

Statistical tools and approaches to validate analytical methods: methodology and practical examples[★]

Soumia Belouafa^{*}, Fatima Habti, Saïd Benhar, Bouchaïb Belafkih, Souad Tayane, Souad Hamdouch, Ahmed Bennamara, and Abdelmjid Abourriche

Laboratory of Biomolecules and Organic Synthesis, Department of Chemistry, Faculty of Sciences Ben M'Sik, University Hassan II of Casablanca, Avenue Driss El Harti B.P 7955, Sidi Othmane Casablanca, Morocco

Received: 21 September 2016 / Accepted: 26 December 2016

Abstract. The analyst needs to know whether the result of measurement can be accepted with confidence or, on the contrary, rejected because it is wrong. It is essential, also, to know whether an analysis method is suitable for the intended use. Likewise, it is more important for the researcher to know if he can trust a new developed method and what are the criteria to respect to ensure its validity. The statistical tools allow us to address all these points. The experimental protocol applied in this work is based on a common methodology, inspired by regulatory guidelines regarding statistical data analysis in analytical method validation, to optimize the number of assays and satisfy the study of validation criteria. In order to better understand the statistical analysis of raw data, practical examples are treated for quantify: an active ingredient in pharmaceutical drugs, a heavy metal in fishery products, and a drug in seizures.

Keywords: statistical tests / analytical validation / quality control / measurement / stability

1 Introduction

The principle of validation of quantitative analytical procedures is widely spread today in all domains of activities where measures are made. The objective of validation of an analytical method is to demonstrate that the method is suitable for the intended use, such as evaluation of a known product for potency, and impurities. The intent of method validation is to provide scientific evidence that the analytical method is reliable and consistent before it can be used in routine analysis of product. The analytical method validation is governed by the International Conference on Harmonization (ICH) [1,2].

The key criteria for evaluation of an analytical method are: specificity, accuracy, precision, detection limit, quantitation limit, sensitivity, working range and linearity, robustness and recovery [3,4]. The ICH definitions for validation characteristics are indexed in Table 1.

The type of method and analytical technique used will determine the nature and extent of the validation studies required. The most common methods for validation are

identification, assay and impurities determination [5–8] (Tab. 2). Results for each applicable validation characteristic are compared against the selected acceptance criteria.

Besides ICH, other guidance can be referred for detailed information on the international requirements such as the US Food and Drug Administration guidance (FDA) [9–11] and the United States Pharmacopoeia (USP) [12].

The aim of this work is to furnish a conspectus of regulatory expectations related to statistical analysis and the survey of mutual statistical techniques used to analyze analytical method validation data with some examples.

The following two points are given:

- Statistics to analyze data for analytical method validation such as mean, standard deviation, confidence intervals, and linear regression.
- Regulatory guidelines regarding statistical data analysis in analytical method validation.

2 Statistical tools in analytical method validation

Statistical analysis of data obtained during a method validation should be performed to demonstrate validity of the analytical method. The primary parameters used for

[★] Paper was presented at the 6th International Metrology Conference–CAFMET 2016.

^{*} Corresponding author: sbelouafa@yahoo.fr

Table 1. Performance criteria for analytical method validation.

Validation characteristics	ICH definitions
Specificity	Specificity is the ability to assess unequivocally the analyte in the presence of components which may be expected to be present. Typically these might include impurities, degradants, matrix, etc.
Accuracy	The accuracy of an analytical procedure expresses the closeness of agreement between the value that is accepted either as a conventional true value or an accepted reference value and the value found. This is sometimes termed trueness.
Precision	The precision of an analytical procedure expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogenous sample under the prescribed conditions. Precision may be considered at three levels: repeatability, intermediate precision, and reproducibility.
Repeatability	Repeatability expresses the precision under the same operating conditions over a short interval of time. Repeatability is also termed intra-assay precision.
Intermediate precision	Intermediate precision expresses within-laboratories variations: different days, different analysts, different equipments, etc.
Reproducibility	Reproducibility expresses the precision between laboratories (collaborative studies usually applied to standardization of methodology).
Detection limit	The detection limit of an individual analytical procedure is the lowest amount of analyte in a sample that can be detected but not necessarily quantitated as an exact value.
Quantitation limit	The quantitation limit of an individual analytical procedure is the lowest amount of analyte in a sample that can be quantitatively determined with suitable precision and accuracy. The quantitation limit is a parameter of quantitative assays for low levels of compounds in sample matrices and is used particularly for the determination of impurities and/or degradation products.
Linearity	The linearity of an analytical procedure is its ability (within a given range) to obtain test results that are directly proportional to the concentration (amount) of analyte in the sample.
Range	The range of an analytical procedure is the interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy, and linearity.
Robustness	The robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate, variations in method parameters and also provides an indication of its reliability during normal usage.

Table 2. Validation characteristics according to the nature of the test.

Type of analytical procedure characteristics	Identification	Testing for impurities		Assay dissolution (measurement only) Content/potency
		Quantitative	Limit	
Accuracy	–	+	–	+
Precision				
Repeatability	–	+	–	+
Intermediate precision	–	+	–	+
Specificity (2)	+	+	+	+
Detection limit	–	–	+	–
Quantitation limit	–	+	–	–
Linearity	–	+	–	+
Range	–	+	–	+
Robustness	–	+	–	+

– signifies that this characteristic is not normally evaluated; + signifies that this characteristic is normally evaluated; (1) in cases where reproducibility (see glossary) has been performed, intermediate precision is not needed [8]; (2) lack of specificity of one analytical procedure could be compensated by other supporting analytical procedure(s); (3) may be needed in some cases.

the interpretation of analytical method validation results are the calculation of the mean (or average), standard deviation, relative standard deviation, confidence intervals, and regression analysis and the main tools are the F -test, t -test, and regression and correlation analysis. These calculations are characteristically performed using statistical software packages such as SPSS, R, and Minitab. The goal of statistical analysis is to summarize a collection of data that provides an understanding of the examined method characteristic. The statistical tools are described in the following.

2.1 Mean

The mean or average of a data set is the basic and the most common statistics used. The mean is calculated by adding all data points and dividing the sum by the number of samples. It is typically denoted by \bar{X} (X bar) and is computed using the following formula:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n},$$

where X_i are individual values and n is the number of individual data points.

2.2 Standard deviation

The standard deviation of a data set is the measure of the spread of the values in the sample set and is computed by measuring the difference between the mean and the individual values in a set. It is computed using the following formula:

$$S = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n - 1}},$$

where X_i is individual value, \bar{X} is the sample mean, and n is the number of individual data points.

2.3 Relative standard deviation

The relative standard deviation is computed by taking the standard deviation of the sample set multiplied by 100% and dividing it by the sample set average. The relative standard deviation is expressed as percent. Generally, the acceptance criteria for accuracy, precision, and repeatability of data is expressed in %RSD:

$$\%RSD = \frac{s}{\bar{X}} \times 100\%.$$

Note

When needed (e.g. for the F -test, $F = S_1^2/S_2^2$) the variance can, of course, be calculated by squaring the standard deviation:

$$V = s^2.$$

2.4 Confidence interval

Confidence intervals are used to indicate the reliability of an estimate. Confidence intervals provide limits around the sample mean to predict the range of the true population of the mean. The prediction is usually based on probability of 95%. The confidence interval depends on the sample standard deviation and the sample mean.

$$\text{Confidence interval for: } \mu = \bar{X} \pm \frac{zs}{\sqrt{n}},$$

where s is the sample deviation, \bar{X} is the sample mean, n is the number of individual data points, and z is constant obtained from statistical tables for z .

The value of z depends on the confidence level listed in statistical tables for z . For 95%, z is 1.96 [13]. For small samples, z can be replaced by t -value obtained from the Student's t -distribution tables [14]. The value of t corresponds to $n-1$.

For example, for:

- $n = 6$,
- mean = 101.82,
- standard deviation = 1.2,
- 95% confidence interval = 0.49.

So, this calculated confidence interval indicates that the range of the true population of the mean is between 101.33 and 102.31.

In brief, confidence interval can be applied in various ways to determine the size of errors (confidence) in analytical work or measurements: single determinations in routine work, certain calibrations, determinations for which no previous data exist, etc.

Note

Student's t -distribution becomes closer and closer then the standard normal distribution as the degrees of freedom get larger. The t -distribution will be indistinguishable from the standard normal distribution, as can be seen in Figure 1.

On the other hand, we can use the table for the standard normal distribution instead of the table for the t -distribution (Fig. 2).

Figure 2 shows that 68% of the data fall in the range $\bar{X} \pm s$, 95% in the range $\bar{X} \pm 2s$, and 99.7% in the range $\bar{X} \pm 3s$.

When the distribution is skewed statistical treatment is more complicated.

2.5 Regression analysis

Regression analysis is used to evaluate a linear relationship between test results. A linear relationship is, in general, evaluated over the range of the analytical procedure. The data obtained from analysis of the solutions prepared at a range of different concentration levels is habitually investigated by plotting on a graph.

Linear regression evaluates the relationship between two variables by fitting a linear equation to observed data. A linear regression line has an equation of the form $Y = b_0 + b_1 \times X$, where X is the independent variable and Y is the dependent variable. The slope of the line is b_1 ,

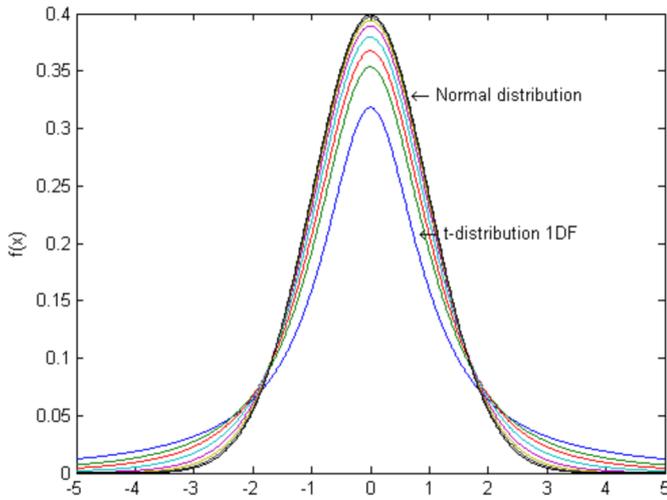


Fig. 1. Student's *t*-distribution.

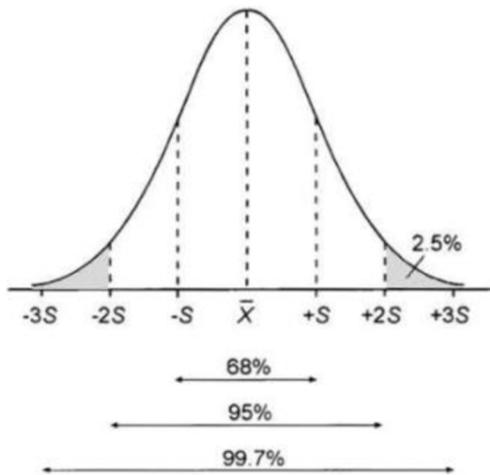


Fig. 2. A Gaussian or normal distribution.

and b_0 is the intercept (the value of y when $x=0$). The statistical procedure of finding the “best-fitting” straight line is to obtain a line through the points to minimize the deviations of the points from the prospective line. The best-fit criterion of goodness of the fit is known as the principle of least squares. In mathematical terms, the best fitting line is the line that minimizes the sum of squares of the deviations of the observed values of Y from those predicted.

In Figure 3, the data clearly shows a linear relationship.

Calculations used to compute y -intercept and the line slope are as follows:

$$b_1 = \frac{S_{xy}}{S_{xx}}, \quad b_0 = \bar{Y} - b_1 \times \bar{X},$$

$$S_{xx} = \sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}, \quad S_{xy} = \sum_{i=1}^n (X_i Y_i) - \frac{\sum_{i=1}^n X_i \times \sum_{i=1}^n Y_i}{n}.$$

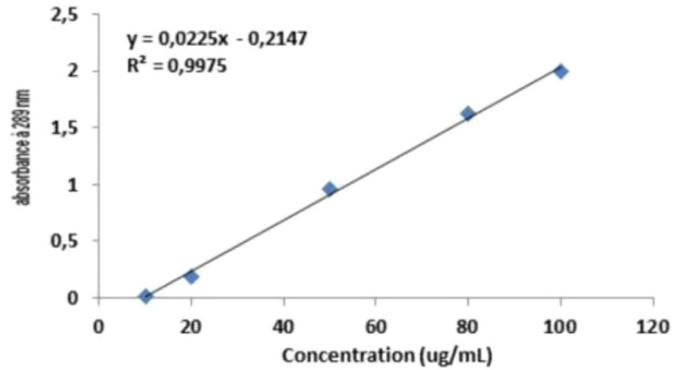


Fig. 3. Linear regression analysis of the calibration curve of Ecstasy in a seizure by UV-Vis spectrophotometer. Each point is the mean \mp SD of three experiments [15].

So, the equation of the line, for data listed in Figure 2, is $Y = -0.2147 + 0.0225 X$.

Once a regression model has been fit to a group of data, examination of the residuals (the deviations from the fitted line to the observed values) allows investigation of the validity of the assumption that a linear relationship exists. Plotting the residuals on the y -axis against the independent variable on the x -axis reveals any possible non-linear relationship among the variables or might alert to investigate outliers.

The other important calculations that are ordinarily reported are the coefficient of determination (r^2) and linear correlation coefficient (r). The coefficient of determination (r^2) measures the proportion of variation that is explained by the model. Ideally, r^2 should be equal to one, which would indicate zero error. The correlation coefficient (r) is the correlation between the predicted and observed values. This will have a value between 0 and 1; the closer the value is to 1, the better the correlation. Any data that form a straight line will give high correlation coefficient; therefore, extra caution should be taken when interpreting correlation coefficient. Additional statistical analysis is recommended to provide estimates of systematic errors, not just the correlation or results. For example, in method comparison studies, if one method gives consistently higher results than the other method, the results would show linear correlation and have a high correlation coefficient, although a difference between the two methods.

Equations used to determine the coefficient of determination (r^2) and the correlation coefficient (r) are listed in Table 3.

2.6 The hypothesis tests

The hypothesis tests are intended to verify if the experimental data are consistent with certain theoretical hypothesis.

- The null hypothesis symbolized by H_0 considers that the two elements or series of elements are equal.
- The second step consists in measuring the deviation between different characteristics.
- The third step is to calculate the probability P to have this deviation if H_0 is true.

Table 3. Regression summary.

Data	Descriptions	Mathematical expressions
Equation of the line	The relationship between the concentration (X) and the response (Y)	$Y = b_0 + b_1 \times X$
Intercept (b_0)	The value of Y when X equals zero	$b_0 = y - b_1 \times X$
Slope (b_1)	The slope of the line relate to the relationship between concentration and response	$b_1 = \frac{s_{xy}}{s_{xx}}$
Standard error (b_0) (SE intercept)	The standard error of the intercept can be used to calculate the required confidence interval	$SE_{b_0} = s \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i + \bar{X})^2}}$
	95% confidence interval	–
Standard error (b_1) (SE slope)	The standard error of the slope can be used to calculate the required confidence interval	$SE_{b_1} = \frac{s}{\sqrt{\sum (X_i + \bar{X})^2}}$
	95% confidence interval	–
Coefficient of determination (r^2)	The square of the correlation coefficient	$r^2 = \frac{SSM}{SST}$
Correlation coefficient (r)	The correlation between the predicted and observed values. This will have a value between 0 and 1; the closer the value is to 1, the better the correlation	$\sqrt{r^2}$
Regression SS	The regression sum of squares is the variability in the response that is accounted for by the regression line	$SS \text{ total} - \sum (X_i)^2$
Residual SS (the error sum of squares)	The residual sum of squares is the variability about the regression line (the amount of uncertainty that remains)	$SS \text{ total} - SS \text{ regression}$
Total SS	The total sum of squares is the total amount of variability in the response	$\sum (Y_i - Y)^2$

– The fourth step is to draw conclusions that are required:

If P is large, we admit that H_0 is plausible, on the other side if P is small, the deviation is incompatible with H_0 . The value limit of P that is fixed to determine if P is large or small is the level of confidence or significance level (usually we chose $P = 0.95$ as level of confidence ($\alpha = 0.05$ as significance level)).

Four situations are possible:

- Acceptance of H_0 true.
- Rejecting true H_0 : first error species (α).
- Acceptance false H_0 : second error species (β).
- Rejecting false H_0 .

2.7 Other statistical tools

Other statistical tools used in method validation include comparative studies using Student's t -test, Fisher's test, analysis of variation (ANVA), design of experiments, and assessment of outliers. Information on these statistical tools can be obtained from references on statistics suggested in the reference section.

3 Validation characteristics

Statistics recommended by ICH, USP, and FDA to evaluate method suitability are presented below.

3.1 Specificity/selectivity

Specificity is a quantitative indication of the extent to which a method can distinguish between the analyte of interest and interfering substances on the basis of signals produced under actual experimental conditions. Random interferences should be determined using representative blank samples. Table 4 presents an example of specificity of a developed method for determination of ornidazole in pharmaceutical formulations [16].

The % recovery was found to be in the range of 99.4–100.4%, hence there were no interferences of the excipients and additives which indicate the selectivity of the developed method.

3.2 Accuracy

Accuracy refers to closeness of agreement between the true value of the analyte concentration and the mean result obtained by applying experimental procedure to a large number of homogeneous samples. It is related to systematic error and analyte recovery. Systematic errors can be established by the use of appropriate certified reference materials (matrix-matched) or by applying alternative analytical techniques. Table 5 provides an example of

accuracy data assessment of an analytical method for atomic absorption spectrometry analysis of Pb in bivalve molluscs [17].

In this example, the calculated accuracy error is less than 2. So, it is considered insignificant [18]. Therefore, the uncertainty associated to accuracy of the method is equal to the uncertainty of the reference material used for testing accuracy study.

Table 4. Determination of ornidazole in presence of excipients and additives [16].

Excipients and additives	Amount (mg)	%Recovery of ornidazol \mp RSD*
Titanium dioxide	40	100.1 \mp 0.76
Talc	50	100.4 \mp 0.81
Hydroxypropyl methylcellulose	50	100.14 \mp 0.9
Corn starch	50	99.98 \mp 0.93
Methylhydroxyethylcellulose	50	100.03 \mp 0.94
Magnesium stearate	40	99.4 \mp 0.84
Microcrystalline cellulose	30	99.66 \mp 0.89

%Recovery = Mean value/Added amount \times 100; RSD: relative standard deviation ($\%RSD = (s/\bar{X}) \times 100\%$).

* Average of five determination.

Table 5. Result of an accuracy study [17].

Number (n)	10
Mean (d)	0.989
Standard deviation (Sd)	0.072
Concentration of reference material (C) ($\mu\text{g/g}$)	1.19
Uncertainty of reference materials	0.18
Relative error (%)	16.879
Accuracy (%)	83.121
Accuracy error	1.036

Table 6. Example of results obtained for a precision study [15].

	Repeatability		Replicability
	1 st day	2 nd day	
Number (n)	10	10	10
Mean (d)	48.2531	49.05252	50
Standard deviation (Sd)	0.0264673	0.056026	0.0168736
CV%	0.054851	0.11421617	0.03356
Precision	96.38%	98.068%	99.45%
T_{crit} (95;9)	2.262	2.262	
Error	3.62	1.931562	
Confidence interval	48.253 \mp 0.019	49.288 \mp 0.196	
Mean of means	48.65281		
Mean of standard deviation	0.04124665		

CV% = %RSD; Precision% = $2.8 \times s_r$, with s_r : standard deviation of these results.

3.3 Precision

Comparison of results obtained from samples prepared to test the following conditions:

- Repeatability expresses the precision under the same operating conditions over a short interval of time. Repeatability is also termed intra-assay precision.
- Intermediate precision expresses within-laboratories variations: different days, different analysts, different equipments, etc.
- Reproducibility expresses the precision between laboratories (collaborative studies, usually applied to standardization of methodology).

Both repeatability and the reproducibility are expressed in terms of standard deviation and are generally dependent on analyte concentration. It is thus recommended that the repeatability and within-laboratory reproducibility are determined at different concentrations across the working range, by carrying out 10 repeated determinations at each concentration level. As stipulated by Horwitz and Albert [19], the variability between laboratories is the dominating error component in the world of practical ultratrace analysis. They conclude that a single laboratory cannot determine its own error structure, except in the context of certified reference materials or consensus results from other laboratories.

Table 6 provides an example of a typical data analysis summary for the evaluation of a precision study for an analytical method for the quantitation of Ecstasy in seizures by UV-Vis spectrophotometer [15]. In this example, the method was tested in two different laboratories by two different analysts on two different instruments.

The standard deviations and the percentage recoveries (not more than 2%) indicate good precision of the method [20].

In the example provided in Table 6, precision is determined for a number of different levels during validation, which include system precision, repeatability, intermediate precision, and reproducibility. The system precision is evaluated by comparing the means and relative standard deviations. Reproducibility is assessed by means of an inter-laboratory assay. The intermediate precision is

established by comparing analytical results obtained when using different analysts and instruments and performing the analysis on different days. The repeatability is assessed by measuring the variability in the results obtained when using the analytical method in a single determination. In each case, the mean and % of RSD is calculated and compared to the established acceptance criteria.

3.4 Detection limit

The ICH guideline mentions several approaches for determining the detection limit: visual inspection, signal-to-noise, and using the standard deviation of the response and the slope. The detection limit and the method used for determining the detection limit should be presented. If visual evaluation is used, the detection limit is determined by the analysis of samples with known concentration of analyte and by establishing the minimum level at which the analyte can be reliably detected. The signal-to-noise ratio is performed by comparing measured signals from samples with known low concentrations of analyte with those of blank. When the detection limit is based on the standard deviation of the response and the slope, it is calculated using the following equation.

$$\text{LDM} = \frac{3.3\sigma}{S},$$

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

The limit of detection is usually expressed as the analyte concentration corresponding to the sample blank plus three sample standard deviations, based on 10 independent analyses of sample blanks.

3.5 Quantitation limit

The ICH guideline mentions several approaches for determining the quantitation limit: an approach based on visual evaluation, an approach based on signal-to-noise and an approach based on the standard deviation of the response and the slope. The quantitation limit and the method used for determining the quantitation limit should be presented. When the quantitation limit is based on the standard deviation of the response and the slope, it is calculated using the equation below:

$$\text{LQM} = \frac{10\sigma}{S},$$

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

Limit of quantitation is set by various conventions to be five, six or ten standard deviations of the blank mean. It is also sometimes known as the limit of determination.

The LDM defined as the lowest detectable concentration on the calibration curve where both accuracy and precision should be within the maximum tolerable CV of 5.32%, was deemed to be 0.406 $\mu\text{g/mL}$ (Tab. 7). This LDM is adequate for the analysis of forensic samples, as this value falls within the concentration range of MDMA in many ecstasy tablets analyzed. Furthermore, Ration of Conformity (6.26) is between 4 and 10 [21], so LDM is validated.

Table 7. Limit of detection and limit of quantification data of a method for the quantitation of Ecstasy in seizures by UV-Vis spectrophotometer [15].

Number (n)	10
Mean (d)	2.54293
Standard deviation (Sd)	0.1352983
CV%	5.32
Precision	82.02
LDM	0.406
LQM	1.35
R (ratio of conformity)	6.26
$R = \bar{X}/\text{LDM}$.	

Table 8. Example of linear regression analysis of standard solutions for cocaine by a HPLC/MS/MS method [22].

Range (ng/mL)	Equation	r
0.1–500	$Y = 5.43 \times 10^3 X - 1.05 \times 10^5$	0.983
0.1–250	$Y = 7.82 \times 10^3 X + 7.27 \times 10^4$	0.9905
1–100	$Y = 2.71 \times 10^4 X + 4.59 \times 10^4$	0.9966

3.6 Working and linear ranges

For any quantitative method, there is a range of analyte concentrations over which the method may be applied (example in Tab. 8). At the lower end of the concentration range the limiting factor is the value of the limit of detection and/or limit of quantification. At the upper end of the concentration range limitations will be imposed by various effects depending on the detection mechanism.

Within this working range there may exist a linear range, within which the detection response will have a sufficiently linear relation to analyte concentration. The working and linear range may differ in different sample types according to the effect of interferences arising from the sample matrix. It is recommended that, in the first instance, the response relationship should be examined over the working range by carrying out a single assessment of the response levels to at least six concentration levels. To determine the response relationship within the linear range, it is recommended that three replicates (examples in Tab. 9) are carried out at each of at least six concentration levels.

If there is a linear relationship, test results should be evaluated by linear regression analysis. The correlation coefficient, y -intercept, and slope of the regression line and residual sum of squares should be submitted with a plot of data.

3.7 Sensitivity

Sensitivity is the measure of the change in instrument response which corresponds to a change in analyte concentration. Where the response has been established as linear with respect to concentration, sensitivity corresponds to the gradient of the response curve.

Table 9. Example of linear regression analysis of standard solutions for a 1–100 ng/mL concentration range, performed in 3 different days for cocaine by a HPLC/MS/MS method [22].

Day	Equation	Range (ng/mL)	r
1	$Y = 1.54 \times 10^4 X + 6.55 \times 10^3$	1–100	0.9919
2	$Y = 1.53 \times 10^4 X$	1–100	0.9921
3	$Y = 2.71 \times 10^4 X + 4.59 \times 10^4$	1–100	0.9966

Table 10. Result of recovery study of cadmium quantification in bivalve molluscs using an atomic absorption spectrometry with graphite furnace method [23].

C	C_f	$C_f - C$	C_a	%Recovery
21.5380	31.8610	10.323	10	103.23
25.5430	33.4810	7.938	10	79.38
16.9935	28.7975	11.804	10	118.04
15.5160	23.6830	8.167	7	116.67
9.8015	16.3830	6.581	5	131.62
			Mean	109.788
			Sd	19.747
			α	0.05
			T_{crit}	2.78
			T_{obs}	12.432

C : concentration before standard additions; C_f : concentration after standard additions; C_a : concentration of standard additions.

Recovery study reported in Table 10 shows that the calculated percentage recovery varies between 79.38% and 131.62%. These percentages were validated by Student's t -test [24].

3.8 Robustness

This is a measure of how effectively the performance of the analytical method stands up to less than perfect implementation. In any method there will be certain parts which will severely affect the method performance, unless they are carried out with sufficient care. These aspects should be identified and, if possible, their influence on the method performance should be evaluated using the robustness tests. These tests provide important information for the evaluation of the measurement uncertainty. The methodology for evaluating uncertainty given in the IS Guide [25] relies on identifying all parameters that may affect the result and on quantifying the uncertainty contribution from each source. This is very similar to procedures used in robustness tests which identify all the parameters likely to influence the result and determine the acceptability of their influence through control. If carried out with this in mind, the robustness tests can provide information on the contribution to the overall uncertainty

from each of the parameters studied. Mean and %RSDs are compared against the acceptance criteria to evaluate impact of changing experimental parameters.

The ICH guidelines suggest detailed validation schemes relative to the purpose of the methods. It lists recommended data to report for each validation parameter. Acceptance criteria for validation must be based on the previous performances of the method, the product specifications and the phase of development.

The path to validation forms a continuum. It begins in the early phases of development as a set of informal experiments that establishes the soundness of the method for its intended purpose. It is expanded throughout the regulatory submission process into a fully-documented report that is required for commercial production. It is repeated whenever there is a significant change in instrumentation, method, specifications and process [1–12].

4 Uncertainty components

All the features that are used to describe the performance of a method of analysis do not directly impact on the evaluation of the uncertainty of the analytical result. Table 11 summarizes the impact of different characteristics on the evaluation of uncertainty.

5 Challenges

To succeed development and validation of a method, a total approach is recommended. A common challenge encountered during methods development and validation is that methods are generally developed by the research and development department while validation is generally the responsibility of a quality assurance and quality control. It is important that all the groups work as one staff. The transfer of analytical methods from one group to another so becomes an important step for ensuring that the proper validation is in place to justify its intended use. In order that the method will be run by several groups during its progress from development to validation, it must be robust. A shared weakness in development and validation of methods is that the methods are not robust quite. If robustness is not built into methods early in development then the results are likely to lack efficiency in quality testing and encounter lengthy and complicated validation process. The design and execution of the studies requires thorough knowledge of the product being tested as well as a good understanding of the analysis technique. New regulatory guidelines are being published [26–30] that govern the expectations of regulatory agencies throughout the world for methods development and validation. There is need to meet current regulatory standards. From a simple method improvement to a complete redevelopment and subsequent implementation is tedious task. For this reason, one must be alert to current trends in regulatory guidelines and to adopt a proactive approach to changes that may affect development and validation programs. Finally, one of the key requirements for methods validation

Table 11. Impacts of characteristics on uncertainty.

Characteristics	Impacts on uncertainty
Specificity/selectivity	Effects related to the selectivity and specificity are overall measured when the evaluation of the uncertainty component associated with accuracy.
Precision	These variances are among the most significant elements for the assessment of uncertainty.
Linearity, sensitivity	1 st case: If the statistical test shows that the model is linear, no effect on the uncertainty. 2 nd case: If the test shows that the model is not linear, then we may either reduce the measurement field or add a component (of non-linearity) in the evaluation of uncertainty. This component can be the maximum deviation to the model.
Limit of detection	This feature does no effect on the uncertainty assessment; it serves to define the field of use of the method.
Robustness	For a factor whose change scope is limited to $\mp a$, then the corresponding uncertainty is: $u_a = c(a/\sqrt{3})$, where c is the sensitivity coefficient of the response due to variations of the factor a .
Accuracy	In chemical analysis, it is not applied corrections of accuracy. The goal during the development of the method is to check that there is no significant systematic error. If: $E_n \leq 2$; $u_{\text{justesse}} = u_{\text{Ref}}$ or if: $ \bar{x}_i - x_{\text{Ref}} \leq L$; $u_{\text{justesse}} = \frac{L}{\sqrt{3}}$

is that only well-characterized reference materials with proper documented purities should be used during method validation.

6 Conclusion

This work discusses applications of statistics in analytical method validation. Method validation helps to validate the analytical method for a range of concentrations so that the change in formulation or concentration do not require additional validation. Once the methods have been developed, qualified and validated the impact they have on out-of-specification rates and process capability needs to be quantified and evaluated to determine its effectiveness for future use.

The statistical methods used during the analytical method validation involve the basic knowledge of statistics. Even though there are statistical packages available to perform statistical calculations, it is important to understand the mathematical basis behind these calculations. It is essential for the analysts to be familiar with the basic statistical elements. Statistics used for validation data interpretations should be incorporated into the company's standard procedure and specified in the validation protocol and report.

This work is dedicated to the memory of Professor M'Hamed CHARROUF, died on November 29, 2014.

References

1. International Conference on Harmonization (ICH), *Q2A: Text on validation of analytical procedures*, March (1995)
2. International Conference on Harmonization (ICH), *Q2B: Validation of analytical procedures: methodology*, May (1997)
3. H. Holcombe, ed., The fitness for purpose of analytical methods, in *A laboratory guide to method validation and related topics* (EURACHEM Guide, LGC, Teddington, 1998)
4. Food and Agriculture Organization (FAO) and International Atomic Energy Agency (IAEA), *Food and nutrition paper 68: Validation of analytical methods for food control* (FAO/UN, Rome, 1998)
5. K.S. Bansal, T. Layloff, E.D. Bush, M. Hamilton, E.A. Hankinson et al., Qualification of analytical instruments for use in the pharmaceutical industry: a scientific approach, *AAPS PharmSciTech* **5**, 1–8 (2004)
6. P. Bedson, M. Sargent, The development and application of guidance on equipment qualification of analytical instruments, *Accredit. Qual. Assur.* **1**, 265–274 (1996)
7. O. McPolin, *The validation of analytical methods for pharmaceutical analysis* (Mourne Training Services, UK, 2009)
8. FDA, *Guidance for industry: analytical procedures and method validation, chemistry, manufacturing, and controls documentation* (U.S. Department of Health and Human Services, 2000)
9. U.S. FDA, *Guidance for submitting samples and analytical data for methods validation* (Center for Drugs and Biologics, Department of Health and Human Services, Rockville, MD, USA, 1987)
10. U.S. Center for Drug Evaluation and Research, *Reviewer guidance: validation of chromatographic methods* (1994)
11. U.S. FDA DHHS, *21 CFR Parts 210 and 211, Current good manufacturing practice of certain requirements for finished pharmaceuticals, proposed rule* (1996)
12. <1225> Validation of Compendial Methods, U.S. Pharmacopoeia 26-National Formulary 21 (United States Pharmacopoeial Convention, Rockville, MD, 2003)
13. S. Bolton, *Pharmaceutical statistics practical and clinical applications* (Marcel Decker, Inc., New York, NY, 2004), 5th ed., p. 558, Table IV.2

14. S. Bolton, *Pharmaceutical statistics practical and clinical applications* (Marcel Decker, Inc., New York, NY, 2004), 5th ed., p. 561, Table IV.4
15. B. Belafkih, S. Belouafa, M. Charrouf, A. Bennamara, A. Skalli, F. Slaoui, A. Abourriche, *Anal. Chem.* **15** (6), 219–224 (2015)
16. F. Saadoune, S. Belouafa, T. Ainane, M. Charrouf, A. Bennamara, A. Abourriche, *Anal. Chem.* **14** (10), 404–408 (2014)
17. S. Benhar, S. Belouafa, A. Chafik, M. Charrouf, R. Chfiri, A. Semlali, M. Maghraoui, A. Bennamara, A. Abourriche, *Anal. Chem.* **15** (6), 213–218 (2015)
18. LNE C370 X18: Guide méthodologique pour l'estimation des incertitudes en analyse chimique du laboratoire national d'essais LNE/France
19. W. Horwitz, R. Albert, Reliability of the determinations of polychlorinated contaminants (biphenyls, dioxins, furans), *J. AOAC Int.* **79**, 589–621 (1996)
20. Rapport de Synthèse : étude de la validation AFAQ AFNOR Certification du test REBECCA + EB pour le dénombrement des Entérobactéries selon la norme NF ISO16140
21. The Commission of the European Communities, Commission decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation, *Off. J. Eur. Comm.* **L221**, 8–36 (2002)
22. B. Belafkih, S. Belouafa, M. Charrouf, A. Bennamara, A. Skalli, A. Abourriche, *J. Chem. Pharm. Res.* **7** (2), 940–947 (2015)
23. S. Benhar, S. Belouafa, A. Chafik, M. Charrouf, R. Chfiri, A. Semlali, M. Maghraoui, A. Bennamara, A. Abourriche, *ScienceLib Editions Mersenne* **6** (2014), No. 141101
24. Guide pour la validation des méthodes d'essai chimico-physiques et l'évaluation de l'incertitude de mesure ; Document N° 324, Edition février, rév. 02 (2013)
25. International Organization for standardization (ISO), *ISO 13752: Air Quality – Assessment of uncertainty of a measurement method under field conditions using a second method as reference* (ISO, Geneva, 1998)
26. M.L.J. Weitzel, The estimation and use of measurement uncertainty for a drug substance test procedure validated according to USP 1225, *Accredit. Qual. Assur.* **17** (2), 139–146 (2012)
27. R.K. Burdick, D. LeBlond, D. Sandell, H. Yang, Statistical methods for validation of procedure accuracy and precision, *Pharmaceut. Forum* **39** (3) (2013)
28. P. Nethercote, J. Ermer, Quality by design for analytical methods: implications for method validation and transfer, *Pharm. Technol.* **36** (10), 74–79 (2013)
29. T.A. Little, Design of experiments for analytical method development and validation, *BioPharm Int.* **27** (3) (2014)
30. N.V.V.S.S. Raman, U.R. Mallu, H.R. Bapatu, Analytical quality by design approach to test method development and validation in drug substance manufacturing, *J. Chem.* 1–8 (2015), Article ID: 435129

Cite this article as: Soumia Belouafa, Fatima Habti, Saïd Benhar, Bouchaïb Belafkih, Souad Tayane, Souad Hamdouch, Ahmed Bennamara, Abdelmjid Abourriche, Statistical tools and approaches to validate analytical methods: methodology and practical examples, *Int. J. Metrol. Qual. Eng.* **8**, 9 (2017)