

Validation of Analytical Methods

By Ricard Boqué, Alicia Maroto, Jordi Riu and F. Xavier Rius*

Department of Analytical Chemistry and Organic Chemistry
University Rovira i Virgili
Pl. Imperial Tàrraco, 1, 43005-Tarragona, Spain

CONTENTS

1. Introduction
2. Validation and fitness-for-purpose
3. Analytical requirements and performance criteria
 - 3.1. Trueness
 - 3.1.1. Certified reference materials (CRM), reference materials and in-house materials.
 - 3.1.2. Reference methods.
 - 3.1.3. Proficiency testing
 - 3.1.4. Spiked samples
 - 3.1.5. Assessment of trueness in restricted concentration ranges
 - 3.1.6. Assessment of trueness in large concentration ranges. Recovery studies
 - 3.2. Precision and uncertainty
 - 3.2.1. Calculation of the within-laboratory precision
 - 3.2.2. Calculation of the repeatability limit
 - 3.2.3. Calculation of the "internal reproducibility" limit
 - 3.3. Uncertainty
 - 3.3.1. Uncertainty of the assessment of trueness, U_{trueness}
 - 3.3.2. Uncertainty of sample pretreatments, $U_{\text{pretreatments}}$
 - 3.3.3. Uncertainty of other terms
 - 3.4. Linear and working ranges
 - 3.5. Limit of detection (LOD)
 - 3.5.1. Estimating the standard deviations of the net concentrations
 - 3.5.2. Alternatives for LOD calculation
 - 3.6. Limit of quantification (LOQ)
 - 3.7. Selectivity (specificity)
 - 3.8. Sensitivity
 - 3.9. Ruggedness (or robustness)
4. The basic principles of method validation
5. Performing the method validation
6. Inter-laboratory comparison as a means for method validation
7. In-house validation
8. Reporting method validation
- References

RESUMEN

Validación de métodos analíticos.

En este artículo se discute el concepto de validación del método, se describen los elementos que la componen y se explica la fuerte relación entre la validación y las características de ajuste. El método de validación se basa en el cumplimiento de una serie de requerimientos, se explica como seleccionar esos requerimientos, la forma en que se suministran evidencias, y que trabajo se debe llevar a cabo en el laboratorio. También se describen, los principios básicos del método de validación y los diferentes caminos para validar una metodología, tanto en la comparación entre laboratorios o como cuando se lleva a cabo una validación dentro del laboratorio.

PALABRAS-CLAVE: Características de ajuste - Criterios de funcionamiento - Validación del método - Validación dentro del laboratorio.

SUMMARY

Validation of analytical methods.

In this paper we shall discuss the concept of method validation, describe the various elements and explain its close relationship with fitness for purpose. Method validation is based on the assumption that a series of requirements are fulfilled and we shall explain how these requirements are selected, the way in which evidence is supplied and what work has to be carried out in the laboratory. The basic principles of method validation and the different ways to validate a methodology, by inter-laboratory comparison or performing an in-house validation, are also described.

KEY-WORDS: *Fitness-for-purpose - In-house validation - Method validation - Performance criteria.*

1. INTRODUCTION

Analytical information can be used for a variety of purposes: to take decisions involving the control of the manufacturing process of a product, to assess whether a product complies with regulatory limits, to take decisions about legal affairs, international trade, health problems or the environment. Consequently, the producer of analytical results must not only ensure that they are reliable but also that they have all the elements that give confidence to the user.

Quantitative analytical results are expressed as "estimated value \pm U" where "estimated value" is the estimate of the true value obtained when the analytical method is applied to the test sample. "U" represents the uncertainty associated to the estimated value. Nowadays, the users of test results are also interested in other characteristics. They also require, for example, that the estimated value be repeatable under certain conditions, that the analytical method provide similar results when the experimental conditions are slightly modified and that the test method be capable of quantifying very low concentrations of the analyte.

Since it is materially impossible to guarantee these characteristics for every single test result, the international analytical institutions have developed a system that ensures the quality of the results if they have been obtained under certain conditions. This quality assurance system, although highly developed nowadays, is in continuous evolution so that it can adapt to the new requirements of society.

This system is based on a few fundamental concepts:

1. The laboratories that produce analytical information must operate under quality assurance principles. The best way to guarantee that they do is to ensure that they are accredited by the ISO/IEC Guide 17025 (1999).

2. The laboratory must participate in proficiency testing schemes, which should be designed and conducted in accordance with the "International harmonized protocol for proficiency testing of analytical laboratories" (Thompson 1993).

3. The laboratory must use internal quality control procedures which comply with the "Harmonized guidelines for internal quality control in analytical chemistry" (Thompson 1995).

4. The laboratory must use validated methods of analysis.

In other words, if the laboratories that produce the results are compared to other similar laboratories and can show that they use good tools properly (validated methods under quality assurance conditions), that they internally control their processes and demonstrate proficiency by comparing their performance to similar laboratories, there are reasons to think that the general quality of the laboratories is translated to their individual results.

In this paper, we shall discuss the concept of validated methods of analysis and explain their different elements after we establish the close relationship between validation and fitness for purpose. We shall also describe the analytical requirements and performance characteristics and provide useful practical information such as which basic norms to follow in a validation process or how to conduct validation studies in house or by using inter-laboratory comparison studies.

2. VALIDATION AND FITNESS-FOR-PURPOSE

Validation is defined as the confirmation by examination and provision of objective evidence that the particular requirements for a specified intended use are fulfilled (ISO 1994b).

This definition implies that analytical methods should be validated taking into account the requirements of specific applications. Therefore, it is misleading to think that there is a single process of method validation that will demonstrate that a method fulfils a list of requirements and that it can be used in a general way for many different applications. On the contrary, the definition implies that the specific requirements for an intended use should be defined before the performance capabilities of the method are specified.

Fitness for purpose, defined by IUPAC in the Orange Book as "degree to which data produced by

a measurement process enables a user to make technically and administratively correct decisions for a stated purpose", is more related to the results of the analytical method than to the method itself. In a very recent joint project between ISO and AOAC International, IUPAC states that fitness-for-purpose is not only related to statistically-based performance criteria but also to practicability and suitability criteria (IUPAC 1999a). So, it is important to evaluate, among other characteristics, the ease of operation or the cost. In fact, method validation is the means that analysts have of demonstrating that the method is fit-for-purpose, showing to the customers that they can trust the reported results.

This present concept of method validation introduces the need for versatility in the laboratory and, therefore, compels analysts to adapt the analytical methodologies to the different problems to be solved. For instance, there is no need to calculate the limit of detection for a method used to determine oleic acid in olive oils but there is a need when another method is used to determine trace residues in the same oil sample. On the other hand, the need for versatility introduces some complexity. The requirements should be defined by first taking into account the needs of the end user (something that is not evident in all cases) and they should then be confirmed by using calibrated instruments, trained analysts and all other quality assurance conditions.

3. ANALYTICAL REQUIREMENTS AND PERFORMANCE CRITERIA

Once the specific analytical problem has been clearly defined, the producer of the analytical information should ideally be able to agree with the customer about the specific analytical requirements.

For instance, if a laboratory carries out the quality control of a product, legislation often determines particular limits for certain parameters. The analyst must ensure that the test results will reflect the characteristics of the analysed sample and not those of the performance method used to obtain the results. Therefore, the estimated value must not only be traceable but also be associated to an appropriate level of uncertainty so as to ensure the compliance or not of that product with the regulations. Analysts also know that they must use a method whose limit of quantification for that parameter is well below the legal limit. At the same time, the rapidity with which results are provided, the cost of the analysis and the availability of trained staff should also be evaluated.

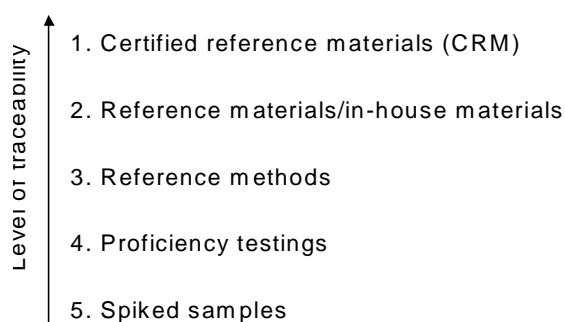
In fact, the end user is often not well defined. The customer may take different forms: it could be an internal or external client to the own company, it could be the administration that is usually guided by the current legislation or even a written standard that must be followed. It is not unusual to meet customers

who are unable to define the requirements for a specific analysis. They generally require that the cost of the analysis be reduced to a minimum and that the results be available very quickly but they find it difficult to understand the actual significance of the uncertainty values. In these cases, analysts must take decisions for the end users and, simultaneously, try to raise their metrological knowledge. Since characterising the method performance is not a simple task—it takes time and considerable resources—there must always be a balance between the different requirements. So the performance criteria must be carefully selected.

The essential parameters that analysts need to assess in order to check whether a method satisfies previously defined analytical requirements are the performance criteria or performance characteristics. In fact, method validation consists of deriving experimental values for the selected performance criteria. There are several types of performance criteria. The basic parameters usually refer to the reliability of the method and are commonly derived by using statistical procedures. The trueness, precision, selectivity, sensitivity, range, ruggedness, limit of detection and limit of quantification are the commonest. These will be studied individually in this section.

Other complementary criteria refer to whether the method can be applied to a wide range of sample types and whether it is practical (e.g. cost, ease of use, availability of material, instruments and trained staff, sample throughput, etc.)

Whenever the sampling and subsampling steps are included in the analytical procedure, it is apparent that the test sample must be representative of the analytical problem. Method validation should also include operations that ensure the representativity of the sample submitted to the test procedure in the initial stages of the analysis.



Scheme 1
References commonly used to assess trueness in chemical measurements.

3.1. Trueness

Trueness is defined as “the closeness of agreement between the average value obtained from a large set of test results and an accepted reference value” ISO 3534 (1993). Trueness should be evaluated, in terms of bias, through the analysis of reference samples. However, not all the references have the same level of traceability. Therefore, the reference selected should be the one that has the suitable level of traceability for our purpose. The references commonly used in chemical analysis are listed in the scheme 1, ordered according to their level of traceability.

3.1.1. *Certified reference materials (CRM), reference materials and in-house materials*

These reference materials can be used as long as they are matched, as far as possible, to the routine samples to be analysed in terms of matrix and analyte concentration. The CRMs should be used whenever possible because they have the highest level of traceability. However, this may be not possible because of the limited availability of CRMs that are very similar to the routine samples. On the other hand, reference materials or in-house materials can also be used as long as the content of the analyte has been thoroughly investigated. Preferably, the analyte should be determined using two different methods based on different physical chemical principles and, if possible, based on determinations carried out in different laboratories (IUPAC 1999a).

3.1.2. *Reference methods*

Trueness can be assessed using a reference method provided that it is well known by the laboratory. Here, analysing the same representative test sample with the reference method and the test method to be validated is the way to assess the trueness of the test method. It is important to note that the test samples analysed should be homogeneous, stable, and as similar as possible to the routine samples (in terms of matrix and analyte concentration).

3.1.3. *Proficiency testing*

Trueness can also be assessed when the laboratory takes part in a proficiency testing scheme. In this case, the reference value corresponds to the consensus value obtained by the participating laboratories.

3.1.4. *Spiked samples*

These references have the lowest level of traceability. However, the analyst usually has to resort to spiked samples when the other references

are not available. This has the disadvantage that bias may be incorrectly estimated if the spiked analyte has a different behaviour than the native analyte. On the other hand, a spike can be a good representation of the native analyte in case of liquid samples and in analytical methods that involve the total dissolution or destruction of the samples (IUPAC, 1999b). Otherwise, the different behaviour of the native analyte can be investigated using a less representative reference material (Barwick 1999).

The assessment of trueness depends on whether the method is intended to be used in a restricted concentration range or in large concentration ranges. In the following sections, we describe how to assess trueness in both situations.

3.1.5. Assessment of trueness in restricted concentration ranges

Trueness is assessed by assuming that the bias is the same in the whole concentration range. This bias is calculated using a reference sample similar to the routine samples. This reference sample should be analysed n times with the method to be validated in different runs, i.e. in different days, with different operators, etc. It is recommended to analyse the reference sample at least 7 times (preferably $n \geq 10$) to obtain an acceptable reliability (IUPAC 1999a). An average value, \bar{c}_{method} , and a standard deviation, s , can be calculated from these n measurements.

Trueness is then assessed by statistically comparing the mean value obtained with the analytical method, \bar{c}_{method} , with the reference value, c_{ref} . This comparison is done with a t test. The type of t test applied depends on the reference used. We can distinguish two different cases:

a) Trueness is assessed against a reference value (here we consider CRMs, reference materials, proficiency testing and spiked samples)

b) Trueness is assessed against a reference method.

Case a. Trueness is assessed against a reference value

The following t test is used to assess trueness:

$$t = \frac{|c_{\text{ref}} - \bar{c}|}{\sqrt{u(c_{\text{ref}})^2 + \frac{s^2}{n}}} \quad (1)$$

where $u(c_{\text{ref}})$ is the standard uncertainty of the reference value. For instance, if the certified value of a CRM is expressed as $c_{\text{ref}} \pm U(c_{\text{ref}})$ (with $k=2$), $u(c_{\text{ref}})$ is calculated as $U(c_{\text{ref}})/k$ (e.g. if the certified concentration of copper is 10.1 ± 0.1 mg/l (with $k=2$), $u(c_{\text{ref}})=0.1/2=0.05$ mg/l).

The t value calculated is then compared with the two-sided t tabulated value, t_{tab} , for $\alpha=0.05$ and the effective degrees of freedom, v_{eff} , obtained with the Welch-Satterthwaite approach (1941).

$$v_{\text{eff}} = \frac{\left(\frac{u(c_{\text{ref}})^2 + \frac{s^2}{n}}{u(c_{\text{ref}})^4 + \frac{s^4}{n-1}} \right)^2}{\frac{u(c_{\text{ref}})^4}{v_{\text{eff}}} + \frac{s^4}{n-1}} \quad (2)$$

where v_{ref} are the degrees of freedom associated with $u(c_{\text{ref}})$. If this information is not available, we can assume that $v=n-1$. The trueness of the method is assessed if $t \leq t_{\text{tab}}$.

Case b. Trueness is assessed against a reference method

Trueness is assessed by analysing a test sample both with the reference method and the method to be validated. This test sample should be analysed with both methods in different runs (i.e. in different days, different operators, etc.). An average value, c_{ref} , and a variance, s_{ref}^2 , can be obtained from the n_{ref} results obtained with the reference method. Likewise, an average value, \bar{c} , and a variance, s^2 , can be obtained from the n results obtained with the method to be validated. The t test used to assess trueness depends on whether the difference between the variances of both methods is statistically significant. This is checked with an F test:

$$F = \frac{s_1^2}{s_2^2} \quad (3)$$

in which s_1^2 is the larger of the two variances. The F value calculated is compared with the two-sided F tabulated value, $F(\alpha, v_1, v_2)$, for $\alpha=0.05$ and the degrees of freedom $v_1=n_1-1$ and $v_2=n_2-1$.

If $F \leq F(\alpha, v_1, v_2)$, the difference between both variances is not statistically significant. Therefore, trueness is verified with the following t test:

$$t = \frac{|c_{\text{ref}} - \bar{c}|}{\sqrt{\frac{s^2 \cdot (n-1) + s_{\text{ref}}^2 \cdot (n_{\text{ref}}-1)}{n + n_{\text{ref}} - 2} \cdot \left(\frac{1}{n} + \frac{1}{n_{\text{ref}}} \right)}} \quad (4)$$

The t value calculated is then compared with the two-sided t tabulated value, t_{tab} , for $\alpha=0.05$ and the degrees of freedom $n+n_{\text{ref}}-2$. The trueness is assessed if $t \leq t_{\text{tab}}$.

If $F > F(\alpha, v_1, v_2)$, the difference between both variances is statistically significant. Therefore, trueness is assessed with the following t test:

$$t = \frac{|c_{\text{ref}} - c|}{\sqrt{\frac{s^2}{n} + \frac{s_{\text{ref}}^2}{n_{\text{ref}}}}} \quad (5)$$

the t value calculated is then compared with the two-sided t tabulated value, t_{tab} , for $\alpha=0.05$ and the degrees of freedom v_{eff} , obtained with the Welch-Satterthwaite approach (Satterthwaite 1941).

$$v_{\text{eff}} = \frac{\left(\frac{s_{\text{ref}}^2}{n_{\text{ref}}} + \frac{s^2}{n} \right)^2}{\frac{s_{\text{ref}}^4}{n_{\text{ref}} - 1} + \frac{s^4}{n - 1}} \quad (6)$$

The trueness is assessed if $t \leq t_{\text{tab}}$.

3.1.6 Assessment of trueness in large concentration ranges. Recovery studies

Trueness can be assessed in large concentration ranges with recovery studies. Recovery is defined as the “proportion of the amount of analyte, present or added to the analytical portion of test material, which is extracted and presented for measurement (IUPAC, 1999b). At present the term recovery is used in two different contexts (IUPAC 1999a):

- To express the yield of an analyte in a preconcentration or extraction stage in an analytical method.
- To denote the ratio of the concentration found, c , obtained from an analytical process via a calibration graph compared to the reference value, c_{ref} , i.e. $R=c/c_{\text{ref}}$.

The first use of recovery should be clearly distinguished from the second one. This is because a

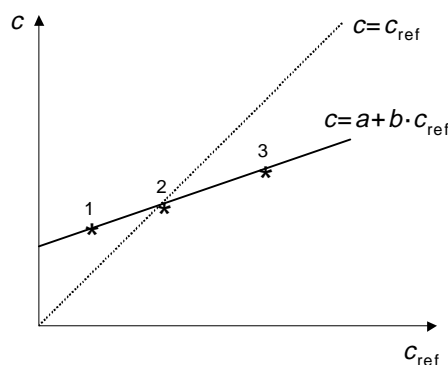


Figure 1

Recovery. The dotted line shows the validation function obtained with an unbiased method. The solid line shows the validation function obtained with a method that has a constant bias, a , and a proportional bias, b .

100% of recovery does not necessarily require a 100% yield for any separation or preconcentration stage. Hence, the IUPAC recommends using two different terms to distinguish between the two uses of recovery. The term “recovery” should be used for yield whereas the term “apparent recovery” should be used to express the ratio of the concentration found versus the reference value (IUPAC, 2001).

Trueness is then assessed in terms of “apparent recovery”. This “apparent recovery” should be calculated for different reference samples that cover representatively the variability of routine samples (in terms of matrix and analyte concentration). Figure 1 shows the necessity of calculating the “apparent recovery” at several concentration levels. We can see that the apparent recovery may fortuitously be 100% in the presence of proportional and constant bias (see point (2) of Figure 1) (IUPAC 2001).

Ideally, the “apparent recovery” should be estimated using several reference samples. However, these references are seldom available and often the analyst has to resort to spiked samples. Such a simple procedure is usually effective in the determination of trace elements, but, for instance, it might not be applicable to a pesticide residue. In this latter case, vigorous reagents can not be used to destroy the matrix. As a result, the “apparent recovery” of the spiked is likely to be larger than the one corresponding to the native analyte. In this case, less representative reference materials can be used to investigate the different behaviour of the native and the spiked analyte (Barwick, 1999).

The “apparent recovery” is calculated as:

$$R = \frac{c}{c_{\text{ref}}} \cdot 100 \quad (7)$$

where c is the concentration found with the method to be validated and c_{ref} is the reference concentration. This expression is valid when reference samples are available or when the analyte is spiked to matrix blanks. However, if the analyte is spiked to samples that already contain the analyte, the “apparent recovery” is calculated as:

$$R = \frac{c_{\text{O+S}} - c_{\text{O}}}{c_{\text{ref}}} \cdot 100 \quad (8)$$

where $c_{\text{O+S}}$ is the concentration obtained for the spiked sample and c_{O} is the concentration obtained for the unspiked sample. The disadvantage of spiking an analyte to a sample which already contains the analyte is that recovery could be 100% and, even so, a constant bias may be present (Parkany 1996). The absence of constant bias can be checked using the

Youden Method (Youden 1947). This method consists of estimating constant bias by analysing different amounts of a sample.

For each reference sample, the “apparent recovery” should be calculated at least 7 times (preferably $n \geq 10$). An average apparent recovery, \bar{R} , and a standard deviation, $s(R)$, can be obtained from the n apparent recoveries calculated. Trueness is then assessed by statistically comparing the mean apparent recovery, \bar{R} , with the 100%. This comparison can be done with the following t test:

$$t = \frac{|100 - \bar{R}|}{\frac{s(R)}{\sqrt{n}}} \quad (9)$$

This t test considers that recoveries are expressed as a percentage and that the uncertainty of the reference value is negligible. The t value calculated is then compared with the two-sided t tabulated value, t_{tab} , for $\alpha=0.05$ and $n-1$ degrees of freedom. There can be two possibilities:

a) $t \leq t_{\text{tab}}$: the apparent recovery does not differ significantly from 100%. The trueness of the method is assessed. Therefore, there is no reason to correct results by the apparent recovery.

b) $t > t_{\text{tab}}$: the apparent recovery differs significantly from 100%. Therefore, the trueness of the method can not be assessed. In this situation, there is a current debate focused on whether results should be corrected by the apparent recovery. IUPAC, ISO and EURACHEM recommend to correct the results. However, AOAC international does not agree that results should be corrected as a general policy (IUPAC 1999b).

3.2. Precision and uncertainty

According to ISO 3534, “precision is the closeness of agreement between independent test results obtained under stipulated conditions” (ISO, 1993). These conditions depend on the different factors that may be changed between each test result. For instance: the laboratory, the operator, the equipment, the calibration and the day in which the test result is obtained. Depending on the factors changed, three types of precision can be obtained: the repeatability, the intermediate precision and the reproducibility (ISO 5725 1994a).

The two extreme precision measures are the reproducibility and the repeatability. The reproducibility gives the largest expected precision because it is obtained by varying all the factors that may affect to the results, i.e. results are obtained in different laboratories, with different operators and using different equipment. This precision can only be

estimated by interlaboratory studies (ISO 5725 1994a). On the contrary, repeatability gives the smallest value of precision because the results are obtained by the same operator, with the same equipment and within short intervals of time. Intermediate precision relates to the variation in results when one or more factors, such as time, equipment and operator, are varied within a laboratory. Different types of intermediate precision estimates are obtained depending on factors changed between the measurements (ISO 5725 1994a). For instance, the time-different intermediate precision is obtained when the test results are obtained in different days. If results are obtained in different days and by different operators, we obtain the [time+operator]-different intermediate precision. Finally, if the laboratory obtains the results by changing all the factors that may affect the results (i.e. day, operator, instrument, calibration, etc.) we obtain the [time + operator + instrument + calibration +...] different intermediate precision. This intermediate precision is also known as the run-different intermediate precision. For a laboratory, this is the most useful precision because it gives an idea of the sort of variability that the laboratory can expect of its results and, also, because it is an essential component of the overall uncertainty of the results.

The within-laboratory precision of an analytical method should be characterised by the repeatability and the run-different intermediate precision. This latter precision is an estimate of the “internal reproducibility” of the laboratory. Precision should be calculated with test samples that have been taken through the whole analytical procedure. These test samples may be reference materials, control materials or actual test samples. It is important to

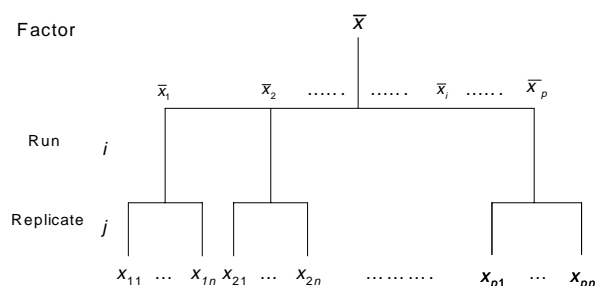


Figure 2

Two-factor fully nested design. The factors studied are the runs and the replicates. p is the number of runs on which the measurement is carried out and n is the number of replicates performed at each run. \bar{x}_i is the mean of the j replicate measurements performed in run i . \bar{x} is calculated as the mean of the mean values obtained for the different runs.

Table I
ANOVA table and calculation of variances for the design proposed in Figure 2

| Source | Mean squares | Degrees of freedom | Expected mean square |
|-----------|---|--------------------|---------------------------------------|
| Run | $MS_{run} = \frac{n \cdot \sum_{i=1}^p (\bar{x}_i - \bar{x})^2}{p-1}$ | $p-1$ | $n \cdot \sigma_{run}^2 + \sigma_e^2$ |
| Replicate | $MS_e = \frac{\sum_{i=1}^p \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{p \cdot (n-1)}$ | $p \cdot (n-1)$ | σ_e^2 |

Table II
Calculation of the precision estimates for the experimental design proposed in Figure 2

| Variance | Expression | Degrees of freedom |
|--|-----------------------------|--------------------|
| Repeatability variance, s_r^2 | MS_e | $p \cdot (n-1)$ |
| Between-run variance, s_{run}^2 | $\frac{MS_{run} - MS_e}{n}$ | |
| Run-different intermediate variance, $s_{(run)}$ | $s_{run}^2 + s_r^2$ | |

note that we should not estimate precision using synthetic samples that do not represent the test samples (IUPAC 1999a). Moreover, if the analytical method is intended to be used over a large concentration range, precision should be estimated at several concentration levels across the working range. For instance, precision can be estimated at low, medium and high concentration levels.

3.2.1. Calculation of the within-laboratory precision

The simplest method for estimating precision within one laboratory consists of calculating the standard deviation of a series of n measurements. It is recommended that at least 7 measurements (preferably $n \geq 10$) should be carried out to obtain good precision estimates (IUPAC 1999a). The repeatability standard deviation is obtained when the n measurements are obtained under repeatability conditions and the run-different intermediate precision is obtained when the measurements are obtained by changing the factors that may affect the analytical results (i.e. operator, day, equipment, etc.).

An alternative procedure consists of performing n replicates in each of the p runs in which the test sample is analysed. This experimental design corresponds to a two-factor fully nested design (ISO 5725 1994a). Here the factors studied are the run and the replicate (Figure 2). The use of the Analysis of the Variance (ANOVA) provides the repeatability,

the run-different intermediate precision and the between-run precision (Table I). Table II shows the expressions for calculating the precision estimates.

It is recommended that the number of replicates per day be equal to 2 ($n=2$) and to focus the effort on the number of runs in which the test sample is analysed (e.g. if we want to estimate precision using a series of 20 measurements, it is better to analyse twice the sample in 10 runs than to analyse 4 times the sample in 5 runs) (Kuttatharmakul 1999). It is also recommended to analyse the sample in at least 7 different runs (preferably $p \geq 10$) to obtain good precision estimates. The advantage of this methodology is that it provides better precision estimates than the simplest method.

3.2.2. Calculation of the repeatability limit

The repeatability limit, r , is the value below which the absolute difference between two single test results obtained under repeatability conditions may be expected to lie with a probability of 95%. This limit is obtained as (ISO 5725 1994a):

$$r = 2.8 \cdot s_r \quad (10)$$

3.2.3. Calculation of the "internal reproducibility" limit

An "internal reproducibility" limit, R , can be calculated from the run-different intermediate standard deviation. This limit is the value below which

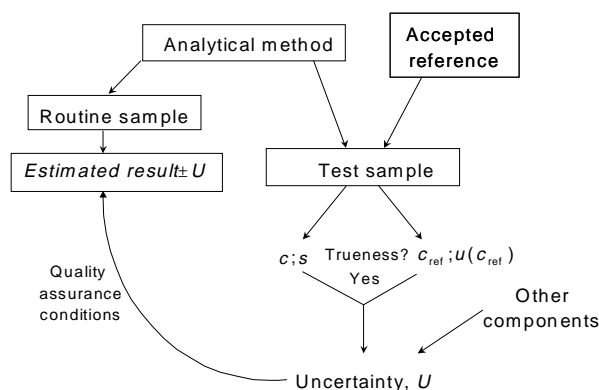


Figure 3

Assessment of trueness of an analytical method against an accepted reference. Once the trueness is assessed, the uncertainty of routine samples can be calculated using the information generated in the validation process.

the absolute difference between two single test results obtained under run-different intermediate conditions may be expected to lie with a probability of 95%. This limit is obtained as:

$$R = 2.8 \cdot s_{I(\text{run})} \quad (11)$$

3.3. Uncertainty

It is widely recognised that the evaluation of the uncertainty associated with a result is an essential part of any quantitative analysis (ISO 17025 1999). Uncertainty is defined as "a parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand" (BIPM 1993).

Uncertainty and trueness are very related concepts. This is because we can not guarantee the correctness of all the possible systematic errors if we have not previously assessed the trueness of the analytical method and, consequently, it is impossible to ensure that the true value is included within the interval "estimated value $\pm U$ " (where U is the uncertainty of the estimated result). Therefore, as Figure 3 shows, every analyst should verify the trueness of the method before calculating uncertainty. Uncertainty can then be calculated using the information generated in the assessment of trueness. Moreover, precision and robustness studies give also useful information for calculating uncertainty (Barwick 2000; Maroto 1999; IUPAC 1999a; EURACHEM 2000). The uncertainty calculated using this information can then be related to the routine test samples as long as the reference sample is representative of these test samples, and the quality assurance conditions are implemented in the laboratory effectively.

Uncertainty can be calculated as the sum of four terms (Maroto 1999):

$$U = 2 \cdot \sqrt{s_{I(\text{run})}^2 + u_{\text{trueness}}^2 + u_{\text{pretreatments}}^2 + u_{\text{other terms}}^2} \quad (12)$$

The first component of uncertainty, $s_{I(\text{run})}$, corresponds to the run-different intermediate precision and considers the experimental variation due to the conditions of the measurement (i.e. day, operator, calibration, etc.). The second component, u_{trueness} , considers the uncertainty of the assessment of trueness, the third component considers the uncertainty of subsampling and/or sample pretreatments not carried out in the assessment of trueness. Finally, the fourth component, $u_{\text{other terms}}$, contains all the sources of uncertainty not considered in the former terms.

3.3.1. Uncertainty of the assessment of trueness, u_{trueness}

In the assessment of trueness it is checked whether the bias of the analytical method is significant. If bias is not statistically significant, the trueness of the method is assessed. However, it remains an uncertainty due to the estimation of the

$$u_{\text{trueness}} = \sqrt{\frac{s_{I(\text{run})}^2}{n} + u(c_{\text{ref}})^2} \quad (13)$$

bias itself. This uncertainty is calculated as:

where n is the number of times that the reference sample is analysed in the assessment of trueness. $u(c_{\text{ref}})$ is the standard uncertainty of the reference sample. The calculation of this uncertainty depends on the reference used in the assessment of trueness. For instance, if a CRM is used, $u(c_{\text{ref}}) = U(c_{\text{ref}})/2$ (where $U(c_{\text{ref}})$ is the uncertainty provided by the manufacturer). If a reference method is used, $u(c_{\text{ref}}) = s_{\text{ref}}/\sqrt{n_{\text{ref}}}$ (where s_{ref} is the standard deviation of the n_{ref} results obtained when the test sample is analysed with the reference method).

3.3.2. Uncertainty of sample pretreatments, $u_{\text{pretreatments}}$

This component considers the uncertainty due to the heterogeneity of the sample and/or to sample pretreatments (filtration, weighing, drying, etc.) not carried out in the assessment of trueness. This uncertainty can be estimated using a sample with the same characteristics as the routine samples. The replicates incorporating subsampling and/or preprocessing steps should be analysed by changing all the factors that may affect them (e.g.

operator, day, material, etc.). The uncertainty can then be calculated as:

$$u_{\text{pretreatments}} = \sqrt{s_{\text{pretreatments}}^2 - s_{\text{conditions}}^2} \quad (14)$$

where $s_{\text{pretreatments}}^2$ is the variance of the results obtained after analysing the different portions. $s_{\text{conditions}}^2$ depends on the conditions in which we analyse the different portions subsampled and/or pre-treated. If they are analysed under repeatability conditions, it corresponds to the variance of the repeatability, i.e. $s_{\text{conditions}}^2 = s_r^2$. If they are analysed under intermediate precision conditions, it corresponds to the run-different intermediate precision, i.e. $s_{\text{conditions}}^2 = s_{l(\text{run})}^2$.

3.3.3. Uncertainty of other terms

This term contains all the sources of uncertainty that have not yet been considered in the former terms. Most of these sources correspond to factors not representatively varied in the estimation of the intermediate precision. The uncertainty of these factors can be calculated using information from robustness studies (Barwick 2000), from bibliography or from other sources such as previous knowledge.

3.4. Linear and working ranges

In all the quantitative methods it is necessary to determine the range of analyte concentrations or property values for which the analytical method can be applied. The lower end of the working range is limited by the values of the limits of detection and quantification. At the upper end, several effects limit the working range.

Within the working range there may exist a linear response range in which the method gives results that are proportional to the concentration of analyte. Obviously, the linear range can never be wider than

the working range. For instance, in the determination of calcium in edible oils using atomic absorption spectroscopy, the working range comprises any oil containing calcium while the linear range is normally between 1 and 10 ppm.

An initial estimation of the working and linear ranges may be obtained using blank and reference materials (or a blank and fortified sample blanks) at several concentration levels. Ideally, at least six concentration levels plus blank are needed. One replicate at every concentration level is enough to visually identify the linear range and the boundaries of the working range. After this primary estimation, a more accurate plot can be obtained using reference materials or fortified sample blanks. At least, six different concentration levels, using at least three replicates of each level, are necessary. The linearity of the calibration line can be assessed using the correlation coefficient and the plot of the residual values. The residual values (e_i , with $i=1 \dots n$, where n is the number of concentration levels used) in a straight-line model are calculated by subtracting the model estimates from the experimental data ($e_i = y_i - b_0 - b_1 x_i$, where b_0 and b_1 are respectively the intercept and the slope of the straight line).

A high value of the correlation coefficient and a good plot of the residual values are normally enough to assess linearity. It is important to point it out that the sole use of the correlation coefficient does not always guarantee the linearity of the calibration line, even if the correlation coefficient has value near to one. A good plot of the residual values has to fulfil the following conditions:

- the residual values must not show tendencies
- the residual values must be more or less randomly distributed
- the number of negative and positive residual values has to be approximately the same
- the residual values should have approximately the same absolute value.

Figure 4 shows some examples of plots of residual values. Figure 4a shows a plot with a good distribution of residual values, where all the residual values fulfil the aforementioned requisites. Figure 4b shows a plot where the residuals do not have the same absolute value, probably showing the need for a weighted calibration method. In Figure 4c the residual values clearly show tendencies (probably indicating the presence of non-linearities). Figure 4d shows the presence of a probable outlier in the data set.

Experimental points with an abnormally high residual value may be outliers, and they should be carefully examined using a test for the detection of outliers (for instance, the test of Cook (Meloun 2002)).

If a more accurate check of the linearity is desired, a test for the lack of fit based on the analysis of variance (ANOVA) can be used alongside with the

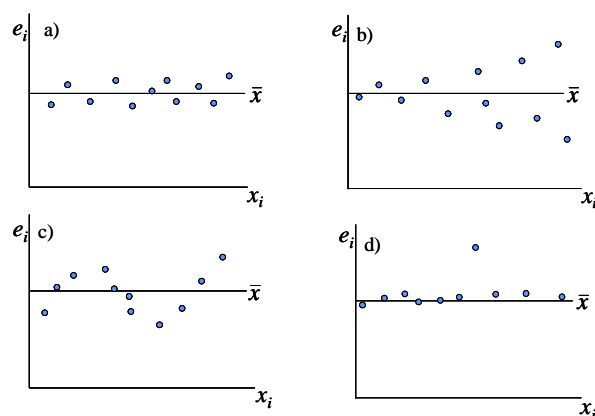


Figure 4
Different plots of residual values.

correlation coefficient and the plot of the residual values (Massart 1997).

It is important to note that the calibration line is usually found using the ordinary least squares method, but if the variance of the replicates at each concentration level is not constant through all the linear range (giving rise for instance to the plot of residual values shown in Figure 4b), then a better option is to use the weighted least squares method, which takes into account the individual variance values in each calibration point (Massart 1997).

3.5. Limit of detection (LOD)

The limit of detection, LOD, is commonly defined as the minimum amount or concentration of substance that can be *reliably* detected by a given analytical method. In a more recent definition, ISO 11843-1 (1997) avoids the term 'limit of detection' and introduces the general term 'minimum detectable net concentration', as the true net concentration or amount of the analyte in the material to be analysed which will lead, with probability $(1-\beta)$, to the conclusion that the concentration or amount of the analyte in the analysed material is larger than that in the blank material. IUPAC (1995), in a previous document, provided a similar definition and adopted the term 'minimum detectable (true) value'.

When an analytical method is used to provide results above the quantification limit, but still an estimation of the detection limit is required, it is usually sufficient to estimate this detection limit as three times the standard deviation of the concentration in a matrix blank or in a sample containing the analyte at a low concentration level. The number of independent determinations is recommended to be at least 10.

When the analytical method is used for trace level analysis or in cases (e.g. food or pharmaceutical analysis) where the absence of certain analytes is required by regulatory provisions, the limit of detection must be estimated by a more rigorous approach, such as the one described by IUPAC (IUPAC 1995, IUPAC 1999a) and must encompass the whole analytical method and consider the probabilities α and β of taking false decisions.

According to the ISO and IUPAC definitions, the LOD is an *a priori* defined parameter of an analytical method, because it is fixed *before* the measurements are made. The LOD is essentially different from a detection decision, because the latter is taken once the result of the measurement is known. In other words: *a posteriori*. The decision of whether a given analyte is present or not in a sample is based on a comparison with the *critical level*, L_C , which is defined as

$$L_C = z_{1-\alpha} \sigma_0 \quad (15)$$

where $z_{1-\alpha}$ is the upper- α percentage point of a normal distribution and σ_0 is the standard deviation of the net concentration when the analyte is not present in the sample (i.e. the true value is zero). The L_C is defined in order to mark a minimum value for which a predicted concentration is considered as being caused by the analyte. By doing so, there exists a risk α of committing a type I error, i.e., a false positive decision, that means stating that the analyte is present when in fact it is not. However, if we want to keep the risk of a false negative decision (called β) low, the LOD of the method, L_D , must be higher by taking into account both probabilities of error:

$$L_D = z_{1-\alpha} \sigma_0 + z_{1-\beta} \sigma_D \quad (16)$$

where $z_{1-\beta}$ is the upper- β percentage point of a normal distribution and σ_D is the standard deviation of the net concentration when the analyte is present in the sample at the level of the LOD. It has been assumed in equations (15) and (16) that the concentrations are normally distributed with known variance. Taking the default values for $\alpha = \beta = 0.05$, and assuming constant variance between $c = 0$ and $c = L_D$, equation (16) becomes:

$$L_D = 3.3 \sigma_0 \quad (17)$$

If variances are not known and have to be estimated based on replicates, then the σ_0 and σ_D values in equations (15) and (16) have to be replaced by their corresponding estimates, s_0 and s_D . Accordingly, the z -values, based on normal distributions must be replaced by the corresponding t -values from a Student t -distribution with degrees of freedom. Taking $\alpha = \beta$ the appropriate expressions for L_C and L_D (assuming constant variance) are:

$$L_C = t_{1-\alpha, v} s_0 \quad (18)$$

$$L_D \approx 2t_{1-\alpha, v} s_0 \quad (19)$$

Equation (19) is an approximation that approaches the true L_D as far as the number of degrees of freedom is larger. When the number of replicates to estimate s_0 is low, the term $2t_{1-\alpha, v}$ must be corrected for the degrees of freedom (see Currie, 1997). Additionally, if s_0 is used instead of σ_0 the LOD is uncertain by the ratio (s_0/σ_0) and an upper bound for L_D has to be calculated (Currie 1997, IUPAC 1995).

3.5.1. Estimating the standard deviations of the net concentrations

It has already been mentioned that the decision of whether a given analyte is present or not in a sample is based on a comparison with the critical level. As in any other situation, this process of comparison

implies experimental errors (in the measurement of the blank sample and the analysed sample), which are assumed to be random and normally distributed. At zero concentration level, the standard deviation of the net concentration is expressed in a general way as:

$$\sigma_0 = \sigma_B \phi \quad (20)$$

σ_B is the standard deviation of the blank and $\phi = \sqrt{1/m + 1/n}$, where m and n are the number of replicates on the analysed sample and on the blank sample, respectively. $\phi = 1$ ($\sigma_0 = \sigma_B$) in the special case when $m=1$ and n is high. When the net concentration is calculated in a paired experiment (i.e. as the concentration in the sample minus the blank) then $\phi = \sqrt{2}$. If σ_B is not known, then it has to be replaced by its corresponding estimate, s_B , in Eq. (20). For a reliable estimation of the detection limit, a minimum suggested number of 10 determinations should be carried out on the blank sample. The precision conditions (repeatability, intermediate precision) in which the blank is analysed should be clearly specified.

3.5.2. Alternatives for LOD calculation

If the calibration line used to check the linearity of the analytical method has been built in a restricted concentration interval close to the expected limit of detection, then this curve can be used to calculate the LOD of the method. The expression for LOD in this case is:

$$L_D = t_{1-\alpha, \nu} \frac{s_{y/x}}{b_1} \sqrt{\frac{1}{m} + \frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^N (x_i - \bar{x})^2}} + t_{1-\beta, \nu} \frac{s_{y/x}}{b_1} \sqrt{\frac{1}{m} + \frac{1}{N} + \frac{(L_D - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}} \quad (21)$$

where:

$s_{y/x}$: is the standard deviation of the residuals of the calibration line

N : is the number of calibration standards

x_i : is each of the concentrations of the calibration standards

\bar{x} : is the mean value of the concentrations of the calibration standards.

The LOD can be calculated by solving Eq (21) or by an iterative procedure.

In chromatographic methods of analysis, in which the peak corresponding to the analyte has to be distinguished from the chromatographic baseline, the detection limit can be calculated by measuring the noise of a chromatogram of several (usually three) sample blanks, which have been submitted to the whole analytical process. The procedure recommended by the SFSTP (STP Pharma Practiques, 1992), consists on determining the maximum amplitude of the baseline in a time interval

equivalent to twenty times the width at half weight of the peak of the analyte. The LOD is expressed as:

$$L_D = 2z_{1-\alpha} h_{\text{noise}} R \quad (22)$$

where h_{noise} is the average of the maximum amplitudes of the noise and R is the response factor (concentration/peak height). This procedure is valid only when the chromatographic peak heights are used for quantification purposes. If peak areas are used instead, then the LOD has to be estimated by other means, such as the described above.

Finally, it is highly recommended when reporting the detection limit of a given analytical method, to specify the approach used for its calculation.

3.6. Limit of quantification (LOQ)

The limit of quantification, LOQ, is a performance characteristic that marks the ability of a chemical measurement process to adequately *quantify* an analyte and it is defined as the lowest amount or concentration of analyte that can be determined with an acceptable level of precision and accuracy (EURACHEM 1998). In practice, the quantification limit is generally expressed as the concentration that can be determined with an specified relative standard deviation (usually 10%). Thus:

$$L_Q = k_Q \sigma_Q \quad (23)$$

where $k_Q = 10$ if the required RSD=10% (this is the recommended value by IUPAC, and σ_Q is the standard deviation of the measurements at the level of the limit of quantification. To estimate σ_Q a number of independent determinations ($n \geq 10$) must be carried out on a sample which is known to contain the analyte at a concentration close to L_Q . Since L_Q is not known, some guides recommend to analyse a sample with a concentration between 2 and 5 times the estimated detection limit (IUPAC 2001). Other guidelines (EURACHEM 1998) suggest to perform the determinations on a blank sample. Such a procedure, however, is discouraged, unless there is a strong evidence that the precision is constant between $c = 0$ and $c = L_Q$.

When determining σ_Q , the sample has to be taken to the whole analytical procedure. Also, the precision conditions in which the sample is analysed must be clearly specified.

Usually, the LOQ estimation is carried out as a part of the study to determine the working range of the analytical method and it is a common practice to fix the LOQ as the lowest standard concentration of the calibration range.

3.7. Selectivity (specificity)

In the majority of analytical methods it is necessary to ensure that the signal produced in the measurement stage is only due to the analyte of

interest, and not to the presence of interferences in the sample. Hence it is necessary to test the selectivity (or specificity) of the analytical method.

The terms selectivity and specificity are often interchanged. Selectivity (or specificity) can be defined as the ability of a method to accurately and specifically determine the analyte of interest in the presence of other components in a sample matrix under the stated conditions of the test (EURACHEM 1998). Some authors consider specificity to be 100% selectivity, and hence make a difference between the two terms: specificity is referred to a method that only gives response for a single analyte, while selectivity refers to a method that gives responses for a number of chemical entities that may or may not be distinguished from each other. This situation creates unnecessary confusions and can be avoided by authors by giving preference for the use of selectivity (IUPAC 2001). IUPAC clarified this overlap expressing the view that 'specificity is the ultimate of selectivity' (den Boef 1983).

From a practical point of view, the selectivity of a method may be checked studying its ability to measure the analyte of interest when other specific interferences have been introduced in the sample. In a similar way, taking into account the matrix effect, which can enhance or suppress the sensitivity of the method. This is usually carried out analysing reference materials and samples containing the analyte of interest and various suspected interferences with both the candidate and other independent analytical methods. From the comparison of the results using all the methods, we can assess the ability of the candidate method to analyse the required analyte in presence of other substances.

Another option, if no independent analytical methods are available, is to analyse a series of three matrix blanks (or samples containing the analyte of interest if matrix blanks are not available), three method blanks and the lowest concentration standard (or a standard corresponding to the analyte concentration in the positive sample if matrix blanks are not available), and assessing their background signals and the significance of any interferences relative to the lowest analyte level specified in the working range (IUPAC, 1999a). Responses from interfering substances should be less than 1% of the response of the lowest concentration measured.

3.8. Sensitivity

Sensitivity is defined as the change in the response of a measuring instrument divided by the corresponding change in the stimulus (being the stimulus for instance the amount of the measurand) (EURACHEM 1988). Although it clearly applies to the measuring instrument, the sensitivity can also be applied to the method as a whole. From a practical point of view, it corresponds to the gradient of the response curve. Inside the linear range, the

sensitivity is simply the slope of the calibration straight-line.

3.9. Ruggedness (or robustness)

The robustness/ruggedness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage (ICH 1994). ISO and IUPAC have not yet provided definitions and recommendations to evaluate ruggedness, but Vander Heyden et al (2001) has extensively covered the subject.

Where different laboratories use the same method, they inevitably introduce small variations in the procedure, which may or may not have a significant influence on the performance of the method. The ruggedness of a method is tested by deliberately introducing small changes to a number of parameters of the method and examining the effect of these changes on the accuracy and precision of the final results.

The evaluation of ruggedness is normally carried out by the individual laboratory, before this participates in a collaborative trial, and should be considered at an earlier stage during the development and/or optimisation of the analytical method.

A large number of parameters may need to be considered, but because most of these will have a negligible effect, it will normally be possible to vary several at once. An established technique for ruggedness testing is covered in detail by Youden and Steiner (1975) and uses Plackett-Burman experimental designs. Such designs allow the investigation of the influence of several parameters in a limited number of experiments. Typical parameters can be subdivided in continuous (i.e. extraction time, mobile phase composition, temperature, flow rate, etc.) or non-continuous (i.e. type of chromatographic column, brands, chemicals, etc.).

The results from ruggedness testing can also be used to estimate components of uncertainty which have not been considered in the assessment of trueness and precision of the method (Barwick 2000).

Table III
Desing to calculate the ruggedness of a method

| Experiment | Parameter | | | Result |
|------------|-----------|---|---|----------------|
| | A | B | C | |
| 1 | + | + | + | y ₁ |
| 2 | - | + | - | y ₂ |
| 3 | + | - | - | y ₃ |
| 4 | - | - | + | y ₄ |

The ruggedness test should be carried out on a representative sample, preferably a reference material. The first step consists on identifying and selecting the parameters that are going to be tested and defining their nominal and extreme levels (i.e. flow rate at 1.0 ± 0.2 ml/min). After this, the experimental design is has to be selected; for instance, if three parameters were to be evaluated, a design with only four experiments would be sufficient as indicated in Table III. The + and – signs denote the coded levels for the maximum and minimum values of the parameter under study. For each parameter, there are two experiments at each level. To study the effect of a given parameter, the mean of the results obtained at the - level have to be subtracted from the mean of the results at the + level. In the example, the effect of parameter A would be calculated as:

$$\text{Effect A} = [(y_1 + y_3)/2 - (y_2 + y_4)/2]$$

Once the effects have been calculated for every parameter, a statistical and/or graphical analysis of the effects must be carried out in order to draw chemically relevant conclusions and, if necessary, to take actions to improve the performance of the method.

4. THE BASIC PRINCIPLES OF METHOD VALIDATION

There are three basic principles of any process of method validation:

1. The validation must include the complete analytical procedure. The measurement of the instrumental signal is often considered to be the main step in the analytical procedure and considerable effort is made to establish the linear range and to calculate the uncertainty derived from this step. However, nowadays, due to advances in instrumentation, this step usually contributes little to the bias or to the expanded uncertainty of the final result. Special care has to be taken with the sample pre-treatment process where the intervention of the analysts and the non-automated steps has a crucial effect on the performance parameters.

2. The validation must comprise the whole range of concentrations in which the analyte may be present in the test samples.

3. Representativity is an essential factor. The method should be validated by considering the range of matrices in which the analyte has to be determined. Moreover, the different effects influencing the results during the normal application of the method should be considered during the method validation process. This concept will be further developed when the uncertainty values are calculated.

5. PERFORMING THE METHOD VALIDATION

When and who should validate the analytical methodology to be used are also aspects of method validation that are of practical importance.

There are many laboratory situations in which a method needs to be validated. For instance, the validation required for a new method is clearly different from the validation required for a method that has previously been validated by a method performance study. Validating modifications introduced into a test procedure is different from validating a well-known method for testing samples that incorporate new matrices. In all cases, some sort of method validation should be performed. However, the extent of the validation varies from case to case. This topic is further developed in section 12.7

Clearly, the laboratory that is going to use a method for analysing test samples is responsible for guaranteeing that the method is fit-for-purpose. Therefore, this laboratory should demonstrate that is valid to analyse the test samples adequately.

6. INTER-LABORATORY COMPARISON AS A MEANS OF METHOD VALIDATION

When a method is going to be used by many laboratories throughout the world, it is advisable that it be validated by means of a collaborative or method performance study involving a group of laboratories.

In this type of studies, all the laboratories follow the same written analytical protocol and analyse the same samples in order to estimate the performance criteria of the method being studied. Usually the within-laboratory and among-laboratory precision are estimated and, whenever necessary and possible, so are other criteria such as the systematic error, recovery, sensitivity or limit of detection.

After the consensus reached under the auspices of the IUPAC, William Horwitz prepared the "Protocol for the Design, Conduct and Interpretation of Method-Performance Studies" (Horwitz 1995). In this document, the design of the method performance study is described in considerable detail. For instance, at least five different materials are needed for a single type of substance, a minimum of eight laboratories should participate in the study and the designs used to estimate the repeatability and reproducibility are described. The statistical analysis, the final estimation of the parameters and the final report are also described.

However, collaborative studies have not been undertaken for many analytical methods. Among other problems, they are too time-consuming and require expensive resources, particularly for the laboratory in charge of the organisation. Moreover, the general guidelines for method validation are too

impractical to be used in industrial laboratories or laboratories that monitor, for instance, trace-level organic compounds.

Because of these unfavourable factors, the Association of Official Analytical Chemists, AOAC International, has introduced the "Peer Verified Method Program" (AOAC 1998) so that methods used by laboratories working with only one or two others can be validated. The aim of this program is to provide a class of tested methods whose performance has been checked in at least one independent laboratory.

7. IN-HOUSE VALIDATION

We have seen that collaborative trials give rise to a sound validation of methodologies. But they also present some drawbacks:

- They require to gather enough laboratories that can devote the necessary resources to achieve proper validation.
- There are not enough scientists willing to undertake the organisation of the very large number of method trials that are required.
- Problems with shipping regulations, import restrictions, analyte stability.
- This type of studies are uneconomic, too slow, restricted in scope

For these and other reasons, there is currently a move towards giving more importance to the use of methods of analysis that have only been validated 'in-house', provided that the laboratory implements appropriate quality assurance procedures. In-house validation is the process of checking, inside a laboratory, that a method gives correct results. It can provide a practical and cost-effective alternative approach to other validation systems.

Many analysts think that a previously validated standard or reference method may be directly applied in their laboratory. The previous validation of the method is erroneously considered as a sufficient warranty to achieve quality results. This is not true, and when a new method is introduced in a laboratory (even if it is a previously validated one by an international organism) it has always to be validated, although in this latter case a full validation is not needed. As there is some parallelism between the laboratory of analysis and the cuisine, we can compare a validated analytical procedure and a recipe in a good cookbook. There are many good previously tested recipes that are very carefully described. Is it enough to get a good dish in our cuisine? We all know that unfortunately this is not enough. The same happens with validated methods. We can get good results using them, but we need to prove that they adequately work in our laboratory. Hence a method always has to be validated to check that their quality parameters are valid for the

particular analytical problem that we have in our laboratory. (EURACHEM 1998) It is always needed some degree of validation, although this degree may significantly change depending in each particular situation. Some of these particular situations are (AOAC/FAO/IAEA/IUPAC 2000):

- *The laboratory wants to use as a routine method a previously validated one in another laboratory or group of laboratories:* Here a full validation is not necessary, since the absence of the method bias has already been assessed. The laboratory only needs to verify that it is capable of achieving the published performance characteristics of the method by checking the absence of laboratory bias, and by calculating its own precision (repeatability and intermediate precision) estimates. In case of measuring at low concentration values, the laboratory also needs to calculate the limits of detection and quantification, since these values depend, among others, on the laboratory instrumentation and on the analysts.
- *The laboratory wants to use a validated method, but the analytes to be determined either are new or they are found in new sample matrices:* In this case the validation has to be extended to the new analytes or matrices. Parameters like trueness, precision, calibration, analytical ranges, and selectivity have to be checked for every new analyte or matrix.
- *The method is published in the scientific literature together with some analytical characteristics:* The laboratory has to undertake verification and a limited validation in order to meet the requirements.
- *The method is published in the scientific literature but not characteristics given:* The laboratory has to undertake a full validation of the method and a self-verification in order to achieve the stated characteristics.
- *The method has been developed in-house:* The laboratory has to arrange a full validation of the method.

Apart from these situations, some other checks are routinely used to test the whole analytical procedure before starting a new set of analysis. For instance, checking the sensitivity of the method (slope of the calibration straight line), or the selectivity in some chromatographic assays. These latter are often referred as to suitability checks.

Another different kind of in-house validation is the retrospective validation. The retrospective validation is the validation of a method already in use based upon accumulated production, testing and control data. In some cases an analytical method may have been in use without sufficient validation. In these cases, it may be possible to validate, to some extent, the adequacy of the method by examination of

accumulated analysis. Retrospective validation allows, for instance, estimating precision through a period of time.

8. REPORTING METHOD VALIDATION

Once an analytical method has been validated, all the procedures must be documented to ensure that the method will always be used in the same conditions. The values of the performance criteria are quite sensitive to any variations in the procedure due to the lack of proper documentation. The uncertainty usually increases and bias may appear. Inconsistencies when the validated method is used to test samples can only be avoided, therefore, by proper reporting and proper application of the validation process.

Several documents recommend the way in which a validated method should be reported. Among others, ISO 78/2 provides advice on how to document general chemical methods. It is advisable that the standardised written procedures be revised by an experienced independent analyst, so that all possible misinterpretations are detected and avoided.

Following the general procedures for managing written documents, the authorised person should check that only the up-to-date documents are used. Similarly, documents should only be modified by the person responsible, who should also ensure that obsolete documents are withdrawn and the revised methodologies put into practice.

REFERENCES

- AOAC (1998) *Peer Verified Programs*. [http://aoac.org/vmeth/peerverified\[1\].htm](http://aoac.org/vmeth/peerverified[1].htm)
- AOAC/FAO/IAEA/IUPAC. (2000) Guidelines for Single-Laboratory Validation of Analytical Methods for Trace-Level Concentrations of Organic Chemicals. In: *Principles and Practices of Method Validation*. Edited by A. Fajgelj and A. Ambrus. Royal Society of Chemistry, Cambridge.
- Barwick, V.J., Ellison, S.L.R. (1999) Measurement uncertainty: Approaches to the evaluation of uncertainties associated with recovery. *Analyst*, **124**, 981-990.
- Barwick, V.J., Ellison, S.L.R. (2000) Development and Harmonisation of Measurement Uncertainty Principles. Part (d): Protocol for uncertainty evaluation from validation data, VAM Project 3.2.1, Report n° LGC/VAM/1998/088. LGC, Teddington.
- BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML (1993) *International Vocabulary of basic and general terms in Metrology*, VIM, ISO, Geneva
- Currie, L.A. (1997) Detection: International update, and some emerging dilemmas involving calibration, the blank, and multiple detection decisions. *Chemom. Intell. Lab. Syst.*, **37**, 151-181.
- EURACHEM (1998) *The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics*. EURACHEM Secretariat, Teddington, Middlesex, <http://www.vtt.fi/ket/eurachem>
- EURACHEM/CITAC (2000) *EURACHEM/CITAC Guide: Quantifying uncertainty in analytical measurement*, 2nd Edition, Teddington, Middlesex.
- den Boef, G., Hulanicki, A. (1983) Recommendations for the usage of selective, selectivity and related terms in analytical chemistry. *Pure Appl. Chem.*, **55**, 553-556.
- Horwitz, W., (1995) Protocol for the Design, Conduct and Interpretation of Method-Performance Studies. *Pure Appl. Chem.*, **67**, 331-343.
- ICH (1994) Harmonised Tripartite Guideline prepared within the Third International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH), *Text on Validation of Analytical Procedures*, 1994, <http://www.ifpma.org/ich1.html>.
- ISO, *International Organization for Standardization* (1993) ISO 3534-1, Statistics, Vocabulary and symbols, ISO, Geneva.
- ISO, *International Organization for Standardization* (1994a) ISO Guide 5725, Accuracy (Trueness and Precision) of Measurement Methods and Results, ISO, Geneva.
- ISO, *International Organization for Standardization*, (1994b) ISO Guide 8402: Quality- Vocabulary, Geneva.
- ISO, *International Organization for Standardization*, (1997) ISO 11843-1, Capability of detection. Part 1: Terms and definitions. ISO, Geneva.
- ISO, *International Organization for Standardization*, (1999) ISO/IEC 17025. General requirements for the competence of calibration and testing laboratories. ISO, Geneva Ibid: UNE-EN ISO/IEC 17025. Requisitos generales relativos a la competencia de los laboratorios de ensayo y calibración. AENOR, Madrid, 2000.
- IUPAC. *International Union of Pure and Applied Chemistry* (1995) Nomenclature in Evaluation of Analytical Methods including Detection and Quantification Capabilities, *Pure Appl. Chem.*, **67**, 1699-1723.
- IUPAC. *International Union of Pure and Applied Chemistry* (1999a) Harmonised Guidelines for In-House Validation of Methods of Analysis (Technical report) http://www.iupac.org/projects/1997/501_8_97.html.
- IUPAC, *International Union of Pure and Applied Chemistry* (1999b) Harmonised guidelines for the use of recovery information in analytical measurement, *Pure Appl. Chem.*, **71**, 337-348.
- IUPAC, *International Union of Pure and Applied Chemistry* (2001) Selectivity in Analytical Chemistry Technical Draft <http://www.iupac.org/publications/pac/2001/>.
- IUPAC, International Union of Pure and Applied Chemistry (2002) Recommendations for the use of the term "recovery" in analytical procedures, Draft <http://www.iupac.org/>.
- Kuttatharmakul, S., Massart, D.L., Smeyers-Verbeke, J. (1999) Comparison of alternative measurement methods. *Anal. Chim. Acta*, **391**, 203-225
- Maroto, A., Riu, J., Boqué, R., Rius, F.X. (1999) Estimating uncertainties of analytical results using information from the validation process. *Anal. Chim. Acta*, **391**, 173-185.
- Massart, D.L., Vandeginste, B.M.G., Buydens, L.M.C., de Jong, S., Lewi, P.J., Smeyers-Verbeke, J., (1997) *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier, Amsterdam.
- Meloun, M., Miitky, J., Hill, M., Brereton, R.G. (2002) Problems in regression modelling and their solutions. *The Analyst*. In press.
- Parkany, M. (1996) *The use of recovery factors in trace analysis*. Royal Society of Chemistry, Geneva
- Satterthwaite, F.E. (1941) Synthesis of variance. *Psychometrika*, **6**, 309-316.

- STP Pharma Practiques (1992) Guide de validation analytique: Report SFTP. *Methodologie and examples* **2** (4), 205-226.
- Thompson, M. Wood, R. (1993) The International Harmonised Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories. *Pure Appl. Chem.*, **65**, 2123-2144. Also published in J. AOAC Intl. **76**, 926-940.
- Thompson, M. Wood, R., (1995) Harmonised Guidelines for Internal Quality Control in Analytical Chemistry Laboratories. *Pure Appl. Chem.*, **67**, 49-56.
- Vander-Heyden, Y., Nijhuis A., Smeyers-Verbeke, J., Vandeginste, B.G.M., Massart, D.L. (2001) Guidance for robustness/ruggedness tests in method validation. *J. Pharmaceutical and Biomedical Analysis* **24**, 723-753.
- Youden, W.J. (1947) Technique for testing the accuracy of analytical data. *Anal. Chem.* **19**, 946-950.
- Youden, W.J., Steiner, E.H. (1975) *Statistical Manual of the Association of Official Analytical Chemists*. AOAC, Arlington, VA.