



HAL
open science

Risk management for analytical methods based on the total error concept: Conciliating the objectives of the pre-study and in-study validation phases

Bruno Boulanger, Walthère Dewé, Aurélie Gilbert, Bernadette Govaerts,
Myriam Maumy

► To cite this version:

Bruno Boulanger, Walthère Dewé, Aurélie Gilbert, Bernadette Govaerts, Myriam Maumy. Risk management for analytical methods based on the total error concept: Conciliating the objectives of the pre-study and in-study validation phases. *Chemometrics and Intelligent Laboratory Systems*, 2007, 86, pp.198-207. 10.1016/j.chemolab.2006.06.008 . hal-00141709

HAL Id: hal-00141709

<https://hal.science/hal-00141709>

Submitted on 11 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Risk management for analytical methods: Conciliating objectives of validation phase and routine decision rules.

Bruno Boulanger^a, Walthère Dewé^a, Aurélie Gilbert^b,
Bernadette Govaerts^{b, 1} and Myriam Maumy^c.

^a*Eli Lilly, European Early Phase Statistics, Belgium.*

^b*Université Catholique de Louvain, Institut de Statistique, Belgium.*

^c*Université Louis Pasteur, Laboratoire de Statistique, France.*

SUMMARY

In industries that involve either chemistry or biology, the analytical methods are the necessary eyes of all the material produced. If the quality of an analytical method is doubtful, then the whole set of decisions that will be based on those measures is questionable. For those reasons, being able to assess the quality of an analytical method is far more than a statistical challenge; it's a matter of ethic and good business practices.

The validity of an analytical method must be assessed at two levels. The “pre-study” validation aims at proving, by an appropriate set of designed experiments, that the method is able to achieve its objectives. The “in-study” validation is intended to verify, by inserting QC samples in routine, that the method remains valid over time.

This paper discusses and compares two methods, based on the total error concept, to check the validity of a measurement method at a pre-study level. The first checks if a tolerance interval on hypothetical future measurements lies within given acceptance limits and the second calculates the probability to lie within these limits and verifies if it is greater than a given acceptance level. For the “in-study” validation, the paper assesses the properties of the $s-n-\lambda$ rule recommended by FDA.

A crucial point is also to ensure that the decisions taken at the pre-study stage and in routine are compatible. More precisely, a laboratory should not see its method rejected in routine when it has been proven to be valid and remains so. This goal may be achieved by an appropriate choice of validation parameters at both pre- and in-study levels.

¹Correspondance to : Bernadette Govaerts, Institut de Statistique, 20 voie du roman pays, 1348 Louvain-la-Neuve, Belgium, Govaerts@stat.ucl.ac.be, Phone : +32-10-47.43.13.

1. INTRODUCTION

In industries that involve either chemistry or biology, such as pharmaceutical industries, chemical industries or food industry, the analytical methods are the necessary eyes and hear of all the material produced or used. If the quality of an analytical method is doubtful, then the decisions and possibly the product released based on measures obtained with this procedure may become questionable. For those reasons, being able to assess the quality of an analytical method is far more than a statistical challenge; it's a matter of ethic and good business practices. Many regulatory documents have been released to address that issue primarily ICH and FDA documents in the pharmaceutical industry [1,2,3].

The objective of validation is to give to the laboratory as well as to the regulatory bodies guarantee that every single measurement that will be performed in routine will be close enough to the unknown "true" value of the sample [4]. The conformity of a given analytical method to this objective is usually assed in two stages [5, 6, 7, 8]. First, a "pre-study" phase is conducted to proof, on the basis of a designed experiment, that the method is able to deliver results of quality. Then, at a routine level, the laboratory must verify that the analytical method of interest remains valid over time and that each run performed provides trustful measures. This is usually achieved by inserting QC samples in the unknown samples runs.

At these two stages, one needs then to have a way to quantify the quality of a measure in terms of its closeness to the "true" value of the property of interest. Traditionally this quantity is assessed by examining the two main performance criteria of an analytical method: the bias or "trueness" and precision of the method. Both should be small enough and are usually quantified separately [3, 9, 10, 11]. This approach focuses the method itself in assuming that if the method is "good" then the measures it will provide are also "good". As already show, this is not always the case [12]. The concept of "total error" [8, 13, 14, 15, 16, 17] however puts the emphasis on the results themselves and tackles the problem globally by estimating the proportion π of measurements expected to lie within a fixed interval ($\pm \lambda$) around their true value. The correct underlying assumption behind this approach is that, if the results produced are "good", then the method that produces them is necessarily "good". This paper presents procedures to check the pre- and in-study validity of an analytical method based on this total error concept, i.e. by examining the quality of results it produces.

At the pre-study level, the validation procedure consists of measuring a given set of samples for which the nominal values are known and arranged according to an experiment adequately designed. The design should be able to estimate measurement bias and precision for different nominal levels and, if necessary, provide a decomposition of the global precision in various components of variances (repeatability, between-run, and between laboratory). Two statistical procedures are discussed to assess the method validity on the basis of such an experiment. The first consists in estimating a tolerance interval in which "future" measurements are expected to lie and verify if this interval is included in predefined acceptance limits. The second estimates directly the probability to get a measure in this acceptance limits and verifies if this estimated probability is greater than a given minimal acceptance level on the basis of a lower bound of a maximum likelihood confidence interval on this probability. Simulations have been conducted to study the laboratory and client risks for these two procedures and show that the first method is particularly efficient when the measurement process is well centered and that

the probability approach gains in power for reasonably biased situations, a common situation in practice.

In routine, budget and simplicity requirements usually lead to use validation rules that do not fully protect both the client and the laboratory. This paper studies the properties of the 4-6-15 rule recommended by FDA [3] in this context, generalized here as s-n- λ rule. It consists of inserting a set of n QC samples within routine unknown samples and check if at least s of the measurements obtained with those samples are not more distant than λ from their nominal (true) value. Power functions of this generalized rule show that it is very difficult to protect simultaneously the client and laboratory interest at a reasonable cost. It is also shown that the FDA 4-6-15 rule protects mainly the laboratory, a surprising result since the spirit of those regulations are precisely to protect the client.

In the practical organization of an industrial laboratory, the pre- and in-study validation studies are often conducted separately by different persons especially if the method is developed and validated in one place (e.g. research laboratory) and is used in routine in another one (e.g. production plan). The compatibility of the decisions taken at both stages is not obvious and even not well understood by the analysts. A laboratory that has declared the validity of a method in a pre-study phase would not appreciate (economically speaking) to see its method rejected in routine if it is still valid. On the other hand, if a valid method is subject to a significant total error increase, the in-study validation rule should be able to detect it rapidly. Conciliating pre- and in-study objectives is then crucial and may be achieved through an appropriate choice of validation rules parameters in order to align the risks associated.

This paper is organized as follows: Section 2 gives a precise definition of method validation based on the concept of total error and introduces related notations. Section 3 introduces two procedures for “pre-study” validation, the β -expectation tolerance interval approach and a maximum likelihood approach aimed at estimating the probability to be within acceptance limits using the delta method. Those two procedures are illustrated on a real example and their performances are then compared using simulations. Section 4 discusses the properties of the s-n- λ rule in terms of client and laboratory risks. Finally, Section 5 shows how to conciliate pre- and in-study validation parameters to attain coherent properties for the validation decisions.

2. ANALYTICAL METHOD EVALUATION BASED ON TOTAL ERROR

The objective of a good analytical method is to be able to quantify accurately each of the unknown quantities that the laboratory will have to determine. In other terms, the analytical method is expected to give results for which the difference with the unknown “true value” (μ_T) of the sample is small or inferior to an acceptance limit, i.e.:

$$-\lambda < X - \mu_T < \lambda \Leftrightarrow |X - \mu_T| < \lambda$$

Two components may influence this difference: the bias or trueness of the method and its precision. As illustrated in Figure 1, a biased method provides results that deviate “in mean” or systematically from the true value μ_T : $\delta = E(X) - \mu_T = \mu - \mu_T$. The precision expresses how results vary around the mean value $\mu = E(X)$ when the measure is repeated. Let’s note σ the standard deviation available to quantify this precision. A “good” analytical method should ideally give results close from the unknown true value of the

sample, i.e. within some acceptance limits. This “closeness” is directly linked to the size of the bias δ and precision σ of the method.

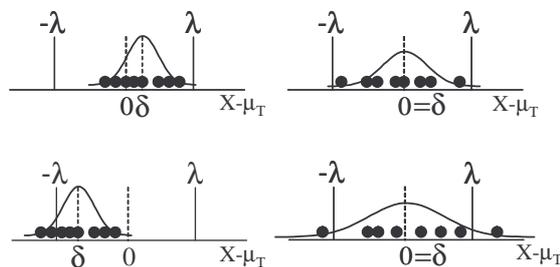


Figure 1 : Comparison of four possible validation situations

Classical method validation and quality control tools usually check the size of these two components separately (t and χ^2 tests in validation or \bar{X} - R control charts in routine) but this approach has the drawback that a very small value of one component may not compensate a weakness of the other.

The total error approach [4, 8, 13, 14, 15, 16, 17] suggests a global approach in considering a procedure acceptable if it is “very likely” that the difference between each measurement X of a sample and its "true value" (μ_T) is inside the acceptance limits $[-\lambda, +\lambda]$ predefined by the analyst. The notion of "very likely" can be translated to the following probabilistic equation:

$$\pi = P(|X - \mu_T| < \lambda) \geq \pi_{\min}$$

where π_{\min} is called the acceptance level and π the quality level. The acceptance limit λ can be expressed either in absolute or in relative value (%). In this later case, the equation is redefined as:

$$\pi = P\left(\left|\frac{X - \mu_T}{\mu_T}\right| < \lambda\right) \geq \pi_{\min}$$

All results presented in this paper are applicable to both cases. Below, we will then use the first formulation without loss of generality.

The value of λ must be chosen according to intended use of the results. The objective is linked to the requirements usually admitted by the practice (e.g. 1% or 2% on bulk, 5% on pharmaceutical specialties, 15% for biological samples, 30% for ligand binding assays such as RIA or ELISA, etc.).

The probability π_{\min} must also be fixed by the analyst according to cost, consumer and analytical domain requirements. The key aspect is to ensure the coherence between the π_{\min} and λ values targeted in the pre-study and in-study phases. This issue is discussed in more detail in section 5.

Under normality assumption for the measurement results, it is easy to establish the relationship between the quality level π and the bias δ (systematic error), and precision σ (random error):

$$\pi = P(|X - \mu_T| < \lambda) = P\left(\frac{-\lambda - \delta}{\sigma} < Z < \frac{\lambda - \delta}{\sigma}\right)$$

where Z is a standard normal random variable. This leads to define the “acceptance region”, i.e. the set of (δ, σ) 's such that the quality level π is greater than π_{\min} . Figure 2 shows, below the curves, the acceptance region for various values of π_{\min} (99%, 95%, 90%, 80% and 66.7%) when acceptance limits are fixed to $[-15\%, +15\%]$ as recommended by FDA [3] for bioanalytical methods. Logically, as it can be seen on Figure 1, the greater the variance of measure or the greater the bias, the less likely a measure will fall within the acceptance limits.

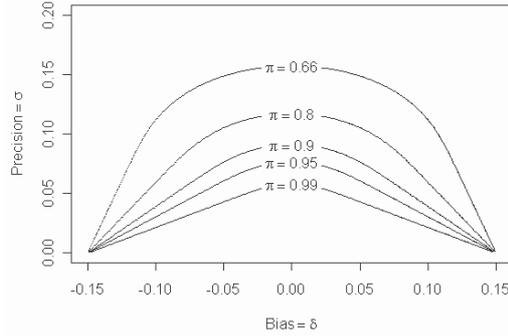


Figure 2: Acceptance region of analytical methods as a function of the method bias and precision when $\lambda = 15\%$.

Note that, in this graph, δ and σ must be interpreted as relative bias and relative standard deviation.

3. PRE-STUDY METHOD VALIDATION

Before an analytical method can be used in routine for qualifying unknown samples, it's the practice to perform a more or less extensive set of experiments to evaluate if the analytical method will be able to achieve the objective stated above. Those experiments are usually called “pre-study validation” as opposed to the “in-study validation” experiments.

Since the bias δ and the precision σ , the intrinsic performance parameters of the analytical procedure, are unknown, experiments are required before using the method in routine to allow the user to obtain estimates of these quantities. These estimates of bias ($\hat{\delta}$) and standard deviation ($\hat{\sigma}$) are intermediary but obligatory steps to evaluate if the analytical procedure is likely to provide accurate measures of the unknown samples to be measured in routine.

The objective of the pre-study validation phase is to evaluate whether, given the estimates of bias $\hat{\delta}$ and standard-deviation $\hat{\sigma}$ obtained, the proportion of measures of new unknown samples, that will fall within the acceptance limits, is greater than a predefined acceptance level, say π_{\min} .

$$P(|X - \mu_T| < \lambda) \geq \pi_{\min} \quad (1)$$

However, there exists neither no exact solution nor easy way to answer this question, even for very easy validation experimental designs. Two methods are discussed in this paper. The first one is based on the notion of tolerance interval and the second consists of calculating the lower bound of a maximum likelihood one side confidence interval on this quality level π using the delta method.

3.1 β -expectation tolerance interval method

A first method to take a decision has already been proposed by other authors [14,4] and consists of computing the β -expectation tolerance interval [18]:

$$E_{\hat{\delta}, \hat{\sigma}} \left\{ P_X \left(\hat{\delta} - k \hat{\sigma} < X - \mu_T < \hat{\delta} + k \hat{\sigma} \mid \hat{\delta}, \hat{\sigma} \right) \right\} = \beta$$

where the factor k is determined so that the expected proportion of the population falling within the interval is equal to β . β is defined as the acceptance level π_{\min} in this context and the value of k depends on the experimental design used for validation. In the simplest case, where a same sample with nominal value μ_T is measured n times in repeatability conditions, $\hat{\delta}, \hat{\sigma}$ and k are calculated as follows from the measurement results X_1, X_2, \dots, X_n if a normality assumption can be postulated on X :

$$\hat{\delta} = \frac{1}{n} \sum_{i=1}^n X_i - \mu_T = \bar{X} - \mu_T \quad \hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{and} \quad k = t_{n-1; (1+\pi_{\min})/2} \sqrt{1 + \frac{1}{n}}$$

where $t_{n-1; (1+\pi_{\min})/2}$ is the quantile $(1+\pi_{\min})/2$ of a $(n-1)$ t distribution.

Mee [19] discussed how to calculate $\hat{\sigma}$ and k in a balanced one-way ANOVA random model where the within and between-run variabilities are taken into account. Hoffman and Kringle [17] generalize it to more general random effect models for the β -content tolerance intervals instead of the β -expectation tolerance intervals as suggested here.

The decision rule proposed is then the following: if the β -expectation tolerance interval defined above is included within the acceptance limits $[-\lambda, +\lambda]$, i.e. if $(\hat{\delta} - k \hat{\sigma} > -\lambda$ and $\hat{\delta} + k \hat{\sigma} < +\lambda)$ then there is high evidence that the method is valid. As a matter of fact, if this condition is verified, the expected proportion of measurements within the acceptance limits is greater or equal to π_{\min} , i.e. equation (1) is also verified, on average, in that case. Note that the opposite statement is not true, i.e. either $\hat{\delta} - k \hat{\sigma} < -\lambda$ or $\hat{\delta} + k \hat{\sigma} > +\lambda$ doesn't imply that the expected proportion is smaller than π_{\min} . This is illustrated on simulations below.

3.2 Maximum likelihood one side confidence interval on π using the delta method.

Another approach proposed to validate the method consists in deriving, from $\hat{\delta}$ and $\hat{\sigma}$, the lower bound of an one side confidence interval on the quality level π and verifying if it is bigger than π_{\min} or not. This is not easy as far as the estimation of a probability is concerned. No exact solution exists even in the simple sampling scheme case but different statistical approaches may be envisaged to attack the problem: search for a mathematical approximation to the exact solution, asymptotic approximation by maximum likelihood, bootstrap or Bayesian modeling. We present here a maximum likelihood (ML) solution for the simple sampling scheme, under normality assumptions.

Let X_1, X_2, \dots, X_n be the measurement results of the validation experiment and suppose that X_i is distributed as a normal distribution with unknown bias δ and variance σ^2 . The maximum likelihood estimators of these parameters are given by :

$$\hat{\delta} = \frac{1}{n} \sum_{i=1}^n X_i - \mu_T = \bar{X} - \mu_T \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} \hat{\sigma}^2 = w \hat{\sigma}^2$$

By the invariance property, a maximum likelihood estimator of π can be defined as:

$$\hat{\pi} = \phi\left(\frac{\lambda - \hat{\delta}}{\tilde{\sigma}}\right) - \phi\left(\frac{-\lambda - \hat{\delta}}{\tilde{\sigma}}\right)$$

where $\phi(\cdot)$ is the distribution function of the standard normal distribution [19]. The delta method can be used to derive an asymptotic approximation for the variance of the estimator $\hat{\pi}$:

$$\text{Var}(\hat{\pi}) \cong \frac{1}{n} (\varphi_U - \varphi_L)^2 + \frac{1}{2nw\sigma^2} ((-\lambda - \delta)\varphi_L - (\lambda - \delta)\varphi_U)^2$$

where φ_U and φ_L are defined as :

$$\varphi_U = \varphi\left(\frac{\lambda - \delta}{\sigma}\right) \quad \text{and} \quad \varphi_L = \varphi\left(\frac{-\lambda - \delta}{\sigma}\right)$$

and $\varphi(\cdot)$ is the density of the standard normal distribution. The asymptotic lower bound of a one side $1-\alpha$ confidence interval on the quality level π can then be calculated as:

$$\hat{\pi}_{\text{inf}} = \hat{\pi} - z_{1-\alpha} \hat{\sigma}_{\hat{\pi}}$$

where $z_{1-\alpha}$ is the quantile $1-\alpha$ of a standard normal variable and $\hat{\sigma}_{\hat{\pi}}$ is calculated by replacing δ by $\hat{\delta}$ and σ by $\tilde{\sigma} = \sqrt{w}\hat{\sigma}$ in $\text{Var}(\hat{\pi})$ above.

The analytical method is then declared valid if $\hat{\pi}_{\text{inf}} > \pi_{\text{min}}$. The generalization of this approach to more general variance components models is possible.

3.3 Example

To illustrate the methodology expressed here, pre-study validation data from a bioanalytical procedure [20, 21] are used to illustrate both statistical methods described in section 3.1 and 3.2.

The design consisted in 3 runs of 4 replicates per run at each concentration level but the run factor was ignored in this illustration in order to align with formulas detailed in section 3.1 and 3.2. This leads to an overall estimate of the total variance without estimating its within within-run and between-run components. Therefore the sample size n in this example is $n = 12$ at each level of concentration. In order to answer to the objectives, the acceptance limit λ was fixed at 0.15, i.e. [-15%, 15%], the acceptance level π_{min} at 80%, and the confidence level for the ML unilateral confidence interval on π at 90%.

The results are summarized in Table 1. The β -expectation tolerance interval limits are all included within the acceptance limits whatever the concentration level. The bioanalytical method of interest can therefore be considered as valid over the range investigated using the tolerance interval rule. Also, the 90% lower limits of the unilateral confidence interval on π estimated by the maximum likelihood method are larger than 0.80 over the range of concentration levels investigated. This lead to the same conclusion as the one obtained with the tolerance interval method.

Table 1: Validation results obtained by β -expectation tolerance interval method and ML method.

μ_T	$\hat{\delta}$	$\hat{\sigma}$	Tolerance Interval	Acceptance Limits	$\hat{\pi}$	$\hat{\pi}_{inf}$
25.4	0	1.4	[-2.1, 2]	[-3.8, 3.8]	0.994	0.982
48.2	-2.7	2.7	[-6.5, 1.2]	[-7.2, 7.2]	0.959	0.908
437.8	-9.3	19.7	[-37.2, 18.6]	[-65.7, 65.7]	0.999	0.995
838.6	11.8	45	[-52, 75.7]	[-125.8, 125.8]	0.995	0.984

3.4 Validation method comparison on the basis of simulations

This section compares the two pre-study validation procedures on the basis of simulations. Four valid and two non valid hypothetical measurement processes or analytical methods have been chosen as shown in Table 2 and Figure 3, and sets of normally distributed data of size ranging from 5 to 200 have randomly been generated. A measurement process was considered as valid if the proportion of measurements within a range of $\pm 15\%$ ($\lambda=0.15$) around the target value μ_T was greater than 80% (acceptance level $\pi_{min}=0.8$). For the ML confidence interval on π using the delta method, a 90% unilateral confidence interval was calculated to get a lower bound for π .

Figure 4 presents, for the 6 possible scenarios and the two validation methods, the proportion γ of cases for which the validation has been accepted over 5000 simulated samples.

Table 2: Six scenarios chosen to compare the performances of the two validation methods

δ_T	σ_T	π_T		δ_T	σ_T	π_T	
0	0.0765	0.95	Valid	0.05	0.0605	0.95	Valid
0	0.104	0.85	Valid	0.05	0.091	0.85	Valid
0	0.13	0.75	Non Valid	0.05	0.12	0.75	Non Valid

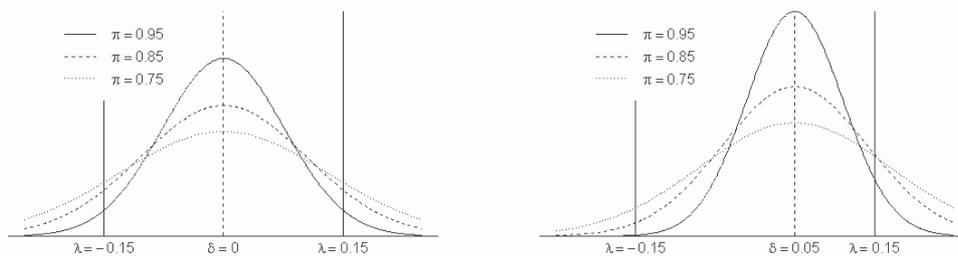


Figure 3 : Six scenarios chosen to compare the performances of the two validation methods

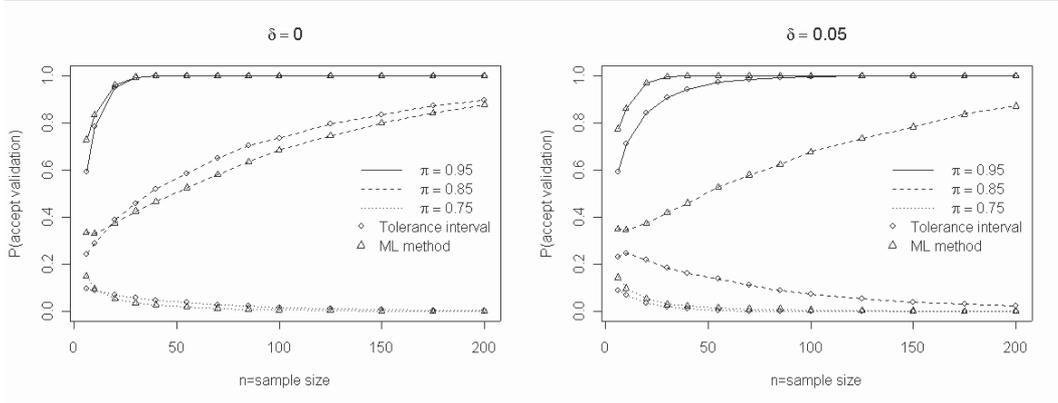


Figure 4 : Simulation results: $\delta = 0$ on the left and $\delta = 0.05$ on the right

This graph leads to several observations:

- When the measurement process is not valid ($\pi = 0.75$), the behaviors of the two methods are quite similar and protect very well the client since the probability to accept the method is very small. The ML method is slightly better than the tolerance interval method for centered processes, the reverse being observed for bias processes.
- When the measurement process is well centered ($\delta=0$) and valid, the probability to accept the method follows an expected behavior as it increases with n and π . For $\pi = 0.85$, the tolerance interval method is more powerful than the maximum likelihood method.
- However, when the process is biased, the behavior of the tolerance interval method is less attractive: for $\pi = 0.95$, the measurement process is less often accepted as valid than for the ML method.
- In addition, for $\pi = 0.85$, the probability to accept the measurement process with the tolerance interval method tends to 0 when n increases. This is not a desirable result, from a statistical point of view, as for $\pi = 0.85$ the process is valid. It can be shown that this arises (asymptotically) for the points in the acceptance region such that :

$$\lambda - |\delta| < \sigma z_{\frac{1+\pi_{\min}}{2}}$$

This result has already been emphasized roughly in [22].

This last point merits to be investigated a little further. Figure 5 illustrates it graphically: the area between the triangle and the acceptance region is a zone where, asymptotically, the measurement process is rejected by the tolerance interval method when the process is indeed valid. Note that asymptotically this acceptance region coincides with the region where the process is asymptotically accepted by the ML method so all valid method should be accepted. The litigious point ($\delta=0.05$, $\sigma=0.091$) is in this zone.

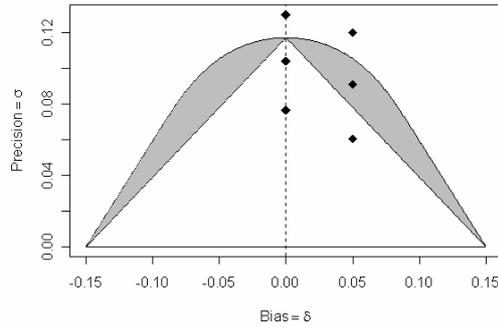


Figure 5 : Representation of the acceptance region area where the measurement method is (asymptotically) wrongly rejected with the tolerance interval method while it is valid and accepted with the ML method. The 6 simulated scenarios are marked.

For small samples, the scenario is different and the “gap” between the tolerance interval and ML methods is less important. Figure 6 presents results of new simulations of the two validation methods behaviors in the δ, σ space. The three curves in this figure delimit the “real” acceptance region for $\pi_{\min}=0.8$ and two regions (one for each method) for which the power is higher or equal to $\gamma=0.75$. Two sample sizes, $n=12$ and $n=200$, are envisaged.

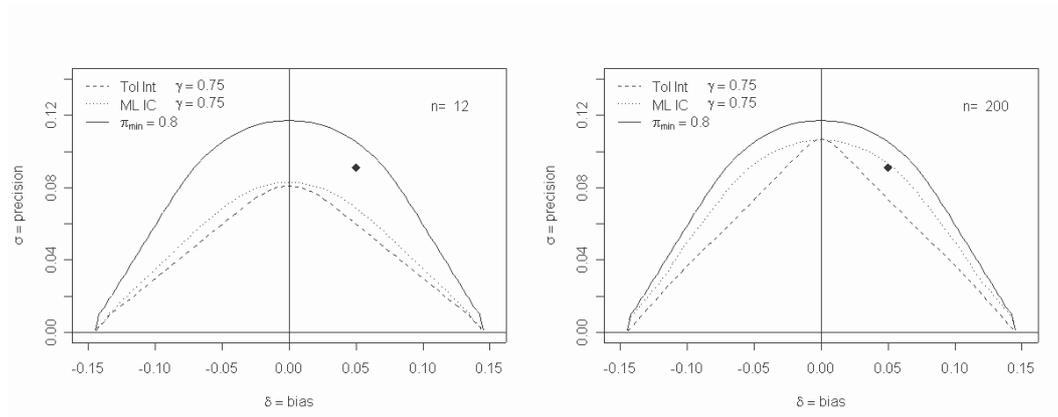


Figure 6: Iso-power line in the δ, σ space of the tolerance interval and maximum likelihood validation methods for two samples sizes: $n=12$ and $n=200$. The outer plain line represent the true acceptance region for $\pi_{\min}=0.8$.

As it can be seen on those figures, for small sample size, there is very little difference of power between the two validation methods and the differences mainly occurs for analytical methods slightly biased. The practical consequence being that with limited sample size, there is a small region of good measurement processes that may be rejected by the tolerance interval procedure while they will be accepted more often with the other approach. This can increase very slightly the cost for the labs but can not increase the risk for the client since this phenomenon does not occur outside the acceptance region where both methods are very conservative.

Once the sample size increases ($n=200$) then, it can be observed that the likelihood method converges to the “true” acceptance region while the tolerance interval acceptance

region keeps its “triangular” shape. The gap between the two procedures -the difference between the “bell” shape and the “triangular” shape- then becomes larger. This highlights the fact that the use of the “two one-sided” approach using the tolerance interval tends to ignore the other side of the distribution once a tolerance limit goes outside one acceptance limit. This doesn’t occur when the bias=0.

4. (S-N-λ) METHOD FOR ROUTINE FOLLOW UP

Once a method has been validated and is used in routine analysis, it should be monitored regularly to check if the method remains valid over time. As opposed to the validation phase where expensive and cautious practices are usually envisaged, the rule used for QC in routine must be simple and cheap. An in-study rule that is largely accepted in the bioanalytical community, is called “4-6-15” rule and is defined in the FDA guidance [3] as : “...At least four of every six QC samples should be within "15% of their respective nominal value...”. This rule is discussed below as a simple and practical solution for routine follow up and its properties are analyzed.

In general terms, this “s-n-λ” rule is applied as follows:

1. n QC samples with known nominal values are integrated in a daily run.
2. The number Y of samples such that the (absolute or relative) difference between the measured value X_i and the nominal value μ_T is lower than λ is counted.
3. If $Y \geq s$, the run is accepted and can be delivered to the laboratory client.

The properties of such decision rule with respect to the laboratory and client interests depend crucially on the choice of s and n [21]. They are well represented by a power function which gives the probability γ to accept a run with respect to the state of the measurement system expressed in term of quality level π or bias and precision δ and σ .

As Y is a binomial distribution with parameters n and π , the power γ is calculated as follows:

$$\gamma = P(\text{accept the run} | \pi) = P(Y \geq s | \pi) = P(Bi(n, \pi) \geq s) = \sum_{i=s}^n C_n^i \pi^i (1 - \pi)^{n-i}$$

Figure 7 presents this power function for the 4-6-λ rule with respect to π and iso-power curves for the 4-6-15 rule in the δ and σ space. These graphics show (as expected) that the more the method is in a valid state (π large or δ and/or σ small) the higher is the probability γ to accept a run.

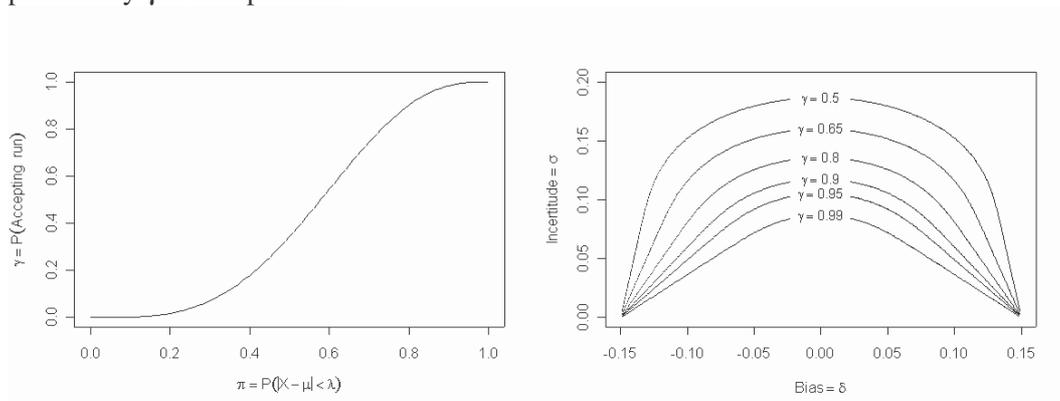


Figure 7 : Power of 4-6-15 decision rule

It is instructive to study the power of s - n - λ rules for other values of s and n . Figure 8 shows the evolution of the power of s - n - λ rule for $s = 2, 4$ and 6 and of s - n - λ rule for $n = 6, 12, 24$ and 96 with $s/n = 2/3$. These results are intuitively meaningful: increasing s (keeping n constant) decreases the client risk while increasing n (keeping s/n constant) decreases simultaneously the client and laboratory risks with respect to the compromise $\pi^* = s/n$ value. When n increases, the laboratory will have a high probability to see accepted a run for which $\pi > \pi^*$ and the run will have a high probability to be rejected if $\pi < \pi^*$. The simultaneous protection of the client and laboratory interests has of course a cost: the number n of QC samples required.

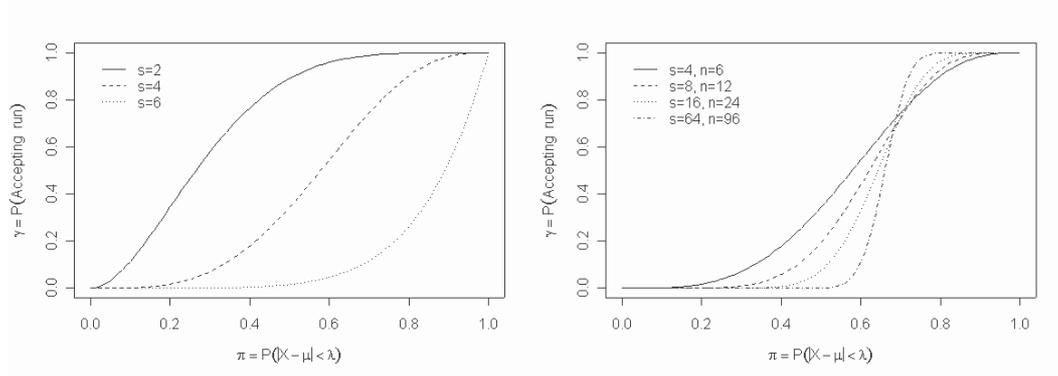


Figure 8 : Power of n - s - λ as a function of s and n .

Such discussion is common in the context of lot acceptance sampling plans as discussed in [24] or in ISO norms [25]. In this framework, the problem is treated in the opposite direction, i.e. the requirements of both client and supplier are first fixed and then the optimal values for n and s are calculated to meet their requirements. In practice, the client has first to choice a quality level π_C under which the probability of accepting a run (or lot) is small (γ_C say). On the other side, the supplier (laboratory) has to choose a quality level π_L over which the probability of accepting a run (or lot) is high enough (γ_L say). This leads to a system of two inequalities to be solved with two unknown values n and s .

$$P(Y \geq s | \pi < \pi_C) < \gamma_C \quad \text{Client equation}$$

$$P(Y \geq s | \pi > \pi_L) > \gamma_L \quad \text{Supplier equation}$$

This is illustrated in Figure 9 for 4 - 6 - λ and 6 - 6 - λ rules. It shows that the 4 - 6 - λ rule is highly protective for the laboratory and favors the laboratory over the client.

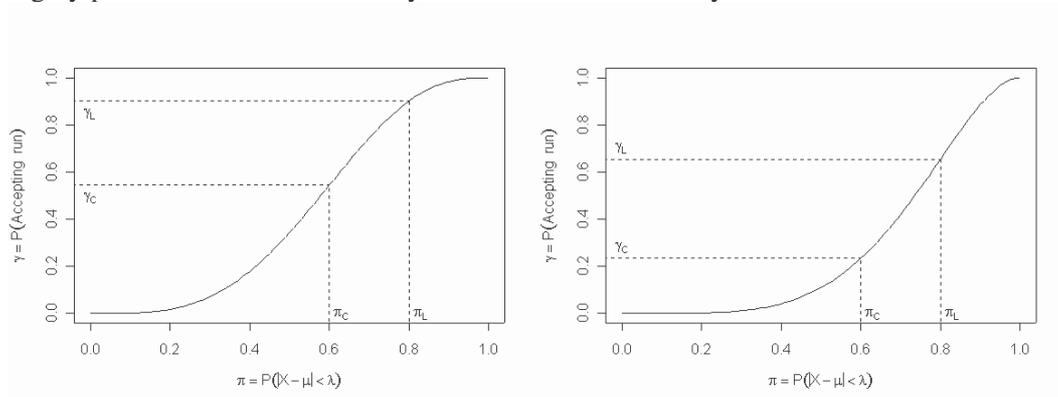


Figure 9 : Comparison of client and laboratory risks for 4 - 6 - λ and 6 - 6 - λ rules.

Table 3 finally presents optimal n and s values for different combinations of client and supplier parameters.

Table 3: Optimal sampling plans for different combinations of client and laboratory requirements.

π_C	γ_C	π_L	γ_L	n	s	π_C	γ_C	π_L	γ_L	n	s
0.6	0.2	0.8	0.8	19	14	0.6	0.1	0.8	0.9	36	26
0.7	0.2	0.8	0.8	55	42	0.7	0.1	0.8	0.9	127	96
0.7	0.2	0.9	0.8	14	12	0.7	0.1	0.9	0.9	25	21
0.8	0.2	0.9	0.8	39	34	0.8	0.1	0.9	0.9	86	74

This table illustrates clearly the impossibility to conciliate acceptable client and supplier risks with a reasonable cost when using classical s - n - λ rule. In consequence, the FDA recommendations should be interpreted as a practice that favors laboratory risks over consumer risks.

5. CONCILIATING VALIDATION AND ROUTINE DECISION RULES

The central question, when applying pre-study and in-study validation procedures to a measurement method, is to conciliate the objectives pursued by both validation procedures.

When the total error approach is used in pre-study and s - n - λ in method in routine, a common objective is already pursued: to control the parameter π or the proportion of measurement results ($X - \mu_T$) expected to lie within the acceptance limit $[-\lambda, \lambda]$. The shape of the acceptance region of both methods is then equivalent as shown in Figures 2 and 5.

The ability to conciliate the pre-study procedure, tolerance interval or maximum likelihood estimate, and s - n - λ validation procedure depends on the adequacy of the parameters chosen for these two. Indeed, a laboratory, who has proven a method to be valid in a pre-study experiment, can expect to see most of the runs produced by this analytical method accepted in routine if the method performance remains stable over time. This is essential; otherwise it would become counterproductive to maintain an analytical method that leads frequently to reject runs.

Let's reformulate this requirement in terms of the test methods parameters in two ways:

1. If the parameters n and s of the s - n - λ rule are fixed, the value of π_{\min} should be chosen to ensure that, if the method remains valid, the s - n - λ rule is accepted in most cases (e.g. with a minimum probability γ_{\min}).
2. For a pre-study validation scheme (π_{\min} and λ), the value of s (for a given n) should guarantee that most of the runs will be accepted if the method remains truly valid.

On the client side, as seen, in section 4, it is difficult with an s - n - λ rule to protect really the client at a reasonable cost. The parameters should then protect the client as much as possible given the budget available and the economical rationale for the laboratory. It exists, however, other routine tests which, for a same cost, may protect the client better like tests directly based on quantitative results (instead of dichotomous response) [26] or moving type statistics like the Scan statistic [27]. This is however not the purpose of this paper to discuss such methods.

Let's consider now the particular case of the 4-6- λ rule. As stated above, a good pre-study validation rule should work with a π_{\min} value which ensures that the routine test will be accepted in most of the cases (let's take $\gamma_{\min}=90\%$) if the method is still valid:

$$P(Y \geq s | \pi > \pi_{\min}) > \gamma_{\min}$$

For a given s and n , this π_{\min} is simply obtained by inverting, in π , the Binomial(n, π) distribution function as shown in Figure 10.

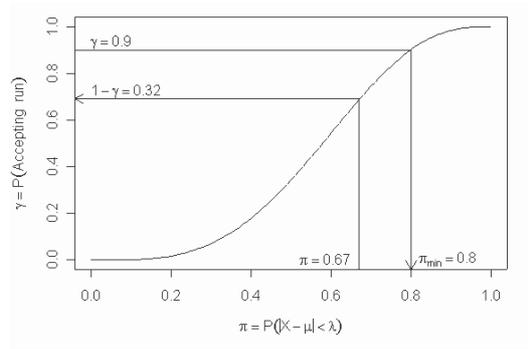


Figure 10 : Conciliating pre-study validation π_{\min} value with 4-6- λ rule

For the 4-6- λ rule we have $\pi_{\min} = 0.8$. This means that, in the pre-study validation experiment, the laboratory should demonstrate that, using validation experiment, it is expected that at least 80% of the measurements $X-\mu_T$ will likely lie within the acceptance limits $[-\lambda, \lambda]$ to ensure to accept the 4-6- λ rule in routine in 90% of the cases if the process remains valid ($\pi > \pi_{\min}$). This contrasts with the (intuitive) proposal frequently encountered in the literature [8, 17] that 4/6 or 66.7% of the measures must lie within the acceptance limits. Proposing 66.7% as value for π_{\min} as suggested by those authors can lead to reject up to 32% of the “valid” runs as can be seen on Figure 9. This was certainly not the intent of the authors and rather results from an erroneous interpretation of the Binomial distribution.

On the other hand, when the s - n - λ rule is not fixed in advance, as we recommend, it is easy to calculate the most adequate value of s for given π_{\min} , n and γ_{\min} . It consists of finding the maximal value of s ($0 \leq s \leq n$) such that:

$$P(Y \geq s | \pi = \pi_{\min}) > \gamma_{\min}$$

Figure 11 gives this optimal value of s for different values of n and π_{\min} when γ_{\min} is fixed at 0.9.

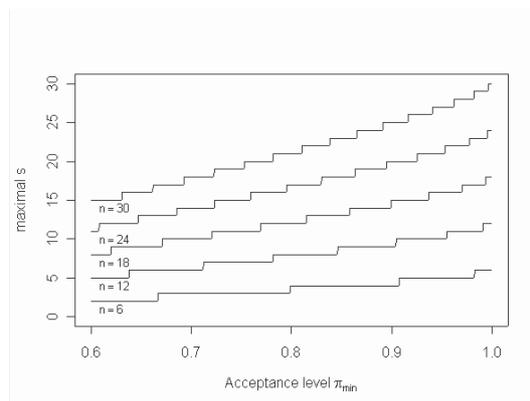


Figure 11 : Choice an optimal s value for a fixed n and pre-study acceptance level π_{\min} .

6. CONCLUSION

The introduction of the acceptance limits to evaluate the quantitative capability of an analytical method has triggered, on purpose, discussions on the values for the acceptance limits $[-\lambda, \lambda]$ in pre-study validation. Some authors [6] wrongly argued that if 15% was acceptable for the bias (trueness) and 15% was acceptable for the precision, then 30% was a reasonable limits for the total error, i.e. the sum of bias and precision. This is a doubtful interpretation of the regulatory requirements that was in Findlay's case duly motivated by the fact that ligand binding assays can hardly have a total error below 30%. However the very spirit of the FDA text [3] for bioanalytical method, was to expect that future measurements will fall within the $[-15\%, 15\%]$ limits and not the $[-30\%, 30\%]$ limits as clearly suggested by the 4-6-15 rule to be applied in routine use. Naively, many thought that having a (relative) bias smaller than 15% and precision better than 15% as observed during the pre-study validation would guarantee measures within $[-15\%, 15\%]$ limits during routine, as evaluated by the 4-6-15 rule. This is not the case and doesn't hold from a statistical point de view. In the case of bioanalytical method, it's clear today that acceptance limits during pre-study validation and during routine must be the same and then should be $[-15\%, 15\%]$ in this context. For other types of analytical methods, the limits $[-\lambda, \lambda]$ should be established depending on the intended use of the results.

Throughout this paper two test methods have been proposed to make a decision, after the pre-study validation, about the acceptability or rejection of an analytical method in order to minimize later both the consumer and laboratory risks. For limited sample size, as often encountered in laboratories, both approaches, the tolerance interval and the maximum likelihood method, behave similarly in the sense that they have comparable power curves over the (δ, σ) space. The tolerance interval approach however has asymptotical unsuitable behavior because the acceptance region of this test doesn't converge, as do the maximum likelihood method, to the true region of "good" methods. Practically for validation of methods, this difference is negligible, but the maximum likelihood should be preferred nevertheless, not only for its nice statistical properties, but also because it directly answers the very fundamental question: what is the expected proportion of measures that will fall within the acceptance limits in the future.

This paper, also, opens and proposes a solution to another necessary connection needed between the pre-study validation criteria and routine run acceptance rule: conciliating the probability levels or risks in order to minimize both consumer and laboratory risks. Stated differently, depending on the rule that will be used in routine for accepting runs, what

should be the minimal probability π_{\min} , as estimated from validation experiments, such that in routine most valid runs will be accepted ensuring that a very limited proportion of measurements will fall outside the acceptance limits.

The results show, that if we keep the 4-6- λ rule, the expected minimum proportion π_{\min} of measures that should fall within the acceptance limits $[-\lambda, \lambda]$ must be at least equal to 80% to guarantee that at least 90% of the runs will be accepted with the 4-6- λ rule when the measurement process remains valid. Taking 80% as π_{\min} value allows making pre-study and in-study decision rules consistent. But, as also shown, the 4-6-15 rule lacks of power and favors the laboratory over the consumer. The only solution to minimize both the consumer and the laboratory risks is to improve the in-study rule, by increasing both the number of QC samples (n) and increasing the number successful QC samples (s) and to adapt the π_{\min} value accordingly for the pre-study validation. But as shown in Table 3, this lead to very large and impracticable number of QC samples. More advanced techniques based directly on the quantitative measures and/or taking into account the history of the method's results are then recommended. Sampling plans for measures [26], scan statistics [27], moving type control charts (Cusum, Ewma) [28] are attractive solutions.

REFERENCES

- [1] International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Validation of analytical procedures, ICH-Q2A, Geneva 1995
- [2] International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Validation of analytical procedures: Methodology, ICH-Q2B, Geneva 1996
- [3] FDA Guidance for Industry: Bioanalytical Methods Validation, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), May 2001.
- [4] Hubert Ph, Nguyen J.J., Boulanger B., Chapuzet E., Chiap P., Cohen N., Compagnon P.A., Dewe W., Feinberg M., Lallier M., Laurentie M., Mercier N., Muzard G., Nivet C., Valat L. Harmonization of approaches for validation of quantitative analytical procedures: a SFSTP proposal. *Journal of Pharmaceutical and Biomedical Analysis* (2004) 36, 579-586
- [5] Smith W.C., Sittampalam G.S. Conceptual and Statistical Issues in the Validation of Analytical Dilution Assays for Pharmaceutical Applications. *Journal of Biopharmaceutical Statistics* (1998), 8(4), 509-532
- [6] Findlay J.W.A., Smith W.C., Lee J.W., Nordblom G.D., Das I., Desilva B.S., Khan M.N., Bowsher R.R. Validation of Immunoassays for Bioanalysis: a Pharmaceutical Industry Perspective. *Journal of Pharmaceutical and Biomedical Analysis* (2000), 21, 1249-1273
- [7] Miller K.J., Bowsher R.R., Celniker A., Gibbons J., Gupta S., Lee J.W., SWANSON S.J., Smith W.C., Weiner R. Workshop on Bioanalytical Methods Validation for Macromolecule: Summary report. *Pharmaceutical Research* (2001), 18(9), 1373-1383
- [8] Desilva B., Smith W., Weiner R., Kelley M., Smolec J, Lee B., Khan M., Tacey R., Hill H., Celniker A. Recommendations for Bioanalytical Method Validation of Ligand-binding Assays to Support Pharmacokinetic Assessments of Macromolecules, *Pharmaceutical Research* (2003), 20, 1885-1900

- [9] Caporal-Gautier J., Nivet J.M., Algranti P., Guilloteau M., Hlste M., Lallier M., N'guyen-Huu J.J., Russotto R. Guide de Validation Analytique – Rapport d'une Commission SFSTP – Méthodologie. STP Pharma Pratiques (1992), 2(4), 205-226
- [10] Hartmann C., Smeyers-Verbeke J., Massart D.L., McDowall R.D. Validation of Bioanalytical Chromatographic Methods. Journal of Pharmaceutical and Biomedical Analysis (1998), 17, 193-218
- [11] Shah V.P., Midha K.K., Findlay J.W.A., Hill H.M., Hulse J.D., McGilveray I.J., McKay G., Miller K.J., Patnaik R.N., Powell M.L., Tonelli A. VISWANATHAN C.T., Yacobi A. Bioanalytical Method Validation – A revisit with a Decade of Progress. Pharmaceutical Research (2000), 17(12), 1551-1557
- [12] B. Boulanger, W. Dewe, Ph. Hubert, E. Rozet, F. Moonen, B. Govaerts, M. Maumy. Conciliating objectives of analytical methods and objectives of validation: A statistical perspective. Inst. Validation Tech. December 7, 2005, Philadelphia
- [13] Hubert Ph., Chiap P., Crommen J., Boulanger B., Chapuzet E., Mercier N., Bervoas-Martin S., Chevalier P. Grandjean D., Lagorce P., LALLIER M., LAPARRA M.C., Laurentie M., Nivet C. The SFSTP Guide on the Validation of Chromatographic Methods for drug Analysis : from the Washington Conference to the Laboratory. Analytica Chimica Acta (1999), 391, 135–148
- [14] Boulanger B., Hubert PH., Chiap P., Dewe W. Objectives of Pre-Study Validation and Decision Rules. AAPS APQ Open forum, Washington, 2000
- [15] Boulanger B., Hubert Ph., Chiap P., Dewe W., Crommen J. Analyse Statistique des Résultats de Validation de Méthodes Chromatographiques, Journées GMP, Bordeaux, 2000.
- [16] Boulanger B., Chiap P., Dewe W., Crommen J., Hubert Ph. An analysis of the SFSTP guide on validation of chromatographic bioanalytical methods: progresses and limitations. Journal of Pharmaceutical and Biomedical Analysis (2003) 32, 753-765
- [17] Hoffman D. and Kringle R. Two-sided tolerance intervals for balanced and unbalanced random effects models. Journal of Biopharmaceutical Statistics (2005), 15 (2) 283 - 293.
- [18] Mee R., W. β -expectation and β -content tolerance limits for balanced one-way ANOVA random model. Technometrics (1984), 26(3), 251-254
- [19] Mee R., W. Estimation of the percentage of a normal distribution lying outside a specified interval. Communication in Statistics – Theory and Methods (1988), 17(5), 1465-1479
- [20] Chapuzet E., Mercier N., Bervioas-Martin S., Boulanger B., Chevalier P., Chiap P., Grandjean D., Hubert Ph., Lagorce P., Lallier M., Laparra M.C., Laurentie M., Nivet J.C., Méthodes chromatographiques de dosage dans les milieux biologiques : stratégie de validation. Rapport d'une commission SFSTP. STP Pharma Pratiques (1997), 7, 169-194.
- [21] Chapuzet E., Mercier N., Bervioas-Martin S., Boulanger B., Chevalier P., Chiap P., Grandjean D., Hubert Ph., Lagorce P., Lallier M., Laparra M.C., Laurentie M., Nivet J.C. Méthodes chromatographiques de dosage dans les milieux biologiques : stratégie de validation, exemple d'application de la stratégie de validation. STP Pharma Pratiques (1997), 8, 81-107.
- [22] Castaneda-Mendez K., Medical utility frequency, Clinical Chemistry, 33/2 (1987) 212-222.
- [23] Kringle R.O. An Assessment of the 4-6-20 Rule for Acceptance of Analytical Runs in Bioavailability, Bioequivalence, and Pharmacokinetic Studies. Pharmaceutical Research (1994), 11, 556-560
- [24] Duncan A., Quality control and industrial statistics, Fifth Edition, Irwin, Boston, 1986.

- [25] International Organization for Standardization. Sampling procedures for inspection by attributes. ISO/DIS (1995) 2859 (Part 0)
- [26] International Organization for Standardization. Sampling procedures for inspection by variables. ISO/DIS (2005) 3951 (Part 0)
- [27] Glaz J., Naus J., Wallenstein S. Scan Statistics, Springer, New York, 2001.
- [28] Montgomery D.C., Introduction to Statistical Quality Control, 4th ed., Wiley, New York, NY, 2004.