R. Causon, Validation of chromatographic methods in biomedical analysis. Viewpoint and discussion, J. Chromatogr. B 689 (1997) 175–180.
[35] W. Lindner, I.W. Wainer, Requirements for initial assay validation, J. Chromatogr. B 707 (1998) 1–2.
[36] S. Bansal, A. DeStefano, Key elements of bioanalytical method validation for small molecules, AAPS J. 9 (2007) E109–E114.
[37] T.M. Annesley, Ion suppression in mass spectrometry, Clin. Chem. 49 (2003) 1041–1044.
[38] P.J. Taylor, Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray-tandem mass spectrometry, Clin. Biochem. 38 (2005) 328–334.
[39] F. Gosetti, E. Mazzucco, D. Zampieri, M.C. Gennaro, Signal suppression/enhancement in high-performance liquid chromatography tandem mass spectrometry, J. Chromatogr. A 1217 (2010) 3929–3937.
[40] A. Van Eeckhaut, K. Lanckmans, S. Sarre, I. Smolders, Y. Michotte, Validation of bioanalytical LC–MS/MS assays: evaluation of matrix effects, J. Chromatogr. B 877 (2009) 2198–2207.
[41] S. Souverain, S. Rudaz, J.-L. Veuthey, Matrix effect in LC–ESI–MS and LC–APCI–MS with off-line and on-line extraction procedures, J. Chromatogr. A 1058 (2004) 61–66.
[42] C.R. Mallet, Z. Lu, J.R. Mazzeo, A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts, Rapid Commun. Mass Spectrom. 18 (2004) 49–58.
[43] H.R. Liang, R.L. Foltz, M. Meng, P. Bennett, Ionization enhancement in atmospheric pressure chemical ionization and suppression in electrospray ionization between target drugs and stable-isotope-labeled internal stan-

dards in quantitative liquid chromatography/tandem mass spectrometry, Rapid Commun. Mass Spectrom. 17 (2003) 2815–2821.
[44] R. Dams, M.A. Huestis, W.E. Lambert, C.M. Murphy, Matrix effect in bioanalysis of illicit drugs with LC–MS/MS: influence of ionization type, sample preparation, and biofluid, J. Am. Soc. Mass Spectrom. 14 (2003) 1290–1294.
[45] C. Muller, P. Schafer, M. Stortzel, S. Vogt, W. Weinmann, Ion suppression effects in liquid chromatography–electrospray-ionisation transport-region collision induced dissociation mass spectrometry with different serum extraction methods for systematic toxicological analysis with mass spectra libraries, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 773 (2002) 47–52.
[46] W. Naidong, H. Bu, Y.L. Chen, W.Z. Shou, X. Jiang, T.D. Halls, Simultaneous development of six LC–MS–MS methods for the determination of multiple analytes in human plasma, J. Pharm. Biomed. Anal. 28 (2002) 1115–1126.
[47] F. Streit, M. Shipkova, V.W. Armstrong, M. Oellerich, Validation of a rapid and sensitive liquid chromatography-tandem mass spectrometry method for free and total mycophenolic acid, Clin. Chem. 50 (2004) 152–159.
[48] B. Fan, M.G. Bartlett, J.T. Stewart, Determination of lamivudine/stavudine/efavirenz in human serum using liquid chromatography/electrospray tandem mass spectrometry with ionization polarity switch, Biomed. Chromatogr. 16 (2002) 383–389.
[49] I. Marchi, V. Viette, F. Badoud, M. Fathi, M. Saugy, S. Rudaz, J.-L. Veuthey, Characterization and classification of matrix effects in biological samples analyses, J. Chromatogr. A 1217 (2010) 4071–4078.
[50] B.K. Matuszewski, Standard line slopes as a measure of a matrix effect in quantitative HPLC–MS bioanalysis, J. Chromatogr. B 830 (2006) 293–300.
[51] J. Lee, H. Ma, Specificity and selectivity evaluations of ligand binding assay of protein therapeutics against concomitant drugs and related endogenous proteins, AAPS J. 9 (2007) E164–E170.
[52] M. Kelley, B. DeSilva, Key elements of bioanalytical method validation for macromolecules, AAPS J. 9 (2007) E156–E163.
[53] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tacey, H. Hill, A. Celniker, Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules, Pharm. Res. 20 (2003) 1885–1900.
[54] J.W. Lee, V. Devanarayan, Y.C. Barrett, R. Weiner, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J.A. Rogers, R. Millham, P.J. O'Brian, J. Sailstad, M. Khan, C. Ray, J.A. Wagner, Fit-for-purpose method development and validation for successful biomarker measurement, Pharm. Res. 23 (2006) 312–328.
[55] UK Department of Trade and Industry, Manager's Guide to VAM, Valid Analytical Measurement Programme, Laboratory of the Government Chemist (LGC), Teddington, UK, 1998, http://www.vam.org.uk.
[56] T. Singtoroj, J. Tarning, A. Annerberg, M. Ashton, Y. Bergqvist, N.J. White, N. Lindegardh, N.P.J. Day, A new approach to evaluate regression models during validation of bioanalytical assays, J. Pharm. Biomed. Anal. 41 (2006) 219–227.
[57] F.T. Peters, H.H. Maurer, Systematic comparison of bias and precision data obtained with multiple-point and one-point calibration in six validated multianalyte assays for quantification of drugs in human plasma, Anal. Chem. 79 (2007) 4967–4976.
[58] A.M. Almeida, M.M. Castel-Branco, A.C. Falcão, Linear regression for calibration lines revisited: weighting schemes for bioanalytical methods, J. Chromatogr. B 774 (2002) 215–222.
[59] J. Tellinghuisen, Weighted least squares in calibration: the problem with using "quality coefficients" to select weighting formulas, J. Chromatogr. B 872 (2008) 162–166.
[60] 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, The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: from the Washington Conference to the laboratory, Anal. Chim. Acta 391 (1999) 135–148.
[61] P. Chiap, P. Hubert, B. Boulanger, J. Crommen, Validation of an automated method for the liquid chromatographic determination of atenolol in plasma: application of a new validation strategy, Anal. Chim. Acta 391 (1999) 227–238.
[62] G.K. Szabo, H.K. Browne, A. Ajami, E.G. Josephs, Alternatives to least squares linear regression analysis for computation of standard curves for quantitation by high performance liquid chromatography: applications to clinical pharmacology, J. Clin. Pharmacol. 34 (1994) 242–249.
[63] M.A. Castillo, R.C. Castells, Initial evaluation of quantitative performance of chromatographic methods using replicates at multiple concentrations, J. Chromatogr. A 921 (2001) 121–133.
[64] Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet, Harmonization of strategies for the validation of quantitative analytical procedures A SFSTP proposal. Part III, J. Pharm. Biomed. Anal. 45 (2007) 82–96.
[65] G.E.P. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. B 26 (1964) 211–246.
[66] E.K. Kimanani, Bioanalytical calibration curves: proposal for statistical criteria, J. Pharm. Biomed. Anal. 16 (1998) 1117–1124.
[67] E.K. Kimanani, J. Lavigne, Bioanalytical calibration curves: variability of optimal powers between and within analytical methods, J. Pharm. Biomed. Anal. 16 (1998) 1107–1115.
[68] J.W.A. Findlay, R.F. Dillard, Appropriate calibration curve fitting in ligand binding assays, AAPS J. 9 (2007) E260–E267.
[69] J.W.A. Findlay, Some important considerations for validation of ligand-binding assays, J. Chromatogr. B 877 (2009) 2191–2197.
[70] P.G. Gottschalk, J.R. Dunn, The five-parameter logistic: a characterization and comparison with the four-parameter logistic, Anal. Biochem. 343 (2005) 54–65.
[71] R.J. Carroll, D. Ruppert (Eds.), Transformation and Weighting in Regression, Chapman Hall, London, UK, 1988.
[72] M. Thompson, S.L.R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC technical report), Pure Appl. Chem. 74 (2002) 835–855.
[73] F.T. Peters, H.H. Maurer, Bioanalytical method validation and its implications for forensic and clinical toxicology—a review, Accred. Qual. Assur. 7 (2002) 441–449.
[74] B. Boulanger, W. Dewé, A. Gilbert, B. Govaerts, M. Maumy-Bertrand, Risk management for analytical methods based on the total error concept: conciliating the objectives of the pre-study and in-study validation phases, Chemometr. Intell. Lab. Syst. 86 (2007) 198–207.
[75] R. Singer, D.M. Lansky, W.W. Hauck, Bioassay glossary, Pharmacopeial Forum 32 (2006) 1359–1365.
[76] Analytical Methods Committee, Recommendations for the definition, estimation and use of the detection limits, Analyst 113 (1988) 1469–1471.
[77] S.V.C. de Souza, R.G. Junqueira, A procedure to assess linearity by ordinary least squares method, Anal. Chim. Acta 552 (2005) 25–35.
[78] J. Ermer, H.-J. Ploss, Validation in pharmaceutical analysis. Part II. Central importance of precision to establish acceptance criteria and for verifying and improving the quality of analytical data, J. Pharm. Biomed. Anal. 37 (2005) 859–870.
[79] Ph. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.-A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP Proposal—Part I, J. Pharm. Biomed. Anal. 36 (2004) 579–586.
[80] D. Hoffman, R. Kringle, A total error approach for the validation of quantitative analytical methods, J. Biopharm. Stat. 15 (2005) 283–293.
[81] D. Hoffman, R. Kringle, A total error approach for the validation of quantitative analytical methods, Pharm. Res. 24 (2007) 1157–1164.
[82] ISO 5725. Application of the Statistics-accuracy (Trueness and Precision) of the Results and Methods of Measurement—Part 1 to 6, International Organization for Standardization (ISO), Geneva, 1994.
[83] ISO 3534-1: Statistics—Vocabulary and Symbols, International Organization for Standardization (ISO), Geneva, 2006.
[84] Analytical Methods Committee, AMC Technical Brief 13, Royal Society of Chemistry, London, 2003, http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp.
[85] H. Rosing, W.Y. Man, E. Doyle, A. Bult, J.H. Beijnen, Bioanalytical liquid chromatographic method validation. A review of current practices and procedures, J. Liq. Chromatogr. Relat. Technol. 23 (2000) 329–354.
[86] C.R. Jensen, Variance component calculations: common methods and misapplications in the semiconductor industry, Qual. Eng. 14 (2002) 647–657.
[87] C. Hartmann, D.L. Massart, R.D. McDowall, An analysis of the Washington Conference Report on bioanalytical method validation, J. Pharm. Biomed. Anal. 12 (1994) 1337–1343.
[88] J.W.A. Findlay, W.C. Smith, J.W. Lee, G.D. Nordblom, I. Das, B.S. DeSilva, M.N. Khan, R.R. Bowsher, Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective, J. Pharm. Biomed. Anal. 21 (2000) 1249–1273.
[89] A. Bouabidi, E. Rozet, M. Fillet, E. Ziemons, E. Chapuzet, B. Mertens, R. Klinkenberg, A. Ceccato, M. Talbi, B. Streel, A. Bouklouze, B. Boulanger, Ph. Hubert, Critical analysis of several analytical method validation strategies in the framework of the fit for purpose concept, J. Chromatogr. A 1217 (2010) 3180–3192.
[90] NCCLS, Evaluation of Precision Performance of Clinical Chemistry Devices. Approved Guideline, NCCLS, Wayne, PA, USA, 1999.
[91] 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.
[92] The Fitness for Purpose of Analytical Methods, Eurachem, Teddington, 1998.
[93] J. Vial, A. Jardy, Experimental comparison of the different approaches to estimate LOD and LOQ of an HPLC method, Anal. Chem. 71 (1999) 2672–2677.
[94] J. Vial, K. Le Mapihan, A. Jardy, What is the best means of estimating the detection and quantification limits of a chromatographic method? Chromatographia 57 (2003), S-303–S-306.
[95] M. Feinberg, B. Boulanger, W. Dewe, Ph. Hubert, New advances in chemical data quality: method validation and measurement uncertainty, Anal. Bioanal. Chem. 380 (2004) 502–514.
[96] A.G. Gonzalez, M.A. Herrador, Accuracy profiles from uncertainty measurements, Talanta 70 (2006) 896–901.
[97] D. Hoffman, R. Kringle, J. Singer, S. McDougall, Statistical methods for assessing long-term analyte stability in biological matrices, J. Chromatogr. B 877 (2009) 2262–2269.
[98] P. Timmerman, S. Luedtke, P. van Amsterdam, M. Brudny-Kloeppel, B. Lausecker, S. Fischmann, S. Globig, C.-J. Sennbro, J.M. Jansat, H. Mulder, E. Thomas, M. Knutsson, D. Kasel, S.A. White, M. Anders Kall, N. Mokrzycki-Issartel, A. Freisleben, F. Romero, M. Pilgård Andersen, N. Knebel, M. de Zwart, S. Laakso, R.S.S. Hucker, D. Schmidt, B. Gordon, R. Abbott, P. Boulanger, Incurred sample reproducibility: views and recommendations by the European Bioanalysis Forum, Bioanalysis 1 (2009) 1049–1056.
[99] D.M. Fast, M. Kelley, C.T. Viswanathan, J. O'Shaughnessy, S.P. King, A. Chaudhary, R. Weiner, A.J. DeStefano, D. Tang, Workshop report and

follow-up—AAPS workshop on current topics in GLP bioanalysis: assay reproducibility for incurred samples—implications of Crystal City recommendations, AAPS J. 11 (2009) 238–241.
[100] D. Hoffman, Statistical considerations for assessment of bioanalytical incurred sample reproducibility, AAPS J. 11 (2009) 570–580.
[101] M. Rocci, V. Devanarayan, D. Haughey, P. Jardieu, Confirmatory reanalysis of incurred bioanalytical samples, AAPS J. 9 (2007) E336–E343.
[102] R. Kringle, An assessment of the 4–6–20 rule for acceptance of analytical runs in bioavailability, bioequivalence, and pharmacokinetic studies, Pharm. Res. 11 (1994) 556–560.
[103] E. Rozet, C. Hubert, A. Ceccato, W. Dewé, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, Ph. Hubert, Using tolerance intervals in pre-study validation of analytical methods to predict in-study results. The fit-for-future-purpose concept, J. Chromatogr. A 1158 (2007) 126–137.
[104] W. Dewé, B. Govaerts, B. Boulanger, E. Rozet, P. Chiap, Ph. Hubert, Using total error as decision criterion in analytical method transfer, Chemometr. Intell. Lab. Syst. 85 (2007) 262–268.
[105] B. Govaerts, W. Dewé, M. Maumy, B. Boulanger, Pre-study analytical method validation: comparison of four alternative approaches based on quality-level estimation and tolerance intervals, Qual. Reliab. Eng. Int. 24 (2008) 667–680.
[106] ISO/IEC Guide 99:2007, International Vocabulary of Metrology—Basic and general concepts and associated terms (VIM).
[107] E. Hund, D.L. Massart, J. Smeyers-Verbeke, Operational definitions of uncertainty, Trends Anal. Chem. 20 (2001) 394–406.
[108] J. Mandel, The Statistical Analysis of Experimental Data, Dover, New York, 1964, pp. 104–105.
[109] S.L.R. Ellison, M. Rösslein, A. Williams (Eds.), EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical Measurement, Second ed., 2000.
[110] X. Fuentes-Arderiu, Uncertainty of measurement in clinical laboratory sciences, Clin. Chem. 46 (2000) 1437–1438.
[111] A. Maroto, R. Boqué, J. Riu, F.X. Rius, Evaluating uncertainty in routine analysis, Trends Anal. Chem. 18 (1999) 577–584.
[112] E. Hund, D.L. Massart, J. Smeyers-Verbeke, Comparison of different approaches to estimate the uncertainty of a liquid chromatographic assay, Anal. Chim. Acta 480 (2003) 39–52.
[113] A.G. Gonzalez, M.A. Herrador, A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles, Trends Anal. Chem. 26 (2007) 227–238.
[114] S.D. Phillips, K.R. Eberhardt, Guidelines for expressing the uncertainty of measurement results containing uncorrected bias, J. Res. Natl. Inst. Stand. Technol. 102 (1997) 577–585.
[115] V. Synek, Attempts to include uncorrected bias in the measurement uncertainty, Talanta 65 (2005) 829–837.
[116] P.H. Petersen, D. Stöckl, J.O. Westgard, S. Sandberg, K. Linnet, L. Thienpont, Models for combining random and systematic errors. Assumptions and consequences for different models, Clin. Chem. Lab. Med. 39 (2001) 589–595.
[117] ILAC, ILAC-G8:03/2009, Guidelines on the reporting of compliance with specification.
[118] ASME B89.7.3.1-2001, Guidelines for decision rules: considering measurement uncertainty in determining conformance with specifications.
[119] S.L.R. Ellison, A. Williams (Eds.), EURACHEM/CITAC Guide, Use of Uncertainty Information in Compliance Assessment, First ed., 2007.
[120] E. Desimoni, B. Brunetti, About acceptance and rejection zones as defined in the EURACHEM/CITAC Guide (2007) "Use of uncertainty information in compliance assessment", Accred. Qual. Assur. 15 (2010) 45–47.
[121] ISO/CEI 17025: General Requirements for the Competence of Testing and Calibration Laboratories, International Organization for Standardization (ISO), Geneva, 2005.
[122] ISO 15189: Medical Laboratories—Particular Requirements for Quality and Competence, International Organization for Standardization (ISO), Geneva, 2007.