Validation of Qualitative Analytical Methods

Doctoral thesis

UNIVERSITAT ROVIRA I VIRGILI

UNIVERSITAT ROVIRA I VIRGILI
Department of Analytical Chemistry and Organic Chemistry

Validation of Qualitative Analytical Methods

Thesis submitted by
ESTHER TRULLOLS SOLER
to obtain the degree of
Doctor from the Universitat Rovira i Virgili
Tarragona, April 2006

UNIVERSITAT
ROVIRA I VIRGILI

DEPARTAMENT DE QUIMICA
ANALITICA
I QUIMICA ORGANICA

C/ Marcel·lí Domingo, s/n
43007 Tarragona
Tel. 34 977 55 97 69
Fax 34 977 55 84 46
e-mail: secqaqo@quimica.urv.es

Prof. F. XAVIER RIUS FERRÚS and Dr. ITZIAR RUISÁNCHEZ CAPELÁSTEGUI, Professor and Associate professor of the Department of Analytical Chemistry and Organic Chemistry at the Universitat Rovira i Virgili,

CERTIFY: That the Doctoral thesis entitled: “**VALIDATION OF QUALITATIVE ANALYTICAL METHODS**”, submitted by ESTHER TRULLOLS SOLER to obtain the degree of Doctor from the Universitat Rovira i Virgili, has been carried out under our supervision in the Analytical Chemistry Area of the Department of Analytical Chemistry and Organic Chemistry at the Universitat Rovira i Virgili, and all the results presented in this thesis were obtained in experiments conducted by the above mentioned student.

Tarragona, April 2006

Prof. F. Xavier Rius Ferrús   Dr. Itziar Ruisánchez Capelástegui

Aquest treball s'ha dut a terme amb la inestimable col·laboració de tots els membres del Grup de Quimiometria, Qualimetria i Nanosensors: Pili, Marisol, Ricard, Joan, Jordi, Santi, Alicia i Jaume. Molt especialment, vull expressar el meu sincer agraïment als meus directors de tesi Itziar i Xavier, per haver-me deixat fer recerca al seu costat, per haver-me ensenyat tantes coses i per deixar que jo n'hagi après d'altres tota sola, donant-me llibertat per a proposar, decidir i ser creativa... Puc dir que tots els objectius han estat assolits.

TABLE OF CONTENTS

1. OBJECTIVES

1.1 SCOPE AND OBJECTIVES

Validation is an important feature in any method of measurement because it is closely related to the quality of the results. A method of analysis is characterised by its performance parameters, which have to be assessed if they are to provide the correct performance values. These performance values must be in accordance with previously defined requirements that the method of analysis should satisfy. But above all, the performance parameters depend on the type of method and its intrinsic characteristics. So depending on what is needed, the user must choose which method of analysis will best solve the analytical problem.

Of all the different methods of analysis, conventional classification differentiates between qualitative and quantitative methods, although semi-quantitative methods can also be considered to be a group apart.

Qualitative methods of analysis provide basic information about the composition of a sample and perform quite simple chemical reactions to identify the analytes it contains [1, 2]. Quantitative methods of analysis provide information not only about the composition but also about the concentration of the analytes present in the sample and, generally speaking, they often require more complex analytical techniques to obtain more accurate and reliable information about the sample. Semi-quantitative methods of analysis lie between the qualitative and the quantitative methods because they assign samples to different classes which delimitate specific ranges after measuring the corresponding property. These different categories are defined by a particular criterion: concentration of a compound, index value, etc. [3, 4]. One example of this sort of method is how the acid-base character of a sample is determined by means of the pH measurement: different colours mean different pH values. These are the semi-quantitative methods of analysis [5].

For various reasons —the need for reliable quantitative results, the greater development of instrumental techniques, etc— research effort and investment has mainly focused on quantitative methods of analysis. As a consequence, validation procedures have been developed almost exclusively for quantitative methods of analysis.

The aim of this doctoral thesis is to study validation processes in qualitative methods of analysis. In particular, it reviews the state of the art as far as the validation of qualitative methods of analysis is concerned. It also proposes classifying these methods of analysis according to their characteristics. And, finally, it defines the qualitative performance parameters that are so important to the establishment of the final validation procedures.

These procedures are addressed to those analytical methods that provide binary results of the type YES/NO, POSITIVE/NEGATIVE or ABOVE/BELOW a certain limit. They are often used as screening methods of analysis, which separate samples according to one or more criteria and then often submit them to the appropriate quantitative analytical method. Or, as is becoming increasingly common nowadays, they are used as routine methods of analysis in fields like environmental, clinical or food analysis.

Considering the applicability of qualitative methods of analysis and the importance of the fields in which they are used, method validation is fundamental to the quality of the final results. Bearing all this in mind, and not forgetting that the area is largely unexplored and that many aspects of the validation of qualitative methods of analysis have yet to be defined, this thesis has been structured in the following way.

One of the main focuses is the theoretical part which studies and defines the performance parameters of the methods of analysis. Several theoretical bases have been discussed and studied in depth, and then applied to practical cases. In these cases, the performance parameters have been defined and estimated.

The objectives of the thesis are the following:

1) To review several aspects of the validation of qualitative methods of analysis, to revise performance parameters and to define more appropriate ones when necessary. These issues are discussed in two papers entitled *Validation of qualitative analytical methods* and *Validation of qualitative methods of analysis that use control samples*. Both articles were the starting point of subsequent practical applications.

2) To establish the performance parameters of a commercial test kit used in food analysis, which provides a sensorial response. This is the central theme of the paper *Qualitative Method for determination of Aflatoxin B_1 in nuts.* The validation procedure is based on the use of Performance Characteristic Curves.

3) To establish the performance parameters of a commercial test kit, which uses control samples in clinical analysis, and provides

instrumental detection but final binary results. This is discussed in the paper *Validation of qualitative test kits with instrumental responses. Detection of Varicella -Zoster Virus IgG antibodies in human serum.* The validation procedure uses the statistical characterisation of the control sample distribution.

4) To establish the performance parameters of a homemade autoanalyzer with instrumental response that combines the measurement of two analytes using Hypotheses Testing. This topic is dealt with in *Statistical intervals to validate an autoanalyzer for monitoring the exhaustion of alkaline degreasing baths.*

5) Robustness is presented separately as a performance parameter. Despite its considerable importance, it is generally not considered in validation procedures. *Robustness in qualitative analysis: a practical approach* presents practical aspects regarding robustness in qualitative methods of analysis.

1.2 STRUCTURE OF THE THESIS

The thesis has been structured in the following chapters:

Chapter 1 briefly introduces the framework of the thesis. Then the objectives are described and justified, and, finally, the structure is outlined.

Chapter 2 deals with the concept of method validation in general. The aim is to present not only the state of the art but also future trends in the field of method validation. This chapter serves as an introduction to the in-depth study of the validation of qualitative methods of analysis in the following Chapter.

Chapter 3 discusses the validation of qualitative methods of analysis. Several prestigious regulatory bodies have prepared validation proposals for these methods, which are summarized together with several classifications of qualitative methods of analysis and an in depth study of performance parameters. In addition to this summary, another paper adds some more general information about qualitative methods that use instrumental detection.

Chapter 4 describes the main experimental applications carried out during this thesis. First, a commercial test kit used in food analysis and based on colour development is validated. Secondly, a commercial test kit used in clinical analysis is validated. In this case the final binary result is obtained by measuring an instrumental response. And finally, the validation procedure is performed on a homemade autoanalyzer used in the environmental field. This method of analysis uses instrumental detection but also gives a final YES/NO result.

Chapter 5 focuses on robustness. In the framework of method validation, both quantitative and qualitative robustness is an important performance parameter. Therefore, a brief summary of the state of the art of this feature and an application are presented.

Chapter 6 presents the general conclusions of the work. Various suggestions for future research, in relation to the applications presented in this thesis, are also made.

The *Appendix* contains the list of papers and meeting presentations given during the period of development of this thesis.

1.3 REFERENCES

[1] F. Burriel, F. Lucena, S. Arribas and J. Hernández, *Química Analítica Cualitativa*, Paraninfo, 13rd ed., Madrid, Spain, 1989.

[2] *Aspectos cualitativos de la Química Analítica*, in M. Valcárcel; *Principios de Química Analítica*, Springer-Verlag, Ibérica, Barcelona, Spain 1999.

[3] C. Heiss, M. G. Weller and R. Niessner, *Anal. Chim Acta 396*, 1999, 309.

[4] R. W. Gerlach, R. J. White, N. F. D. O'Leary and J. Van Emon, *Water Res. 31*, 1997, 941.

[5] H. F. De Brabander, P. Batjoens, K. De Wasch, D. Courtheyn, G. Pottie and F. Smets, *Trends Anal. Chem. 16*, 1997, 485.

2. METHOD VALIDATION

2.1 INTRODUCTION

For several years now, method validation studies, guidelines and procedures have focused mainly on quantitative methods of analysis. As a result, a large bibliography has grown up which defines performance parameters, discusses procedures and describes theoretical studies. If the validation of qualitative methods is to be analysed appropriately, the concept and main topics must be reviewed. The present chapter, then, discusses what method validation is and how it can be used. The different aspects of the validation process, the types of validation and the usefulness of the information gathered are also presented. It should be borne in mind that only quantitative methods are dealt with. In the following Chapter, we will move on to qualitative methods of analysis.

2.2 METHOD VALIDATION

During method development, analysts establish the most suitable steps of the analytical process that will lead to the information required: sample pre-treatment, when necessary, separation technique and the detection system, among others. The best analytical conditions for obtaining good results are also considered. The information gathered after the analysis may have several goals: 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, etc. Therefore, the analytical information must be of sufficient quality, which means that it must be reliable and match the purposes of the analysis. To meet these premises, analysts must define the purposes of the analysis and the requirements that the method should fulfil. Therefore, the validation of the method of analysis will provide, according to the ISO definition [1] the " confirmation by examination and provision of evidences that the particular requirements for a specified intended use are fulfilled" . Another definition given in the Handbook for the Quality Assurance of Metrological Measurements [2] states that " method validation consists of documenting the quality of an analytical procedure, by establishing adequate requirements for performance criteria, such as accuracy, precision, detection limit, etc. and by measuring the values of these criteria" . In general terms, then, the requirements and performance parameters must first be defined for every analytical method and purpose of analysis; and second, the value for these parameters must be estimated and checked to see if they really meet the criteria. This is an essential condition if the results provided are to be used.

The process of assessing the performance criteria is closely related to the concept of ' fitness-for-purpose' , which is defined by IUPAC in the Orange Book [3] as the " degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose" . Hence, it is important, first, to consider the necessary conditions related to the problem at hand, second to choose the method of analysis that best fits the necessities, and, finally, to validate it as is shown in Figure 1.

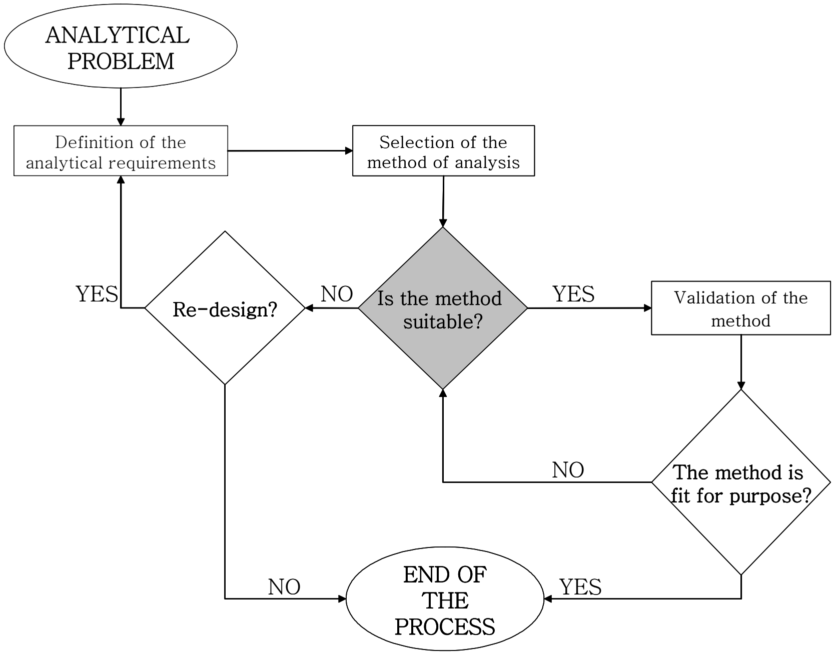


Figure 1. Fitness for purpose concept. Adapted from the EURACHEM *The Fitness for Purpose of Analytical Methods* [4].

The EURACHEM Guide *The Fitness for Purpose of Analytical Methods* [4] also describes how important it is for the analytical performance and the analytical problem to be suited. It also describes the importance of method validation, and indicates when, how and who should perform the validation, among other equally relevant statements. Fitness for purpose also involves practicability and suitability criteria [5], which entail evaluating operational and time constraints, as well as such other parameters as reusability or possibilities of automation.

Although the users of the method of analysis will focus the validation process on their own needs, there are some common features that all validation procedures must have. The validation process must satisfy three requirements [4]:

1) The whole method must be validated. It is quite usual to focus on the detection technique or the instrumental measurement, which often means that just this stage is validated. However, the previous steps of sample pre-treatment, extraction or pre-concentration also belong to the method of analysis and are of utmost importance. So they must all be validated.

2) The whole range of concentrations must be validated. It is difficult to comply with this condition because a method may work very well in one particular concentration range but not in others.

3) The whole range of matrices must be validated. It is well known that the matrix can have a decisive effect on the analysis. Therefore, and for the sake of representativeness, several matrices must be submitted to method validation.

In addition to the conditions mentioned above, it should also be pointed out that the method developed, before it is validated, should include the various types of equipment and the locations where it will be run. That is to say, if the analysis is always to be performed with the same equipment and in the same laboratory, then other equipment and other laboratories need not be taken into account. Before the equipment is used, its performance must be checked with generic standards.

The analytical requirements that the analyst has defined are translated to the performance criteria of the method of analysis. So one of the stages of method validation is to estimate and assess the values of the quality parameters. In general terms, performance criteria can be divided into two main categories [6] although some authors may suggest other classifications. The basic parameters usually refer to the reliability of the method and are commonly derived with statistical procedures. Some examples are trueness, precision, selectivity, sensitivity, limit of detection and quantification. Criteria such as cost, ease of use, rapidity, etc. are considered to be complements of these.

In the *Handbook of Chemometrics and Qualimetrics* [7], Massart et al. state that there are two types of performance criteria: primary and secondary. Precision, bias, accuracy, trueness and the detection limit belong to the first group while the other parameters that can influence these primary criteria belong to the second (eg. linearity, the range of linearity, the quantification limit, selectivity, and sensitivity or ruggedness, etc.)

2.2.1 Types of method validation

Because methods of analysis are designed for different uses, not all validation procedures are equal. Some examples of factors that can influence the definition of these procedures and which must be carefully considered are the quality of the final results, the consequences in terms of economy and time, whether the method has been developed recently or whether it is an adaptation of a previously adapted one. Depending on these factors, different method validation will be carried out in a different manner.

A validation procedure cannot be performed if the validation level required is not taken into account. The validation level is the degree of effort invested in the validation process, so a high validation level requires greater effort. On the other hand, if the validation level is low, the effort investment will also be low. In both cases, the quality of the results obtained by the validated method of analysis will be rather different. Figure 2 shows the different levels of method validation.

So establishing the most suitable validation level is fundamental because the definition of the process depends on it and the results after the validation will also be of a different quality. To correctly choose the most appropriate validation level, operational, economic and material resources or the requirements the method must fulfil must be considered. The analyst can then choose to perform either an internal method validation or an interlaboratory validation [8, 9]. A recently accepted alternative is for a third laboratory to make an assessment of the properties claimed.

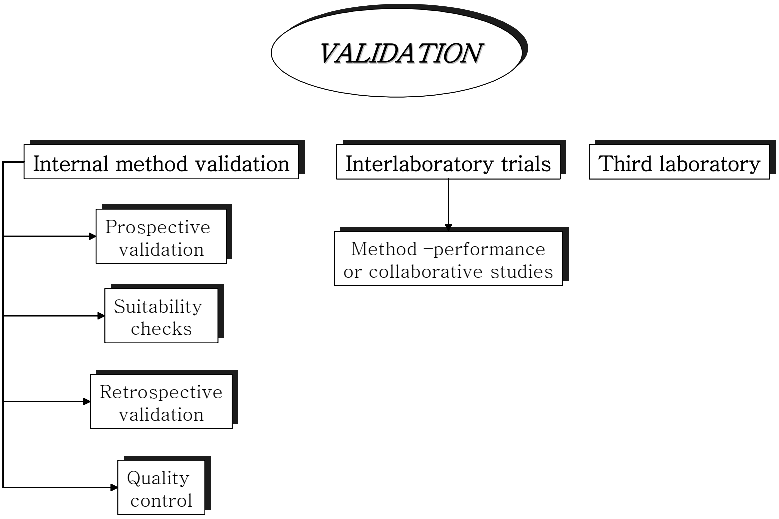


Figure 2. The types of method validation that involve different validation levels

Internal method validation is the lowest validation level [10]. The laboratory that incorporates a new method of analysis that has been developed internally or externally tests the quality of both the method and the results. Internal method validation is mainly carried out in three cases: to assess new methods developed in-house, to assess methods transferred from other laboratories and, for instance, to estimate long-term precision. Routine internal quality control is also considered as internal method validation.

Each of the above mentioned situations requires a particular validation scheme because the requirements of every individual case are different. As a general philosophy, fitness-for-purpose is also applicable here. The main types of internal method validation are briefly described below.

A full validation process is undertaken when the laboratory develops a new method and has to be used in routine control. Again, before carrying out the full validation process, the most appropriate performance parameters must be considered. If there is no information about the method' s performance characteristics, it is recommended first to check if it is suitable for the intended purpose with several samples: for example if the method is selective enough, if the sensitivity is tolerable or if the matrix will not interfere excessively. If the results are favourable, then the subsequent quality parameters are determined. If not, the method itself, the equipment, the analysis technique or the acceptance limits should be changed. Method development and validation, then, is an iterative process. This is so-called prospective validation.

Transferring analytical methods from one laboratory to another is quite a common situation. Because the transferred method must be fully validated in the source laboratory, the receiving laboratory does not need to undergo another complete validation process. However, it must assess whether the methods of analysis perform correctly under the corresponding conditions. This is called suitability checks.

Retrospective validation is performed on validated methods that are already being used. It may be necessary to examine accumulated results to assess whether the method keeps on performing appropriately. Likewise, long-term precision can also be assessed by collecting data over a long period of time.

Once the method is in normal use, a quality control program should be run. Control charts [11] are a very useful tool for this purpose.

On the other hand, ***interlaboratory trials*** provide the highest validation level because several laboratories assess one property of a sample, usually the concentration of one or more analytes. Depending on the aim, any one of three main trial types can be used. Method performance or collaborative studies are performed on analytical methods that will be extensively used and which must provide high quality results. In these cases, several laboratories participate in validating the analytical method. The participating laboratories have been inspected, they are known to perform well and it is assumed that their results are highly reliable. They follow the same analytical procedure, which is described in detail, and they analyse the same samples to establish the performance criteria. After all the results have been reviewed, the final values of the quality parameters defined are calculated.

To perform a collaboratory trial, either the ISO guideline 5725-2 [8] or the IUPAC technical report [12] are good starting points because they define all necessary terms, they specify the optimum number of participating laboratories and samples analysed, and they describe how the study must be performed and how the data must be treated if the method is to be validated.

A laboratory proficiency study tests the performance of the laboratory itself. Though it is not always possible, it is advisable to analyse a material, whose true concentration is known, by using the method of analysis that each laboratory considers most suitable for the problem at hand. When the results are compared, appropriate conclusions about the individual performance of each laboratory can be inferred. The ISO/IEC Guide 43-1: 1997 [13] reports a procedure for performing proficiency tests.

The last objective when performing an interlaboratory trial is to certify a material. The group of participating laboratories have been proven to be good and reliable, so they analyse a material containing one or more analytes using several methods of analysis to determine the most probable concentration value/s with the minimum uncertainty. Although these studies are not the most commonly used ones, there is an ISO guideline that describes the suitable protocol [14].

Interlaboratory trials are not easy whatever their purpose is. Collaborative studies need to find enough laboratories that have been proven to perform well. Economical investment is also important so that samples and materials can be shipped. And the samples themselves can be problematical: despite having the ideal composition they are often not stable. And finally, the trials are time-consuming for the organizing laboratory.

Because of these drawbacks in interlaboratory trials, the alternative of a third laboratory to test method performance is an interesting one. To be more precise, the laboratory which verifies the quality parameters of the method under examination belongs to an institution or has the competence to assess the quality of other laboratories.

This option consists of providing the examining laboratory with the quality parameters claimed by the method developer. Then, the examining laboratory must verify if the values provided are correct or if, on the contrary, they must be estimated again. The best example of this in operation is the *Peer Verified Methods* [15] program of the AOAC International. The International Seed Testing Association

(ISTA) [16] also provides a program called *Performance Validated Method*, which has similar characteristics.

Reporting method validation correctly is also an important issue. After the validation procedure, all the actions taken must be clearly and orderly documented. In the same way, the values of the performance criteria must be documented so that any change or variation due to different laboratory conditions can be easily avoided. As is usual in these cases, the ISO has a guideline [17] that describes how standards should be laid out. Written documents also need to be revised: all copies must be up-to-date and any uncontrolled copy must be withdrawn.

2.3 REFERENCES

[1] UNE-EN ISO 9000, Sistemas de gestión de la calidad. Fundamentos y Vocabulario, AENOR, Madrid, 2005

[2] J. K. Taylor and H. V. Opperman, *Handbook for the Quality Assurance of Metrological Measurements,* Lewis Publ., Chelsea, 1988.

[3] J. Inczédy, T. Lendyel and A. Ure, *Compendium of Analytical Nomenclature (The IUPAC 'Orange Book')*, M. Blackwell Science, 3rd ed., Oxford, UK, 1998.

[4] Eurachem, The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics, Eurachem, 1998. Available at http://www.eurachem.ul.pt

[5] International Union of Pure and Applied Chemistry, IUPAC, Harmonized Guidelines for Single-Laboratory Validation of Methods of Analysis, (IUPAC Technical report), *Pure Appl. Chem.*, *74*, 2002, 835.

[6] R. Boqué, A. Maroto, J. Riu and F. X. Rius, *Grasas y Aceites 53*, 2002, 128.

[7] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke, *Data Handling in Science and Technology 20A. Handbook of Chemometrics and Qualimetrics: Part A.* Elsevier Science, Amsterdam, The Netherlands, 1997.

[8] International Organisation for Standardization, ISO 5725-2, Accuracy (trueness and precision) of measurement methods and results. Part 2: Basic method for the determination of

repeatability and reproducibility of a standard measurement method, Geneva, Switzerland, 1994.

[9] E. Hund, D. L. Massart and J. Smeyers-Verbeke, *Anal. Chim. Acta 423*, 2000, 145.

[10] A. M. García-Campaña, J. M. Bosque-Sendra, L. Cuadros Rodríguez and E. Almansa López, *Biomed. Chrom. 14*, 2000, 27.

[11] E. L. Grant, R. S. Leavenworth, *Statistical Quality Control*, McGraw-Hill, Inc., 6th ed., New York, 1988.

[12] W. Horwitz, Protocol for the Design, Conduct and Interpretation of method-Performance Studies, *Pure Appl. Chem. 67*, 1995, 331.

[13] International Organisation for Standardization, ISO/IEC Guide 43-1: 1997, Proficiency testing by interlaboratory comparison. Part 1: Development and operation of proficiency testing schemes, Geneva, Switzerland, 1997.

[14] International Organisation for Standardization, ISO Guide 35: 1989, Certification of reference materials-General and statistical principles, Geneva, Switzerland, 1989

[15] AOAC International, *Method Validation Programs. Peer Verified Programs.* Gaithersburg, Maryland, USA, 2002. Available at http://aoac.org/vmeth/peerverimtd[1].htm.

[16] International Seed Testing Association. Available at http://www.seedtest.org

[17] International Organization for Standardization, ISO 78-2: 1999, Layouts for standards -- Part 2: Methods of chemical analysis. Geneva, Switzerland, 1999.

3. VALIDATION OF QUALITATIVE ANALYTICAL METHODS

3.1 INTRODUCTION

Method validation, as it has been presented in the previous chapter, is a step that must be carried out whenever a new method of analysis is going to be used in a laboratory or in field analysis. This means that before a quantitative, qualitative or semi-quantitative method of analysis is to be used, its performance values must be estimated and checked. So, there is a higher level of assurance in the quality of the results.

However, almost all the guidelines discussed in the previous chapter are for the validation of quantitative methods of analysis. This means that the end user of a quantitative method of analysis has the essential tools to perform a proper validation procedure.

Qualitative methods of analysis have been applied for a long time. However, recently they have been arousing increasingly greater interest, like quantitative methods. Unfortunately, they have not been widely studied yet. Therefore, the end user of a qualitative method of analysis does not have the suitable guidance to submit a method to complete validation. Recently, some concepts have been clarified and some terms defined. This is no more than a starting point but it can be a helpful tool to plan a validation procedure.

Basically, the concepts concerning the reliability of the results have been quite well established in recent decades. Reliability involves studying other quality parameters such as sensitivity, specificity and false results rates. The first article in this chapter contains some bibliographic references which define and study these parameters. They deal mainly with clinical, pharmaceutical and microbiological analysis since qualitative methods of analysis (either in test kit format or in classical reactions) were largely developed for these disciplines. However, these parameters were not estimated as part of what nowadays is considered to be a validation procedure: they were estimated individually and often not all of them were evaluated.

Considering the growing interest in qualitative methods of analysis, the concepts behind the above mentioned parameters have recently been summarized in a document published in the Official Journal of the European Communities [1]. This document deals with the performance of confirmatory analytical and screening methods and the interpretation of results.

Although considerable headway has been made in qualitative method validation, there is still some work to do as far as the important performance parameters are concerned. In this respect it is important that they also be estimated as part of the same validation procedure.

The aim of the present chapter is to review the state of the art in the validation of qualitative methods. Several concepts regarding the validation of qualitative analysis are presented in two papers:

1) *Validation of qualitative analytical methods* published in Trends in Analytical Chemistry. This contribution is a general review of qualitative method validation. First, it defines and classifies qualitative

methods. Then it presents the organizations that deal with qualitative method validation and their proposals. Finally, it briefly describes the most common quality parameters for qualitative methods and the possible alternatives by which they can be estimated.

2) *Validation of qualitative methods of analysis that use control samples* published in Trends in Analytical Chemistry. This paper is an extension of the first one and describes a particular case of qualitative methods. It focuses on the validation of test kits that use control samples and, basically, presents the implications of using control samples, from two points of view: the experimental one and the estimation of the quality parameters. To conclude, it provides a brief example of the validation procedure for a test kit that gives instrumental responses in the clinical context.

3.2 VALIDATION OF QUALITATIVE ANALYTICAL METHODS

*Trends in Anal. Chem., 23, **2004**, 137*

E. Trullols, I. Ruisánchez and F. Xavier Rius.

Universitat Rovira i Virgili. Departament de Química Analítica i Química Orgànica. Plaça Imperial Tàrraco 1. 43005 Tarragona (Spain)

Abstract

This article reviews the state of the art in validating qualitative analytical methods. After introducing the scope of these qualitative methods, their main characteristics and how they differ from quantitative analytical methods, we propose a classification according to the detection system. The institutions, programmes and documents dealing with the validation of qualitative methods are discussed and the performance parameters ┌ false positive and negative, sensitivity and specificity rate, cut-off, unreliability region, ruggedness and cross-reactivity are presented. The various strategies used to validate qualitative analytical methods contingency tables, Bayes' theorem, statistical hypothesis tests and performance characteristic curves┌ are also briefly described.

Keywords: Binary results; Performance parameters; Qualitative analysis; Validation

1. Introduction

One of the trends in modern analytical chemistry is the development of new analytical techniques and methods that can reliably identify and quantify the components in complicated samples such as those related to environmental problems or food protection. Hyphenated techniques such as the combination of chromatography with mass spectrometry or various spectroscopic techniques are just some of the examples of these developments. These powerful tools have involved a considerable investment in expensive instruments and require analysts to be properly trained.

However, from a practical point of view, many users find it increasingly important to reconsider whether quantitative results are really necessary. In routine laboratories, for example, it is quite usual for the first stage to determine whether one or more analytes are present/absent in a sample and, if so, for the second step to estimate their concentration level. For example, to assess if a sample of drinking water is free from pollutants. Therefore, instead of trying to quantify the pollutants in the sample as the first goal, it could be enough just to assure if they are present above or below the permitted concentration level. Qualitative methods are used in these cases. They are commonly used as screening techniques before quantification with the routine method, which enables both the time and cost of analysis to be reduced.

The quality of the results provided by these qualitative methods is of utmost importance. The users of these analytical methods must make sure that the results obtained in their laboratory are fit for their purpose. This means that the analytical requirements must be defined and the values of the performance parameters assessed before they are used as routine methods in the laboratory. In other words, qualitative methods must also be validated [1]. Usually, validation of analytical methods has been developed and applied to quantitative methods. As a consequence, nowadays there are many validation guidelines that are either accepted by regulatory bodies or by communities of practitioners in specific fields. There is, however, no general validation guideline available for qualitative analytical methods.

This review discusses the state of the art of validation in qualitative methods. We try to fill a gap by clarifying the concepts related to qualitative analytical methods. First we review the various programs provided by the organisms that deal with qualitative method validation, and then we define and discuss some terms. Then we go on to explain some performance parameters and how they are calculated, and finally we describe the strategies used to validate qualitative analytical methods.

2. Qualitative Methods of Analysis

The idea of qualitative method is by no means new. In fact, it has been defined by the European Community as " the assessment of the presence or absence of one or more analytes in a sample due to its physical and chemical properties" [2].

Association of Official Analytical Chemists (AOAC) defines qualitative methods as a " method of analysis whose response is either the presence or absence of the analyte, detected either directly or indirectly in a certain amount of a sample" [3].

It can be concluded from the definitions that a qualitative analytical method is used to find out if a sample contains one or more specific analytes. In these cases, the result of the analysis can only be of the binary type: presence/absence or YES/NO.

As can be easily inferred, presence/absence is not considered to be an absolute measure related to a concentration level of zero but to a specific concentration level. Below this limiting level, the concentration of analyte is considered not significant. The detection of the analyte may require either an instrument or the human senses, but whatever the way the response is recorded, it is converted into a YES/NO result.

It is well known that quantitative methods make it possible to quantify one or more analytes in a sample by using calibration curves that transform the instrumental response into the measurand, often expressed as the concentration of analyte. Between qualitative and quantitative methods, there is still room for semi-quantitative methods of analysis. These methods provide an approximate response that enables the analyte to be roughly quantified, and they usually assign the test sample to a given class (e.g. the concentration could be high, medium, low or very low). This means that the estimate of the true concentration has a large associated uncertainty. Even so they are useful because quantification does not always have to be accurate. A representative example would be the test stripes for pH measurements. These methods usually cost less than quantitative methodologies, they are easier to handle, and have other practical performance parameters.

One of the main drawbacks when dealing with qualitative methods is the terminology used because there is no internationally accepted vocabulary so several names are commonly used in the bibliography. Although terms such as screening systems, test kits, field tests or immunoassays are traditionally used when referring to qualitative methods, they could also be used when dealing with quantitative and semi-quantitative methods. Consequently, here we shall try to put into context the terms that are usually found in the literature.

To start with, it is interesting to consider the term “ screening” in this regard. In an analytical problem, a screening analysis separates or discriminates samples from a large group that contain, e. g., one or more analytes above or below a pre-set value (Fig. 1). This value is often expressed as a concentration level, and can be set by an official agency, internal quality control or a client, among other possibilities. This pre-set concentration is also called specification limit, threshold value or maximum permitted level, among other names.

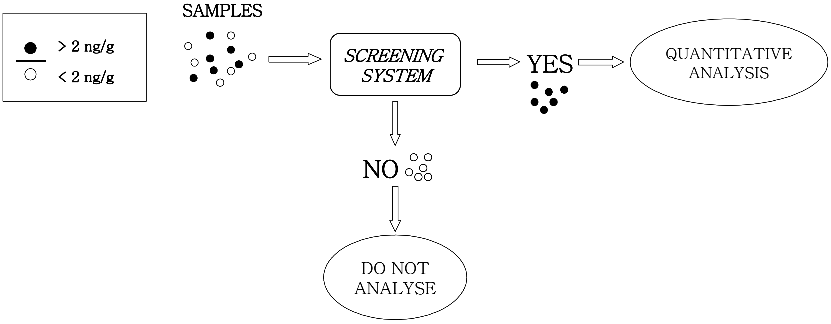


Figure 1. Scheme for a screening system of samples. ● Samples containing more than 2 ng/g of analyte. ○ Samples containing less than 2 ng/g of analyte.

Nowadays, it is quite usual for the term “ screening method” to be used as a synonym for “ qualitative method” [4]. However, often the term “ screening” is also used to describe a step that comes before the calibration stage in a quantitative method. Therefore, screening is not always related to qualitative but also to quantitative analysis [5].

Another similar term is “ screening test” that gives a reliable indication that the analytes of interest are present/absent in the sample at a level that is hazardous or not permitted [6]. Usually, screening tests are commercially available in a package containing all the reagents and sometimes the instrumentation for the analysis, and they are also known as “ test kits” [7]. These kits are used for “ rapid and direct analyses” because they are easy to handle, cheap to purchase and to run, and quick. They also provide results on site.

Another widely used, synonymous term in some fields is “ immunoassay” [8], an analytical technique that uses an antibody molecule as a binding agent to detect and quantify substances in a sample. Immunoassays have been shown to detect and to quantify many compounds of environmental interest such as pesticides, industrial chemicals or drug residues, so some specific forms of immunoassay [9] can be considered as quantitative methods. Some of the most important advantages of immunoassays are their rapidity, sensitivity, specificity and cost-effectiveness; they can be designed as rapid field-portable, qualitative methods or as standard quantitative laboratory procedures; and, they can also be used as screening methods to identify samples that need to be analyzed further by classical analytical methods.

2.1 Classification of qualitative methods.

As often happens in many disciplines, there is no generally accepted classification of qualitative methods, although several schemes with a diversity of criteria have been proposed by various authors.

Valcárcel [4] et al. suggest quite a broad classification based on a variety of criteria: the physical state of the sample (i. e., whether it is solid or liquid); the detection system (either sensorial or instrumental); etc. The authors discuss the integration of the chromatographic techniques and the qualitative methods, so the resulting analytical systems can be classified as sensors, as systems that use separate laboratory steps or as methods that integrate the body of operations.

More intuitive sorting exists, e. g., Unger-Heumann [7] considers test kits as adaptations of well-known analytical methods, so the classification takes into account if test kits are based on chemical, physical-chemical, biochemical or biological methods.

Throughout this article, we have classified qualitative methods of analysis according to the type of detection system so as to differentiate between sensorial and instrumental detection.

2.2 Qualitative methods based on sensorial detection.

The main feature of these qualitative methods is that human senses are used to record and interpret the response. As might be expected, vision is the sense that is most used (e. g., the response can be a signal, such as a coloured solution, a spot on a test strip or the appearance of turbidity). In order to obtain this response, these methods are based on the reaction between the analyte of interest in the sample and specific reagents involved in the procedure. The magnitude of this response can be either directly or indirectly related to the concentration of the analyte. The reaction follows different principles, mainly chemical and immunological. The most commonly used chemical reactions are complexation and precipitation. However, in immunological methods, in particular those of the ELISA (enzyme-linked immunosorbent assay) type, the appearance of the coloured spot requires the addition of an enzyme that recognizes the analyte-antibody binding.

In addition to visual inspection, colour development can be measured and colour intensity related to analyte concentration. One way of doing so is to compare the colour to a colour card or wheel with a predefined correspondence between colour intensity, either in solution [10] or test strip [11], and concentration.

2.3 Qualitative methods based on instrumental detection.

These methods provide an instrumental response, which, in many cases, measures absorbance, although in principle any instrument can be used. There are considerable differences between the way instruments are used in qualitative and quantitative analysis. The final decision is made by comparing the response of a test sample and the response of a sample containing the target analyte at the specification level. We call this the reference sample. Instead of working in the concentration domain, these methods work in the response domain. They can also be used to quantify the analyte in the sample if necessary.

Their basis is that an instrumental response is used to decide whether the analyte is above or below a specific concentration level. No calibration curve is prepared, however; the test-sample response is simply compared to the response provided by the reference sample, so this reference sample, which should ideally be a reference material, is measured and its response (r_{SL}) recorded. Subsequently, the recorded test sample response (r_i) is compared to r_{SL}. If r_i is larger than r_{SL}, it can be concluded that the test sample contains the analyte at a concentration level higher than the reference sample. However, if r_i is lower than r_{SL}, then the conclusion is that the test sample contains less analyte than the reference. Thus the instrumental response is converted into a binary response of the type YES/NO.

Using this procedure, Waters et al. [12] compared the test-sample response with the reference-sample response but did not consider either probabilities of type α or type β errors. These probabilities of error are used by Pulido et al. [13] to calculate the so-called cut-off value, a limiting value in the response domain, at which the decision about whether the analyte is above or below the specific concentration level must be taken.

As in the previous case (sensorial detection), chemical and immunological based reactions are commonly used. ELISA-based methods can be considered to be special cases because a specific detection tool is sometimes required (e. g., when a 96-microtiter-plate format is used). This tool enables the calibration standards and some samples to be measured simultaneously. Although the calibration curve can be computed, it need not be used if the only thing required is a comparison between the response of the reference sample and the test sample.

3. Method Validation in Qualitative Analysis

As is well known, before any analytical method is applied to test samples on a routine basis, it should be validated, so its performance characteristics should be defined and properly assessed. The ISO/IEC 17025 standard [14] describes the importance of method validation and its application in the analytical laboratory.

There is general agreement about the concept of method validation. The ISO defines method validation as a " confirmation with an examination and provision of objective evidences that particular requirements for a specified use are met" [1], so the first thing to be done is to define these particular requirements that depend on the specific determination ahead and are, therefore, particular to each case. This is very much related to the concept of " fitness-for-purpose" [15] and can also be applied to qualitative analytical methods.

The validation of these methods must follow the same philosophy as that of quantitative methods, although there are some differences in the methodology, as described below. In recent years, some organizations have published guidelines or documents about the validation of qualitative analytical methods. The aim of the next section is to give an overall view of the institutions involved in this subject.

3.1 Organisms that deal with qualitative method validation

All organizations that deal with qualitative method validation focus on the concept of fitness for purpose, and therefore on evaluating the relevant performance parameters. Among the different possibilities, the general recommendation is that participation in collaborative studies is the preferred way of validating methods. The strongest exponent of this idea is AOAC International [16]. Like the " Peer-Verified Methods Program" for quantitative in-house methods [17, 18], AOAC International has the " Performance Tested Methods Program" [19] specifically addressing test kits. This validation program makes it possible for the quality parameters claimed by the manufacturer or end user to be assessed by a third laboratory. Similarly, the “ International Seed Testing Association” (ISTA) [20] has a program called “ Performance Validated Method” in which a third laboratory proves the quality parameters of the test kits based on immunological reactions.

The US Environmental Protection Agency (EPA) [21] also has a specific document called “ Guidance for Methods Development and Methods Validation for the RCRA Program” [22]. This ensures that established, validated immunoassays are available for measuring and monitoring needed for the RCRA (Resource Conservation and Recovery Act) Program and it is addressed to developers of qualitative and quantitative methods in general.

In “ The Fitness for Purpose of Analytical Methods” document [15], EURACHEM specifies that the qualitative performance parameters that should be evaluated are: confirmation of identity; sensitivity; selectivity/specificity; and precision. Precision may be expressed as true and false positive (and negative) rates and it has to be taken into account that these rates are related to sensitivity and specificity. To avoid problems of nomenclature, the same guide clarifies the meaning of these two parameters in chemical usage. AOAC International also proposes and defines what it calls the four performance indicators: sensitivity, specificity, false negative and positive rates [3].

Similarly, in its official bulletin [2], the European Union (EU) defines and proposes the evaluation of the following qualitative parameters: limit of detection (CCβ); selectivity/specificity; stability; applicability; and robustness. The EU also states that screening methods can be used as long as they are properly validated and the percentage of false complaints (probability of β error) is lower than 5% at the concentration level of interest.

Finally, the European Cooperation for Accreditation of Laboratories (EAL) has a guide entitled " Validation of test methods" [23], which emphasizes that the uncertainty associated with the method is the most important quality parameter. This guide also makes specific reference to qualitative methods that deal with sensorial responses, in the sense that not all known validation procedures are applicable. It has to be clarified that, in this guide, " test methods" refers to any analytical method (quantitative and qualitative).

According to the above, the definition of method validation is applicable to both quantitative and qualitative methods of analysis, although there are differences in the validation process. The different meanings of the performance parameters used in qualitative and quantitative methods and the disparity in their definitions require changes in the ways that they are calculated.

3.2 Use of references

References are essential in method validation, as trueness has to be assessed, so, if we try to use references from quantitative analysis in a qualitative method, we can follow an established hierarchical order.

The hierarchy ranges from primary methods to recovery studies, and it includes certified reference materials (CRM), participation in collaborative studies and the use of confirmatory methods.

Unfortunately, there are considerably fewer possibilities for qualitative analytical methods. For these cases, there is still no primary method. Moreover, CRMs are rather complicated to use. It should be emphasized that any qualitative method claimed to work at the specification level will provide positive and negative results about the test samples. But, as a result of experimental or random error, false rates (either positive or negative) are obtained close to this concentration level, so the CRM should contain the analyte at a concentration level that is near to the specification limit. If the concentration level is either far below or far above the specification limit, we will be able to check only if the method correctly classifies the samples as negative or positive. For CRM concentrations close to this concentration level, we have to compute the probabilities of false positive and negative responses, so the comparison with a CRM has to be in terms of probabilities, and cannot be in terms of concentration.

As a result, whenever possible, comparison with a reference method is the best option. The analysis must be made using both the reference method (usually quantitative) and the qualitative method [24, 25]. To assess whether the qualitative method is performing well, the proportions of positive results obtained by both methods have to be compared by means of a suitable hypothesis test such as the Chi-square test (χ^2) [3].

Participation in collaborative studies is also recommended. However, as with CRMs, basic statistics, such as mean and standard deviation, cannot be computed. Each laboratory will report its own results (positive and negative test samples). The positive or negative rates can be computed both individually, for each participating laboratory, and globally, for the study as a whole [26, 27]. Again the probabilities obtained by each laboratory can be compared by means of the Chi-square test. If any one of these possibilities is impracticable, spiked samples can be used as a first approximation for the validation process.

3.3 Qualitative performance parameters

The definition of the performance parameters is an important aspect to consider when dealing with qualitative analysis. Table 1 shows some of the most common parameters according to whether the type of analytical method chosen is quantitative or qualitative.

Table 1. Quality parameters for both quantitative and qualitative analytical methods

Quantitative method	Qualitative method
Accuracy: trueness, precision	Sensitivity and specificity
Uncertainty	Unreliability region
Sensitivity and specificity	False positive and negative rates
Selectivity: interferences	Selectivity: interferences
Range and linearity	Cut-off limit
Detection limit	Detection limit
Ruggedness or robustness	Ruggedness or robustness

Although some performance parameters have the same name, the concepts attached to them and their evaluation can be different, e. g., sensitivity can be differently considered depending on the analytical method. If a quantitative method is used, sensitivity should be a numerical value that indicates how the response changes whenever there is a variation in the concentration of the analyte. However, this parameter will be evaluated in a different way if a qualitative method is used. The same occurs with the specificity, detection limit, cut-off value and uncertainty or unreliability region.

The following parameters have to be considered when dealing with qualitative responses.

3.3.1. False positive and negative rates. The false positive rate is " the probability that a test sample is a known negative, given that the test sample has been classified as positive by the method" [3].

$$False\ positive\ rate = \frac{fp}{tn + fp} \qquad (1)$$

where *fp* are false positive test samples and *tn* are known true negative test samples.

Similarly, the false negative rate is " the probability that a test sample is a known positive, given that the test sample has been classified as negative by the method" [3].

$$False\ negative\ rate = \frac{fn}{tp + fn} \qquad (2)$$

where *fn* are false negatives samples and *tp* known true positive test samples.

3.3.2. Sensitivity and specificity. Generally speaking, when dealing with qualitative methods, sensitivity is " the ability of a method to detect truly positive samples as positive" [6], so the sensitivity rate " is the probability, for a given concentration, that the method will classify the test sample as positive, given that the test sample is a ' known' positive" [28]. It can be calculated as:

$$Sensitivity\,rate = \frac{test\ positives}{total\,number\,of\ known\ \ positives} = \frac{tp}{tp + fn} \qquad (3)$$

where *tp* are truly positive test samples and *fn* are false negative test samples.

The same occurs with specificity, which is defined as " the ability of a method to detect truly negative samples as negative" [6]. In the same way, the specificity rate " is the probability, for a given concentration, that the method will classify the test sample as negative, given that the test sample is a ' known' negative" [28], so it can be expressed as

$$Specificity\ rate = \frac{test\ \ negatives}{total\,number\,of\ known\ negatives} = \frac{tn}{tn + fp} \qquad (4)$$

where *tn* are truly negative test samples and *fp* are false positive test samples.

3.3.3. Unreliability region. In quantitative analysis, the uncertainty is the numerical value related to the interval in which the measurand may be found with a given probability. However, for qualitative methods, having binary responses of the YES/NO type, there is no meaning for a number associated with the result and expressed as a semi-interval that is attached to it, so uncertainty is expressed not as a numerical value but as a region of probabilities of committing error. Moreover, following the nomenclature used until now, it corresponds to the region in which false responses are obtained (either false positive or negative).

As we are dealing with a region where there are certain probabilities of error, some authors prefer to call it an unreliability region rather than an uncertainty region [29]. This region is defined by an upper and a lower concentration limit [30], between which the qualitative method can provide false responses. As these false responses can be either positive or negative, the upper and lower limits that define this unreliability region depend on the probability of obtaining these false responses, which is fixed by the analyst.

3.3.4. Detection limit and cut-off value. The term detection limit was defined by the IUPAC [31] in 1995 for quantitative analysis. According to this definition, it can be calculated when the response is a numerical value and when a value is assigned to the two probabilities of α -and β -type errors. When the response is of the binary-sensorial type, however, the standard deviation of the blank samples cannot be calculated, and the probabilities of α -and β -type errors cannot be considered at the same time, although they are both set by the analyst. Depending on the interest of the analyst and the problem in hand, either the probability of committing an α type error or that of committing a β -type error will be considered.

The detection limit has also been defined as " the lowest concentration of the analyte which the test can reliably detect as positive in the given matrix" [6]. This implies that we should consider only the probability of a β -type error or false negative rate, usually at 5%. This definition is presented in the context of assessing a maximum permitted concentration level, but, if it is extrapolated to the case of assessing a minimum concentration level, we should consider only the probability of an α -type error or false positive rate, also at 5%. Therefore, both probabilities of committing error cannot be considered simultaneously. In the first case, the limit of detection coincides with the upper limit of the unreliability region, where the sensitivity rate is 95% and it also coincides with the cut-off value. However, in the second case, the limit of detection coincides with the lower limit of the unreliability region.

The cut-off value is a special performance parameter, since it has been widely studied and used in qualitative analytical methods that use instrumental responses [13]. Regarding the qualitative methods with sensorial responses, this value means the concentration level where the qualitative method differentiates the samples with a certain probability of error, usually of 5%. In the particular case of problems related to the maximum permitted level, the cut-off value is related to the sensitivity, as it corresponds to the concentration level at which the sensitivity rate is 95%, when the β-type error probability has been set at 5%.

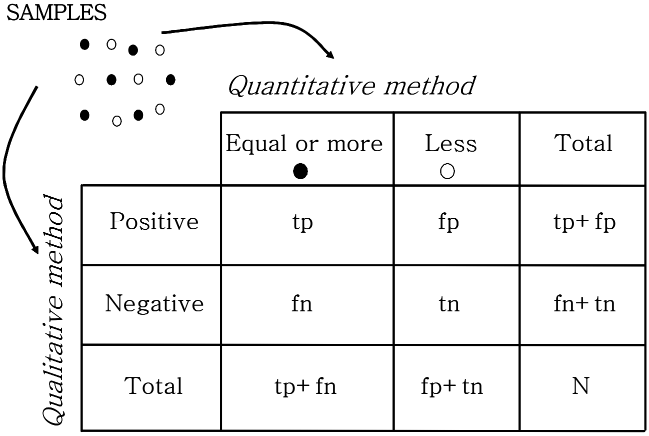
Other parameters should also be considered. Ruggedness is an important parameter related to how the method performs under variations in the operational, environmental, etc. conditions. In quantitative methods, it must be evaluated [2, 15], but in qualitative methods it need not be. According to some authors [3], it is not a “ formal part of the validation protocol” , and “ it is not a submission requirement” when submitting a method for evaluation.

Another parameter to be considered is cross-reactivity or the presence of interferences. For test kits, in particular, it is recommended to check whether the presence of analytes of the same family as the one under study might modify the result of the analysis. These checks are mandatory for manufacturers of the test kits.

3.4 Evaluation of the qualitative performance parameters

There are various ways of evaluating the performance parameters in qualitative analysis. Recently, Pulido et al. [32] showed that Contingency Tables [33], Bayes’ Theorem [34], Statistical Hypothesis tests [13] and Performance Characteristic Curves [35] are the four main ones, each of which has advantages and drawbacks. However, depending on whether or not the type of response obtained is instrumental and on the number of analyses that the analyst wants to perform, etc. we will have to choose one methodology or another.

3.4.1. Contingency tables. Contingency tables have been widely used in bioassays [36, 37]. They are based on the calculation of probability. Although other formats are possible, the simplest and most commonly used are those that give a two-category classification: positive or negative, above or below a regulatory concentration level, etc. Then, the qualitative method result is compared with the results obtained using the confirmatory method (see Fig. 2). From this table, it is possible to calculate only four performance parameters (false positive, false negative, sensitivity and specificity rates) and two predictive values (positive, PPV, and negative, NPV).



	Equal or more ●	Less ○	Total
Positive	tp	fp	tp+ fp
Negative	fn	tn	fn+ tn
Total	tp+ fn	fp+ tn	N

Figure 2. Example of a 2x2 contingency table. ‘ tp’ are true positive samples, ‘ fp’ are false positive samples, ‘ fn’ are false negative samples and ‘ tn’ are true negative samples.

One of the main features of this approach is that it gives an overall vision of how the qualitative method performs, but it does not give individual information, as a probability of error for each sample is not computed. This means that it is assumed that the unknown sample has the same statistical behaviour as the samples used to build the Contingency Table. One of the drawbacks is that the capacity of the Contingency Table depends on the total number of analyzed samples used to build it and the experimental design. It should also be pointed out that all samples must be analysed using both the qualitative and confirmatory methods.

3.4.2. Bayes' Theorem. This methodology is based on the well-known Bayes' Theory of Probability. Several intermediate probabilities must be computed and evaluated. Bayes' Theorem calculates the probability of giving a correct result (either positive or negative) when it is indeed correct, P(a/p). This probability is called conditional probability, so many analyses are required in order to achieve a good uncertainty estimation or a better error probability. The main feature of this methodology is that, unlike Contingency Tables, the probability of giving a wrong result is estimated individually, because the conditional probability is calculated for each analysed sample. And, again, only the same four parameters can be calculated: false positive, false negative, sensitivity and specificity rates.

3.4.3 Statistical Hypothesis Tests. These Hypothesis Tests compare the response of the sample with that of a pre-set reference [13] (Fig. 3). As was said above, this reference sample contains the analyte at a specific concentration level.

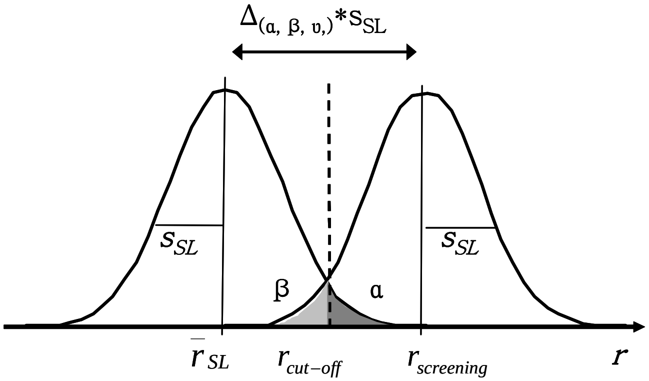


Figure 3. Statistical hypothesis tests for qualitative analytical methods providing instrumental response

The main advantages of these Hypothesis Tests derive from the use of the well-known probability of an α -type error (the probability of committing false positives) and the increasingly used probability of a β -type error (the probability of committing false negatives). This method makes it easy to evaluate uncertainty when using qualitative methods that provide an instrumental response. Traceability can also be verified and the detection limit computed. However, if the test kit does not provide an instrumental response, or if the response is based on a visual observation that cannot be quantified, Hypothesis Tests cannot be used.

3.4.4. Performance characteristic curves. Performance Characteristic Curves are a plot of the probability of having a positive result versus the concentration level of the analyte. The result is a sigmoidal type of curve the slope and the amplitude of which are particular for each qualitative method (see Fig. 4).

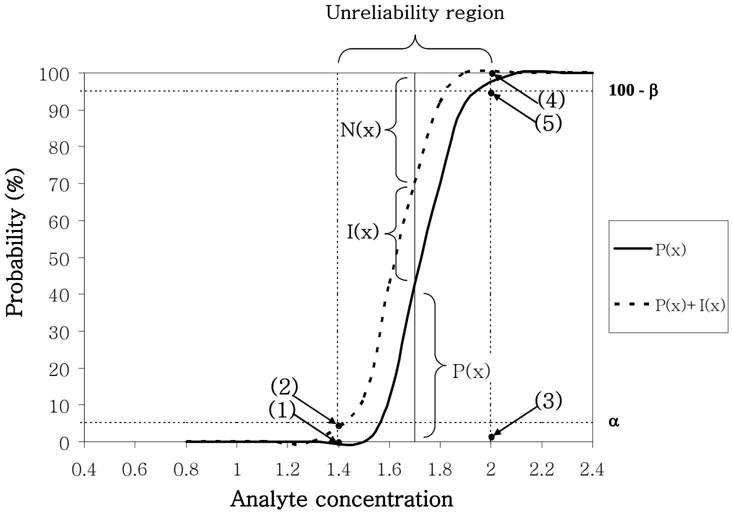


Figure 4. Performance Characteristic Curve. Probability of positive responses, P(x), and probability of positive plus inconclusive responses, P(x)+I(x), were plotted versus concentration levels tested. (1) FP=P(x); (2) $X_{0,05}$ where specificity=N(x)=100-(P(x)+I(x)), (3) $X_{0,95}$, Cut-off limit, detection limit; (4) FN=100-(P(x)+I(x)), (5) Sensitivity =P(x)=100-β

The main advantage is that considerable information is provided. In addition to false positive and negative rates, these Curves make it possible to calculate sensitivity and specificity rates and other performance characteristics of qualitative methods, such as the detection limit and the cut-off limit or the unreliability region. The main drawback is that it is necessary to perform several analyses for each concentration level.

4. Conclusions

Demand for qualitative analytical methods is increasing and they are becoming more and more important. However, some aspects still need to be developed and clarified. For users, one of the most confusing is the nomenclature used to refer to qualitative analysis, since there are many different terms that often have different meanings. Similar confusion occurs with the classification of qualitative methods, where there are several possibilities, according to different authors. Although this may be of no practical importance for many users, some work should be done to structure the criteria for classification.

Validation of qualitative analytical methods is an important issue to consider so as to provide confidence to the analysts. Although several organizations are working on this task, very few of them have defined validation protocols and their own validation programs for method developers. It has to be said that there is still confusion regarding how this validation process should be generally performed. Performance parameters are quite well defined, but, even so, a way of evaluating them has yet to be established. In this article, we have briefly described some possibilities. As far as the use of references in qualitative analytical methods is concerned, the possibilities are considerably fewer compared with quantitative analytical methods. Consequently, the references available should be examined more intensively.

Acknowledgments

The authors acknowledge the economic support from the MCyT (project No BQU2000-1256)

References

[1] International Organisation for Standardisation, ISO 8402, Quality management and quality assurance. Vocabulary, ISO, Geneva, Switzerland, 1994.

[2] /657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results.

[3] P. Feldsine, C. Abeyta and W. Andrews, J. of AOAC Int. 85 (2002) 1187.

[4] M. Valcárcel, S. Cárdenas and M. Gallego, Trends Anal. Chem. 18 (1999) 685.

[5] A. Sanz-Medel, B. San Vicente de la Riva, J. M. Costa-Fernández, R. Pereiro, Anal. Chim. Acta, 451 (2002) 203.

[6] J.J. O' Rangers, R.J. Condon, in J.F. Kay, J.D. MacNeil,
J.J. O' Rangers (Editors), Current Issues in Regulatory Chemistry, AOAC Int., Gaithersburg, Maryland, USA, 2000, p. 207.

[7] M. Unger-Heumann, Fresenius' J. Anal. Chem. 354 (1996) 803.

[8] D. Barceló, M.-C. Hennion, Anal. Chim. Acta, 362 (1998) 3.

[9] D. Barceló, A. Oubiña, J.S. Salau, S. Pérez, Anal. Chim. Acta, 376 (1998) 49.

[10] MERCK Farma y Química, S. A. (http://www.merck.es)

[11] C. Heiss, M. G. Weller, R. Niessner, Anal. Chim. Acta, 396 (1999) 309.

[12] L.C. Waters, R.R. Smith, J.H. Stewart, A. Jenkins, J. AOAC Int., 77 (1994) 1664.

[13] A. Pulido, I. Ruisánchez, R. Boqué, F. Xavier Rius, Anal. Chim. Acta, 455 (2002) 267.

[14] International Organization for Standardization, ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories, ISO, Geneva, Switzerland, 1999.

[15] EURACHEM, The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics, EURACHEM Secretariat, Teddington, Middlesex, UK, 1998 (http://www.eurachem.ul.pt).

[16] AOAC International, The cornerstone for online analytical methods (http://www.aoac.org).

[17] AOAC International, Peer Verified Methods Program – Manual on policies and procedures, AOAC Int. Gaithersburg, Maryland, USA.

[18] AOAC International. Method Validation Programs. (http://www.aoac.org/vmeth/page1.htm).

[19] AOAC International, Rapid test Kits/Performance Tested Methods (http://www.aoac.org/testkits/perftestedmtd.html)

[20] International seed Testing Association. (http://www.seedtest.org)

[21] Environmental Protection Agency. (http://www.epa.gov/)

[22] Test Methods: Methods Development and Approval Process. (http://www.epa.gov/epaoswer/hazwaste/test/methdev.htm)

[23] EA (European co-operation for Accreditation) (http//:www.european-accreditation.org)

[24] D. Barceló, B. Ballesteros, A. Dankwardt, P. Schneider, M.P. Marco, Anal. Chim. Acta, 475 (2003) 105.

[25] R. W. Sheets, Sci. Total Env. 219 (1998) 13.

[26] F.D. McClure, J. Assoc. Off. Anal. Chem. 73 (1990) 953.

[27] S. De Saeger, L. Sobanda, A. Desmet, C. Van Peteghem, Int. J. Food Microbiology, 75 (2002) 135.

[28] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Data Handling in Science and Technology 20A. Handbook of Chemometrics and Qualimetrics: Part A, Elsevier Science, Amsterdam, The Netherlands, 1997, p. 436.

[29] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhardt, R. W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel, Accred. Qual. Assur. 8 (2003) 68.

[30] E. Trullols, I. Ruisánchez, F.X. Rius, J. AOAC Int. In press

[31] L. Currie, Pure Appl. Chem. 67 (1995) 1699.

[32] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, Trends Anal. Chem. 22, (10) (2003) 647.

[33] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Data Handling in Science and Technology 20A. Handbook of Chemometrics and Qualimetrics: Part A, Elsevier Science, Amsterdam, The Netherlands, 1997, p. 475.

[34] R.M. McFall, T.A. Treat. Ann. Rev. Psychol. 50 (1999) 215.

[35] R. Song, P.C. Schlecht, K. Ashley, J. Hazard. Mater. 83 (2001) 29.

[36] B.J. Neil, E. Keeler, S.J. Adelstein, New Engl. J. Med. 293 (1975) 267.

[37] N.E. Hawass, Brit. J. Radiol. 70 (1997) 360.

3.3 QUALITATIVE METHODS OF ANALYSIS THAT USE CONTROL SAMPLES

Now that a general overview has been given of qualitative methods of analysis validation, it must be stressed that every qualitative method has special features that must be taken into account before the validation process is designed. One example is the case of qualitative methods that use control samples.

Some qualitative analytical methods provide the final result by recording the signal obtained from the sample measurement after the necessary pre-treatment steps. This final result is obtained by comparing the response or decision value, among other possibilities, of the analyte with an accepted reference. Depending on the response of this reference or the previously established decision value (using this reference or not), the sample can be classified (YES/NO) appropriately.

The reference mentioned above can be either external or internal. External references are usually well-characterized samples (e. g. Certified Reference Materials, working reference materials or spiked samples). In some situations a suitable Certified Reference Material cannot be obtained (e. g. the matrix is not stable enough or it is too complicated) or spiking a sample is not viable. In such cases, the analytical method often uses internal references.

Internal references are well-characterized samples, as well, but they are intrinsic to the qualitative method, which usually has a commercial format. These internal references are called controls and they are supplied with each specific unit of the test kit. So they must be used as long as the specific unit of the test kit is in use.

The analytical methods which provide internal references or control samples are mainly used in the field of clinical analysis, where the sample response is usually compared to a cut-off control or calibrator.

In the next section, a review of these methods of analysis is presented. The main groups, their characteristics and performance are described. A validation procedure is also briefly described. The example given is for a method of analysis used in the clinical context. It requires control samples to calculate the decision value.

3.3.1 VALIDATION OF QUALITATIVE METHODS OF ANALYSIS THAT USE CONTROL SAMPLES

Trends in Anal. Chem. 24, 2005, 516.

E. Trullols, I. Ruisánchez, F.X. Rius and J. Huguet[a].

Universitat Rovira i Virgili. Departament de Química Analítica i Química Orgànica. C/ Marcel·lí Domingo s/n. 43007 Tarragona (Spain)
[a]Laboratorio de análisis Dr. Echevarne. C/ Provença 312. 08037 Barcelona (Spain)

Abstract

Qualitative methods are frequently used for screening. In some applications, the resulting positive samples are subsequently analyzed by a suitable quantitative confirmatory method, so it is important that the qualitative assay provides reliable results. Although some validation procedures have been reported in this area, much work is still required because there are many different qualitative methods with many different characteristics. In this report, we examine the different types of control samples used in qualitative analysis that provide instrumental responses, we review the most important quality parameter in the validation process, we propose a procedure for estimating the selected quality parameters-traceability, the unreliability region, sensitivity and specificity rates, and false positive and negative rates- and we show how their values can be calculated in a case study: an ELISA method used in a clinical context.

Keywords: Binary results; Control samples; Performance quality parameters; Qualitative analysis; Validation

1. Introduction

During the last decade, qualitative methods have been widely developed and, as a result, some of them are now used as routine laboratory methods. However, the range of applications is not as wide as in quantitative analysis. They are mainly used as screening methods, selecting the positive samples and considerably reducing the time and cost of the confirmatory analyses.

As is well known, a key point when dealing with either quantitative or qualitative analytical methods is their validation. Method validation was defined some time ago by ISO [1] and, from the practical point of view, it can be considered as the definition and the estimation of the performance parameters necessary to match the analytical requirements. The validation procedure should always take into account the intended use of the analytical method. The validation of qualitative methods is not as developed as the validation of quantitative methods, which have been the subject of numerous studies [2-4]. Some guidelines are therefore already available and accepted by either regulatory bodies or practitioners in specific fields. At present, the situation is changing, because recent studies have focused on the validation of qualitative methods. This means that some documents and guidelines are available, although they are still not generally accepted [2, 3, 5-7].

There are numerous qualitative methods and their validation methodology depends on their specific nature. In this article, we focus on those methods that use an instrumental response (e. g., absorbance, current intensity, and peak area) to classify the test sample into two different categories: positive/yes and negative/no. More specifically, we focus on those methods that use control samples. Controls are commonly used to establish a limit value known as a “ cut-off value” (COV) (i. e., the limit at which the samples can be assigned to one of the two different categories).

As a case study, we discuss the validation of an immunoassay-based test kit that measures immunoglobulin G class antibodies to Varicella-Zoster Virus in human serum (i. e., it is used in the clinical context). On the one hand, the kit uses controls to establish the COV and, on the other, it provides a final YES/NO result based on absorbance measurements as the instrumental response. The test samples are therefore classified according to the established COV. The test kit is an enzyme-linked immunosorbent assay (ELISA) [8], based on the antigen-antibody reaction. Some ELISA methods use control samples to calculate a reference value that is necessary for classifying the samples into different categories according to the property measured, whereas other test kits use different types of calibration samples [9].

First, we briefly describe the different types of test kits that use control samples, either in the same way as the kit selected for the case study or not. Then, we propose and define the quality parameters for such test kits, and, finally we report the validation of the specific test kit.

2. Qualitative methods that use control samples

Numerous test kits use control samples, supplied by the manufacturer, as part of their methodology for classifying test samples. Within this format, there is a wide variety of possibilities. Some test kits just use positive and negative controls, which are usually used to establish the COV and to validate the analysis internally. Others provide the end user with a solution at the activity level of the COV. And yet others, in addition to the positive and the negative controls and to the cut-off control sample, also require intermediate levels of positive controls.

Because the key point of this paper is the validation of test kits that use controls, let us first define the various controls used:

- Negative control is a blank sample (i. e., a sample that is known to be free of the target analyte). In the framework of clinical chemistry, it is a real serum sample from a patient (or a pool of patients) that it is proved not to have the antibodies against a specific antigen.
- Positive control is a sample containing a perfectly known amount of the target analyte. In the framework of clinical chemistry, it is a pool of positive real serum samples (i. e., samples from patients that have been proved beyond all doubt to have the antibodies against a specific antigen).
- Cut-off control or cut-off calibrator is a sample containing the amount of analyte corresponding to the cut-off level. For clinical chemistry, it is a sample of human serum that has been prepared to provide a limit value of activity.

- Intermediate controls can be used depending on the application and the test kit (e. g., samples that are positive even though they do not contain a considerable amount of the target analyte and are considered as low positive controls). In some cases, they provide an inconclusive result. Some samples are high positive samples, which mean that they contain a considerable amount of the target analyte (antibodies).

In most cases, these controls are required for estimating the COV. Generally speaking, the COV is the value from which the decision about the test sample must be taken and it refers either to the response domain or to the concentration or activity domain. This value can be set by legislation when dealing with the maximum contents of some contaminants in food, as is the case of Aflatoxin B_1 in nuts, the maximum content of which is regulated by the European Commission [10] or when dealing with drinking water pollutants [11], which are strictly controlled by several regulation bodies [12].

It is quite common, mainly in the context of clinical analysis, to estimate this COV using a mathematical expression provided by the manufacturer' s test kits [13]. Alternatively, it is recorded as an instrumental response, when measuring a specific sample (' cut-off control or calibrator'), also provided with the test kit [14].

Although attempting to classify the different test kits is always risky, for the sake of clarity we have decided to differentiate between them by the presence or absence of a cut-off control sample and specially those cases where the COV refers to the response domain.

2.1. Test kits without a cut-off control sample

These test kits measure only the positive and the negative control samples to estimate the COV that is usually calculated by means of a mathematical expression that combines the response values from both control samples. Once the cut-off has been obtained, and, always in terms of instrumental response, it is compared to the value obtained for the test sample.

In some cases, the control samples are measured every day, so the COV is also obtained daily. Additional information about the day-to-day variation in the COV is therefore also possible. These control values can also be used as internal validation for the assay; since they are different every day, they must comply with some requirements. Usually, it must be ensured that the positive and negative controls fall into a specific range of instrumental response values.

2.2. Test kits with a cut-off control sample

As well as the positive and negative control samples in their test kits, some manufacturers provide an extra sample named the “ cut-off control” or “ cut-off calibrator” . In these cases, the measured responses from the test samples are directly compared with the measured response from the cut-off control. This can be done by directly comparing instrumental responses or sometimes by establishing a function between both responses and comparing this value with a preset range of values. In addition to the cut-off control calibrator, other manufacturers provide a low and a high positive control so that the range near the COV and the upper positive range can be controlled. This does not affect the way the COV is established and subsequently compared to the test sample value.

We would like to emphasize the importance of the COV, although it is not considered a quality parameter in the validation process. This importance is illustrated by the fact that the COV directly defines the regions where actual negative and positive responses are obtained. Moreover, the limits that define the region where inconclusive sample results are obtained depend on the error associated to this COV. Significant information about the performance of the test kit can therefore be inferred from the COV.

3. Identification of the relevant Quality Parameters

The quality parameters must be carefully identified and estimated according to the requirements that the analysis should fulfil [15]. These requirements normally involve a wide variety of items related to the information we want to obtain: verification of traceability, estimation of the uncertainty associated to the results, cost and time constraints, and practical parameters, such as reusability or possibilities of automation, to give just a few examples. The " fitness-for-purpose" [16] approach is used to identify, estimate and finally validate the quality parameters depending on the requirements to be fulfilled.

Concerning the quality parameters that have a statistical character, as in any qualitative method, in addition to the traceability and to the estimation of the uncertainty of the results, it is important to consider the probabilities of providing false positive and false negative results. It is also important to properly define the region that provides inconclusive results. This region is around the COV so, as we will see later on, most of the quality parameters are related to the lack of precision associated to the COV. The most important quality parameters are described below.

3.1. Traceability

According to a recent definition of traceability [17] and from the practical point of view, we assume that there is an unbroken chain of calibrations of a measuring system or comparisons. Among other possibilities, traceability can be assessed by comparing the results obtained from the method to be validated with those obtained by a reference method, or by using a certified reference material [18].

The control samples provided by the manufacturer can be considered as secondary references, since there is a formal statement that they have been compared to an in-house serum preparation and that the whole test kit has been compared to another commercially available ELISA [19].

3.2. Sensitivity and specificity rates

In the framework of qualitative analysis, sensitivity and specificity refer to the ability of the test kit to classify positive samples (sensitivity) or negative samples (specificity) when indeed they are positive or negative [20]. Both parameters therefore give an idea of how good the test kit classifies positive and negative samples. They are closely related to the rates of false results. It is of utmost importance to assess that the test kit has a high sensitivity and a high specificity in order to avoid any false result.

Closely related to the occurrence of interferences, to specificity and to false positive rate, selectivity [21] must be also taken into account. A test kit lacks selectivity if a set of substances, or the matrix as a whole, has an effect on the signal of the analyte measured. The manufacturer assesses the general absence of cross reactivity [19]. However, in the cases where the test kit is used as a routine method, the manufacturer also suggests ruling out some infections before interpreting the result of the Varicella-Zoster Virus (VZV) test, due to expected cross reactivity, since the VZV is related to other viruses of herpes viridae family.

3.3. Unreliability region

When dealing with binary responses (YES/NO), it is not meaningful to consider the classical definition of the uncertainty of the final results [22]. The term unreliability region better describes the idea of a region in which there is a certain probability of error, and therefore a region in which false results may be obtained [6, 23]. In the particular case of test kits that provide an instrumental response (numerical value), the unreliability region can be defined by the range of instrumental responses that provide inconclusive results. The unreliability region is a key point in the validation process because of the considerable amount of information that it provides.

If the measurement of control samples to establish the COV is needed, the definition of the unreliability region takes into account the precision associated with this COV. When a mathematical expression is used to calculate the COV, its precision is easily determined by using the error-propagation law. The unreliability region is defined by an upper and a lower limit, which make it possible to estimate the sensitivity and the specificity, and the false positive and false negative rates (see Fig. 1).

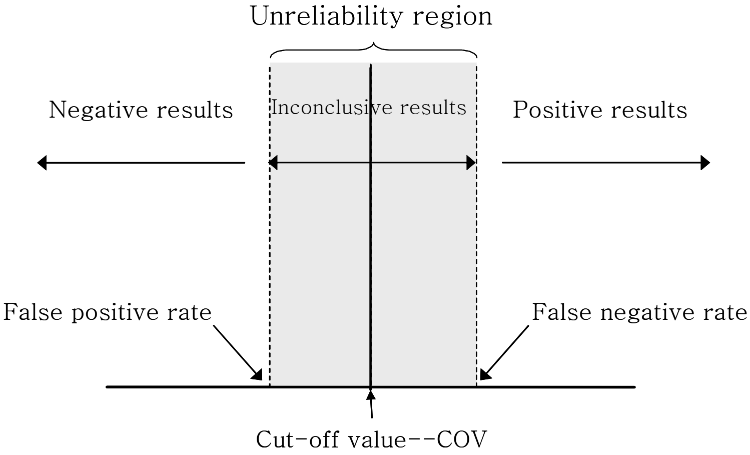


Figure 1. Definition of the unreliability region and the information that it provides: region of positive, negative and inconclusive results, false positive and negative rates in the response domain

Test-kit manufacturers usually provide an error associated with the COV (e. g., as a percentage in terms of relative standard deviation). This means that the samples with response values higher than the COV plus the specific percentage of this COV will be positive and the probability of error will theoretically be very small. However, the manufacturer does not provide this information. The same occurs with the samples that give rise to response values smaller than the COV minus the specific percentage of this COV: they will be negative with a very small probability of error. The samples that give rise to response values within this interval will be classified as inconclusive.

The information related to the error associated to the COV, which is provided by the manufacturer, should be validated by comparison with the experimental results. This means that the error of the cut-off should be experimentally evaluated, the unreliability region should be defined and they should then be compared with the values claimed by the manufacturer.

3.4. False positive and false negative rates

False positive and false negative are the probability that the test kit will classify the samples as positive when they are in fact negative (false positives), or as negative when they are positive (false negative) [3]. Closely related to sensitivity and specificity, these false rates also give an idea of how well the test kit classifies, although in the sense that it estimates the probability of giving results that are false. The false rates are closely related to the unreliability region because they are inferred from its lower (false positive) and upper limits (false negative).

In many cases it is a challenge to keep both rates (probabilities of error) to nearly zero. In such situations, one should evaluate the consequences of either providing false positive results or false negative results. Depending on that, the approach would be either setting the probability of committing:

- α type error (false positives) as small as possible, if the consequences of considering a not immunized patient (without the antibodies) as immunized are worse than considering an immunized patient as not immunized; or,

- β type error (false negatives) as small as possible, if the consequences of considering an immunized patient as not immunized are worse than considering a not immunized patient as immunized.

Though it is not a rule, in a wide range of clinical analysis, false negative results are more critical since positive results are checked using other analytical methods, either test kits or not. In these cases, and particularly if the test kit is used as a routine method in the laboratory, deeper studies concerning the occurrence of false negative results using different matrixes and involving a wider range of possible cross-reactants should be carried out.

When the test kit also provides inconclusive results, the lower limit of the unreliability region is related to the false results, but in the sense that it gives the percentage of negative samples that will give an inconclusive result. The percentage of negative samples that give a positive result will be always much lower (or nearly zero) than those that give an inconclusive result. For the upper limit of the unreliability region, the situation is very similar. The upper limit is related to the probability of giving false results but in the sense that it provides the rate of positive samples that will give an inconclusive result.

Quality parameters other than those defined in this article can also characterize a test kit. Parameters such as robustness may be important when the same assay is to be used in different conditions (e. g., in different laboratories). The detection limit is also an important quality parameter because, in some qualitative assays, it is given by the lowest concentration of the analyte that the kit can reliably detect as positive in the sample matrix [6]. Finally, in relation to the unreliability region, prediction intervals for future samples can be estimated, as they are directly related to the error associated with the COV.

To improve the characterization of the test kit, statistical tools, such as control charts, can also provide valuable information (e. g., whether the instrumental responses of the control samples are within or beyond the accepted range of values). These control charts and the information they provide, discussed in another contribution in the present issue, together with the last mentioned quality parameters, will be studied in a future paper.

4. Estimation of the Quality Parameters: a case study

Once the main quality parameters required to validate a test kit have been defined, we show how they can be estimated in practice with the validation procedure of a particular test kit. We have used a test kit that measures IgG antibodies to Varicella-Zoster Virus in human serum, so the context is a clinical one. First, we will describe the test kit and the experimental work carried out to estimate the quality parameters.

4.1. Test kit performance

The test kit used, VZV IgG [13], is an indirect ELISA that detects the IgG antibodies to Varicella-Zoster Virus in human serum. The microtiter wells are coated with a Varicella-Zoster Virus antigen from a cell culture. After an incubation period, the antibodies in the test sample or the control sample are linked to the antigen coating the microtiter. In a second incubation period, a conjugate anti-IgG (anti-human IgG antibodies traced with peroxidase) binds to the IgG antibodies. When the substrate 3, 3' , 5, 5' -tetramethylbenzidine with hydrogen peroxide is added, it turns blue and finally yellow when the stop solution is added. The intensity of this colour, measured by means of a spectrophotometer at 450nm, is proportional to the concentration of antibodies in the sample.

Once the absorbance value of the test sample (or a related index) has been recorded, it is compared with the absorbance value of the cut-off (or cut-off index, which is always equal to one). As a consequence of this comparison, the test-kit result is transformed into a YES/NO result for the presence or absence of IgG antibodies to Varicella-Zoster Virus.

In the test kit we used, the COV was obtained by combining the absorbance values of the negative and positive controls with a mathematical transformation specified by the manufacturer. This transformation involves two steps: the first is to calculate the COV (Equation (1)), using the mean absorbance value for the control samples (negative and positive) that are measured in the same microtiter plate as the samples:

$$COV = \overline{A}_{-} + 0.1 \times \overline{A}_{+}\ , \qquad (1)$$

where $\overline{A}_{-}$ is the mean value of the absorbance for the negative control; $\overline{A}_{+}$ is the mean value of the absorbance for the positive control.

According to the manufacturer, the COV has an associated variation of 15%. Although this value is provided without units and without information of how it is calculated, we have assumed that it is a coefficient of variation. The following results are therefore derived from the test:

1. If the absorbance measured at 450 nm (serum test sample) is higher than the COV + 15%, the sample is given as positive. This means that the sample serum is considered to have IgG antibodies to Varicella-Zoster Virus.
2. If the absorbance measured at 450 nm (serum test sample) is lower than the COV - 15%, the sample is given as negative. In this case, the sample is considered not to have IgG antibodies to Varicella-Zoster Virus.
3. If the absorbance measured at 450 nm (serum sample) lies between the COV plus and minus 15%, the sample is given as inconclusive. This value is given according to the intrinsic characteristics of the samples and, in addition, because the manufacturers must provide the end users with a range of values that refer to inconclusive samples.

From the practical point of view, it may be more convenient to work with indexes than with raw absorbance values, because all response values refer to the COV. So, the second step is to calculate the sample indexes (Equation (2)):

$$Index = \frac{Sample\ absorbance}{COV} \qquad (2)$$

With these indexes, it is even easier to apply the criteria described above to take the decision about the sample. The cut-off index will be always equal to 1 by definition and, if the 15% is taken into account, the criteria can be stated as:

1. If the sample index is higher than 1.15, the sample is considered positive.
2. If the sample index is lower than 0.85, the sample is considered negative.
3. If the sample index is between 1.15 and 0.85, the sample is considered inconclusive.

According to this description, it can be seen that the COV for this test kit is compared to the sample in the response domain, as no relation is established between the response and the activity of the sample.

4.2. Experimental work

There are several ways of establishing the quality parameters of a test kit [6, 24]. Depending on the option chosen, the experimentation to be carried out should be carefully designed. In the present study, the experimental work is based on characterizing the distribution of the control samples, as we are dealing with an instrumental response (numerical values). The control samples must therefore be analyzed a sufficient number of times for their distributions to be characterized. One of the possible experimental designs considered is the one shown in Fig. 2. The analyses are performed for 30 days and, every day, two replicates of the controls are measured by the same analyst under the same conditions, according to the instructions provided with the kit.

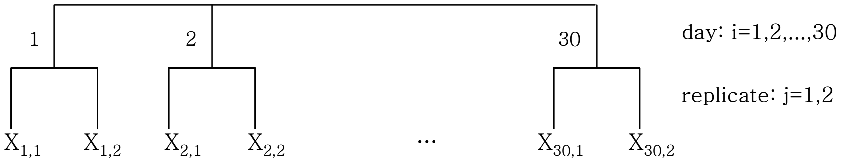


Figure 2. Experimental design used to measure controls (positive and negative) and estimate the cut-off value

4.3. Results and discussion

4.3.1. Traceability In order to assess the traceability of the results, a reference material [18] is measured simultaneously with the control samples for 30 days. The aim is to compare the responses of the reference material and the positive control sample, both of which have the same activity. This reference material is an ampoule containing lyophilized Varicella-Zoster IgG antibodies. If these antibodies are diluted in 1 mL of distilled water, the activity is 4 UI/mL. Once we have this solution with the antibodies, we need to further dilute it by a factor of 1/200 for the activity to be equivalent to the activity of the positive control sample.

The data obtained with the reference material follow a *t*-Student probability-distribution function (Fig. 3). The mean value from the data of the positive control and the mean value of the reference material can therefore be compared using the *t*-Student test. Table 1 shows that the traceability is assessed because the mean values of both distributions do not differ significantly (t_{cal} = 1.89 is lower than t_{tab} = 1.99) at a 5% level of significance.

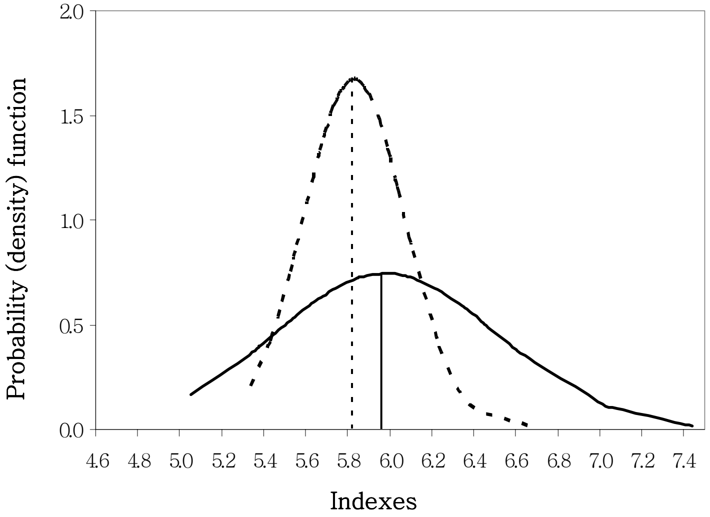


Figure 3. Index distribution obtained for the positive control sample (dotted line) and for the reference material (solid line)

Table 1. Mean values and standard deviations from the positive control sample and the reference material measurements

Positive control sample	Mean value = 5.83	Standard deviation =0.24	n= 60
Reference material sample	Mean value = 5.98	Standard deviation =0.57	n= 60
t_{calc} = 1.89 t_{tab} = 1.99 $t_{calc} < t_{tab}$ no significant differences are detected at α = 5%			
Comparison of the mean value using a t-Student test			

4.3.2. Sensitivity and specificity. According to the manufacturer’ s instructions, a negative and a positive control sample, both of which are provided with the kit, must be measured twice every day so that the daily COV can be calculated. We use these measurements to estimate the sensitivity and the specificity of the test kit by assessing that positive control sample measurements give positive results and that negative control samples provide negative results. Fig. 4 shows the *t*-Student probability-distribution function for both controls and the COV with the upper and lower limit of the unreliability region. As can be clearly seen, all the negative samples measured provide negative results as they are below the lower limit of the unreliability region (0.85) and all the positive results are above the upper limit of the unreliability region (1.15) and the test kit always provides a positive result.

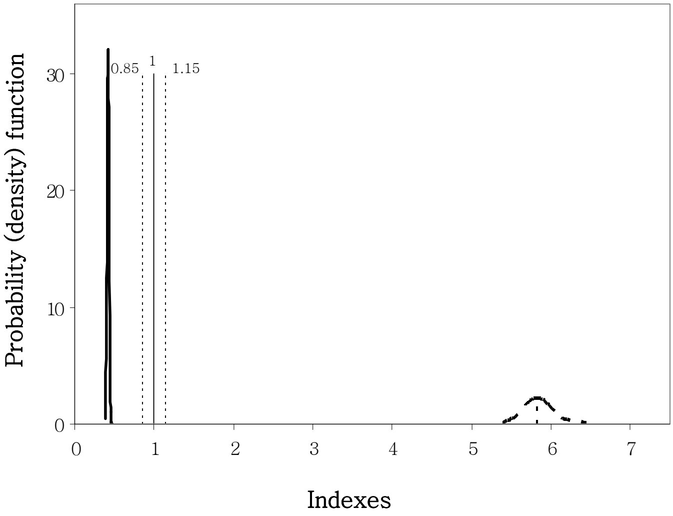


Figure 4. Index distribution obtained for the negative control sample (solid line) and for the positive control sample (dotted line). The variation of 15 % in the cut-off value (0.85 and 1.15) is also plotted

It can therefore be concluded that this test kit is specific because it provides negative results for all the negative control samples measured and that it is also sensitive because it provides positive results for all the positive control samples measured. In this particular case, it is logical, as the positive and negative control distributions are far from the cut-off or unreliability limits.

4.3.3. Unreliability region. The estimation of this region is directly related to the lack of precision associated with the (COV). Indeed, the cut-off precision can be used for two purposes:

1. To estimate the cut-off variation over time. Once this region has been established, future COVs that may be suspected of being wrong can be evaluated by checking whether they belong to this unreliability region or not.
2. To estimate the unreliability region and, therefore, to predict test sample compliance.

In order to estimate the precision of the COV, we can use the information gathered during the analysis of the control samples and apply the error-propagation law to Equation (1). For this particular case, the variables are the mean absorbance value of the negative control measurements ($\overline{A}_{-}$) and the mean absorbance value of the positive control measurements ($\overline{A}_{+}$). The final expression is depicted in Equation (3):

$$s_{COV}^2 = \left[\frac{\partial COV}{\partial \overline{A}_{-}}\right]^2 \times s_{\overline{A}_{-}}^2 + \left[\frac{\partial COV}{\partial \overline{A}_{+}}\right]^2 \times s_{\overline{A}_{+}}^2 \ . \qquad (3)$$

When both variables are partially derived, the result is as expressed in Equation (4):

$$s_{COV}^2 = 1^2 \times s_{\overline{A}_{-}}^2 + 0.1^2 \times s_{\overline{A}_{+}}^2 = s_{\overline{A}_{-}}^2 + 0.1^2 \times s_{\overline{A}_{+}}^2 \ . \qquad (4)$$

The mean values $\overline{A}_{-}$ and $\overline{A}_{+}$ correspond to a set of measurements that were made over 49 days. They also show a *t*-Student probability-distribution function with a mean of 0.145 and a standard deviation of 0.032 for the negative control sample; and a mean of 1.431 and a standard deviation of 0.081 for the positive control sample. The average value for the COV is 0.28.

Finally, s_{COV} = 0.033, which is a relative standard deviation of 12% with respect to the COV. It can therefore be concluded that there is no significant difference between the 12% estimated experimentally and the value given by the manufacturer of 15%. If necessary, a shorter unreliability region could be defined in which results would be inconclusive between the indexes 0.88 and 1.12.

4.3.4. False positive and negative rates. As has been shown in the assessment of sensitivity and specificity, this test kit classifies negative samples and positive samples correctly when control samples are used. The α and β probabilities of error (false positive and negative rates) are therefore nearly zero in the region where these control samples provide their indexes. However, it is advisable to have information in the proximity of the unreliability region, where the probability of obtaining false results is high. The easiest way to obtain samples that elicit indexes close to this region is to dilute the positive control.

Theoretically, the relation between the instrumental response and the activity or the concentration of the analyte when ELISA methods are used is not linear. This relation can be established using such models as the 4-parameter, the logit-log or the cubic spline, among others [9]. In this application, an in-depth study of several dilution factors showed that the relation between the index value and the dilution factor follows a quadratic function (Fig. 5). However, it also shows that the dilution factors that give rise to samples within the target region are 1/8 and 1/12.

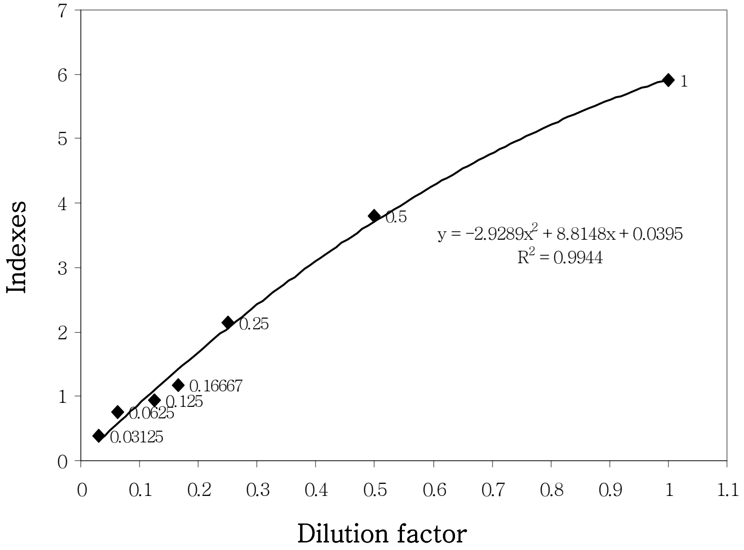


Figure 5. Quadratic relation (solid line) of the indexes (◆) obtained when the positive control sample is diluted by several dilution factors

The diluted positive control samples at 1/8 and at 1/12 were analyzed using the same experimental design described at the end of section 4.2. The results depicted in Fig. 6 show that they also follow a *t*-Student distribution function.

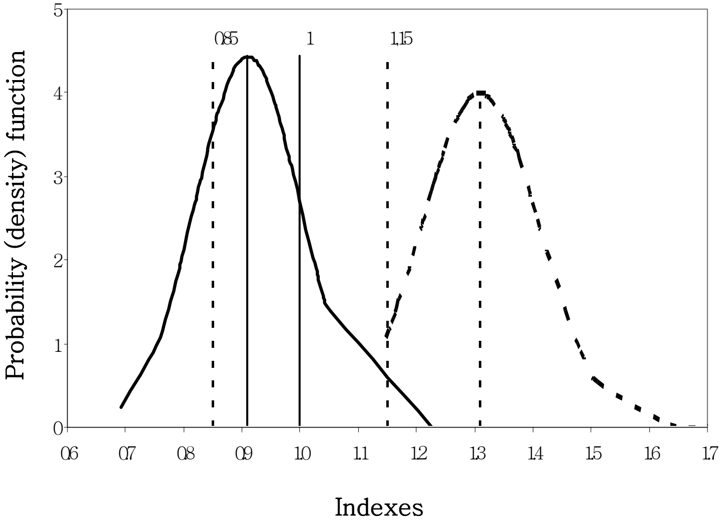


Figure 6. Index distribution obtained for the positive control sample diluted at 1/12 (solid line) and 1/8 (dotted line). The cut-off value and its variation (0.85 and 1.15) are also plotted

The results for the 1/8 dilution factor of the positive control sample have a mean distribution of 1.3 and a standard deviation of 0.10 and the results of the 1/12 dilution factor have a mean distribution and standard deviation of 0.91 and 0.09, respectively. These distributions of the results are used to estimate the theoretical false positive and false negative rates according to the well-known *t*-Student probability-distribution function (Equation (5)).

(a) False positive rate

$$\bar{I}_{-} + t \times S_{-} = 1.15$$

$$t = \frac{1.15 - \bar{I}_{-}}{S_{-}} = \frac{1.15 - 0.91}{0.090} = 2.7$$

$$\alpha = 0.48\%$$

(b) False negative rate

$$\bar{I}_{+} - t \times S_{+} = 0.85$$

$$t = \frac{\bar{I}_{+} - 0.85}{S_{+}} = \frac{1.3 - 0.85}{0.10} = 4.6$$

$$\beta = 0.001\%$$

(5)

where $\bar{I}_-$ is the mean value of the indexes for the negative control sample; s_- is the standard deviation of the indexes for the negative control sample; $\bar{I}_+$ is the mean value of the indexes for the positive control sample, and s_+ is the standard deviation of the indexes for the positive control sample.

These theoretical probabilities of committing errors were compared to the experimental probabilities, which were calculated using the experimental data. To estimate the false positive rate, we consider as false positives those results whose experimental index after the analysis of the positive control sample diluted by a factor of 1/12 was equal to or higher than 1.15. In our case, there was just one measurement out of 60, so the probability of committing a type α error is calculated using Equation (6):

$$\frac{False\ positive\ results}{Total\ number\ of\ analysis}\times 100=\frac{1}{60}\times 100=1.6\% \qquad (6)$$

The difference between the theoretical probability (0.48%) and experimental probability (1.6%) is probably due to the relatively small number of samples analyzed.

The experimental probability of committing false negatives (type β error) is calculated in a similar way. In this case, to estimate the false negative rate, we consider false negatives to be those results whose experimental indexes after the analysis of the positive control sample diluted by a factor of 1/8 are equal to or lower than 0.85. For this group of indexes, we have not obtained a single measurement with these characteristics. The probability is therefore 0%. In this case, there is a good agreement between the theoretical false negative rate (0.001%) and the experimental false negative rate (0%).

5. Conclusions

Without any doubt, qualitative methods should also be validated. Method validation depends on the characteristics of the qualitative methods being used and should be designed according to their particularities. We have identified and defined the most important quality parameters for the different qualitative methods that use control samples and obtain the responses by instrumental analysis. Control samples are mainly used to establish the COV. Because they are standards of a certain metrological level, they can also be used in the validation process. The instrumental responses are transformed to obtain the final binary result (YES/NO) or, as in the case study, inconclusive.

The range of values considered as inconclusive is a key point in the definition of quality parameters, such as the unreliability region around the COV, and the false positive and false negative rates.

We have also estimated other important quality parameters, such as traceability, sensitivity and specificity rates.

As a case study, we validated a commercial test kit that uses control samples and provides instrumental responses with a final result of the positive/yes and negative/no type. The quality parameters - sensitivity, specificity, and rate of false results - are defined and estimated using the statistical distributions of the control samples. Traceability is assessed by using a reference material and the definition of the unreliability region takes into account the precision associated to the COV. This precision is estimated by applying the error-propagation law to the response measurements, which, in this case, were the absorbance values.

The validation process has revealed that the cut-off provided by the manufacturer was accurate but that its associated standard deviation was wider than the experimental value. This meant that the manufacturer had chosen a conservative option when providing the final results in order to avoid false positive and negative results.

Acknowledgments

The authors acknowledge economic support from the MCyT (Project nº BQU2003-01142). Esther Trullols would like to thank the URV for providing a doctoral fellowship.

References

[1] International Organization for Standardization, ISO 8402, Quality Management and Quality Assurance. Vocabulary, ISO, Geneva, Switzerland, 1994.

[2] European Commission, 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC

concerning the performance of analytical methods and the interpretation of results, Off. J. Eur. Commun. L221 (2002) 8 (http://europa.eu.int/eur-lex)

[3] P. Feldsine, C. Abeyta, W. Andrews, J. AOAC Int. 85 (2002) 1187.

[4] R. Song, T.J. Fischbach, K. Ashley, Am. Ind. Hyg. Ass. J. 57 (1996) 161.

[5] M. Valcárcel, S. Cárdenas, M. Gallego, Trends Anal. Chem. 18 (1999) 685.

[6] E. Trullols, I. Ruisánchez, F.X. Rius, Trends Anal. Chem. 23 (2004) 137.

[7] E. Trullols, I. Ruisánchez, F.X. Rius, M. Òdena, M.T. Feliu, J. AOAC Int. 87 (2004) 417.

[8] J.R. Crowther, The ELISA Guidebook, Humana Press, Totowa, NJ, USA, 2001 p. 302.

[9] J.W.A. Findlay, W.C. Smith, J.W. Lee, G.D. Nordblom, I. Das, B.S. DeSilva, M.N. Khan, and R.R. Bowsher, J. Pharm. and Biomed. Anal. 21 (2000) 1249.

[10] European Commission, 2002/257/EC: Commission Regulation of 12 February amending Regulation (EC) No 194/97 setting maximum levels for certain contaminants in food-stuffs and Regulation (EC) No 466/2001 setting maximum levels for certain contaminants in food, Off. J. Eur. Commun. L77 (2001) 1 (http://europa.eu.int/eur-lex/)

[11] Eurpean Commission, 1998/83/EC: Council Directive (and Corrigendum) of 3 November on the quality of water intended for human consumption, L330 (1998) 32 (http://europa.eu.int/eur-lex/)

[12] Environmental Protection Agency (http://www.epa.gov/)

[13] Human Gesellschaft für Biochemica und Diagnostica mbH. Max-Planck-Ring 21, D-65025 Wiesbaden, Germany

[14] DiaSorin srl, Anti-HSV-1/Anti-HSV-2 IgG Enzyme Immunoassay Kit, DiaSorin srl, I-13040 Saluggia, Italy.

[15] Eurachem, Guide to Quality in Analytical Chemistry. An Aid to Accreditation, CITAC/EURACHEM Guide, Eurachem, 2002 (http://www.eurachem.ul.pt/guides)

[16] Eurachem, The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics, Eurachem, 1998 (http://www.eurachem.ul.pt)

[17] International Organization for Standardization, International Vocabulary of Basic and General Terms in Metrology (VIM), Revision of the 1993 edition, International Vocabulary of Basic and General Terms in Metrology (VIM), ISO, Geneva, Switzerland, 2004 (http://www.abnt.org.br/ISO_DGuide_99999_(E).PDF)

[18] National Institute for Biological Standards and Control, ' British Standard Varicella-Zoster antibodies' . (http://www.nibsc.ac.uk)

[19] Human Gesellschaft für Biochemica und Diagnostica mbH. (http://www.human.de/data/gb/vr/el-vzvg.pdf)

[20] J.J. O' Rangers, R.J. Condon, in: J.F.Kay, J.D. MacNeil, J.J. O' Rangers (Editors), Current Issues in Regulatory Chemistry, AOAC International, Gaithersburg, MD, USA, 2000, p. 207.

[21] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, in: Data Handling in Science and Technology, Handbook of Chemometrics and Qualimetrics: Part

A, vol. 20A, Elsevier Science, Amsterdam, The Netherlands, 1997, p. 436.

[22] International Organization for Standardization, ISO 3534-1, Statistics, Vocabulary and Symbols, ISO, Geneva, Switzerland, 1993.

[23] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhardt, R.W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel, Accred. Qual. Assur. 8 (2003) 68.

[24] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, Trends Anal. Chem. 22 (2003) 647.

3.4 TRENDS IN QUALITATIVE METHODS OF ANALYSIS

As is only to be expected, there is considerable agreement among the scientific community about the importance of developing and applying qualitative method validation. This increasing agreement is largely due to the performance characteristics of qualitative methods, particularly those of rapidity and easiness of handling. Therefore, new progress is necessary to satisfy the demand.

In this respect, the European project ‘ MEQUALAN’ focused on the quality assurance of qualitative analysis. The working group reported the main issues that affected the quality principles of qualitative analysis [2]. Such important topics as traceability, reliability or validation were examined to assess the quality of the results and, finally, to incorporate qualitative methods in the laboratory routine with a high degree of confidence.

Recently, a special issue of the journal Trends in Analytical Chemistry focused on modern qualitative analysis. The aim was not only to review the main features of qualitative analysis but also to present the new approach to qualitative analysis. This new approach was first developed in the above mentioned European Project. Such aspects as the reliability of binary analytical responses [3], the identification of chemical compounds [4] or quality control [5] were discussed.

The starting point in this issue was a discussion of the current terminology and the statistics used [6]. The part on terminology describes several problems which can affect qualitative analysis and defines some important quality parameters.

The section on statistics is divided into several parts. Current practice in interlaboratory studies is examined quite intensively. Calculation of error rates, the concepts of accordance and concordance and contingency tables are all used to analyze the data and to extract the maximum amount of information. The modeling of qualitative responses is also discussed and some examples given.

The special issue provides an in-depth definition of analytical features in qualitative analysis [7]. These qualitative methods of analysis and their binary type responses must first be characterized if the analytical properties are then to be defined. On the basis of the classical analytical characteristics in quantitative analysis, relevant performance parameters such as reliability, representativeness and robustness are carefully defined and discussed. Finally, it is stressed that method validation is fundamental to the conjunction of fitness for purpose and the performance parameters (derived from the analytical properties). Validation procedures for qualitative methods of analysis are divided into two groups: methods of identification and methods of classification. The validation procedures presented vary according to the intended use, the quality of the results required and the inherent characteristics of each qualitative method of analysis.

Reliability is one of the most important analytical features [3]. The basic descriptors of reliability are traceability and uncertainty, even when dealing with analytical methods that provide binary responses. This article deals with the problem of defining and applying uncertainty and traceability in qualitative analysis.

The identification of chemical compounds is also discussed [4]. Concepts such as testing hypotheses, the so-called false response rates or the prevalence of the analyte are described. They link qualitative analysis with the identification of chemical compounds.

Once the performance parameters have been well-established and the method of analysis has been validated, the performance of the method needs to be supervised. This point is discussed thoroughly in the paper by Simonet [5]. Once the distinction between *Quality Assurance* and *Quality Control* has been made, the concepts related to quality control in qualitative analysis are systematized. Several proposals to establish an internal quality control are also discussed.

In this special issue there is also a place for an approach involving multivariate-based methods for qualitative analysis [8]. In particular, it focuses on the difficulties that must be faced when the unreliability region is defined in the multivariate analysis methods.

Two articles deal with different practical aspects. Barceló et al. [9] focus on the screening of pollutants in water, sludge and sediment samples. The biological methods used with screening responses are classified according to the technical principles involved and the subcategories are characterized.

The field of clinical analysis is also approached [10]. The authors present a review of the terminology from laboratory medicine which is related to qualitative analysis but not in a classical way.

As well as this special issue, some validation strategies for specific applications other than the ones presented in this doctoral thesis have recently appeared in the bibliography. Nitrite control in water, for example, has been chosen as a case study [11]. The development of a qualitative spot test and its validation involves several steps. First, the preparation of the spot test means that the qualitative method, which is thoroughly described, must be optimized. Second, the reliability of the spot test is determined. Then, the validation is carried out by analyzing synthetic standard samples and screening real samples. A very interesting novelty is that the validation process is integrated with an internal quality control, which is based on qualitative control charts and Youden plots in this case.

There is also another example of the screening of toxic metal ions in water samples [12]. A spectrofluorometric method measures the spectra of the complex resulting from the reaction of 6-mercaptopurine with toxic metal ions (e. g. Pb (II), Hg (II) or Cd (II)). The performance parameters are also evaluated considering the legislation limits for the toxic metal ions under study.

As has been stated, qualitative analysis is becoming an important issue in several fields. The subjects involved are heterogeneous (e. g. performance parameters, statistics, quality control) and method validation is also important. Although there is still room for greater effort, interest in defining flexible and applicable validation procedures is growing [7]. These should be adapted by the end user to the problem at hand, always bearing in mind the requirements that the qualitative method of analysis must fulfill.

3.5 REFERENCES

[1] Decision from the Commission. Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. CO (2002) 3044 final (12.08.02).

[2] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhardt, R.W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel, *Accred. Qual. Assur. 8*, 2003, 68.

[3] A. Ríos and H. Téllez, *Trends in Anal. Chem. 24*, 2005, 509.

[4] B. L. Milman, *Trends in Anal. Chem. 24*, 2005, 493.

[5] B. M. Simonet, *Trends in Anal. Chem. 24*, 2005, 525.

[6] S. L. R. Ellison and T. Fearn, *Trends in Anal. Chem. 24*, 2005, 468.

[7] S. Cárdenas and M. Valcárcel, *Trends in Anal. Chem. 24*, 2005, 477.

[8] B. Lendl and B. Karlberg, *Trends in Anal. Chem. 24*, 2005, 488.

[9] M. Ferré, R. Brix and D. Barceló, *Trends in Anal. Chem. 24*, 2005, 532.

[10] U. Forsum, H. O. Hallander, A. Kallner and D. Karlsson, *Trends in Anal. Chem. 24*, 2005, 546.

[11] M. R. Plata, N. Pérez-Cejuela, J. Rodríguez and A. Ríos. *Anal. Chim. Acta 537*, 2005, 223.

[12] A. Sanz-Medel, B. San Vicente de la Riva, J. M. Costa-Fernández and R. Pereiro, *Anal. Chim. Acta 451*, 2002, 203.

4. VALIDATION OF QUALITATIVE ANALYTICAL METHODS. PARTICULAR APPLICATIONS

4.1 INTRODUCTION

Validation should be the last step in the development of a method before it is applied to actual samples in qualitative analytical methods. However, the only support the end user has in qualitative analysis are the documents summarised in the previous chapter.

The aim of the present chapter is to compensate for this lack of validation procedures by describing the tools required to validate some qualitative analytical methods.

The validation procedures presented have been designed in accordance with the intrinsic characteristics of the qualitative analytical method, and in particular the detection system. Thus, the cases studied were two commercial test kits, one of which provides a sensorial response and the other an instrumental response, and a home-made autoanalyzer with an instrumental outcome, although the final result is also binary. These validation procedures have been published and submitted as articles. They are presented below.

The first contribution describes the validation scheme designed for a commercially available test kit used in the field of food analysis. The test kit detects the presence of aflatoxin B_1 above a certain concentration level in nuts. The detection is visual, so the appearance of a coloured spot on the analysis card means that the analyte does not exceed a particular concentration level and the sample can be said not to contain the analyte. If the spot does not appear, however, then the analyte exceeds a certain concentration and the sample can be said to contain it.

Before the article there is an introduction to aflatoxins: the mycotoxin family, their toxicity, legislation, natural presence and the conditions in which they can be produced, etc. After the paper there are some extended practical aspects that are not included in the article.

The second article describes the procedure for validating a test kit that is also available commercially but which is used in the field of clinical chemistry. The test kit detects the presence of IgG antibodies to Varicella-Zoster Virus in human serum. The response is obtained using a UV-Vis spectrophotometer for 96-well microtiter plates and the instrumental value is transformed into an index value. The final result is a comparison between the index value of the sample and a reference index value.

Finally, the third paper focuses on the validation procedure of a home-made autoanalyzer. The device was designed to analyse samples from the degreasing baths used in the automotive industry. So the field of application is industry although it also has environmental effects.

There is a wealth of qualitative analytical methods, all of which have their intrinsic characteristics. The validation procedures should be designed with these features in mind. The three validation case studies reported here might serve as a guide to validating new methods as long as the differences with the methods to be validated are slight.

4.2 *AFLACARD B_1* : A VISUAL DETECTION TEST KIT

This study was planned and performed in collaboration with the Laboratory of Public Health in Tarragona. The aim was to validate a qualitative method of analysis for detecting the presence of aflatoxin B_1 in nuts. The analysis of aflatoxin B_1 in nut and spice matrixes belongs to the Surveillance Program of Foodstuffs in Catalonia which also includes the detection of Sudan I colouring in spices, the analysis of heavy metals in processed baby food, apple juice and fishing products, or the investigation of Lysteria Monocytogenes in processed salads and milk derivates, among other determinations. The aim of these analyses is to appraise the quality of particular foods because they may contain hazardous substances.

The routine method of analysis for the application chosen is based on chromatography and requires a tedious sample pre-treatment and pre-column derivatization of the analyte. On the other hand, the qualitative method of analysis requires a simple sample pre-treatment and the response is obtained rapidly. Therefore, because of the advantages of the operational performance, the qualitative method of analysis is an excellent candidate to be used as the routine method. Then it needs to be validated so that the basic performance parameters such as traceability and reliability can be verified over time.

4.2.1 Aflatoxins

Aflatoxins are secondary metabolites that belong to the group of mycotoxins, which are toxic metabolites produced by a fungus under special conditions of moisture and temperature. They are potential pathogens for animals and humans as they can cause kidney and liver diseases as well as immunodeficiency and damage to the nervous system.

They are generated by various species of fungi during the biosynthesis of fatty acids. During this process, the reduction of the keto functional groups may be interrupted. If this occurs, condensation reactions can take place and give rise to poliketonic compounds.

Not all fungi can produce mycotoxins. They usually need special conditions such as specific levels of moisture, pH and the correct temperature to produce mycotoxins. However, they might not be produced continuously. The absence of mycotoxins does not necessarily mean the absence of fungal spores, so fungi may be produced when the temperature and humidity are right. In addition to this, mycotoxins are very resistant to temperature treatments and to conventional food processes such as cooking, freezing etc.

Although almost 200 different mycotoxins have been characterised, only a few are often found in food and feed, although they are rather hazardous. These are aflatoxins, trichothecenes, ochratoxins, zearalenone, citrinin and fumonisins, among others.

Mycotoxins can naturally contaminate a wide variety of foodstuff. Table 1 summarizes the most common products that can be contaminated with mycotoxins.

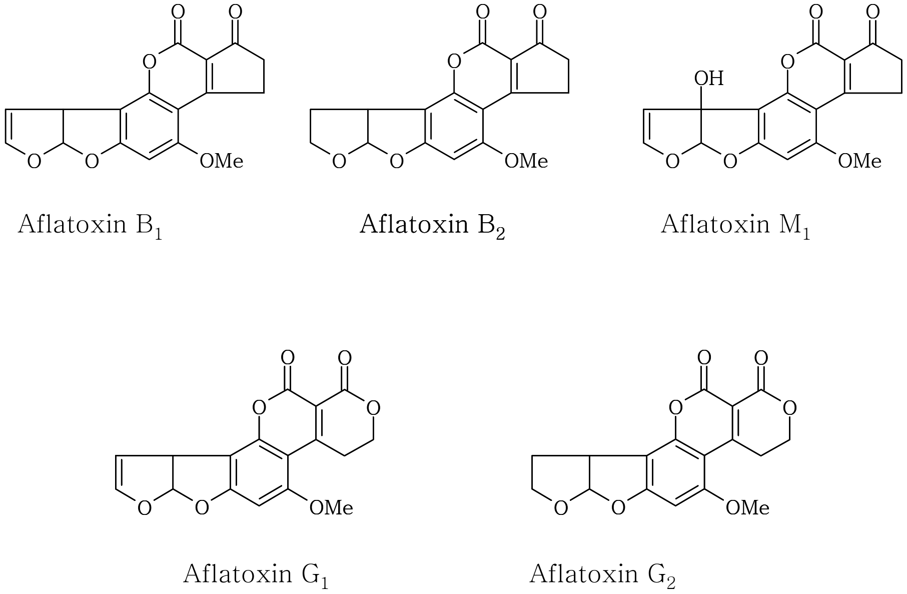
Table 1. Occurrence of natural contamination of some mycotoxins

MYCOTOXIN	MATRIX
AFLATOXINS	Nuts, cereals Dried fruit, milk and derivates Coffee, cacao Spices, feed
OCHRATOXIN A	Green and plain coffee Cereals, spices Wine, feed
ZEARALENONE	Cereals, feed

The fungi Aspergillus Flavus and Aspergillus Parasitivus produce aflatoxins, which are difuran-cumarin derivatives. Aflatoxin B_1, B_2, G_1, G_2, M_1 and M_2 (Figure 1) are the most common. Nevertheless, up to 20 different classes of aflatoxins have been found. Although aflatoxin B_1 is clearly the most toxic, aflatoxins B_2, G_1 and G_2 have considerable carcinogenic, teratogenic and mutagenic activity which mainly affects kidney, liver and brain [1], in the following order: G_1, B_2 and G_2. The term B and G refer to their fluorescent colour (blue and green), when they are exposed to UV-light.

Aflatoxins M_1 and M_2 are hydroxyl derivatives of aflatoxin B_1 and B_2 which are usually found in milk and its derivatives. Although they are not as toxic as the other aflatoxins, their presence in dairy food products is somehow troublesome.

For all these reasons, toxicological studies of aflatoxins tend to deal only with the ones mentioned above. In the 90’ s, immunochemical methods for analysing mycotoxins were introduced [2]. These immunochemical methods have rapidly evolved and are nowadays the basis of many other methods of analysis for mycotoxin determination [3], where method validation is also an important feature [4, 5].



Aflatoxin B_1

Aflatoxin B_2

Aflatoxin M_1

Aflatoxin G_1

Aflatoxin G_2

Figure 1. Chemical structure of aflatoxins B_1, B_2, M_1, G_1, and G_2. As can be seen, they are structurally related.

The production of aflatoxins is affected by physical, chemical and biological factors. The main physical factors are humidity (> 16%), temperature (25-30°C) and healthiness of the grains (broken seeds encourage fungi to develop). The chemical factors are pH (2.5-7.5), substrate composition (greasy seeds undergo more intensive fungi attack) and mineral nutrients in the seeds (iron, zinc and copper). And finally, one of the principal biological factors is that insects can spread the spores, which leads to the rapid development and multiplication of the fungi.

These optimal conditions are met mainly in the tropical and sub-tropical areas because of the considerable amount of humidity stored in the seeds before they are gathered. Therefore, aflatoxins have been proved to cause extensive health damage and important economic losses because of the international trade in products such as nuts, coffee and spices.

Several countries have legislation on the maximum permitted levels of aflatoxins in various foodstuffs. For example, the European Community establishes that the maximum concentrations of aflatoxin B_1, B_2, G_1 and G_2 in several food matrices [6, 7] should range between 2 ng/g for aflatoxin B_1 and 4 ng/g for the total content of aflatoxins. Likewise, the U. S. Food and Drug Administration sets the so-called action level at 20 ng/g for aflatoxin B_1 in several food matrices [8].

The distribution of aflatoxins in the sample is rather heterogeneous. Decontamination and food processing do not eliminate them efficiently. Therefore, sampling techniques must be used to provide quite homogeneous sub-samples [9, 10], and strictly accurate analytical procedures in order to provide high quality and healthy foodstuffs. If the sub-samples analysed are not homogeneous enough, the analytical results may not be representative of the contamination in the food matrix.

Nowadays, modern analytical techniques, which are based on monoclonal antibodies and high performance liquid chromatography, can reliably determine and quantify aflatoxins at rather low detection limits and with excellent specificity. The main drawbacks are the expense and the time of the analysis because it requires rather expensive material, such as immunoaffinity columns for the clean-up, extremely pure solvents and derivatization agents. The analysis also requires a rather tedious procedure, involving an extraction, sample clean-up and pre-column derivatization.

Thin-layer chromatography performs the analyses quicker and simpler but does not provide detection limits as low as the previous one. Moreover, the large volume of solvents used increases the expense.

The number of samples to be analysed and the drawbacks of the quantitative methods have meant that immunochemical techniques are increasingly being used either as qualitative methods of analysis or combined with more sophisticated analytical techniques. This is the case of Aflacard B_1 which is used to discriminate samples of nuts contaminated with 2 ng/g of aflatoxin B_1 from non-contaminated nuts. The contaminated samples are subsequently submitted to confirmatory methods, which are quantitative methods of analysis, and will provide the concentration of aflatoxin B_1. The most commonly used confirmatory method is High Performance Liquid Chromatography with Fluorescence Detection. It is based on the AOAC International Methods of Analysis (see references in the following paper). The sample requires a clean-up step with an immunoaffinity column and aflatoxin B_1 must be derivatized. The derivatization reaction can occur either before or after the separation, depending on the method of analysis used. Non-contaminated samples do not require any special processing.

The validation procedure of Aflacard B_1 is presented in the following paper, but the other relevant, practical aspects not included are listed afterwards.

4.2.2 QUALITATIVE METHOD FOR DETERMINATION OF AFLATOXIN B_1 IN NUTS

J. AOAC Int. 87, ***2004****, 417*

E. Trullols, I. Ruisánchez, F. X. Rius, M. Òdena[a] *and M. T. Feliu*[a]*.*

Universitat Rovira i Virgili, Departament de Química Analítica i Química Orgànica. Plaça Imperial Tàrraco 1. 43005 Tarragona (Spain)
[a]*Public Health Laboratory. C/ M. Cristina nº54, 43002 Tarragona (Spain)*

Abstract

The proper characterization of a commercial qualitative method for determining aflatoxin B_1 in some nuts is described. A qualitative method that provides binary responses of the yes/no type means that the performance parameters have been properly adapted and defined. Performance characteristics such as the cut-off limit, the detection limit, sensitivity, specificity, the false-positive and negative rates, and the unreliability or uncertainty region are defined and then estimated by means of the performance characteristics curves. The commercial test kit showed the cut-off limit at 1.6 ng/g, with a sensitivity rate of 95% and a false-negative rate of zero. A modification can be performed to shift the cut-off to 2.0 ng/g, keeping the same values for the sensitivity and false-negative rate.

In recent years, analytical developments have tended towards fast screening methods, efficient cleanup procedures, and precise but easily applied techniques. Screening test kits, commercial packages containing all the reagents and sometimes the instrumentation for the analysis, are now widely available (1). The main reason for this is that, rather than aiming to quantify a particular concentration, we are often more interested in knowing whether the concentration of a specific analyte is above or below a regulatory value or a threshold value. This value is mostly referred to as a specification limit, although other names, e. g., threshold value, are quite common. The current legislation, a client with specific needs, or even an internal quality control standard may fix that value. Therefore, qualitative methods have been developed to provide binary responses of the ‘ yes/no’ or ‘ positive/negative’ type that are used for making immediate decisions, for instance, of whether the sample complies with a specific regulation.

In order to provide confidence to end users, the test requirements and performance characteristics of any analytical method must be defined and properly validated. Although much work has been done on the definition of the requirements and the validation of quantitative analytical methods (e.g., by the European Committee for Standardization (2) AOAC INTERNATIONAL), less work has been done on qualitative methods.

The present study discusses the characterisation of a commercial test kit, Aflacard B_1 (3). The determination of aflatoxins in some nuts (pistachios, peanuts) is used as a case study. This characterization has meant the definition and, subsequently, the establishment of performance parameters such as sensitivity, specificity, false-positive and negative rates, unreliability or uncertainty region. Although there are several ways of characterizing a qualitative method, we propose to use performance characteristics curves (4).

Aflatoxins are organic compounds that belong to the mycotoxins family and are produced by some fungi. At certain concentrations, they are proven to be toxic compounds. Though there are a wide variety of them, just a few are present in food products like cereals, nuts, or milk. We will focus on aflatoxin B_1 because it is found in daily food and is potentially carcinogenic. According to European Union (EU) legislation (5), the maximum level of aflatoxin B_1 permitted in nuts is 2.0 ng/g. Therefore, samples of nuts whose concentration of aflatoxin B_1 is above this EU regulation limit are considered to be contaminated.

The most common quantitative methods used for determination of aflatoxin B_1 in nut samples are based on liquid chromatography (LC) and thin layer chromatography. However, some new methods are based on an immunoaffinity reaction such as the enzyme-linked immunosorbent assay (ELISA). AOAC (6) proposes a method based on a derivatization of the aflatoxin B_1 and LC with fluorescence detection. We used this confirmatory technique in addition to the qualitative method, the Aflacard B_1.

Experimental

Samples

The raw material used consisted of fried ready salted peanuts sampled according to a European Directive (7). Once the material was homogenized and the absence of aflatoxin B_1 was confirmed by LC, subsample portions of it were spiked with aflatoxin B_1 at different concentration levels. These samples were analyzed with both the test kit and the confirmatory LC method.

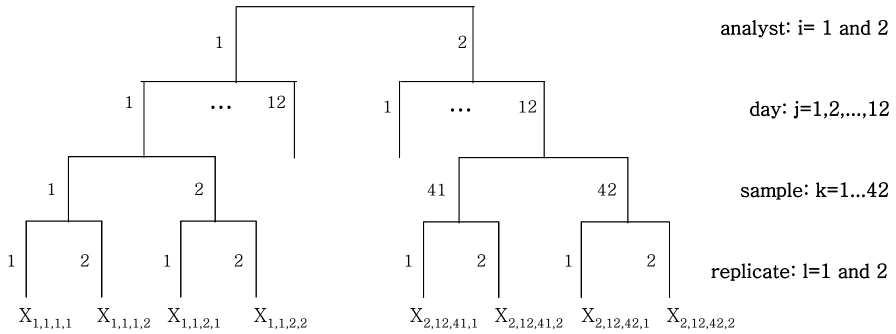


Figure 1. Experimental design followed. Each experiment is represented by Xi,j,k,l where i corresponds to the analyst (1 and 2); j, the day (12 days); k, the sample (42 samples); and l, the 2 replicates

Basis of the Test Kit

The assay is based on a competitive ELISA format, i. e., on the immobilization of monoclonal antibodies attached to a card' s membrane. This monoclonal antibody retains the aflatoxin B_1 present in the sample. The antibody sites that are free because of the absence of enough analyte are then covered by the addition of an aflatoxin B_1- enzyme conjugate. As the amount of aflatoxin B_1 in the sample increases, the number of free antibody sites decreases. The membrane is then washed to remove any unbound conjugate. When substrate is added, the spot on the port' s membrane where the conjugate has bound will turn purple. Any colour development on the sample port indicates a negative result, which means that the sample contains <2.0 ng/g of aflatoxin B_1. On the other hand, if the sample port shows no colour change, the assay is positive, as all the antibody sites are occupied by the analyte.

Table 1. Concentration levels of aflatoxin B_1 tested, with each sample’ s replicate and probabilities of positive, P(X), negative, N(X), and inconclusive, I(X), results calculated for each concentration level

Conc. (ng/g)	Analyst 1	Analyst 2	P(X)	N(X)	I(X)	P(X)+ I(X)
0.6	–[a] –	– –	0	100	0	0
0.6	– –	– –				
0.8	– –	– –				
0.8	– –	– I[b]	0	90	10	10
0.8	– –	NS[c]				
1.0	– –	– –				
1.0	– +[d]	– +	14,3	50	35,7	50
1.0	– I	I I				
1.0	I I	NS				
1.2	– –	I I	0	66,7	33,3	33,3
1.2	NS	– –				
1.4	+ +	I I				
1.4	+ I	+ +	50	0	50	100
1.4	I I	+ I				
1.6	+ +	+ +	100	0	0	100
1.8	+ +	+ +				
1.8	+ +	+ +	100	0	0	100
1.8	+ +	+ +				
2.0	+ +	+ +				
2.0	+ +	+ +	100	0	0	100
2.0	NS	+ +				
2.2	+ +	NS	100	0	0	100
2.4	+ +	+ +	100	0	0	100
2.4	NS	+ +				

[a] Negative sample

[b] Inconclusive sample

[c] No sample analysed

[d] Positive sample

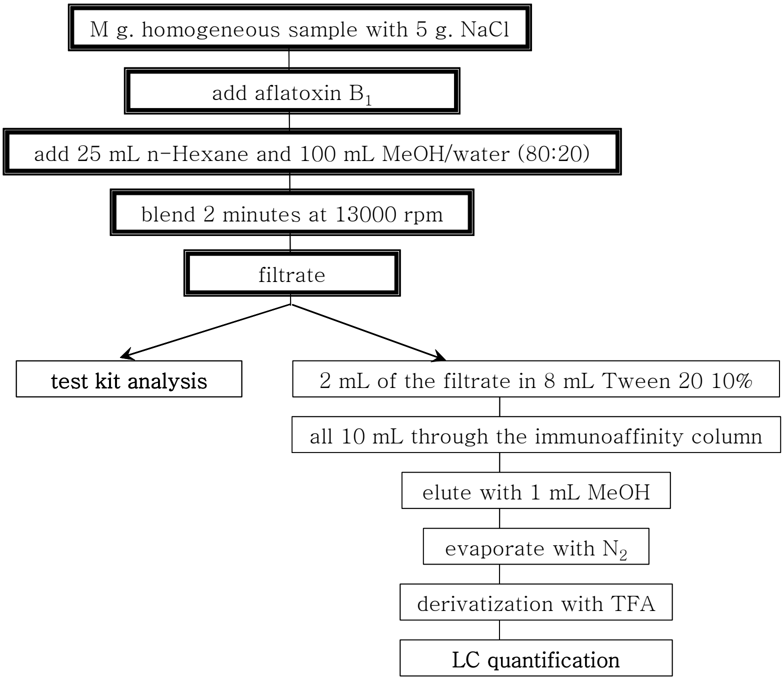

M g. homogeneous sample with 5 g. NaCl
add aflatoxin B_1
add 25 mL n-Hexane and 100 mL MeOH/water (80:20)
blend 2 minutes at 13000 rpm
filtrate
test kit analysis
2 mL of the filtrate in 8 mL Tween 20 10%
all 10 mL through the immunoaffinity column
elute with 1 mL MeOH
evaporate with N_2
derivatization with TFA
LC quantification

Figure 2. Scheme of the intended parallel analysis for screening and confirmatory methods

In the present study, the sample was considered not to be contaminated if the sample port had a clearly visible colour development. On the other hand, the sample was considered to be contaminated when the sample' s port failed to develop a readily detectable color. Some samples were inconclusive when it was not possible to decide if the sample' s port color was readily detectable or not.

Experimental Design

The experiments were undertaken following a predefined nested design (Figure 1) in which the main sources of variation considered were analyst, day, sample and replicate. Each of the two analysts prepared and analyzed 2 independent samples per day over a period of several days, in repeatability conditions. Some days, only one sample was analyzed, and the final number of analyses was 84. Blank samples of peanut butter were spiked at concentration levels ranging from 0.6 to 2.6 ng/g (Table 1).

Sample Preparation and Analysis by the Screening Test Kit

The analysis was performed according to the manufacturer' s instructions: 100 ml methanol-water (80+20, v/v) solution was added to a 50 g portion of ready-salted peanuts containing 5 g of NaCl and the right amount of aflatoxin B_1 standard. This mixture was blended in an UltraTurrax (IKA-WERKE GmbH & Co. KG., Staufen, Germany) for 2 min. The blended extract was filtered through a 2V Whatman (Clifton, NJ) filter paper. Briefly, 1 ml filtrate was added to the tube provided with the kit containing 3 ml of sample solvent. A 4 ml portion of mixture was filtered through the cleanup column, also provided with the kit. And 500 µ l sample was applied to the card' s port and allowed to pass through the membrane. Then, 100 µ l conjugate and 100 µ l wash solution, in this order, were applied to the port. The colour developed after the substrate had been added and after 5 min of waiting time. At the end of the 5 min, 100 µ l stop solution was be added.

As one of our main interests was to make parallel analysis possible (screening method and confirmatory method), we modified the LC Official Method [6] to ensure the concentration of the spiked samples. We mixed 2 ml of MeOH-water (80+20, v/v) with 8 ml 10% Tween 20 (figure 2). This quantity (10 ml) was then used in the immunoaffinity columns.

Performance Characteristic Curves

To build the performance characteristic curves (4), we spiked several blank samples with aflatoxin B_1 at different concentration levels. The highest level should provide only positive results (i.e., the samples are contaminated) while the lowest level should provide only negative responses. At each concentration level, the number of positive results is computed and the positive' s percentage, P(x), is calculated. In the same way, the negative' s percentages, N(x), and the inconclusive ones, I(x), are calculated. With these data, a graph of positive' s percentage versus the concentration levels tested is plotted. As inconclusive results are also obtained, the P(x) + I(x) curve is depicted as well as the P(x) curve. The shape of these performance curves is sigmoidal, the position and the slope being characteristic of each particular qualitative method. Once the curves are drawn, it is reasonably straightforward to calculate the performance parameters of the qualitative method by setting the probabilities of type I (α) and type II (β) errors.

Definition of performance parameters

Performance parameters are perfectly defined for quantitative analytical methods (8), but not for qualitative analytical methods. Table 2 shows some of the most common parameters for both quantitative and qualitative analysis. In this section, we report the definitions of the most usual parameters, some of which have already been defined by the AOAC (9).

Table 2. Some of the most common quality parameters for quantitative and qualitative analysis

Quantitative	Qualitative
Accuracy: trueness, precision	False positive and negative rates
Uncertainty	Sensitivity and Specificity
Sensitivity and Specificity	Selectivity: Interferences
Selectivity: Interferences	Limit of Detection
Range and Linearity	Cut-off limit
Limit of Detection	Unreliability region
Ruggedness or Robustness	Ruggedness or Robustness

As a result of a qualitative method analysis, a positive or negative result is obtained. The positive result that the test kit gives for a sample is considered to be true-positive if a reference method has given the same sample a concentration level higher than a predefined value. Likewise, a true negative is the negative result provided by the test kit for a sample that, according to a reference method, has a concentration level lower than a predefined value. Two related parameters should be defined: a false positive is a positive test kit response to a sample that is a true negative. An associated parameter is the false-positive rate: the ratio between the number of false positive results and the total number of true negatives. The same occurs when the test kit provides negative responses when the samples are true positive. In this latter case, these responses are called false negatives, and the false negative rate is the ratio between the number of false negatives and the total number of true positives.

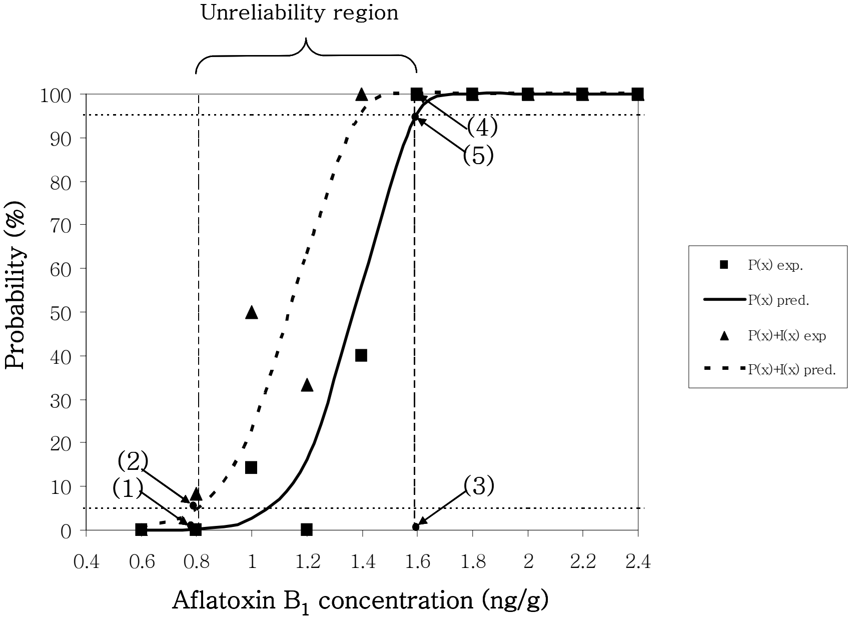


Figure 3. Experimental performance characteristic curves. Probability of positive responses, P(X), and probability of positive plus inconclusive responses, P(X)+ I(X), were plotted versus concentration levels tested. (1) FP=P(X); (2) $X_{0.05}$ where specificity=N(X)=100-(P(X)+ I(X)); (3) $X_{0,95}$, cut-off limit, detection limit; (4) FN=100-(P(X)+ I(X)); (5) sensitivity =P(X)=100-β

When handling samples providing true positive or negative responses, 2 other parameters must be taken into account: sensitivity and specificity. Sensitivity, in the context of qualitative analysis, is the ability of an assay to detect true positive samples as positive. An associated parameter is the sensitivity rate (10), which is defined as the probability, for a given concentration, that the method will classify the test samples as positive given that the test sample is true positive. Specificity (10), again in the context of qualitative analysis, is also defined as the ability of an assay to detect true negative samples as negatives. Again, an associated parameter is the specificity rate, which is the probability, for a given concentration, that the method will classify a test sample as negative given that this test sample is true negative.

A very informative performance parameter is the uncertainty associated to any measurement due to the presence of random errors. In the case of a binary response, the statement of uncertainty of a result cannot be expressed as a standard deviation-related statistical parameter but should be expressed as a probability of obtaining false responses. Therefore, rather than talking about uncertainty, we will talk about unreliability. As false responses can be either positive or negative, this unreliability (11, 12) becomes a region whose limits are the concentration levels where these probabilities of having false responses are set by the analyst. As the aim of a qualitative method is mainly to detect positive samples (that subsequently are usually submitted to quantitative confirmatory analysis), it is important that the unreliability region be placed below the established specification limit or threshold value. The upper limit appears when a given probability of having a positive result is set when it is indeed true positive. Usually, this probability is fixed at 95% and corresponds to a false-negative rate of 5% (β =5%, the probability of calling a positive assay negative). This upper limit corresponds to an analyte concentration symbolized by $X_{0,95}$ and is directly related to the sensitivity rate.

On the other hand, the lower limit is usually set at a given probability of having a negative result when it is indeed true negative. Usually, this probability is fixed at 95% and corresponds to a 5% probability of committing a false positive. Like the previous limit, it corresponds to the concentration $X_{0.05}$ and is directly related to the specificity rate because it indicates the concentration where 5% of the responses are false positive. Another concentration value arises from the middle point of the unreliability region, the $X_{0,5}$, where the probability of a negative result is equal to 50%. Song et al. (13) denotes this point as the ‘ identification limit’ .

Two additional parameters should be defined: the cut-off value and the detection limit. These 2 parameters have different connotations; depending on the type of qualitative analysis, the kit can provide either a numerical continuous value or a binary response. The treatment of the cut-off values for instrumental responses consisting of continuous values has been explained elsewhere (14).

When the qualitative method provides binary responses, such as the one we are using, the cut-off value is defined as the concentration value at which the test kit screens. Any test kit that performs well should ensure that any sample that contains a concentration of the analyte above the cut-off value will provide positive results with a certain and low probability of error. In this particular case, it is much more important to have no false-negative responses, because a contaminated sample must not be falsely considered non-contaminated. In well validated qualitative methods, the cut-off should coincide with the upper limit of the unreliability interval (13) as it corresponds to 95% probability of detecting true positives as positives.

The detection limit was defined by the International Union of Pure and Applied Chemistry in 1995 for quantitative analysis (15), but this definition is not equally applicable to qualitative methods that provide binary responses because this response type prevents the statistical parameters from being calculated by the well established procedure (14). Although, the concept of detection limit as the minimum concentration or amount that can be detected by a chemical procedure considering the 2 probabilities of committing α type and β type errors should be maintained, it is particular from binary responses the possibility to maintain just one probability depending on what one is interested in controlling, either the α or β probability of committing error.

The detection limit has also been defined as “ the lowest concentration of the analyte which the test can reliably detect as positive in the given matrix” [10]. Like the quantitative approach, reliably usually means 5% of false-negative (probability of β -error) responses. Therefore, this parameter is the upper limit of the unreliability interval. It must be stressed, however, that in this case, the false-positive error (probability of α -error) cannot be considered because, at the concentration level of the upper limit, all samples should provide a positive response. If this approach is adopted, it can be concluded that, for a binary response, the cut-off value and the limit of detection coincide with the upper limit which, as stated above, should be placed at the specification limit. Other authors consider the detection limit as the identification limit (13) in the sense that, in quantitative measurement, the detection limit corresponds “ to the values of analyte concentration at which one has a 50% chance of a negative result” . Therefore, the detection limit would now be placed at the centre of the unreliability region, at $X_{0,5}$.

Results and Discussion

The performance characteristic curve is obtained by fitting the experimental results from Table 1 to a sigmoidal function that minimizes the root mean square of the residuals (Figure 3). To obtain the performance parameters discussed in the previous section, the analyst has to fix the probabilities of committing false-positive (α) and false-negative (β) errors that can be accepted. Usually, these values are fixed at 5%, and correspond to the horizontal lines plotted on Figure 3, α =5% and 100-β =95%. Once the probabilities of error are fixed, all performance parameters are calculated. The first is the unreliability region, defined by its upper and lower limits. The upper limit corresponds to the concentration at which the 100-β line crosses the P(X) curve [in our case it is at 1.6 ng/g (dotted vertical line)]. Therefore, 1.6 ng/g is also the cut-off and the detection limit of the test kit under study. Similarly, the lower limit corresponds to the concentration at which the α line crosses the P(X) + I(x) curve [in our case, it is placed at 0.8 ng/g (dotted vertical line)]. Therefore, the unreliability region is between 0.8 and 1.6 ng/g. The sensitivity rate at 1.6 ng/g is equal to 95% as it corresponds to the point where:

$$P(x) = 100 - \beta$$

Also, the false-negative rate at 1.6 ng/g must be known and, in this case, is equal to zero as:

$$FN=100-(P(x)+I(x)) = 0$$

Similarly, the specificity rate at 0.8 ng/g is equal to 95%, as it corresponds to the point where:

$$N(x) =100-(P(x)+I(x))$$

The false-positive rate at this concentration level is also zero because

$$FP=P(x) = 0$$

These values show that the test kit tends to give a positive result, not a negative one, to any sample containing between 1.6 ng/g and 1,9 ng/g of aflatoxin B_1; this corresponds to a region where only false-positive results occur. In order to minimize this tendency, we propose a modification based on the cut-off' s shift from 1.6 to 2.0 ng/g, and maintaining no false negatives at 2.0 ng/g. This involves varying the amount of sample. Working with 40 g instead of 50 g of sample shifts the performance curves to the right by 0.4 ng/g. The experimental design was the same as above (Figure 1), but considering the results obtained in the previous experimentation, fewer samples were analyzed (Table 3). The parallel analysis using the test kit and the LC method was also done. The performance characteristic curves plotted from these experimental results are shown in Figure 4. This curve shows that the unreliability region appears between 1.2 and 2.0 ng/g. As inconclusive results are considered, the sensitivity rate at 2.0 ng/g is equal to 95%, and there is no false-negative response. Similarly, the specificity rate at 1.2 ng/g is equal to 95% and, at this level, there is no false-positive response.

Table 3. Concentration levels of aflatoxin B_1 tested, with each sample' s replicate and probabilities of positive, P(X), negative, N(X), and inconclusive, I(X), results calculated for each concentration level when using 40 g of the sample

Conc. (ng/g)	Analyst 1	Analyst 2	P(X)	N(X)	I(X)	P(X)+I(X)
0.6	–[a] –	– –	0	100	0	0
0.6	– –	– –				
0.6	– –	n. s.[b]				
0.8	– –	– –	0	100	0	0
0.8	– –	– –				
0.8	n. s.	– –				
1.0	– –	– –	0	100	0	0
1.0	– –	n. s.				
1.2	– –	– –	0	90	10	10
1.2	– –	– –				
1.2	n. s.	– I[c]				
1.4	– I	– I	0	50	50	50
1.8	+ +[d]	+ I	56,3	0	43,7	100
1.8	I I	I +				
1.8	+ +	+ I				
1.8	I I	+ +				
2.0	+ +	+ +	100	0	0	100
2.0	n. s.	+ +				
2.2	+ +	+ +	100	0	0	100
2.2	+ +	+ +				

[a] Negative sample

[b] No sample analysed

[c] Inconclusive sample

[d] Positive sample

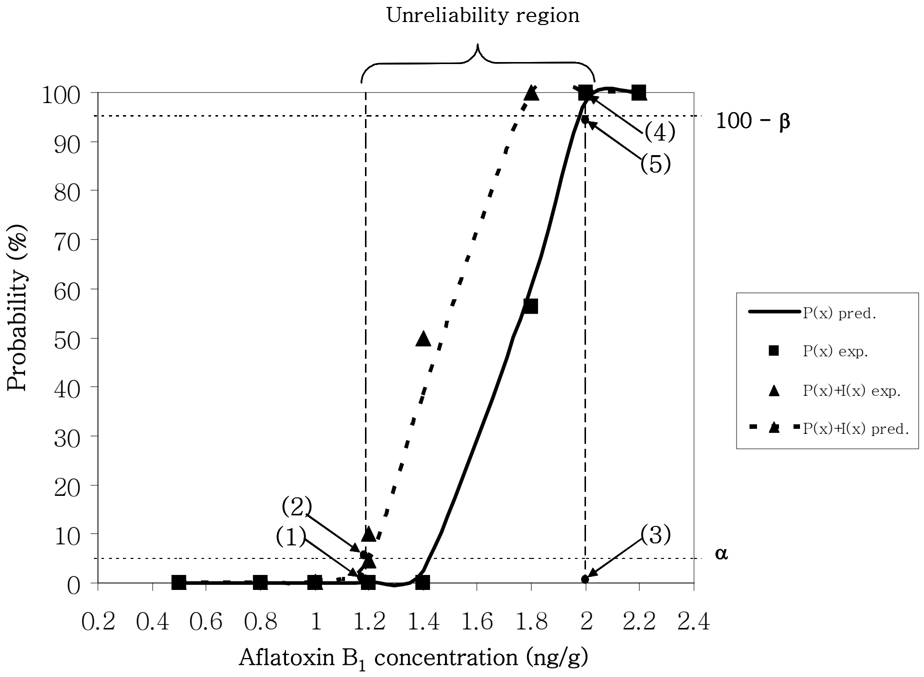


Figure 4. New performance characteristic curve and recalculation of the performance parameters using 40g of raw sample in the experimental procedure. (1) FP=P(x), (2) X0.05 where Specificity=N(x)=100-(P(x)+ I(x)), (3) X0,95, Cut-off limit, detection limit, (4) FN=100-(P(x)+ I(x)), (5) Sensitivity =P(x)=100-β

It is important to state that this modification must take the requirements of the end user into account. This means that in some cases, obtaining a positive sample at 1.6 ng/g is not a hitch: the laboratory can assume the cost (mainly in time and economical terms) of quantifying by the confirmatory LC method all samples containing 1.6 ng/g of aflatoxin B_1 or more.

Conclusions

A commercial test kit that gives binary responses of the positive/negative type was validated and its performance parameters' sensitivity rate, specificity rate, and limit of detection were defined and established by means of performance characteristic curves. When qualitative methods are used, performance parameters are expressed in terms of probability, whereas in the quantitative approach some of them are expressed as quantities having the same dimension as the obtained result. The validation process has shown a tendency towards false-positive samples, as the cut-off value appeared at 1.6 instead of at 2.0 ng/g, which was the value claimed by the kit's manufacturer. These cannot be considered an inconvenience because, in case of hazardous substances, the false-negative rate must be controlled and, in many cases, minimized. In that sense, the users of test kits should take into account that manufacturers tend to provide procedures that reassure the absence of false negatives. This can be acceptable when the cost of this assurance is known; a number of unnecessary confirmatory analyses should be made because of the high number of false-positives reported. Nevertheless, some modification in the Aflacard B_1 analysis procedure can be performed, but always assessing that the false negative rate is equal to zero.

Acknowledgments

We acknowledge economic support from the MCyT (Project No. BQU2000-1256). T 0429 Peanut Butter was supplied by Central

Science Laboratory Proficiency Testing Group, Sand Hutton (York) YO41 1LZ, United Kingdom, http://ptg.csl.gov.uk/fapas.cfm.

References

(1) Stroka, J.,& Anklam, E. (2002) *Trends in Anal. Chem.* **21**, 90-95

(2) EURACHEM (1998) *The Fitness for Purpose of Analytical Methods*, http://www.eurachem.ul.pt

(3) R-Biopharm Rhone Ltd., http://www.r-biopharmrhone.com

(4) Song, R., Schlecht, P.C., & Ashley, K. (2001) *J. Haz. Mat.* **83**, 29-39

(5) Commission Regulation (EC) No 257/2002, *Off. J. Eur. Commun.* 12.2.2002, No. L 041, pp 12-15, http://europa.eu.int/eur-lex/

(6) *Official Methods of Analysis* (2000) 17th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Methods **990.33** and **991.31**

(7) Commission Directive 98/53/CE of 16 July 1998, *Sampling and Analysis Methods for Official Control of the Maximum Content of Some Contaminants in Food*

(8) International Organization for Standardization (1994) ISO 8402, *Quality Management and Quality Assurance-Vocabulary,* ISO, Geneva, Switzerland

(9) Feldsine, P., Abeyta, C., & Andrews, W. (2000) AOAC INTERNATIONAL *Methods Committee Guidelines for validation of Qualitative and Quantitative Microbiological Official Methods of Analysis*, AOAC INTERNATIONAL, Gaithersburg, MD

(10) Boison, J.O. (2000) in *Current Issues in Regulatory Chemistry*, J.F. Kay, J.D. MacNeil, & J.J. O' Rangers (Eds), AOAC INTERNATIONAL, Gaithersburg, MD, pp 159-170

(11) Rios, A., Barcelo, D., Buydens, L., Cárdenas, S., Heydorn, K., Karlberg, B., Klemm, K., Lendl, B., Milman, B., Neidhart, B., Stephany, R. W., Townshend, A., Zschunke, A., & Valcarcel, M. (2003) *Accred. Qual. Assur.* **8** 68-77.

(12) Trullols, E., Ruisánchez, I., & Rius, F.X (2004) *Trends in Anal. Chem.* **23**, 137-145

(13) Song, R., Fischbach, T.J., & Ashley, K., (1996) *Am. Ind. Hyg. Assoc. J.* **57**, 161-165

(14) Pulido, A., Ruisánchez, I., Boqué, R., & Rius, F.X. (2002) *Anal. Chim. Acta* **455,** 267-275

(15) Currie, L., (1995) *Pure Appl. Chem.* **67**, 1699-1723

4.2.3. Practical aspects not discussed in the previous article

The validation process of the qualitative method of analysis has been described. However, one practical aspect requires further consideration: the results obtained with the test kit must be traceable to a validated confirmatory method.

For this particular case, and according to the statement in section 4.2.1, the confirmatory method is based on High Performance Liquid Chromatography with fluorescence detection and pre-column derivatization. There are two official methods of analysis included in the AOAC INTERNATIONAL 17th Official Methods of Analysis (2000) [11]. They both propose a derivatization reaction but at different stages of the procedure: the 990.33 method describes a pre-column derivatization reaction, while the 991.31 method includes a clean-up procedure using an immunoaffinity column and post-column derivatization step. Liquid Chromatography and fluorescence detection is used in both cases.

Considering the current instrumentation and the characteristics that the confirmatory method of analysis should have, the Public Health Laboratory has modified both official methods of analysis. As a result, its confirmation method involves a clean-up procedure that uses an immunoaffinity column but a pre-column derivatization step. Since the new procedure was developed and validated by this laboratory, the only thing we had to do was perform suitability checks before applying it to test samples.

These checks involve such steps as peak identification, and studies of reproducibility, the calibration line (which includes the detection limit) and traceability.

Spiked samples were analysed to verify the results provided by Aflacard B_1 (see the corresponding article).

The confirmatory method of analysis

The quantitative method based on HPLC has been used for years in the Public Health Laboratory to quantify the aflatoxin B_1 present in the supposedly contaminated samples. Depending on the matrix (spices or different kinds of dried fruits), the quantity of sample used may range between 10 and 15 g. (for spices) and between 40 and 60 g. (for dried fruits). In the present application, the matrices are dried fruits: ready-salted fried peanut butter and ready salted roasted pistachio butter. These two matrices are used because of the availability of blank sample that can be spiked with aflatoxin B_1 standard.

In general terms, this method of analysis consists of several steps: first, the extraction of the analyte with n-hexane and methanol/water; then, the sample clean-up using an immunoaffinity column in order to specifically retain the analyte and the following elution with methanol. Finally, the derivatization with trifluoroacetic acid (TFA) provides the appropriate fluorescent signal.

The instrumental components used are a Liquid Chromatograph HP1100 Series including a quaternary pump, a degasser, a Rheodyne

7725i six-port injection valve and a sample loop of 100 μl. The column is a Hypersil-ODS (4.6x200 mm), 5 micron and the mobile phase used is Milli-Q water/acetonitrile/methanol (55:30:15). Both solvents are of HPLC Grade and provided by SDS (Peypin, France). The flow rate is 0.7 ml/min, which provides a working pressure of 62-65 bar. The detector is an FLD G1321A from Agilent Technologies and it is set at an excitation wavelength of 365 nm and an emission wavelength of 430 nm. However, excitation spectra for a specific emission wavelength or emission spectra for a specific excitation wavelength are also possible.

Depending on the type of sample to be analysed (standard or spiked), the preparation process can vary. Unlike spiked samples, standard aflatoxin B_1 samples do not require the analyte to be extracted or the sample to be cleaned up.

Peaks are identified and the reproducibility studied with 10 ml standards of aflatoxin B_1 which are prepared per duplicate at a concentration level of 2.0 ng/g. The calibration line is also designed with duplicated standards in the concentration range between 0.4 and 5.0 ng/g.

The preparation is a rather tedious process because it involves such complicated steps as evaporating the stock solution solvent and derivatizing with trifluoroacetic acid. In addition to this, the dark glass material used must be kept in sulphuric acid 2 N for 24 hours and later rinsed out with Millipore water and acetone.

The stock standard solutions of aflatoxin B_1 are kindly supplied by the Public Health Laboratory. They consist of small, frozen, sealed vials with approximately 2 ml of a solution containing aflatoxin B_1 in chloroform (0.4 µl/ml). A register of the weight of each vial is also supplied with the vials. Every time a standard solution or the standards of the calibration line are prepared, a new vial is used. Before it is unsealed, its weight at room temperature is recorded. Whenever the mass measured varies by more than 0.0009 grams, the vial is discarded.

Once the appropriate volume of stock solution is placed in the 10 ml dark vials, and the chloroform is evaporated under a nitrogen stream, 200 µl of n-hexane is rapidly added to prevent the analyte from oxidizing. After mixing for 30 seconds, 50 µl of trifluoroacetic acid (min. 99%, spectrophotometric grade) is added to carry out the derivatization according to Figure 2. Finally, 10 ml of an acetonitrile/water (10:90) solution is added after 5 minutes of waiting time. These standards must be kept in the freezer and protected from UV-light for one week.

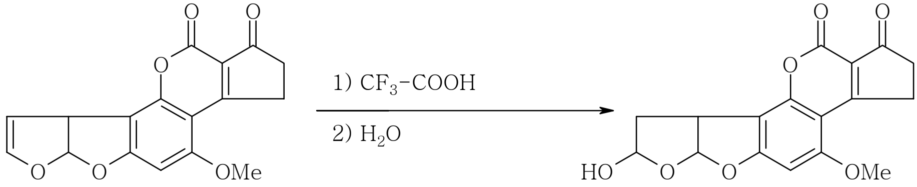


Figure 2. Derivatization reaction of the aflatoxin B_1

Once the standards are ready, the analysis should be performed as soon as possible to prevent the analyte from changing.

The information above shows that only two peaks should appear in the chromatograms. They correspond to the trifluoroacetic acid excess at minute 3, which has a characteristic shape, and to the derivatized aflatoxin B_1 at minute 4.75, approximately. Nevertheless, the first analyses show a third peak at minute 8, the fluorescent intensity of which depends on which duplicate of the standard is being measured (Figure 3). There are various reasons why this third peak should appear (e. g. contamination of the vial containing the stock solution of aflatoxin B_1, an impurity from the trifluoroacetic acid or from the acetonitrile/water (10:90) solution). However, the most probable reason is an inaccuracy during the derivatization. Therefore, the emission spectra of the derivatized aflatoxin B_1 and of the peak at eight minutes are recorded at 365 nm of excitation wavelength (Figure 3). Figure 4 shows the emission spectra of the non-derivatized aflatoxin B_1 (40 ng/g) recorded in the same conditions as the previous two. As is assumed, the peak at minute 8 approximately corresponds to the non-derivatized aflatoxin B_1. Even though there is an excess of trifluoroacetic acid, it must be at room temperature if the reaction is to be completed.

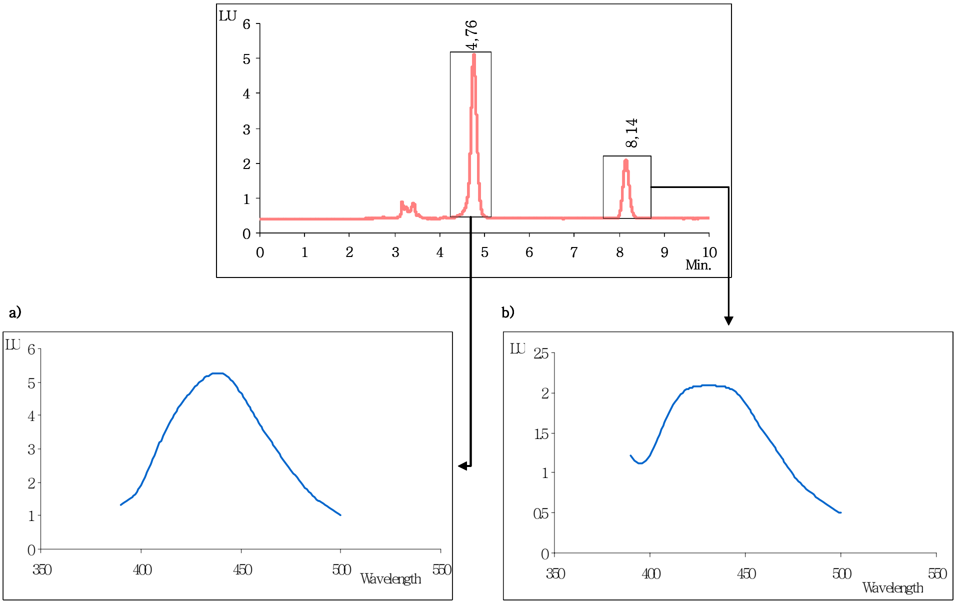


Figure 3. Chromatogram and emission spectra at 365 nm excitation wavelength corresponding to a) the derivatized aflatoxin B_1 and b) to the unknown peak

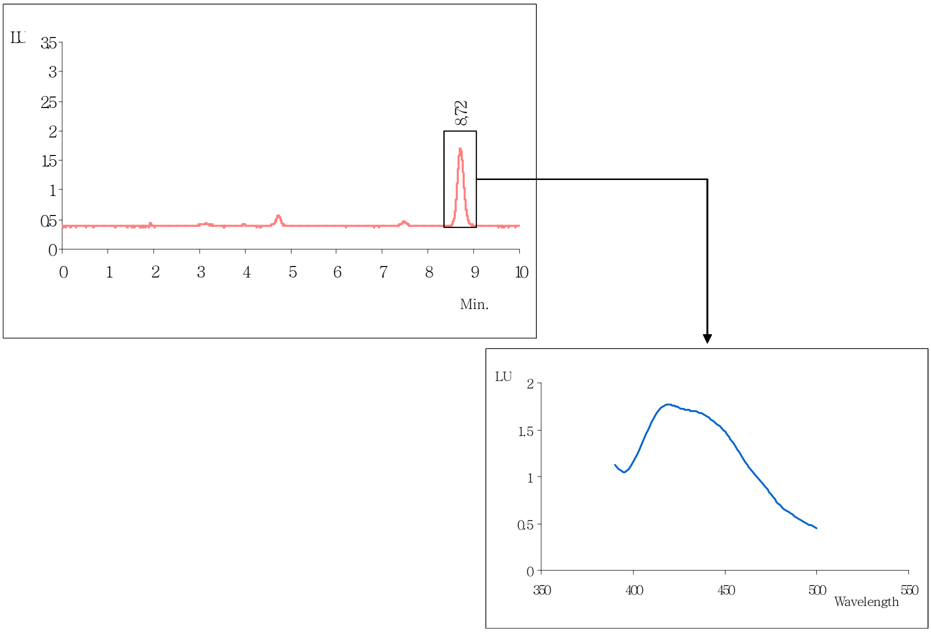


Figure 4. Chromatogram (red solid line) and spectrum (blue solid line) corresponding to a 40 ng/g standard of non-derivatized aflatoxin B_1

The residues from the standards must be placed in a special container because the remaining aflatoxins are submitted to the appropriate waste management. The re-usable glass material must be decontaminated by immersion in a 10%-sodium hypochlorite solution for twenty four hours.

The 2 ng/g aflatoxin B_1 replicated standards are also used in a reproducibility study. The aim is to check the comparability of the results when the analyst or the day of the analysis is changed. To assess any variation in the results, we studied the peak areas obtained when two different analysts measure the standards for four days. On each day, each analyst performed five measurements.

The theoretical approach used is the well-known two-way Analysis of Variance or ANOVA [12]. ANOVA is a statistical tool used to compare the variances of the different sources of error considered as significant in the total variance. The variance of each source of error considered must be compared with the residual variance by means of an F-test. In the present case, the two main sources of error are the two analysts and the day. The expressions used to calculate the F-values are summarized in Table 2.

Table 2. Expressions used to calculate the F-values in the 2-way ANOVA

Source of variation	Sum of squares (SS)	Degrees of freedom	Mean squares (MS)	F-value
Analyst	$SS_{ANALYST}$	p^a-1	$SS_{ANALYST}/(p-1)$	$MS_{ANALYST}/MS_R$
Day	SS_{DAY}	q^b-1	$SS_{DAY}/(q-1)$	MS_{DAY}/MS_R
Interaction	SS_{A-D}	$(p-1)(q-1)$	$SS_{D-A}/[(p-1)(q-1)]$	MS_{D-A}/MS_R
Residual	SS_R	r^c	SS_R/r	
Total	SS_T	$n^d-1=t$		

[a] p is the number of analysts (2); [b] q is the number of days (3); [c] r is calculated as t−[(p−1)+(q−1)−(p−1)(q−1)]; [d] n is the total number of analyses performed (40)

A previous graphical examination of the experimental data may help to detect possible outlier measurements. The peak areas for the analyses performed during the four-day period are plotted in Figure 5.

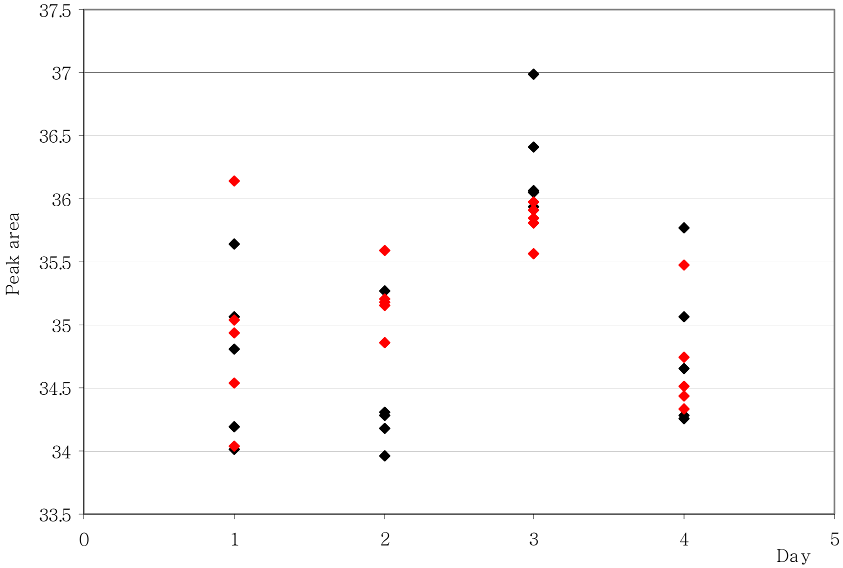


Figure 5. Peak areas obtained during the four-day period. The red dots correspond to analyst 1 and the black ones correspond to analyst 2.

The response values are clearly higher on day three. The Q Dixon statistical test uses the mean values of the days to calculate a Q-value of 0.93. For n equal to 4 at a significance level of 5%, the tabulated Q-value is 0.83. In addition to this, it is found in the laboratory register notebook that the conditions of analysis were rather different from the standard conditions. Therefore, the responses obtained on day three are removed from the data set.

Provided that a) outlier data have been examined; b) the variances studied (i. e. variance caused by the day and variance caused by the analyst) a priori do not differ significantly; and c) the data follow a normal probability distribution function, the ANOVA table is constructed. The F-values shown in Table 3 are compared to the corresponding tabulated F-values at a 5 % significance level and corresponding degrees of freedom. The tabulated F-values for the analyst, for the day and for the interaction are 4.3, 3.4 and 3.4. So it can be concluded that the sources of variation and how they interact make the same contribution to the final variance as the experimental error.

Table 3. ANOVA table computed for the 2 analysts measuring for 3 days

Source of variation	Sum of squares (SS)	Degrees of freedom	Mean squares (MS)	F-value
Analyst	0.66	1	0.66	2.0
Day	0.043	2	0.021	0.065
Interaction	1.1	2	0.53	1.6
Residual	7.9	24	0.33	
Total	9.7	29		

The calibration line is studied by preparing duplicate calibration standards in a range of concentrations between 0.40 and 5.0 ng/g. The calibration standards are prepared following the procedure described above but taking into account the final concentration of the analyte. The same vial of 0.4 μl/ml is used to prepare the necessary standards for a single calibration line. The measurements are also performed under the conditions described above. The resulting calibration line has a slope of 17.8, an intercept of -0.8793 and a determination coefficient of 0.9959. The plot of the residuals in Figure 6 shows that no point is near the so-called Warning Limits (i. e. twice the standard deviation of the points), so a priori, the data set contains no outliers.

In addition to this the residual distribution throughout the concentration range is not heteroscedastic.

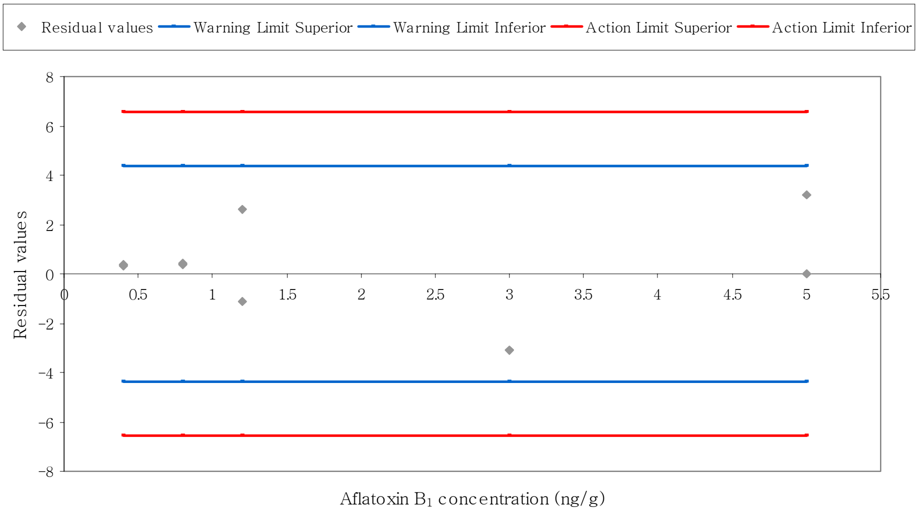


Figure 6. Residual values for the calibration standards measured

The statistical tool ANOVA is used again but to validate the linear model chosen to fit the data. Regression ANOVA (Table 4) compares the variance of the lack of fit with the variance of the experimental error. The calculated F-value is compared with the corresponding tabulated F-value. The linear model is appropriate whenever both variances do not differ significantly.

Table 4. Expressions used to calculate the corresponding variances

Source of variation	Sum of squares (SS)	Degrees of freedom	Mean squares (MS)	F-value
Regression	SS_{REG}	1	MS_{REG}	
Residual	SS_R	n-2	MS_R	
Lack of fit	SS_{LOF}	k-2	$SS_{LOF}/k-2$	MS_{LOF}/MS_{PE}
Experimental error	SS_{PE}	n-k	$SS_{PE}/n-k$	
Total	SS_T	n-1		

In the present case, the data are obtained in repeatability conditions. Provided that a) the series probability distribution function is approximately normal; and b) the variances of the series are homogeneous, the ANOVA Table 5 shows that the linear model is suitable for the data obtained in repeatability conditions, because the calculated F-value is 3.53 and the tabulated F-value is 7.76 at a 5% significance level.

Table 5. ANOVA table computed for the calibration standards measured

Source of variation	Sum of squares (SS)	Degrees of freedom	Mean squares (MS)	F-value
Regression	9256.40	1	9256.40	
Residual	38.09	8	4.76	
Lack of fit	25.89	3	8.63	
Experimental error	12.20	5	2.44	3.53
Total	9294.49	9		

The detection limit has been calculated from the calibration line data. Taking into account the standard error of the residuals (2.18), the number of concentration levels used (5) and the number of measurements on the future sample (usually 1), and for a significance level of 5%, the detection limit is 0.52 ng/g. This value is acceptable as long as the samples with less than 2.0 ng/g of aflatoxin B_1 are considered not to be contaminated.

The following stage is to analyse spiked samples using a real blank matrix. The use of spiked samples involves calculating a value of recovery which will we considered in subsequent analyses. A Certified Reference Material is also analysed in order to assess the traceability of the method when analysing real samples. Next, we describe the preparation and validation of the calibration line used for quantification and the preparation of the spiked samples for estimating the value of the recovery.

The calibration line used for quantification is in the concentration range from 0.80 to 2.4 ng/g. The calibration standards are prepared in duplicate. The plots in Figure 7 and Figure 8 show that there are no outlier and that the residual values are rather homoscedastic. The ANOVA table (Table 6) confirms that the linear method is suitable because the tabulated F-value (4.15) is higher than the calculated F-value (0.09). The detection limit is computed in the same way as for the first calibration line. Its value is 0.29 ng/g.

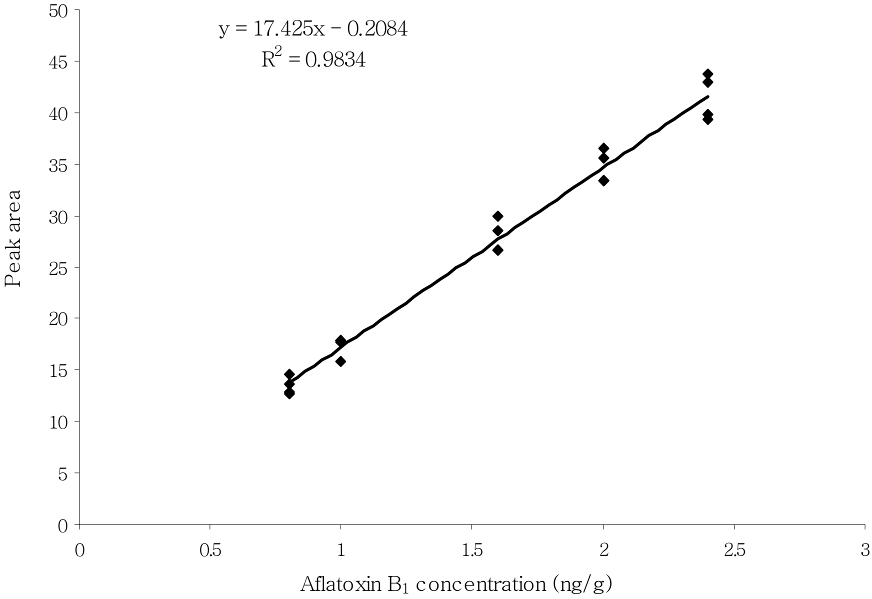


Figure 7. Calibration line used to quantify the spiked samples

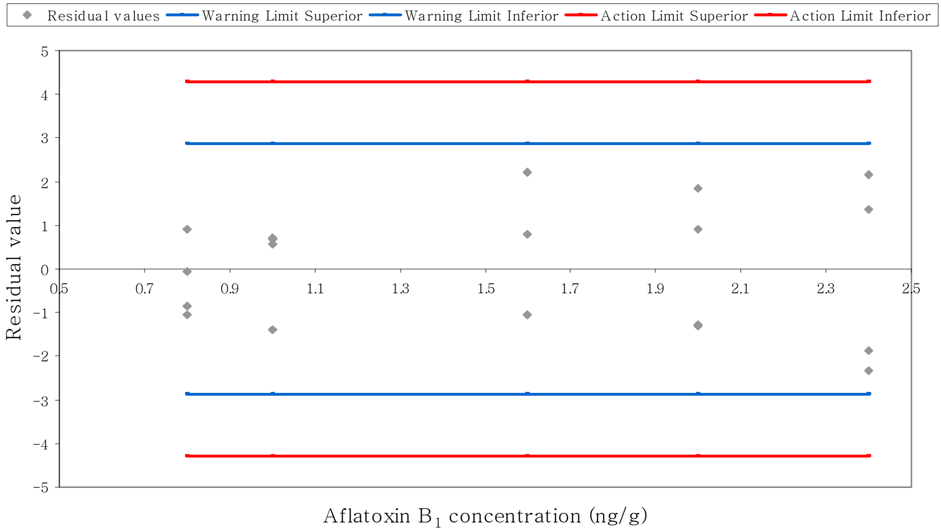


Figure 8. Residual values of the quantification calibration line

Table 6. ANOVA table obtained when measuring the duplicate calibration standards twice

Source of variation	Sum of squares (SS)	Degrees of freedom	Mean squares (MS)	F-value
Regression	2176.48	1	2176.48	
Residual	36.72	18	2.04	
Lack of fit	0.69	3	0.23	
Experimental error	36.03	15	2.40	0.09
Total	2213.21	19		

The preparation of the spiked samples is more complicated than the preparation of the standards because it requires a clean-up step. However, some of the stages are the same for both standards and spiked samples (e. g. the enrichment and derivatization steps).

The weight of the frozen vial containing 0.40 µl/ml of aflatoxin B_1 standard should also be taken into account when spiking a sample. Then, the corresponding volume of 0.40 µl/ml aflatoxin B_1 standard is added to 40 g of the same sample matrix used to validate the test kit (ready salted peanut butter), described in the previous paper, and which also contains 5 g of sodium chloride. A total of 40 g is used because of the intended parallel analysis with the test kit. The results obtained with 50 g were presented in the previous paper.

After a few minutes to favour the contact between the analyte and the matrix, 25 ml n-hexane and 100 ml of a methanol/water (80:20) solution are added. The homogenization is performed according to the process described in the previous article: 2 minutes at 13000 rpm. If the filtrate is not transparent when a Whatman 2V double filter is used, it is exchanged for a glass microfibre filter.

The clean-up stage is critical. Immunoaffinity columns contain a sorbent with the monoclonal antibody for the specific retention of 200 ng of aflatoxins B_1, G_1, B_2 and G_2 that are adsorbed on it [13]. A 10% Tween 20 solution is used to dilute 2 ml of the methanol/water filtrate (80:20) up to 10 ml. This solution is then passed through the immunoaffinity column drop-wise. Air must not be allowed to form so that the uniformity of the analyte retention can be assessed. Once the 10 ml have nearly passed through the column, 20 ml of Millipore water are used to wash the 10-ml Erlenmeyer. This volume is also passed through the immunoaffinity column. Finally, it is dried with air.

The elution of the analyte is also decisive. After the immunoaffinity column has been dried with air, 1 ml methanol is left in contact for 1 minute. Then the analyte is eluted by back flushing the methanol rapidly and repeatedly. The methanol is evaporated under a nitrogen stream. The derivatization process is now the same as for the standards. Instead of adding 10 ml of the acetonitrile/water solution (10:90), the volume added is 950 µl. After a few minutes of waiting time, Millex HV must be used to carry out a filtration before the injection. If the analysis is not performed immediately, the samples must be frozen. The chromatograms obtained (see Figure 9) are very similar to the chromatograms of the standards, with two peaks corresponding to the excess of trifluoroacetic acid and to the derivatized aflatoxin B_1. The emission spectrum of the second peak confirms its identity.

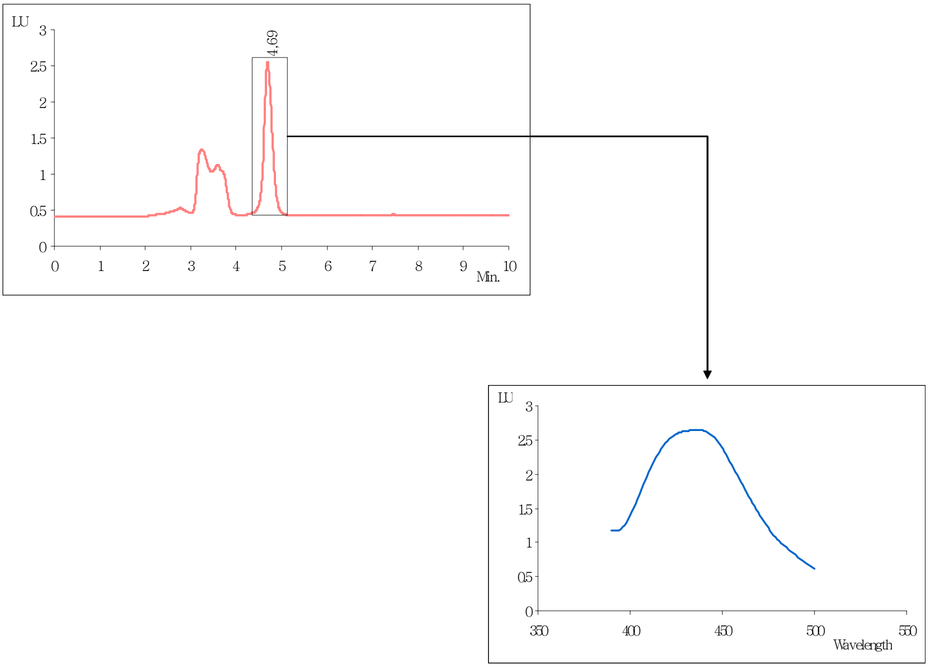


Figure 9. Chromatogram and emission spectrum at an excitation wavelength of 365 nm of a spiked sample with 2.0 ng/g of aflatoxin B_1

The procedure described above consists of several stages, which involve diluting the sample (i. e. the concentration of the analyte at the end of the sample preparation process is lower than the concentration at the beginning). Therefore, this dilution must be taken into account when the recovery of the method. Equation /1/ shows the concentration of the analyte at the end of the preparation process when it is spiked with 2.0 ng/g of aflatoxin B_1.

$$\frac{2\ ng\ analyte}{1\ g\ sample}\times\frac{40\ g\ sample}{100\ ml\ MeOH/water\ (80:20)}\times\frac{2\ ml\ \ filtrate\ MeOH/water}{10\ ml\ Tween20+filtrate\ MeOH/water}\times \qquad /1/$$

$$\times\frac{10\ ml\ Tween+filtrate\ MeOH/water}{1\ ml\ TFA+ACN/water\ (10:90)}=\frac{1.6\ ng}{ml\ sample}$$

Although the range of concentration levels is wider, the spiked samples quantified contain 1.2, 2.0 and 2.2 ng/g of aflatoxin B_1. Table 7 shows the analyte concentration added, the concentration at the end of the sample preparation, the mean value of the calculated concentration after three analyses, the corresponding standard deviation and the recovery.

Table 7. Summary with the concentrations found and the recoveries

Sample ID	Sample weight (G)	Added concentration (ng/g)	Diluted concentration (ng/g)	Mean value of the detected concentration (ng/g)	Standard deviation	Recovery (%)
M23G2(2)	39.98	2.00	1.60	1.65	0.01	103.20
M24G2(1)	39.9	2.00	1.60	1.44	0.00	90.35
M24G2(2)	39.98	2.00	1.60	1.51	0.01	94.42
M28G22(1)	39.68	2.20	1.75	1.4	0.04	80.70
M28G22(2)	40	2.20	1.75	1.58	0.03	89.79
M5F12(1)	40.09	1.20	0.96	0.79	0.03	82.31
M5F12(2)	39.9	1.20	0.96	0.97	0.01	101.07
M6F12(1)	40.09	1.20	0.96	0.78	0.01	81.43
M6F12(2)	39.9	1.20	0.96	0.96	0.01	100.61
M7F12(1)	39.96	1.20	0.96	0.83	0.02	87.47
M7F12(2)	39.98	1.20	0.96	0.92	0.01	96.05

The mean value of these recovery values is 88.7%. The individual values and the mean value lie within the interval of values accepted by the Official Method of the AOAC recovery values (60-110%). An aliquot of these spiked samples was also analysed using the Aflacard B_1test kit, and the results are reported in the previous article.

To conclude this section on the confirmatory method of analysis, the experiments carried out with a Certified Reference Material are presented. The aim of these analyses is to assess the traceability of both the confirmatory method and the Aflacard B_1 test kit. Therefore, the experiments use the quantitative and the qualitative method of analysis. The material is kindly supplied by the Public Health Laboratory although it is purchased at the Central Science Laboratory Proficiency Testing Group [14]. It consists of 50 g of ready salted peanut butter naturally contaminated with aflatoxin B_1. The sample has been measured by seventy laboratories and the resulting concentrations range between 1.67 and 4.30 ng/g. The concentration value of 2.98 ng/g is accepted as the reference value.

Therefore, our aim is to analyse an aliquot of the material with a concentration of aflatoxin B_1 which lies in the unreliability region defined in the previous article (between 1.2 and 2.0 ng/g). As the concentration value of 2.98 ng/g is too high, some modifications must be made if the same sample is to be used for both methods of analysis. If the weight of the sample of reference material is 24.16 g, the theoretical aflatoxin B_1 concentration to be determined is 1.8 ng/g. Taking into account the dilution process during the preparation of the sample, the final aflatoxin B_1 concentration is 1.44 ng/g. Table 8 shows the results of the analysis using both methods:

Table 8. Summary of the results obtained when analysing an aliquot of 24.16 g of the Certified Reference Material

	Aflacard B_1 test kit	Confirmatory method (ng/g)	Reference material (ng/g)
Sample 1	inconclusive	1.16	2.40
Sample 2	inconclusive	1.18	2.43
Sample 3	negative	1.16	2.37
Mean value and uncertainty (ng/g)		1.17±0.133	2.42±0.148

After three instrumental replicates (Sample 1, Sample 2, Sample 3) the analysis with the confirmatory method provides a mean concentration of aflatoxin B_1 of 1.17 ng/g and an uncertainty value of 0.133 ng/g. In order to confirm that the method of analysis performs appropriately, the relative standard deviation is checked throughout the concentration range tested to see that it does not vary significantly (i. e. between 1 and 2%). Then, an extrapolation is carried out to consider a sample size of 50 g. The concentration is 2.42 ng/g and the uncertainty associated to this concentration value 0.148 (see Table 8).

Since we have not found a statistically significant difference between the results of both methods of analysis, the test kit is validated.

Analysis using a different matrix: ready salted roasted pistachio butter

The Aflacard B_1 test kit mainly deals with two different matrix types: nuts and spices. The paper presented deals with ready salted peanut butter. However, just one matrix might not be sufficient to consider that a method of analysis is fully validated for that type of sample. Therefore, the same experiments are carried out with another nut matrix. As has been stated in the section above, the work with nut matrices is directly related to the availability of blank samples. Thus, the Laboratory of Public Health also supplied the ready salted roasted pistachio butter. Other available matrices were not suitable since they were slightly contaminated with the four main aflatoxins or with aflatoxin B_1.

The results of previous analyses of the new matrix are similar to the analyses performed with the peanut butter. At a concentration level of 1.0 ng/g, one result out of six is negative, four are inconclusive and one is positive. At a concentration level of 1.40 ng/g, just one result out of twelve is inconclusive. The rest are positive. Then, the same modification proposed for the peanut butter matrix is made. Although, an in depth study of the raw data may conclude that it is also possible to work with 35 g of sample, we decided to work with 40 g to unify standard operation procedures.

The reproducibility and linearity studies carried out before on aflatoxin B_1 standards are not repeated. In order to periodically check the instrumental responses, 2.0 ng/g aflatoxin B_1 standards are prepared for each duplicate and measured together with the spiked samples.

The spiked samples are also analysed in parallel, using both methods of analysis. The results obtained, shown in Table 9, are represented in Figure 10 where they are fitted to the sigmoidal curve.

Table 9. Results of the 40-g samples of ready salted roasted pistachio butter when using both methods of analysis

Conc. (ng/g)	Analyst 1	Analyst 2	P(X)	N(X)	I(X)	P(X)+I(X)	HPLC mean recovery
0.9	– –	– –		83.3	16.6	16.6	79.6
0.9	– –	NS	0				
1.1	– I	– –	0	60	40	40	64.1
1.1	– –	I –					
1.1	NS	I I					
1.3	– I	I I	30	10	60	90	79.6
1.3	+ I	+ I					
1.3	I +	NS					
1.5	– I	+ I	30	20	50	80	80.3
1.5	– I	I I					
1.5	+ +	NS					
1.7	+ +	+ I	70	0	30	100	76.7
1.7	+ +	+ +					
1.7	I I	NS					
1.9	+ +	+ +	100	0	0	100	77.3
1.9	+ +	+ +					
1.9	+ +	+ +					
1.9	+ +	+ +					
2.1	+ +	+ +	100	0	0	100	79.4
2.1	+ +	+ +					
2.1	+ +	+ +					
2.1	+ +	+ +					

– means a negative result   I means an inconclusive result

NS means that no sample is analysed   + means a positive result

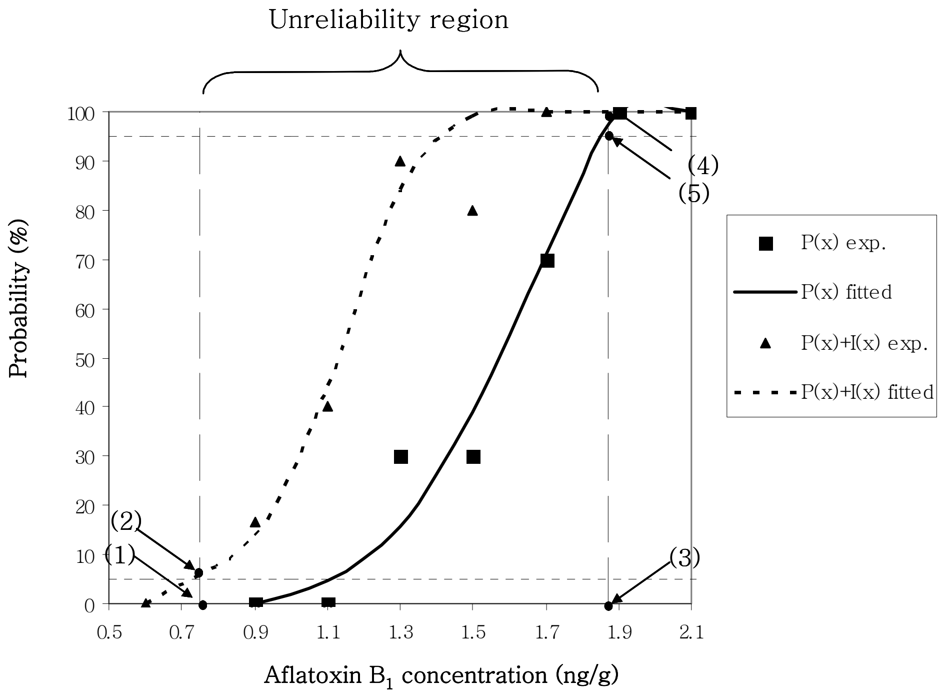


Figure 10. Performance characteristic curves for 40 g of pistachio butter samples

The final results show that the sensitivity rate at 1.9 ng/g is equal to 95% and at this concentration level there are no false negative responses. The specificity rate at 0.75 ng/g is equal to 95% and there are no false positive responses at this concentration level. The cut-off and the detection limit are placed at this concentration level as well. The unreliability region, then, lies in the concentration interval between 0.75 and 1.9 ng/g.

At the time of this experimentation, no Certified Reference Material with the suitable characteristics was available. Therefore, traceability could not be assessed at this level.

The final conclusions are very similar to those in the article. The validation process has confirmed the bias to false positive responses. In general, manufacturers are conservative as far as false negative results are concerned. In applications dealing with food contaminants, for example, a high false negative rate must be avoided because harmful consequences would affect public health.

The first commercial test kit chosen for validation has been characterised. Performance characteristic curves are a very useful and informative tool whenever a visual detection test kit needs to be validated. Although the main drawback might be the considerable number of experiments to be made, the information gathered permits a detailed characterisation. In any case, the number of experiments proposed in several validation documents referenced in the article is higher than the number proposed in the procedure described.

4.3 *VZV IgG:* A BINARY RESULT TEST KIT THAT PROVIDES AN INSTRUMENTAL RESPONSE

Among the great variety of commercial test kits, those that use an instrumental response to provide a final binary result are of special interest. These methods of analysis record a response value for a reference, and is a crucial step in the subsequent comparison with the response of the test sample. Depending on this comparison, the final binary result is YES or NO.

These methods have several advantages. If the decision about the sample is taken just by comparing the responses of the test sample and the reference, there will be no need for a quantification with a fitting model. So a considerable amount of experimental work and its associated costs are avoided. In addition, they are also quick and easy to handle.

Our study is made in close collaboration with the Immunology Department of the Laboratorio de Análisis Dr. Echevarne in Barcelona. The aim of this cooperation is to validate a commercial test kit with the characteristics mentioned above. The test kit studied is the VZV IgG, which measures the presence of IgG antibodies to Varicella-Zoster Virus (VZV) in human serum. This assay is just one of a wide range of clinical analyses performed in the Immunology Department, such as the determination of IgG and IgM antibodies to Helicobacter pylori or to Dengue virus in human serum.

The commercial test kit is an indirect Enzyme Linked Immunosorbent Assay with Varicella-Zoster Virus antigen coating the 96-well microtiter plate. It is used as a routine method in the laboratory according to the habitual standard procedures in clinical analysis. The application does not use a confirmation technique because the standard operational procedure of the Varicella-Zoster Virus does not require one. Therefore, independently of the result obtained, serum samples are not re-analyzed in the laboratory with a different method of analysis. Only in a few extreme situations (i. e. inconclusive results that in a second analysis provide the same results or test samples that first provide a clear positive response and, in a replicate, a clear negative response), is further analysis required, usually with another commercial test kit. In the exceptional case that the new commercial test kit does not clarify the final result, the test for Varicella-Zoster Virus fluorescent antibody to membrane antigen (FAMA) [15] is then carried out. However, this is not a common situation.

4.3.1 Varicella-Zoster Virus

Varicella-Zoster Virus is a human alphaherpesvirus that belongs to the Herpesvirus family. Herpesviruses get their name from the Greek ‘ *herpein* ’ *,* which means ‘ to creep’ , and they can cause chronic, latent and/or recurrent infections. Approximately 100 Herpesviruses have been isolated, at least one for most of the animal species studied. However, to date, just eight human Herpesviruses are known [16]. Infection to humans is nearly universal. That is to say, most adults have the virus in a latent stage and various cell types can be infected (e. g. nerve cells, lymphocytes)

The International Committee on Taxonomy of Viruses (ICTV) classifies human Herpesviruses into three subfamilies (Table 10) according to their biological properties [17]. *Alfaherpesvirinae*, which can affect several host cells, have a relatively short reproductive cycle, and are rather efficient at both destroying host cells and establishing latency in the cells of the dorsal root and sensory ganglia. *Betaherpesvirinae,* which are made up of a narrower range of host cells, have a longer reproductive cycle and establish latency in lymphocytes and kidney tissues, among other possibilities. Finally, *gammaherpesvirinae* affect lymphoblasts and sometimes cause the lysis of fibroblast or epithelial cells.

Table 10. Summary of the eight human Herpesviruses

SUBFAMILY	GENRE	NAME
ALPHAHERPESVIRUS	human herpesvirus 1, 2 human herpesvirus 3	Simplexvirus 1, 2 (HSV-1, HSV-2) Varicella-Zoster virus (VZV)
BETAHERPESVIRUS	human herpesvirus 5 human herpesvirus 6, 7	Cytomegalovirus (CMV) Roseolovirus (HHV-6, HHV-7)
GAMMAHERPESVIRUS	human herpesvirus 4 human herpesvirus 8	Lymphocryptovirus (Epstein-Barr virus-EBV) Rhadinovirus (HHV-8)

Varicella and herpes zoster are caused by the VZV which is a virus transmitted by the respiratory route [18]. The result of primary VZV infection is varicella, also known as chicken pox, which shows symptoms such as fever and pruritic rashes. Although it is a typical childhood illness prevalent in temperate climates [19] it can be found worldwide. However, herpes zoster or shingles can only be caused by a VZV reactivation. The symptoms are a localized, and usually painful, vesicular rash which involves dermatomes. The occurrence of herpes zoster increases with age or immunosuppression, and is particularly prevalent in patients being treated with immunosuppressive drugs for malignant diseases or to prevent the rejection of bone marrow or organ transplants, and in individuals with antibodies against human immunodeficiency virus (HIV).

Primary VZV infection produces IgG, IgM and IgA antibodies that bind to many classes of viral proteins. Antibody production is detectable within three days of the first symptoms. The first antibodies to be produced are IgM, followed immediately by IgG. However, the number of IgM antibodies decreases after a month, while IgG antibodies to many viral proteins remain at acceptable levels for years as part of the long-term immune response to VZV.

The presence of VZV can be detected in two ways depending on the problem at hand. Virological methods based on immunofluorescence with monoclonal or polyclonal antibodies, which are rapid and sensitive [20], can be used to assess VZV infection. Also DNA isolation of the vesicles helps to detect virus infection. These methods are used in quite specific circumstances (i. e. to assess if there is viral infection in the fetus during pregnancy or to check the viral origin of some lesions because the symptoms of Herpes Zoster are not so evident). These analyses can lead to an antiviral therapy being prescribed, which is of outstanding importance for high-risk (that is to say, immunocompromised) patients.

However, if it is the immune status of the individual that needs to be determined, then serological methods are required. Commercial enzyme linked immunosorbent assays are useful in this case for screening purposes. The antibodies used in these methods of analysis are highly specific. The rate of false positive results is low but results are often inconclusive. In these cases, the sample is reanalysed and if the final result is not conclusive or does not agree with the clinical history, another commercial test kit is used. The confirmatory testing (FAMA or latex agglutination [21]) is seldom necessary.

Both the virological and serological methods of analysis are based on procedures that use monoclonal or, more commonly, polyclonal antibodies to the specific VZV antigens. The strength of this bond is a reversible interaction in which non-covalent inter-molecular forces (i. e. hydrogen bonding, electrostatic, van der Waals and hydrophobic) take part. The most important factor is the complementarity between the antigen and the antibody: if they are complementary enough, the reactive parts of the antigen (antigenic determinant) and of the antibody draw closer together because the water molecules in between are removed. The affinity of the antibody for the antigen is the sum of all non-covalent intermolecular binding forces of a single antigenic determinant to an antibody. So, antibody affinity is an expression of the attraction between the molecules of the antibody and the antigen.

Due to the multivalent nature of antigens, the immunological system can produce different types of antibodies (i.e. a single antigenic determinant generates multiple antibodies and a single natural antigen has multiple antigenic determinants). The binding strength of antibodies to multiple antigenic determinants on natural antigens is known as avidity [22]. It is a measure of the multivalent antigen-multivalent antibody stability complex and it depends on the different affinities. This factor is stronger than the sum of the affinities.

An infection with VZV is treated with antiviral agents such as acyclovir, famcyclovir and valacyclovir, which are all licensed. The usual concentrations required to inhibit the virus are about 1.0 to 2.0 mg/ml and the dose is usually 10 mg/kg. However, the virus is not eliminated from the host, so it may reactivate when the treatment is stopped. Prevention is carried out by means of passive immunization with VZV IgG antibodies: that is to say, a preparation of high-titer VZV IgG antibodies is given to susceptible high-risk individuals (i.e. immunocompromised children and pregnant women who have been in contact with VZV), by means of a live attenuated varicella vaccine, although some special cases are treated with antiviral agents (bone marrow transplant).

The prevalence of an illness is the ratio, for a given time period, of the number of occurrences of this illness to the number of units at risk in the population. Varicella and Herpes Zoster are well-known for their rather high prevalence and, considering the negative consequences of infection (either primary [23] or reactivation [24]), it is important for rapid and reliable methods of analysis to be available.

In order to contribute to this task, we describe below the validation of a commercial ELISA used in serology to detect VZV antibodies.

4.3.2 VALIDATION OF QUALITATIVE TEST KITS WITH INSTRUMENTAL RESPONSES. DETECTION OF VARICELLA - ZOSTER VIRUS IgG ANTIBODIES IN HUMAN SERUM

J. Pharm. Biomed. Anal. Submitted

E. Trullols, I. Ruisánchez, F.X. Rius and J. Huguet[a]*.*

Universitat Rovira i Virgili. Departament de Química Analítica i Química Orgànica. C/ Marcel·lí Domingo s/n. 43007 Tarragona (Spain)

[a]*Laboratorio de análisis Dr. Echevarne. C/ Provença 312 08037 Barcelona (Spain)*

Abstract

Qualitative analytical methods are nowadays being widely used as screening methods. The detection of analytes in samples above certain concentration levels is important before their quantification with the routine method. In order to achieve the best quality also in clinical analysis results, all methods must be validated. A relevant issue is to define and to assess the quality parameters, considering the actual sample matrix (urine, serum, etc). A strategy to validate an ELISA that assesses the presence or absence of the IgG antibodies against Varicella-Zoster Virus (VZV) analysing a serum sample is presented. The absorbance is transformed into an index that links the test sample values to those obtained daily by analysing the control samples provided with the test kit. The data obtained during two months from the control samples are examined in detail. The calculated indexes are distributed following normal distributions. Then, performance parameters (traceability, sensitivity and specificity, the unreliability region and the false response rates) are calculated easily.

Keywords: clinical test kits, binary results, control samples, quality assurance, performance parameters

Introduction

Qualitative methods of analysis are becoming increasingly valuable, among other purposes, as screening methods. They are used in different contexts, such as environmental, clinical or food-quality control, but all are related to analytical chemistry. Depending on the application, a positive result should be accompanied by a confirmatory analysis as is the case of environmental or food analysis [1, 2]. However, clinical chemistry is a rather particular case. Confirmatory methods are only used if concrete determinations (e.g. HIV or Hepatitis C) require the corroboration of a positive result. Other cases are those in which inconclusive results are obtained in duplicate, though such cases are even fewer. In all these situations, time and cost are considerably reduced. Like any other method of analysis, those used in the clinical context must be validated. Most determinations in clinical analysis are performed with commercial test kits, so confirmation of the quality parameters claimed by the manufacturer is very important for the quality of the results.

Validation must always be considered at the end of the method development process. Method validation was described some time ago by ISO [3] and, from the practical point of view, it can be considered as the definition and estimation of the performance parameters necessary to fit the analytical requirements. The validation procedure must be carefully defined in order to properly assess the performance characteristics of the analytical method. Validation of quantitative methods has been widely developed and it is a well established process, as is shown by the existence of several validation documents and guidelines addressed to practitioners [4, 5]. In the field of qualitative analytical methods, some regulatory bodies provide validation documents and guidelines, although they are still not generally accepted [6, 7]. Following this trend, several research teams are attempting to establish qualitative validation procedures as shown by recent papers [8-10].

The validation procedure depends not only on the external analytical requirements, but also on the intrinsic characteristics of the qualitative method. The detection system, either sensorial or instrumental, is one of these internal characteristics. And whether or not the method is based on the measurements of control samples should also be taken into account. In all cases, as we are dealing with qualitative methods, the final result is binary (yes/no or positive/negative). Therefore, the validation strategy should be adapted to and consistent with each particular methodology [11, 12].

This paper reports the validation procedure of a test kit based on instrumental detection (absorbance measurement) and control samples measurement. This test kit is used in laboratories that perform clinical analyses and does not require confirmation of positive results. The technique is an Enzyme Linked Immunosorbent Assay (ELISA) [13] based on the antigen-antibody reaction. The control samples must be measured to establish a cut-off value which is a key point in the classification of the test samples. Traceability, and estimate values for the sensitivity and specificity, the unreliability region and the probabilities of giving false results are specifically confirmed.

As we record absorbance values for both the test and the control samples, a statistical characterization of the distributions of the control samples, always in the response domain, is useful to validate the method. The main performance parameters of the test kit are estimated from the positive and negative control samples distribution and also from the information associated to the cut-off value obtained during its estimation. Sample compliance is assessed by means of prediction intervals defined around the cut-off value. Although these intervals have already been used to estimate the uncertainty of a qualitative method of analysis, involving control samples is a novelty. First, the characteristics of the test kit are described, and then the performance parameters and how to estimate them are defined. As a case study, the validation of a specific ELISA test kit that detects IgG antibodies against VZV in human serum is reported.

Materials and methods

Basis of the test kit

The test kit VZV IgG [14] is an indirect Enzyme-Linked Immunosorbent Assay for detecting the presence or absence of IgG antibodies against VZV in human serum. The microtiter wells are coated with VZV antigen from a cellular culture. If the sample analysed contains IgG antibodies to this virus, they link to the antigen coating the microtiter well. Adding the enzyme-labelled anti-antibody provides the coloured solution after a chromophore is added. The colour intensity, which is directly related to the amount of antibodies in the sample, is measured by means of a spectrophotometer at 450 nm. Since an instrumental response is recorded, it is compared to the cut-off value in terms of raw absorbance though, as is shown later, a comparison in terms of index is usual. So the final result is YES, if the test sample response is higher than the cut-off value, meaning that the serum sample contains IgG antibodies to the virus. And it is NO, if it is lower than the cut-off value, meaning that there are no IgG antibodies or that they cannot be detected.

The cut-off value is a key value during the assay as every test sample response has to be compared to it before the final result can be given. In this particular kit, it is obtained by combining the response values of the negative and the positive control samples provided with the kit. The mathematical expression that calculates the cut-off value (COV) is defined by the manufacturer as a combination of the duplicated absorbance values of both control samples (Equation 1):

$$COV = \overline{A}_{-} + 0.1 \times \overline{A}_{+} \qquad /1/$$

Where,

$\overline{A}_{-}$ is the mean absorbance value of the negative control

$\overline{A}_{+}$ is the mean absorbance value of the positive control

The manufacturer also supplies information about the range of variation of this cut-off value. This range refers to inconclusive sample results and its value is calculated according to the previous testing of internal and external quality controls. However, nothing is said about how it is estimated or the units in which it is expressed.

Our own results suggest that this range is derived from the coefficient of variation associated to the COV. In this particular test kit the range of variation of the cut-off value provided by the manufacturer is 15%, so the test samples responses must be compared in the following terms:

1) If the absorbance of the test sample measured at 450 nm is higher than the COV plus 15%, the sample is given as positive. That it is to say, the serum sample is considered to have IgG antibodies to VZV

2) If the absorbance is lower than the COV minus 15%, the sample is given as negative. In this case, the sample is considered not to have enough IgG antibodies to VZV

3) Finally, if the absorbance lies between the COV plus and minus 15%, the sample is given as inconclusive.

Though the manufacturer does not suggest that a mathematical transformation be used for the COV, practitioners usually like to standardise the COV value, and therefore assign it a value of 1. Subsequently, the raw test sample responses are also divided by the COV value giving rise to the index of the cut-off value. Equation /2/ shows how these indexes are obtained. The indexes are compared in the response domain as well. Therefore, though it exists, a relationship between the response and the activity level is not established.

$$Index = \frac{Sample\ absorbance}{COV} \qquad /2/$$

Following the same reasoning as in the previous paragraph, the decision about the test sample also takes into account the lack of precision of the cut-off value. So, 1) the result is positive if its index is higher than 1.15; 2) the result is negative if its index is lower than 0.85 and 3) the result is inconclusive if its index is between 1.15 and 0.85.

Samples

Two types of samples are distinguished: positive and negative control samples and reference material. All tests are performed using test kits from the same batch.

The negative control sample is a pool of different human sera that is proven to be free of VZV IgG antibodies. Similarly, the positive control sample is a pool of human sera that has an activity level of 20 milliInternational Units per millilitre (mIU/ml) of IgG antibodies against the VZV. Both control samples are provided by the manufacturer and are ready to use. The positive control sample is also diluted at two different dilution factors to obtain two samples that provide responses in the region close to the cut-off value. The dilution is made following the manufacturer’ s instructions using the dilution buffer provided in the test kit.

As traceability is a key performance parameter, a reference material is also used [15]. This reference material consists of an ampoule with lyophilised VZV IgG antibodies, which provide an activity level of 4 International Units per millilitre when they are diluted with 1 millilitre of distilled water as it is indicated by the provider.

Experimental procedure

The samples are measured in a 96-microtiter well plate so 96 analyses can be performed simultaneously. Though the negative and the positive control samples are ready to use, the samples resulting from the dilution of the positive control sample are not. They are obtained by using the correct dilution factor to reach responses close to the cut-off value (inconclusive and low positive responses). After testing several dilution factors it is proven that the relation between the index value and the dilution factor follows a quadratic function [16] and that the best dilution factors are 1/8 and 1/12 as they provide low positive and inconclusive responses, respectively (Figure 1).

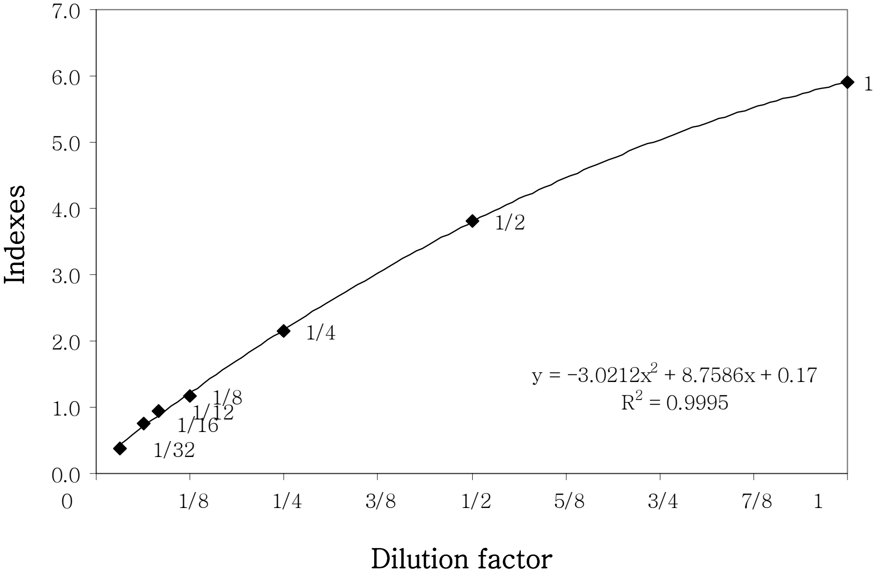


Fig. 1. Quadratic relation (solid line) of the indexes (◆) obtained when the positive control sample is diluted by several dilution factors.

To verify traceability, the reference material should have an activity level equivalent to that of the positive control sample, so in the present case it must be diluted. The activity of the reference material is 200 times that of the positive control sample since the kit insert states that the activity of the latter is 20 mUI per millilitre. So a two-step dilution is performed according to the standard procedures in the laboratory: first, a dilution factor of 1/20 and second, a dilution factor of 1/10. As in the case of the positive control sample, the dilution is made using the buffer provided with the test kit.

The analyses are performed according to the manufacturer' s instructions: 100 microliters of the sample are added to the microtiter wells coated with VZV antigens. After an incubation period of 30 minutes, the microtiter plate is washed four times with 400 microliters of the wash solution provided with the test kit. Then, 100 microliters of the conjugate (enzyme-labelled anti-antibody) are added to the microtiter well and after 30 minutes of incubation time and being washed 5 times with 500 microliters, 100 microliters of the substrate (3, 3’ , 5, 5’ -Tetramethylbenzidine) are added. The reaction finishes when, after 15 minutes of incubation time, 100 microliters of stop solution are added. The absorbance of the yellow coloration shown by the microtiter plate is measured in a spectrophotometer at 450 nm using a reference wavelength of 620 nm.

Validation methodology

The validation methodology does not depend only on the analytical requirements: it is also closely related to the intrinsic characteristics of the analytical method. Pulido et al. describe a validation strategy for qualitative analytical methods that provide an instrumental response [17]. But for the particular case in which control samples are used, no validation procedure has been published. This procedure is based on estimating the cut-off value and its statistical distribution, which, in turn, depends on the statistical distributions of both control samples. The methodology also characterises the statistical distribution of the reference material.

Characterisation of the statistical distributions

First of all, the type of probability distribution function that the test kit responses follow must be checked. If the distribution is to be estimated properly, the experiments must follow a predefined experimental design (Figure 2). The analyses are performed for thirty days, which should be long enough to be able to estimate the theoretical errors. Every day two replicates of each sample (positive and negative control sample, the reference material and the diluted positive control sample) are measured by the same analyst under the same conditions, and according to the instructions provided with the test kit. The analyst is not considered a source of variation because the routine analyses are always performed by the same operator. However, this should be taken into account in future studies.

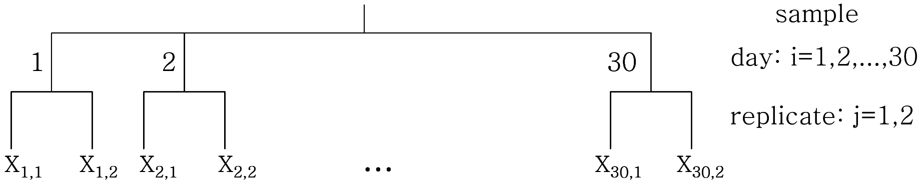


Fig. 2. Experimental design used to analyse the samples (control samples and reference material)

Performance parameters estimation

The validation of an analytical methodology with either a quantitative or a qualitative method must fulfil some requirements [18, 19] related to the quality of the analysis, and time and cost constraints. According to the ‘ fitness-for-purpose’ approach, the best performance parameters for our test kit depending on these requirements are identified and estimated.

As has been stated, traceability is assessed [21] by using a reference material, although other possibilities do exist (reference or alternative methods). In addition to the ‘ British Standard varicella-zoster antibodies’ reference material [15], it is assumed that the control samples submitted by the test kit are secondary references since they have been compared to an in-house serum preparation and that the whole test kit performance has been compared to another commercially available ELISA [22]. The control samples are measured twice daily to establish the cut-off value. Also, as is stated in the test kit instructions, the analysis will be valid if: a) the absorbance value for the blank sample (i. e. the microtiter well that undergoes the analysis without the control or test sample) is lower than 0.150; b) the mean absorbance value of the negative control sample is lower than or equal to 0.250; and c) the mean absorbance value for the positive control is equal to or higher than 0.750. Therefore, traceability is established daily by reference to these control samples.

In the context of clinical analyses, performance parameters such as sensitivity and specificity are also used. On the one hand, clinical assays may refer to analytical sensitivity and analytical specificity, both of which are classical concepts in analytical chemistry [13]. On the other hand, they may also refer to diagnosis sensitivity and diagnosis specificity. Though these latter terms generally agree with those considered for sensitivity and specificity in qualitative methods of analysis [12, 13, 23], some fine distinctions should be highlighted. Diagnosis sensitivity is a measure of the probability of correctly diagnosing a diseased condition. In the present case study, it is the probability of correctly assessing the presence of VZV IgG antibodies. Diagnosis specificity measures the probability of correctly identifying a non-diseased condition. In this case study, it is the probability of correctly assessing the absence of VZV IgG antibodies. The lower these probabilities are, the better the test kit performs.

These two parameters are very important because they show the ability of the test kit to correctly classify positive and negative samples. They are evaluated with the positive and the negative control samples. Sensitivity is also estimated with a sample at another level of activity. This corresponds to the sample obtained by diluting the positive control by a factor of 1/8. Moreover, from their measurement results and subsequent probability distributions, the probability of misclassification is also estimated.

Selectivity is an intrinsic characteristic of ligand-receptor systems such as antigen-antibody. However, unspecific reactions often occur. In this case study, rheumatoid factor may interfere with the binding of IgG-specific conjugates, producing a lower or false reaction in the test kit. It is assumed that the manufacturer has checked and confirmed the absence of these reactions [22]. Also, cross reactivity with Herpes Simplex Virus and Epstein-Barr Virus may be expected. It is necessary, therefore, to rule out such infections before interpreting the results of the test kit. As a first stage, and according to the standard protocol and experience of the laboratory, it is assumed that these two infections are absent. However, sound studies of cross reactivity are left for subsequent stages of the validation process.

In qualitative analysis and particularly when dealing with binary responses, the unreliability region is where inconclusive results are obtained [12, 24]. In the test kit this region of inconclusive responses is placed around the cut-off value.

It is therefore in the range of absorbance or index values where the test kit wrongly determines the presence or absence of IgG antibodies against VZV. This performance parameter is estimated using the definition of the two-sided upper prediction bound [17, 25] and the precision associated to the cut-off value. As will be shown below, the precision of the cut-off value is estimated by means of the propagation error law.

As can be seen in Figure 3, the limits of the unreliability region around the COV (index equal to 1) provide very valuable information because 1) they define when the positive, negative and inconclusive results are obtained; and 2) together with the probability density function, they make it possible to estimate the probability of obtaining false responses (positive and negative). It should be pointed out that if both limits are to be defined a predefined probability of error (α and β probabilities of error) must be considered.

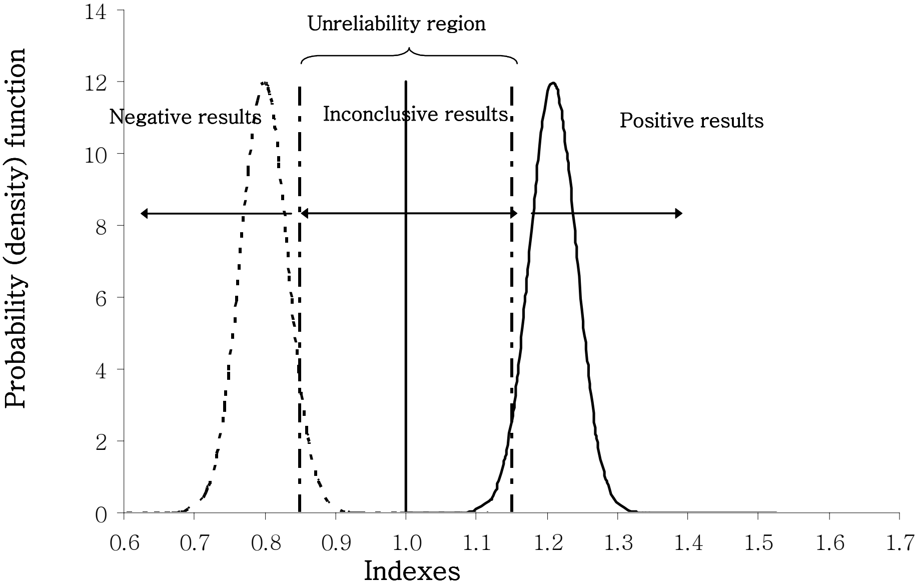
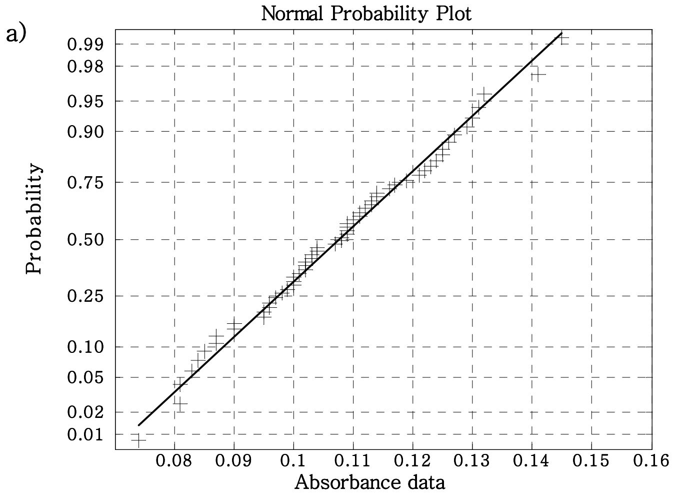


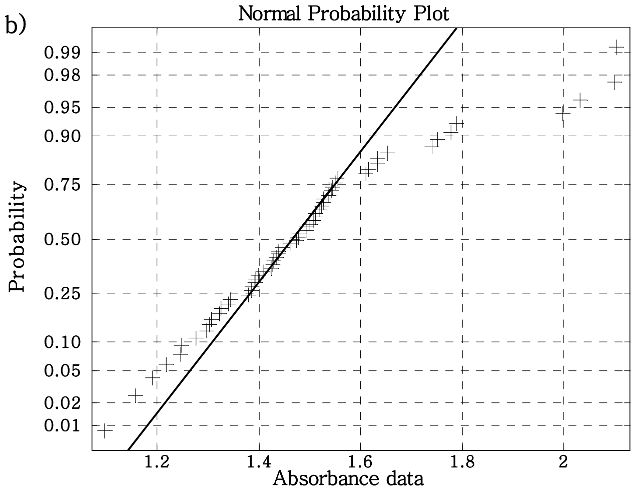
Fig. 3. Information provided when the unreliability region is defined. The dotted vertical lines are the lower and upper unreliability region limits. The solid line corresponds to the distribution function of a hypothetical positive sample. The dotted line corresponds to the distribution function of a hypothetical negative sample.

The probabilities of giving false results are the false results rates. False results can be either false positive (when the test kit classifies a true negative sample as positive, i. e. the presence of IgG antibodies is wrongly assessed) or false negative (when the test kit classifies a true positive sample as negative, i. e. the absence of IgG antibodies is wrongly assessed) [6]. They are closely related to the sensitivity and specificity of the diagnosis and they also give an idea of how good the test kit classifies. Inconclusive results slightly modify this reasoning because, on the one hand, the lower limit of the unreliability region gives the percentage of real negative samples providing an inconclusive result; and on the other hand, the upper limit of the unreliability region gives the percentage of real positive samples providing an inconclusive result. Figure 3 shows that the area of a negative control sample above the index value 0.85 is the probability of providing an inconclusive result, i. e. the false inconclusive rate. The area of the same distribution function above the index value of 1.15 would be the false positive rate. Regarding the distribution function of a positive control sample, the area below the index value of 1.15 is the probability of providing an inconclusive result, i. e. a false inconclusive rate of a positive sample, and the area below the index value of 0.85 would be the false negative rate.

Results and discussion

In theory, the data obtained when the control samples and reference material are analysed in duplicate every day for thirty days should follow a Gaussian distribution. Therefore, once this hypothesis is assessed, we propose a validation procedure based on the properties of the t-Student distribution function, which is a particular case of the normal probability distribution function. First, the expected distribution function of the data must be checked. Figure 4 shows the normal plot of the experimental data. The data of the negative control sample clearly show a normal probability distribution function as the data points fit the straight line. Some data points of the positive control sample and of the reference material slightly deviate from normality. The deviations shown could be due to the occurrence of unspecific reactions. However, it is assumed that the data of the positive control sample and the data of the reference material follow a normal probability distribution function. These assumptions have been assessed by means of the Chi-square normality test [26]. Before estimating the performance parameters, it is systematically checked the presence of abnormal data and/or outliers. The different sources of variation by means of the Analysis of Variance (ANOVA) are also studied.





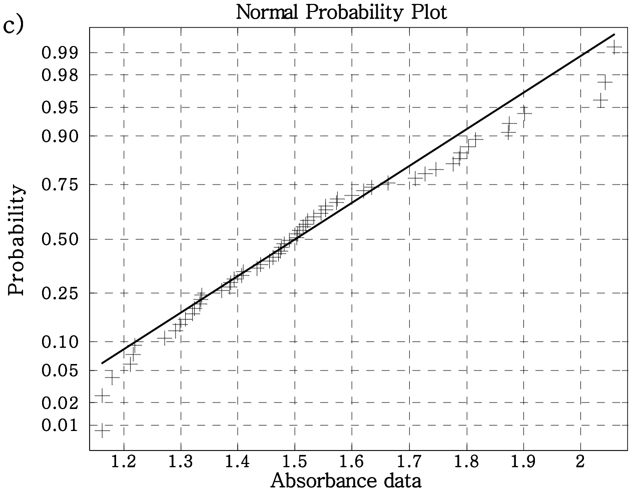


Fig. 4. Normal probability plot for a) negative control sample data, b) positive control sample data and c) reference material data

Data pre-treatment

The measurements are performed in accordance with a specific experimental design (Figure 2). Therefore, information about the sources of variation (day and replicates) that affect the data dispersion can be obtained. Table 1 shows the results of performing an Analysis of Variance (ANOVA) [26].

Table 1. Calculated and tabulated F-values for both control samples, for the reference material and for both diluted samples from the positive control sample

	Neg. control sample	Pos. control sample	Reference material	Pos. control sample: dil. factor 1/8	Pos. control sample: dil. factor 1/12
S^2 day	0.00053	0.053	0.57	0.016	0.0098
S^2 replicate	0.00037	0.025	0.068	0.0025	0.0070
Calculated F-value	1.5	2.1	8.3	6.4	1.4
Tabulated F-value	1.8	1.8	1.8	1.8	1.8
Significant differences?	NO	YES	YES	YES	NO

There is no significant difference between the within-days (replicates) and between-days variance of the negative control sample. On the other hand, there are significant differences between these two variances for the positive control sample and the reference material. With regard to the normality of the positive control sample and the reference material sample, the higher level of activity probably increases the unspecific reactions. The differences between the variances for the reference material are considerable. In this case, two factors must be considered: first, the dilution (200 fold) that decreases the initial activity of 4 IU/ml; and second, the final reference material solution, which is prepared every day as has been described above. This dilution process negatively affects the data dispersion.

Estimation of the quality parameters

Traceability is assessed by statistically comparing the measurement response of the positive control sample with the measurement response of the reference material, both at the same level of activity. In the case under study, the positive control sample has a mean index value of 5.8 and a standard deviation of 0.20, while for the reference material the mean value and standard deviation are 5.9 and 0.50, respectively. The t-value calculated to compare mean values is 1.4 and the tabulated t-value is 2.0 for 74 degrees of freedom and a significance level of 5%. Therefore, it is concluded that the mean values do not statistically differ at the level of significance chosen. Consequently, the results obtained with the test kit are traceable to the reference material.

Sensitivity and specificity are estimated from control samples measurements. As it is shown, all negative samples provided negative results as they are below the range of values that the manufacturer considers to be inconclusive (0.85-1.15) so, the specificity of the diagnosis given by the test kit is 100%; and all the positive results are above the previously mentioned range of values, so sensitivity of the diagnosis given by the test kit is also 100%. Moreover, for the 1/8 factor diluted positive control sample, 59 measurements out of 60 provided a positive result: that is to say 98.3 % sensitivity at that level of activity.

The unreliability region is the range of instrumental responses around the cut-off value that gives rise to inconclusive results. As it has been stated in the section Performance Parameters Estimation, this region can be defined using the two-sided prediction bounds [25]. The cut-off values are derived from absorbance values that follow a t-Student' s probability distribution function and therefore follow the same type of distribution. Because of this, the limits of the unreliability region provided by the manufacturer are checked. To achieve this, a prediction interval around the COV (Equation /3/) is established. The sample is positive or negative if its absorbance value is above or below the limits of this interval (Equation /3/).

$$\textit{Prediction interval} = COV \pm t_{(\alpha,\nu)} \times s_p \qquad /3/$$

s_P is the standard deviation associated to the prediction interval and it is estimated by taking into account the standard deviation of the cut-off value and the number of replicates measured of the test sample (Equation /4/):

$$s_p = \left(\frac{1}{m} + \frac{1}{n_{COV}}\right)^{\frac{1}{2}} \times s_{COV} \qquad /4/$$

And

m is the number of replicates carried out on the future unknown sample. It is usually 1.

n_{COV} is the number of independent analyses performed to calculate the cut-off value. In our case, this is 30 measurements.

s_{COV} is the standard deviation of the cut-off value.

A key point in estimating of the unreliability region is the precision associated to the cut-off value. To experimentally assess its size, the error propagation law to the COV expression (Equation /1/) is applied. As can be observed, it depends on two variables: the mean value of the absorbance for the negative control sample, $\overline{A}_-$, and the mean value of the absorbance for the positive control sample, $\overline{A}_+$. The precision associated to the COV is calculated therefore as a function of the precisions associated to the control samples in terms of variance, $s^2_{\bar{A}_-}$ and $s^2_{\bar{A}_+}$ (Equation /5/).

$$s^2_{COV} = \left[\frac{\partial COV}{\partial \bar{A}_-}\right]^2 \times s^2_{\bar{A}_-} + \left[\frac{\partial COV}{\partial \bar{A}_+}\right]^2 \times s^2_{\bar{A}_+} \qquad /5/$$

This leads to Equation /6/,

$$s^2_{COV} = 1^2 \times s^2_{\bar{A}_-} + 0.1^2 \times s^2_{\bar{A}_+} = s^2_{\bar{A}_-} + 0.1^2 \times s^2_{\bar{A}_+} \qquad /6/$$

As the absorbance mean values and standard deviations are 0.11 and 0.015 for the negative control sample, and 1.5 and 0.21 for the positive control sample, the s_{COV} value is equal to 0.026. Expressed in terms of relative standard deviation, this corresponds to 10% of the mean value of the cut-off value, which is 0.25. This value of 10% is not so different from the variability value provided by the manufacturer (15%), although it is slightly lower.

In our case study, s_P and s_{COV} have about the same value because the factor $\left(\frac{1}{m}+\frac{1}{n_{COV}}\right)^{1/2}$ is 1.016. Therefore, in accordance with Equation /3/, the new unreliability region lies between the absorbance values 0.30 and 0.21 considering the tabulated t-value at a 5% level of significance, or in terms of indexes between 0.83 and 1.17. That it is to say, a sample will be positive when its index is above 1.17, with an error probability of 5%. The same occurs with negative samples whose index is below 0.83. This unreliability region coincides with the

15% of variability provided by the manufacturer (0.85 – 1.15). It should be emphasized that the unreliability region strongly depends on the previously set probability of error. For instance, if an error probability of 1% is considered, then the new unreliability region lies between the absorbance values of 0.19 and 0.32, or in terms of indexes between 0.75 and 1.25.

The advantage of this approach is that the properties of the normal probability distribution function are perfectly established. The analyst has to set a probability of error when setting the confidence intervals, which means that he already knows how many samples will provide an inconclusive result. In addition to this, the probability of error can be adapted to the problem in hand: the smaller the probability is, the larger the confidence interval will be and vice-versa.

The data for the dilution factors 1/8 and 1/12 of the positive control sample are used to estimate the false rates at activity levels close to the COV. Though they do not seem so important, there are some differences between the two data sets. The reason is that the higher the activity level is, the higher probability for unspecific reactions to occur is. So, the probability of having measurements out of range is also higher when using the dilution factor 1/8. As it has been stated before, the same experimental design shown in Figure 2 is used. The resulting distributions are shown in Figure 5.

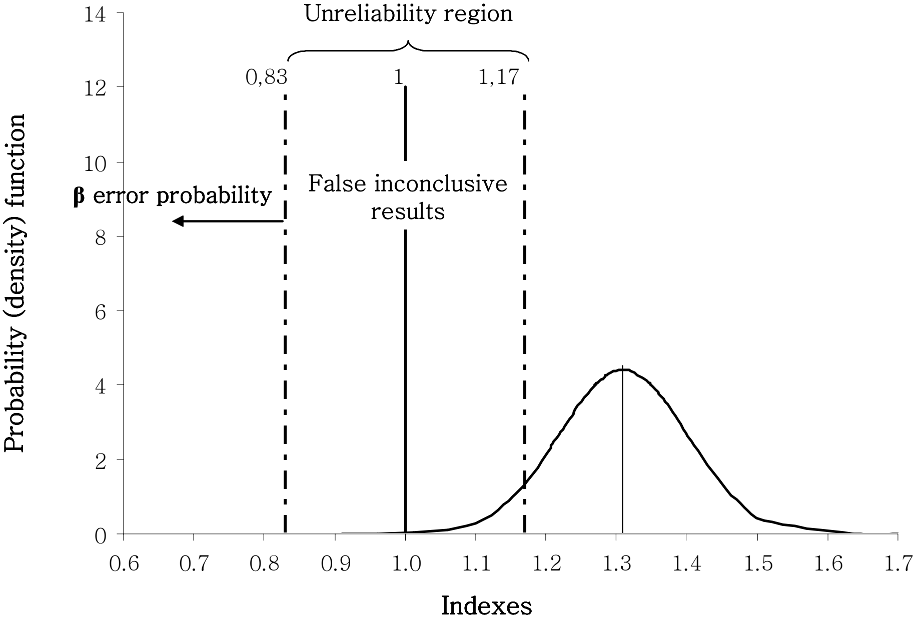


Fig. 5. Index distribution obtained for the positive control sample diluted at 1/8. It is computed following the theoretical t-distribution function. The unreliability region is also plotted.

The false negative rate or probability of committing a β type error is calculated using the data from the 1/8 dilution factor distribution function. The experimental probability of committing error is zero because no samples out of the sixty tested provide an index equal to or lower than 0.83. That is to say that no sample provides a false negative result. However, in this case false inconclusive results arise because samples belonging to the low positive distribution function (1/8 dilution factor) provide inconclusive results instead of positive ones. All samples whose indexes are lower than 1.17 belong to this category. From 60 measurements, 3 provided an inconclusive result. The rate is computed as 3x100/60=5 %.

Similarly, using the distribution function corresponding to the dilution factor 1/12 the false negative rate from an inconclusive sample is estimated. Therefore, this rate is computed by considering negative results: that is to say indexes below 0.83 from among all the replicated measurement samples belonging to the inconclusive distribution function. Of the 60 analyses, 9 were negative, so the probability is calculated as 9/60x100=15 %. No false positive results were obtained from this sample. Since this sample should provide inconclusive results, it makes no sense to determine the sensitivity rate at this activity level.

Finally, the probability of committing a false positive (the probability of committing an α error) from a negative sample is visualized from the distribution function for a sample that provides indexes that are almost below the lower limit of the unreliability region. As this sample is not available, it is indicated how this rate should be computed. Using this distribution function, the positive results obtained should be computed, which means indexes above 1.17. False inconclusive results from a negative sample can also be computed using this distribution function.

Conclusions

In the framework of the recent development of the validation of qualitative methods and of clinical analysis, a methodology for test kit validation that is adapted to the characteristics of the present assay has been reported. As it is based on an instrumental response measurement and on the use of control samples to estimate the cut-off, a methodology that characterizes the statistical distributions of the positive and negative control sample measurements is provided.

The validated immunoassay-based test kit, which is used in some laboratories as a routine method, measures immunoglobulin G antibodies of VZV in human serum. Traceability, sensitivity and specificity of the diagnosis, the unreliability region and the probabilities of giving false results (positive and negative false rates) are estimated. The value of the specificity, agrees with the manufacturer' s specifications of 100%, but the sensitivity differs slightly from the value provided of 92.68%. This difference could arise from the differences in the levels of activity tested in both validation studies.

In spite of the lack of any formal statement, the manufacturer indirectly sets the probability of committing false results at nearly zero because the unreliability region proposed is very wide. The price to be paid is that there are more inconclusive results. In the validation process, this region is re-estimated experimentally by taking into account the precision associated to the cut-off value and a pre-set probability of committing both types of error. Considering this new unreliability region, the probability of obtaining false results is estimated. In this particular case, not only the false negative rate but also the false inconclusive rate is calculated.

The validation methodology presented in this paper has been chosen according to the intrinsic characteristics of the test kit, i. e. the instrumental response and the measurement of control samples to calculate the cut-off value. The statistical intervals derived are used to estimate the performance parameters by taking into account the variability associated with the cut-off value. The analyst can either adopt the usual 5% probability of committing an error or select a certain value for probability of error that suits the problem at hand. Depending on the availability of an appropriate antigen-antibody system, this methodology could be used to validate other ELISA formats, such as direct and sandwich, as well as other analytes in other matrices.

Acknowledgments

The authors acknowledge economic support from the MCyT (projects nº BQU2003-500 and BQU2003-1142). Esther Trullols would also like to thank the URV for providing a doctoral fellowship.

References

[1] P. P. McDonald, R. E. Almond, J. P. Mapes, S. B. Friedman, Journal of AOAC Int. 77 (1994) 466-472.

[2] G. Suhren, K. Knappstein, Anal. Chim. Acta 483 (2003) 363-372.

[3] International Organisation for Standardization (1994), ISO 8402, Quality management and quality assurance. Vocabulary. ISO, Geneva, Switzerland

[4] International Organization for Standardization (1999), ISO/IEC 17025, General requirements for the competence of testing and calibration laboratories. ISO, Geneva, Switzerland

[5] Environmental and Protection Agency (2004) Test Methods: Methods Development and Approval Process.
URL: http://www.epa.gov/epaoswer/hazwaste/test/methdev.htm

[6] 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results.

[7] P. Feldsine, C. Abeyta, W. Andrews, Journal of AOAC International 85 (2002) 1187-1200.

[8] R. Song, T. J. Fischbach, K. Ashley, AIHA Journal 57 (1996) 161-165.

[9] M. Valcárcel, S. Cárdenas, M. Gallego, Trends in Anal. Chem. 18 (1999) 685-694.

[10] E. Trullols, I. Ruisánchez, F. X. Rius, M. Òdena, M. T. Feliu, Journal of AOAC Int. 87 (2004) 417-423.

[11] A. Pulido, I. Ruisánchez, R. Boqué, F. X. Rius, Trends in Anal. Chem. 22 (2003) 647-654.

[12] E. Trullols, I. Ruisánchez, F. X. Rius, Trends Anal. Chem. 23 (2004) 137-145.

[13] J. R. Crowther, The ELISA Guidebook, Humana Press, Totowa, New Jersey, 2001.

[14] Human Gesellschaft für Biochemica und Diagnostica mbH (2004) Max-Planck-Ring 21 - D-65025 Wiesbaden. Germany

[15] British Standard varicella-zoster antibodies' (2004) National Institute for Biological Standards and Control (NIBSC). URL: http://www.nibsc.ac.uk

[16] J. W. A. Findlay, W. C. Smith, J. W. Lee, G. D. Nordblom, B. S. Das I, DeSilva, M. N. Khan, R. R. Bowsher, J. Pharmaceut. Biomed. 21 (2000) 1249-1273.

[17] A. Pulido, I. Ruisánchez, R. Boqué, F. X. Rius, Anal. Chim. Acta 455 (2002) 267-275.

[18] CITAC/EURACHEM (2002) Guide to Quality in Analytical Chemistry. An Aid to Accreditation. CITAC/EURACHEM Guide. URL: http://www.eurachem.ul.pt/guides

[19] EURACHEM (1998) ' The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics' EURACHEM Secretariat Teddington, Middlesex. URL: http://www.eurachem.ul.pt

[20] E. Trullols, I. Ruisánchez, F. X. Rius, J. Huguet, Trends Anal. Chem. 24 (2005) 516-524.

[21] International Vocabulary of Basic and General Terms in Metrology (VIM) (1993) Revision of the 1993 edition, International Vocabulary of Basic and General Terms in Metrology (VIM).
URL:http://www.abnt.org.br/ISO_DGuide_99999_(E).PDF

[22] Verification Report for Varicella-Zoster Virus IgG ELISA Antibody Test (VZV IgG) Human Gesellschaft für Biochemica und Diagnostica mbH (2004).
URL: http://www.human.de/data/gb/vr/el-vzvg.pdf

[23] J. F. Kay, J. D. MacNeil, J. J. O' Rangers (Eds.), Current Issues in Regulatory Chemistry, AOAC International, Gaithersburg, Maryland, 2000.

[24] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhardt, R. W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel, Accred. Qual. Assur. 8 (2003) 68-77.

[25] G. J. Hahn, W. Q. Meeker, Statistical Intervals. A Guide for Practitioners, John Wiley & Sons, New York, 1991.

[26] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke, Data Handling in Science and Technology 20A. Handbook of Chemometrics and Qualimetrics: Part A, Elsevier Science, Amsterdam, 1997.

4.4 A HOMEMADE AUTOANALYZER THAT PROVIDES GLOBAL INDEXES

The last part of this chapter describes the validation of a homemade autoanalyzer that is used to determine whether two analytes in the same solution are simultaneously above or below a certain concentration level.

Like the two previous practical applications, this one is carried out in collaboration with the Department of Analytical Chemistry of the University of Córdoba.

This method of analysis is like the new automated methods of analysis that are to be used as screening sample systems, and which are being developed by the Department of Analytical Chemistry of the University of Córdoba. Other configurations are used for the direct screening of lyophilised fluids for bile acids [25], for the sequential determination of total sugars in soft drinks [26] or for the screening and confirmation of sulphonamide residues in milk [27].

This particular application aims to determine the total content of mineral oil and surfactant in degreasing baths, which are commonly used in the automotive industry. The characteristics of the method of analysis depend on the requirements of the external client, and it is designed and built with these in mind. It is, therefore, a homemade method of analysis unlike the previous practical applications which are commercially available and whose performance parameters have already been defined by the corresponding manufacturer.

Another difference between this application and the ones that have been described so far is its environmental approach. Degreasing baths are rather important contaminants, so they should be used and replaced, and their waste managed with care. Therefore, external clients require a method of analysis which determines when the degreasing bath must be replaced. Our method is quicker and more reliable than current methods, which are rather tedious, involve several steps and take 24 hours to provide the final results.

4.4.1 Degreasing baths

Various finishing processes are carried out in all industries that produce metal parts that must comply with specific conditions before they are subject to any other changes

Production and mechanization (i. e. modification of the physical structure of the metal pieces to give them particular characteristics or specifications) is usually performed with tools which have cutting edges. The friction and deformation have several undesirable effects on both metal surfaces (e. g. local cold welding, swarf, chipping). To minimize these effects, cooling lubricants or cutting fluids are widely used. Cutting fluids dissipate heat, reduce friction and remove swarf during the mechanization. They are responsible for the lubrication that prevents breakage and wear of the cutting tool and protects metal surfaces from oxidation and corrosion [28].

Cutting fluids are essentially of two different types. On the one hand, oil-based systems are composed of either mineral or synthetic oils with some additives which, depending on the subsequent application, can be either organic or inorganic compounds [29]. Due to the oil-based formulation, they considerably reduce friction. On the other hand, water-based emulsions (i. e. 2-10% oils and additives) have a higher cooling efficiency. In the present application, the fluids used basically contain mineral oils.

Once the metal piece has the required physical conformation, the oil and grease must be completely removed because the subsequent processes (i. e. galvanization, painting, passivation, etc.) need the surface to be absolutely free from oil and grease. Therefore, degreasing baths are an essential part of the productive process.

Traditionally, metal components have been degreased with halogenated organic solvents such as trichloroethylene or tetrachloroethylene [30]. Because the volatile organic compounds emitted represent both an environmental hazard (i. e. ozone damage, air and ground water contamination) and a health risk (i. e. damage to the hormonal system and neurotoxicity) [31], they are gradually being substituted by other degreasing modalities. More modern techniques use aqueous cleaners which, in combination with a mechanical action, make cleaning efficient [32]. However, not all aqueous cleaners can be used to degrease all metal components. Such factors as the chemical and physical properties of the metal component or the finishing operation required, as well as the amount of oil to be removed must be considered before it is submitted to a proper degreasing process.

To achieve the best performance, the mechanical action is also very important [33]. Immersion systems (i. e. tanks in which the parts are immersed in the degreasing solution for a certain time) can combine both temperature and ultrasound to improve the separation of the dirt from the substrate. Sprinkling systems are also widely used. They combine the chemical effect of the degreasing solution with the mechanical effect of the high pressure of the flow, and give good results.

The three main types of aqueous degreasers are acid, neutral and alkaline. Acidic solutions are composed of mineral acids (hydrochloric, sulphuric and nitric), chromic acid, carboxylic acids, and other organic acids. They are useful for removing metal oxides before pre-treatment or painting. However, such cleaning solutions generally require more attention due to the aggressive action of the acid on the metal parts. Increased dirt loading and neutralization are common problems that require the cleaning solution to be changed frequently. Acidic cleaners are generally not the best choice as degreasing agents.

Cleaning solutions with a pH from six to eight are considered neutral. They generally include surfactants, which act as wetting and emulsifying agents. Other ingredients, such as corrosion inhibitors and dispersants, are also generally added. These formulations are best suited for removing organic residues (e. g. oil and grease) and many inorganic residues. The key point for best performance is to select the solution that is most appropriate for the targeted dirt; also important is the type of mechanical agitation chosen for the process.

Alkaline cleaning solutions are formulated by adding such materials as sodium or potassium hydroxide, carbonate, bicarbonate, phosphate, silicate, or other similar materials. It is important to keep in mind that sodium hydroxide, like other alkaline metal hydroxides, is very corrosive.

A solution at a pH of 13.5 will remove carbonaceous soils. pHs ranging from 8 to 13 are generally used to remove oils and greases. As a rule, alkaline degreasing solutions do not need the same level of attention as acidic ones. Nonetheless, they must be periodically monitored and adjusted for concentration and soil loading.

Alkaline degreasing solutions use both physical and chemical means to clean the substrate surface. Chemical action can occur via the saponification of certain contaminants. In the saponification process, water-soluble soaps are produced by neutralizing fatty acid soils. Physical cleaning occurs via the wetting and emulsification caused by the addition of surfactants. These are the degreasing baths that are used in the present application.

The surfactants found in alkaline degreasing baths can have different hydrophilic parts and therefore different natures. Anionic surfactants have been the most common because of their optimal detergency properties. However, non-ionic surfactants also provide good detergency but at lower pH values. Amphoteric surfactants, which are used in personal care products and in neutral cleaning solutions, are added to improve the properties of other surfactants. Some examples of the three main types of surfactants are given in Figure 11.

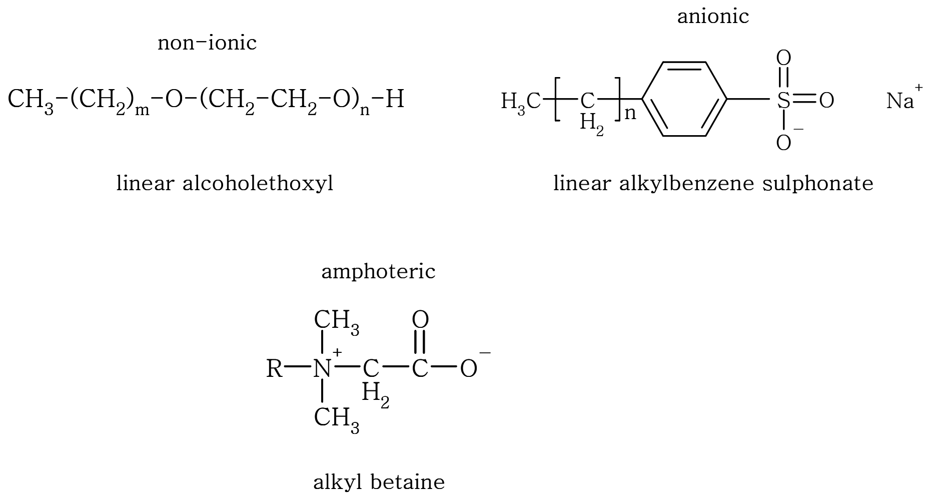


Figure 11. Three examples of non-ionic, anionic and amphoteric surfactants

In addition to the wetting and emulsifying effects of the surfactants, other components are also added to alkaline degreasing baths. The so-called builders or alkaline agents (e. g. sodium and potassium hydroxides, carbonates and phosphates) [34] help to disperse dirt and to enhance the properties of the surfactants. Anti-corrosive and chelating agents also have a key role in the degreasing process.

Decreasing baths cause a variety of environmental problems because they contain oil and grease residues, the management of is strictly legislated [35] and the effluents are alkaline. Therefore, appropriate waste management is required. However, some proposals have been made to prolong the lifetime of aqueous degreasing baths. Ultrafiltration has become an extensively used technique for separating oil and grease from the solid particles which may be in suspension. The separation is performed using a selective membrane which is renewed periodically. Oil skimmers collect the oil and grease accumulated on the surface when the emulsion formed is not highly stable. The filtered residue is then submitted to suitable waste management. A more recent technique uses microorganisms that degrade complex molecules such as oils and greases. The resulting main product is carbon dioxide. These aerobic bacteria need special conditions, which involve relatively low temperatures (40-50°C) (compared to the temperatures used in the subsequent processes), and a pH range between 8.5 and 9.5 to enhance the consumption of the emulsified oil. In these conditions, the degreasing bath achieves optimal performance [33].

The prolongation of the bath' s lifetime is limited; i. e. the oil and grease content and the concentration of surfactant must be controlled. In the application below, the client uses a 24-hour method of analysis which is based on the EPA official method [36]. This involves extracting the analytes with n-hexane to determine the total content of both families of analytes and, then, treatment with silica gel to remove the polar compounds and determine the total content of grease. Therefore, there is a need for a rapid and reliable method of analysis.

The configuration designed for our method is a continuous flow system that can sequentially determine the surfactant and the mineral oil content in a run time of less than ten minutes. The two families of compounds are separated with a silica sorbent column so that the surfactant can be retained [37]. The detection system is an evaporative light scattering detector which has been proven to provide good results in the determination of industrial surfactants determination [38].

The new method of analysis, which determines both the surfactant and grease content, is submitted to a suitable validation procedure that takes into account its particular characteristics. The validation process and the results are presented in the following section.

4.4.2 STATISTICAL INTERVALS TO VALIDATE AN AUTOANALYZER FOR MONITORING THE EXHAUSTION OF ALKALINE DEGREASING BATHS

Anal. Chim. Acta. In press

E. Trullols, I. Ruisánchez, E. Aguilera[a], R. Lucena[a], S. Cárdenas[a] and M. Valcárcel[a].

Universitat Rovira i Virgili. Departament de Química Analítica i Química Orgànica. C/ Marcel·lí Domingo s/n. 43007 Tarragona (Spain)
[a] Universidad de Córdoba. Marie Curie Annex Building, Campus Rabanales. 14071 Córdoba (Spain)

Abstract

We describe how to use the statistical intervals for validating a qualitative method for determining the alkaline degreasing baths exhaustion. A homemade autoanalyzer based on flow injection-evaporative light scattering detector (FI-ELSD) coupling measures two instrumental responses related to the contents of surfactant and mineral oil. These two responses are necessary to decide whether the degreasing bath is exhausted. The instrumental responses r_i are compared to their corresponding decision values i.e. cut-off response ($r_{cut-off}$) and screening response ($r_{screening}$). These decision values are calculated by defining the one-side prediction bound around the specification limit (S_L) of both analytes. The prediction bound of each analyte must be defined differently according to their corresponding specification limit. Performance parameters, such as sensitivity, specificity, false response rates and the unreliability region, are established. The performance of this qualitative method of analysis is checked by analyzing a set of 10 real samples. Our results show that the method is accurate as far as mineral oil content is concerned.

Keywords: statistical intervals; flow-injection-evaporative light scattering detector coupling; degreasing baths; validation

Introduction

The correct performance of an analytical method is important because it implies that it satisfies the requirements for which it was designed. This is part of the validation process, which is carried out at the end of the method development stage. This process must be carefully defined if the method' s performance characteristics are to be accurately assessed. According to the ISO definition [1, 2], validating an analytical method means defining and estimating the performance parameters needed to satisfy the analytical requirements. In a similar way, the EURACHEM defines validation as the confirmation of the method performance capabilities consistency with the requirements of the application [3].

The validation of qualitative analytical methods has not been within the scope of the main regulatory bodies, although some documents and guidelines, which are not generally accepted but valuable nonetheless, can be found in the bibliography [4-6]. Some validation proposals have been published addressing to specific applications. The methodologies used in those cases are different depending on the intrinsic characteristics of the method of analysis. The main approaches are the Performance Characteristic Curves, Bayes' theorem, the Contingency Tables and the Statistical Intervals. As far as the methodology that uses Performance Characteristic Curves [7] is concerned, it is suitable for methods providing sensorial (i.e. visual) detection [8]. They allow the estimation of several performance parameters of the method such as sensitivity and specificity rates, as well as the unreliability region. Bayes' theorem allows the calculation of conditional probabilities referred to just one sample. However, it has been used as an approach to quantify uncertainty [9]. Contingency tables also permit the calculation of predictive values of the method of analysis and they have been widely used in clinical analysis. Recently, statistical intervals, and concretely prediction intervals, have been used to validate a qualitative method providing an instrumental response [10, 11].

Following the last presented approach, in this paper we report the validation procedure for a qualitative method that assesses if an alkaline degreasing bath is to be replaced. The analytical method is not a test kit but a homemade autoanalyzer that uses a high-pressure pump, an injection valve, a silica sorbent column and an evaporative light-scattering detector [12]. It measures simultaneously two analytes and compares their response with the response of their corresponding specification limit (S_L), i. e. the concentration of the mineral oil and of the surfactant at which the bath is exhausted. The decision about the sample is done considering simultaneously the two target analytes. Therefore, statistical intervals are defined around the specification limit of each analyte. New decision values such as cut-off and screening limits are also defined to take into account the different types of error.

In addition to the establishment of the new decision values, performance parameters such as sensitivity and specificity, the unreliability region and false results rates, are also estimated from the statistical intervals defined in response terms.

Degreasing baths are often used as a necessary step before the final processing of some metallic components because a perfectly clean and active surface is required. Previous steps in the manufacturing or processing of these components involve using greases and oils, usually mineral, with cooling and lubricant properties. These are usually removed from the metallic components using an alkaline degreasing bath by simply dipping them or sprinkling them with the cleaning solution. Other possibilities involve electrolytic techniques or ultrasounds. A wide range of alkaline degreasing baths exist because the composition of the bath must suit the problem at hand: e.g. dirt, the cleaning system, the composition of the metallic component or the subsequent process. They all have a similar formulation, which is based on the following main components [10]: surfactants (used as humectants), alkaline salts (used for the saponification of the oil and greases) and chelating agents (used to avoid the precipitation of metallic hydroxides).

As the amount of mineral oil in the degreasing bath increases, the bath becomes less and less efficient until a new one is required. The exhausted bath must then be submitted to proper waste management, which involves both economic and environmental costs. The exhaustion of the degreasing bath must therefore be correctly assessed in order to remove any still usable bath. Exhaustion is defined by the client or by the end user of the metallic components and measured in terms of different indicators—basically, the amount of mineral oil collected and the content of alkaline salts, though the amount of surfactant is an equally useful parameter.

Experimental

Apparatus

The autoanalyzer [9] used as the screening system is shown in Figure 1. It consisted of a Hewlett-Packard 1050 high-pressure quaternary gradient pump, a Rheodyne (Cotati, Ca, USA) 7725 injection valve fitted with a 250 μl PTFE sample loop, a laboratory-made silica column constructed by packing 40 mg of silica sorbent into a 3 cm x 4 mm i.d. PTFE tube using small cotton beads to prevent material losses, and a DDL 31 evaporative light-scattering detector (Eurosep, Cergy- Pontoise, France). The detector used air as nebulizing gas at 1.5 bar, the temperature of the nebulizing chamber was set at 75 ºC and the photomultiplier gain was set at 350V for the mineral oil and at 550V for the surfactant. Signals were acquired using an HPChem software connected to the detector via an HP 35900C (Agilent, Palo Alto, CA) multichannel interface. Peak height was selected as the analytical signal for the measurement of both grease and surfactant.

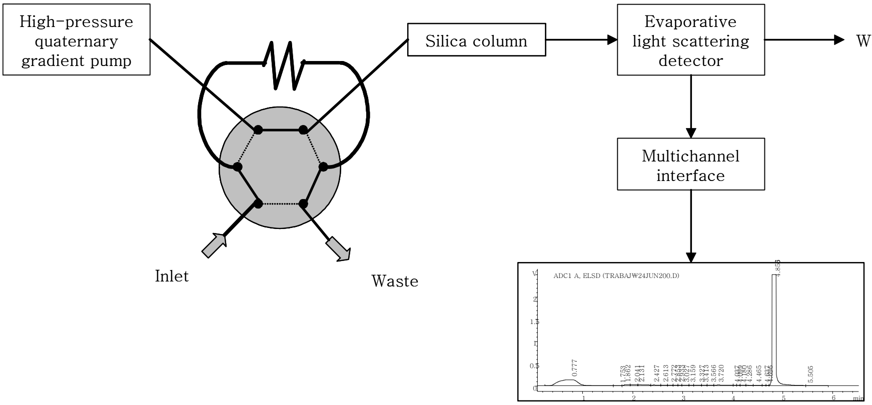


Fig. 1. The autoanalyzer used as the screening method

Reagents and samples

The validation standards contained both surfactant (0.39 g/l) and mineral oil (1g/l). In addition to these, 20 g/l alkaline salts solution was also added. The surfactant was a commercial product called Ridosol®. This, and the alkaline salts solution (Ridoline® 1565/1), were kindly supplied by Henkel Surface Technologies. Ridosol® is a mixture of 4 surfactants (Triton DF-11 (11%), Genapol PN-70 (3%), Lutensol DN-70 (11%), Plurafac LF-431 (5%) in 70% deionised water). Ridoline® is a solution of 48% potassium hydroxide (61.30%), 50% sodium hydroxide (3.20%), 75% phosphoric acid (6%) and boric acid (24.50%) in 5% deionised water. Ethanol 96% and *n*-hexane were obtained from Sharlau (Barcelona, Spain), sulphuric acid, sodium sulphate and light mineral oil were purchased from Sigma-Aldrich (Madrid, Spain).

Sample preparation and analysis by the screening method

The validation standards were prepared by measuring 830 µl of a standard solution of 9.05 g/l surfactant in deionised water and measuring 830 µl of a standard solution of 30.1 g/l mineral oil in n-hexane. Also, 0.5 g of Ridoline® was added to the 25 ml round flask. The organic phase was left to evaporate overnight and the corresponding amount of water was then added. The 25 ml aqueous solution was mixed in a separation funnel with 5 g sodium sulphate and 2 ml concentrated sulphuric acid. It was then extracted with 15 ml n-hexane and the final volume of the organic solution was 25 ml. 250 µl of the extract was injected into the screening system carried by an n-hexane stream at a flow rate of 0.5 ml/min, passed through the silica column for quantitative surfactant fraction retention while the grease was directly driven to the detector and quantified. The flow rate of the n-hexane was raised to 0.8 ml/min (3.5 min) for column clean-up. The surfactant fraction was eluted using an ethanol stream at a flow rate of 1.0 ml/min (4 min). A post-time of 5 min with 0.8 ml/min n-hexane was required as washing step.

The signal recorded was the light scattered by the analyte particles via previous nebulisation and evaporation of the mobile phase. The response is mass dependent [11], so the peak height of both analytes (mV) depends on the concentration of the analyte. The purity of the peaks was corroborated by infrared analysis [9].

Validation methodology

According to the requirements of the end user, the degreasing bath must be replaced when the mineral oil content is above the specification limit, and when the alkaline salts content or the surfactant content are below their corresponding specification limit, both in terms of concentration. Not complying with these values will negatively affect the final quality of the metallic components. Then, just by measuring the response of the mineral oil and of the surfactant, the decision on the bath lifetime can be taken.

In the present case, the final YES/NO result comes from the instrumental signal measured. This means that the decision about the sample, i. e. bath lifetime, is made by comparing the surfactant response and the mineral oil response obtained for a specific bath with the corresponding one-sided prediction bound [12] of the specification limit response ($\bar{r}_{SL}$) obtained when recording several times, in intermediate precision conditions, the response at the concentration specification limit defined for a given standard. The three main steps in this process are:

1) set the specification limit response for each analyte;

2) if error probabilities are considered, estimate new response values ($r_{cut-off}$ or $r_{screening}$) where the right final decision about the sample will be taken; and

3) estimate the performance parameters of the analytical method (sensitivity and specificity rates, the unreliability region and the false result rates).

Cut-off and screening responses

We have introduced the concept of specification limit responses as the responses at the concentration value of the mineral oil and at the concentration value of the surfactant, given by the client, at which we consider that the bath is no longer usable. However, taking the decision at the specification limit level is risky because, due to the associated imprecision of the measurements, the probability of committing an error is 50 %. Therefore, it is useful to take the decision at the cut-off response ($r_{cut-off}$) which is the response value beyond which the sample is positive with a certain probability of committing a type I error. This probability of committing the type I error is defined by taking into account the consequences of having false positive responses. The $r_{cut-off}$ strongly depends on the variability in the response values at the specification limit concentration [8, 12].

Although the type I error is taken into account, this may not be sufficient to make the decision at the $r_{cut-off}$ because the probability of committing a type II error is rather high. The decision is therefore made at the screening response ($r_{screening}$) to also take into account the type II error. This probability of committing a type II error is set considering the consequences of having false negative results. It is therefore also a response value beyond which the sample is positive with certain probabilities of committing type I and type II errors. Similarly, $r_{screening}$ depends on the variability of the response values of both analytes at their corresponding specification limit concentration level.

Both limits ($r_{cut-off}$ and $r_{screening}$) are defined from the statistical distribution of the specification limit. Then, the starting point in the definition of the prediction boundary is the response values corresponding to this limit. In the current application, both specification limits are set by the client at 0.39 g/l surfactant and 1 g/l mineral oil.

As the content of the mineral oil increases and the content of the surfactant decreases, their prediction boundary must be defined differently (Figure 2). This definition is done according to the following hypotheses:

1) Surfactant: H_0: $r_i \geq r_{SL}$ (SL=0.39 g/l) H_1: $r_i < r_{SL}$

A type I error means accepting H_1 when actually H_0 is true. This means affirming that the content of the surfactant is less than 0.39 g/l (the bath is exhausted) when it is not. The probability of making this type of error should, for several reasons, be as low as possible. An exhausted degreasing bath is subjected to a waste management process, which involves both economical and environmental costs. For those involved in waste management, therefore, it is more attractive to replace a truly exhausted degreasing bath that has perhaps been used longer than its shelf-life than to replace a falsely exhausted degreasing bath.

A type II error means accepting H_0 when actually H_1 is true. This means incorrectly affirming that the content of the surfactant is more than 0.39 g/l, i.e. the bath is not exhausted when in fact it is. Since the consequences of this wrong decision are relatively unimportant, it is not necessary to set the probability of this type of error very low.

The prediction boundary is therefore defined as in Equation (1 a):

$$r_{cut-off} = \bar{r}_{SL} - t_{(\alpha,\nu)} \times s_{SL} \qquad (1\ a)$$

To establish the cut-off response value we need to set the probability of committing a type-I error (false positive) as low as possible. Also, though in the present case, it is not so important, defining the β-type probability error would avoid a considerable number of false negative results. Therefore, if we take into account the probabilities of committing both types of error, we obtain the so-called screening response value, which depends on α, β and the bias Δ, which is defined as the difference between the screening response and the response at the specification limit (Δ = ($r_{screening} - \bar{r}_{SL}$)). From Figure 2 we can see that α, β and Δ are closely related. The previous definitions of α and Δ involve a particular β. Similarly, therefore, from pre-defined α and β, the bias is automatically set.

This new decision value is expressed as shown in Equation (1 b), which takes into account the specification limit response of the surfactant:

$$r_{screening} = \bar{r}_{SL} - \Delta(\alpha,\beta,\nu) \times s_{SL} \qquad (1\ b)$$

2) Mineral oil: H_0: $r_i \leq r_{SL}$ (SL=1 g/l) H_1: $r_i > r_{SL}$

A type I error means affirming that the content of mineral oil is higher than 1 g/l (the bath is exhausted) when it is not. For the mineral oil, the probability of a type-I error should be as low as possible.

A type II error means incorrectly stating that the content of mineral oil is equal to or lower than 1 g/l, i.e. the bath is not exhausted when in fact it is. Again, it is not necessary to set this probability of error as low as possible.

The prediction interval is defined as in Equation (2 a):

$$r_{cut-off} = \bar{r}_{SL} + t_{(\alpha,\nu)} \times s_{SL} \qquad (2\ a)$$

The same occurs if we consider the probability of committing a type II error (see Equation (2 b)) and take into account the specification limit response of the mineral oil:

$$r_{screening} = \bar{r}_{SL} + \Delta(\alpha,\beta,\nu) \times s_{SL} \qquad (2\ b)$$

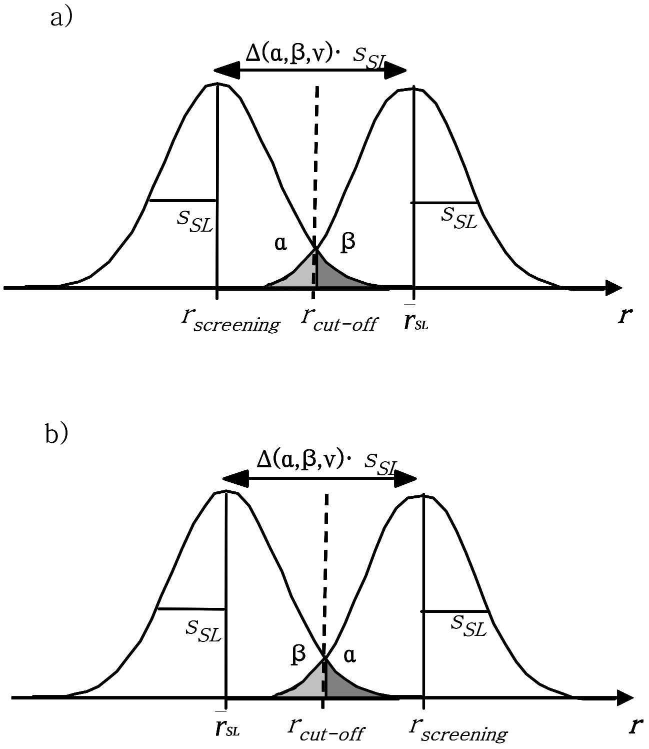


Fig. 2. Specification limit, cut-off and screening response for a) the surfactant and b) the mineral oil

Experimental procedure

The variability of the measured responses needs to be reliably evaluated. The experimental design is crucial to achieving this aim.

A key value in the estimation of the screening response is the standard deviation of the specific limit response s_{SL}. This value must be conveniently calculated using the following experimental design. To calculate the major sources of variability, the experimental design is therefore a 4-factor fully-nested design in which, for 22 days, two operators twice analysed two new and different validation standards (Figure 3).

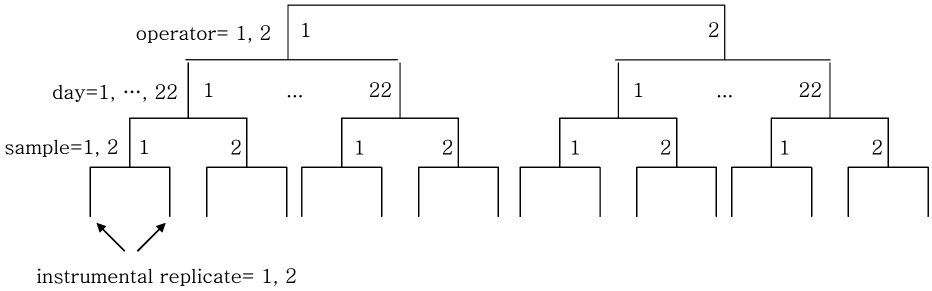


Fig. 3. Experimental design

The variance estimated in intermediate precision conditions contains the variability from the operator, day and sample. It is the estimated variance of an individual measurement made by an arbitrary operator on an arbitrary day. The intermediate precision can easily be estimated [13] by applying ANOVA to the results of this experimental design. However, the ANOVA table for the 4-factor fully-nested design is quite rare and a simpler design can be used if we consider the factors we vary within a run, which in the present case are the operator, the day and the sample. The design therefore becomes a two-factor fully-nested design with two instrumental replicates per run in which the variances are calculated according to Tables 1 and 2:

Table 1. ANOVA for a two-factor fully-nested design

Source	Mean Squares	Degrees of freedom
Run	$MS_{run} = \frac{n\sum_{i}\left(\bar{x}_i - \bar{x}\right)^2}{p-1}$	p-1
Residual	$MS_E = \frac{\sum_{i}\sum_{j}\left(x_{ij} - \bar{x}_i\right)^2}{p(n-1)}$	p(n-1)
Total		(pn)-1

Table 2. Variances for a two-factor fully-nested design

Variance	Expression	Degrees of freedom
Repeatability variance, S^2_r	MS_E	(pn)-1
Between-run variance, S^2_{run}	$\frac{MS_{run} - MS_E}{n}$	
Run-different intermediate variance, S^2_I	$S^2_r + S^2_{run}$	

The s_{SL} or the now called $s_{I_{SL}}$ is then calculated according to equation (3)

$$s_{I_{SL}} = \sqrt{\left(\frac{s_r^2}{n_r p_r} + \frac{s_{run}^2}{p_r}\right) + \left(\frac{s_r^2}{n\,p} + \frac{s_{run}^2}{p}\right)} \qquad (3)$$

As n_r and p_r are the number of replicates and the number of runs performed over the unknown sample, both are usually equal to 1. n and p are the number of replicates and runs used in the experimental design (Figure 3), so it becomes even simpler to calculate $s_{I_{SL}}$ from equation (4):

$$s_{I_{SL}} = \sqrt{s_r^2\left(1+\frac{1}{n\,p}\right)+s_{run}^2\left(1+\frac{1}{p}\right)} \qquad (4)$$

The value obtained is substituted in equations (1) and (2) for each analyte. The effective number of degrees of freedom of the Student-t test must be computed using the Satterthwaite [14] approach.

Results and discussion

Following the experimental design shown in Figure 3, two operators twice analysed two different validation standards for 22 days, thus leading to 88 runs. From these analyses performed at the specification limits of the surfactant and mineral oil, both responses were recorded and, from the standard deviations in intermediate precision conditions, the cut-off and screening responses were calculated (Table 3).

Table 3. Variances, effective degrees of freedom, and for the surfactant and the mineral oil. All values are calculated in response terms.

	Surfactant	Mineral oil
Mean response at the specification limit, $\bar{r}_{SL}$ (mV)	2.6	0.53
Repeatability variance, S^2_r	4.5x10-4	1.3x10-4
Between-run variance, S^2_{run}	3.8x10-2	1.4x10-3
Run-different intermediate variance for the specification limit, $s^2_{I_{SL}}$	3.9 x10-2	1.5 x10-3
Effective degrees of freedom, ν_{eff}	89	103
$r_{cut\text{-}off}$(α=1%) (mV)	2.1	0.62
$r_{screening}$ (α=1%, β=10%) (mV)	1.9	0.67

We can estimate the performance parameters by taking into account the decision values shown in table 3. Sensitivity was assessed by measuring 20 times a sample with a concentration of surfactant below 0.39 g/l (0.099 g/l). All 20 responses recorded were below $r_{screening}$ (i.e. 1.89 mV), so the sensitivity rate at this concentration level was 100%. Similarly, specificity was estimated from a sample with a surfactant concentration above 0.39 g/l (0.619 g/l) and a mineral oil concentration below 1g/l (0.707 g/l). All of the 20 responses recorded showed a response value for the surfactant above $r_{screening}$ (i.e. 1.89 mV) and a response value below $r_{screening}$ (i.e. 0.67 mV) for the mineral oil. This implies a specificity rate of 100% at both levels of concentration.

The unreliability region is the interval of responses or concentrations where the probability of obtaining false responses or results obtained is higher [6]. In the present case, this region is placed between the specification limit response and the screening response of the analyte because is where these probabilities of committing false responses are higher. Once calculated these two response values, (i. e. specification limit response and screening response), the unreliability region is estimated easily. For the surfactant, the unreliability region lies between the response values of 2.61 mV (specification limit) and 1.89 mV (screening response). For the mineral oil content, the unreliability region lies between response values of 0.53 mV (specification limit) and 0.67 mV (screening response). In both cases, within the unreliability region the probability of a type I error is the most important.

False positive and false negative rates are interesting in the present application because the decisions depend on them. For the mineral oil, the false positive rate is assessed using a sample that contains 0.707 g/l of mineral oil. None of the 20 responses recorded provided a value above $r_{screening}$ = 0.67 mV, which is a false positive rate of 0% at this concentration level. In this case, there is just one analyte to provide the false response rate. If we consider both analytes, several situations arise:

a) a false positive result for the mineral oil content but a true negative result for the surfactant content, which means that the bath can still be used since there is enough surfactant,

b) a false positive result for the surfactant but a true negative result for the mineral oil, which means that the bath can still be used if a small amount of surfactant is added,

c) both results are false positives, which means that the degreasing bath must be replaced. This situation will not happen often because the probability of a type I error has been set at 1%,

d) both results are true negative, which means that the bath can still be used.

A similar situation occurs with regard to the false negative rate since it is assessed by measuring a sample with 1.246 g/l mineral oil but a rather low concentration of surfactant. Twenty samples were measured but no response recorded was below $r_{cut-off}$ =0.62 mV, which means that the false negative rate was 0%. Again, if we consider both analytes (mineral oil and surfactant), several situations arise:

a) a false negative result for the mineral oil but true positive result for the surfactant, which means that, even though it is falsely assumed that there is not enough mineral oil, the bath can still be used if a small amount of surfactant is added.
b) a false negative result for the surfactant but true positive result for the mineral oil. We can decide to longer use the bath if removing a small part of the mineral oil on the surface.
c) a true positive result for both analytes, which means that the degreasing bath must be replaced.
d) a false negative result for both, which means that we can continue to use the bath if we add more surfactant. This situation will not happen often because the probability of a type II error is set at 10%.

To properly validate this method, we analyzed ten samples provided by a specialized industry. These samples were collected for 5 days and every 12 hours from a degreasing bath with a lifetime of one week.

Table 4 shows the results for mineral oil content measured with the reference method of analysis [15] and with the qualitative method. We can see that, with the reference method, all the results except one were clearly negative. Note that the mineral oil concentration of the sample with the positive result was close to the one corresponding to $r_{screening}$. When we analyse the samples with the qualitative method, a sample is positive if the instrumental response r_i is higher than the $r_{screening}$. On the other hand, a sample is negative whenever the instrumental response of the mineral oil is lower than the corresponding $r_{screening}$. As we can see, all the samples analysed—including the one that was positive with the reference method—provided negative results. This is therefore a false negative result and is acceptable if we consider that its mineral oil concentration (1.31 g/l) is extremely close to the corresponding concentration of the $r_{screening}$ (1.246 g/l). The method of analysis therefore performed accurately with respect to mineral oil.

Table 4. Results of the analysis of the real samples using the reference method and the qualitative method.

Mineral oil (reference method)	Mineral oil *ri* (qualitative method of analysis)	Final result
Negative	0.25	Negative
Negative	0.22	Negative
Negative	0.37	Negative
Negative	0.40	Negative
Negative	0.52	Negative
Positive	0.61	Negative
Negative	0.51	Negative
Negative	0.46	Negative
No information	0.50	Negative
No information	0.47	Negative

Conclusions

We have described how to use the statistical intervals in the validation procedure of an innovative qualitative method of analysis. The screening method we have validated determines the exhaustion of a degreasing bath, which is used in the automobile industry. Two components were considered (the content of mineral oil and the content of surfactant) in order to decide whether the bath should be replaced.

The statistical intervals are defined in response terms and for both measurands simultaneously. As the specification limit is considered in terms of response, the one-sided prediction bounds are defined around the corresponding responses at the specification limit concentration because the probabilities of committing a type I error and a type II error are considered. On the basis of the $r_{screening}$ responses for the two analytes, the sample is considered positive or negative. When the two responses are combined, however, the considerations may be different.

Our results, obtained with a set of ten real samples, show that the method classified correctly at low concentrations of mineral oil and close to the concentration value for the specification limit. In the region near the concentration of the $r_{screening}$, however, one false negative result was obtained. No information is available on the surfactant content, so this cannot be checked.

Although the validation procedure considered only two components of the degreasing baths, it can be extended to the content of alkaline salts provided the method of analysis is suitable for these analytes. These salts are another valid indicator for the replacement of the degreasing bath.

Future proposals are to perform the validation study at other concentrations of these analytes and to determine robustness and ruggedness. Control charts are also a feature to consider in the future.

Acknowledgments

The authors acknowledge economic support from the MCyT (projects nº BQU2003-500 and BQU2003-1142). Esther Trullols would like also to thank the URV for providing a doctoral fellowship. This work was also supported by grant CTQ2004-01220 from the MCyT. Special thanks should be given to Dr. A. Solans of Henkel Iberica for collaborating on method validation and supplying specific chemical standards and bath samples.

References

[1] International Organisation for Standardization, Quality management systems. Fundamentals and Vocabulary, ISO 9000, Geneva, Switzerland 2005.

[2] International Organisation for Standardization, General requirements for the competence of testing and calibration laboratories, ISO 17025, Switzerland 2005.

[3] EURACHEM, ‘ The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics’ EURACHEM Secretariat, Teddington, Middlesex, 1998.

[4] 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results.

[5] P. Feldsine, C. Abeyta and W. Andrews, J. AOAC Int., 85 (2002) 1187.

[6] A. Rios, D. Barcelo, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhart, R. W. Stephany, A. Townshend, A. Zschunke, M. Valcarcel, Accred. Qual. Assur. 8 (2003) 68.

[7] R. Song, P. C. Schlecht and K. Ashley, J. Hazard. Mater., 83 (2001) 29.

[8] E. Trullols, I. Ruisánchez, F. X. Rius, M. Òdena and M. T. Feliu, J. AOAC Int., 87 (2004) 417.

[9] S. L. R. Ellison, S. Gregory and W. A. Hardcastle, Analyst, 123 (1998) 1155.

[10] E. Trullols, I. Ruisánchez, F. X. Rius and J. Huguet, Trends Anal. Chem., 24 (2005) 516.

[11] A. Pulido, I. Ruisánchez, R. Boqué, and F. Xavier Rius, Anal. Chim. Acta, 455 (2002) 267.

[12] E. Aguilera-Herrador, R. Lucena, E. Trullols, S. Cárdenas and M. Valcárcel, Anal. Chim. Acta, accepted for publication.

[13] J. Lavoué, D. Bégin and M. Gérin, Ann. Occup. Hyg., 2003, 47, 441.

[14] J. M. Charlesworth, Anal. Chem. 50 (1978) 1414.

[15] G. J. Hahn and W. Q Meeker, Statistical Intervals. A Guide for Practitioners, John Wiley & Sons, New York, 1991.

[16] S. Kuttatharmmakul, D. Luc Massart and J. Smeyers-Verbeke, Anal. Chim. Acta, 391 (1999) 203.

[17] F.E. Satterthwaite, Psychometrika, 6 (1941) 309.

[18] EPA Methodd 1664. Available on-line at http://www.epa.gov

4.5 REFERENCES

[1] J. D. Miller, M. E. Savard, A. Sibilia and S. Rapior, Mycologia 85, **1993**, 385.

[2] M. W. Trucksess and D. E. Koeltzow, *Evaluation and application of immunochemical methods for mycotoxins in food*, in J. O. Nelson, A. E. Karu and R. B. Wong (Eds.); *Immunoanalysis of agrochemicals*, *vol. 586* ACS Symposium Series, Washington D. C., **1995**.

[3] E. Anklam and J. Stroka, Trends in Anal. Chem. 21, **2002**, 90.

[4] E. Anklam and J. Gilbert, Trends in Anal. Chem. 21, **2002**, 468.

[5] S. de Saeger, L. Sibanda, A. Desmet and C. Van Peteghem, Int. J. Food Microbiol. 75, **2002**, 135.

[6] Commission Regulation (EC) No 257/2002. Official Journal of the European Community, 12.2.2002, No. L 041, pp12-15. Available at http://europa.eu.int/eur-lex/

[7] Commission Regulation (EC) No 472/2002. Official Journal of the European Community, 12.3.2002, No L 75/18. Available at http://europa.eu.int/eur-lex/

[8] Center for Food Safety and Applied Nutrition. U. S. Food and Drug Administration.
Available at http://www.cfsan.fda.gov/list.html

[9] Food and Agriculture Organization of the United Nations. *Appendix XI: Proposed Draft Revised Sampling Plan for Total Aflatoxins in Peanuts Intended for Further Processing*.
Available at:
http://www.fao.org/docrep/meeting/005/Y0474E/y0474e2i.htm#TopOfPage

[10] Commission Regulation (EC) No 27/2002. Official Journal of the European Community, 13.3.2002, No L 75/44. Available at http://europa.eu.int/eur-lex/

[11] Official Methods of Analysis (2000) 17th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Methods 990.33 and 991.31

[12] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-Verbeke, *Data Handling in Science and Technology, Handbook of Chemometrics and Qualimetrics: Part A, vol. 20A,* Elsevier Science, Amsterdam, 1997

[13] R-Biopharm Rhone-Ltd. Available at http://www.r-biopharmrhone.com/pro/afla/afla1.html

[14] T 0429 Peanut Butter. Central Science Laboratory Proficiency Testing Group. Sand Hutton (York) YO41 1LZ-United Kingdom. Available at http//:www.fapas.com

[15] V. Williams, A. Gershon and P. A. Brunell, *J. Infect. Dis. 130*, 1974, 669.

[16] Microbiology Leicester. Available at http://www-micro.msb.le.ac.uk/

[17] M.H.V. van Regenmortel, C.M. Fauquet, D.H.L. Bishop, E.B. Carstens, M.K. Estes, S.M. Lemon, J. Maniloff, M.A. Mayo, D.J. McGeoch, C.R. Pringle and R.B. Wickner, *Virus Taxonomy: classification and nomenclature of viruses: Seventh Report of the International Committee on Taxonomy of Viruses,* Academic Press, San Diego, USA, 2000.

[18] A. M. Arvin, *Clin. Microbiol. Rev. 9*, 1996, 361.

[19] A. M. Arvin, *Varicella-Zoster virus,* in B. N. Fields, D. M. Knipe and P. M. Howley (Eds.); *Fields virology,* Lippincott-Raven, 3rd. ed., Philadelphia, Pennsylvania, 1996.

[20] J. L. Pérez, A. García, J. Niubo, J. Salva, D. Podzamczer and R. Martin, *J. Clin. Microbiol. 32*, 1994, 1610.

[21] S. P. Steinberg and A. A. Gershon, *J. Clin. Microbiol. 329*, 1991, 1527.

[22] S. Bassion, *Immunological reactions,* in L. A. Kaplan and A. J. Pesce (Eds.); *Clinical Chemistry. Theory, analysis, correlation,* Mosby, 3rd ed., St. Louis, Missouri, 2000.

[23] M. Myers, *J. Infect. Dis. 140*, 1979, 229.

[24] M. Schmidbauer, H. Budka, P. Pilz, T. Kurata and R. Hondo, *Brain 115*, 1992, 383.

[25] A. Criado, S. Cárdenas, M. Gallego and M. Valcárcel, *J. Chromatogr. B 792,* 2003, 299.

[26] R. Lucena, S. Cárdenas, M. Gallego and M. Valcárcel, *Anal. Chim. Acta, 530,* 2005, 283.

[27] B. Santos, A. Lista, B. M. Simonet A. Ríos and M. Valcárcel, *Electrophoresis 26*, 2005, 1567.

[28] B. Haase, M. Stiles, T. Haasner and A. Walter, *Surface Engineering 15*, 1999, 242.

[29] Ingurumen Jarduketarabo Sozietate Publikoa (IHOBE), *Gestión eficaz de aceites lubricantes y fluidos hidráulicos*, 2002. Available at http://www.ihobe.es

[30] R. E. Doherty, *Environmental forensics 1*, 2000, 69.

[31] S. Hellweg, E. Demou, M. Scheringer, T. E. McKone and K. Heingerbühner, *Environ. Sci. Technol., 309*, 2005, 7741.

[32] U. S. Environmental Protection Agency (EPA), *Guide to Industrial Assessment for Pollution Prevention and Energy Efficiency*, EPA/625-/R-99/003, Cincinatti, OH, 2001.

[33] PROQUIMIA.
Available at http://www.proquimia.com/ingl/docs/proquimia1.pdf. 2005.

[34] J. Lavoué, D. Bégin and M. Gérin. *Ann. occup. Hyg.*, *47*, 2003, 441.

[35] Commission Directive 2000/76/EC of 4th. December 2000, *Incineration of Waste*, L332/91, 2000.

[36] Environemtal Protection Agency (EPA), Method 1664, Revision A: n-Hexane Extractable Material (HEM; Oil and Grease) and Silica Gel Treated N-Hexane Extractable Material (SGTHEM; Non-polar Material) by Extraction and Gravimetry.

[37] E. Aguilera-Herrador, R. Lucena, E. Trullols, S. Cárdenas and M. Valcárcel, *Anal. Chim. Acta*, accepted for publication.

[38] H. S. Park and C. K. Rhee, *J. Chromatogr. A, 1046*, 2004, 289.

5. ROBUSTNESS IN QUALITATIVE ANALYSIS

5.1 INTRODUCTION

The practical applications presented in the previous chapter have shown how to estimate some of the most common performance parameters of three particular methods of qualitative analysis. Although the general definition of method validation may involve assessing several performance parameters, the fact is that only a few are evaluated in a first approach to the validation process (e. g. sensitivity, specificity and false response rates, the unreliability region and, in some cases, the detection limit. Thus, parameters that may be important in some cases are left undetermined in others. This is what often happens as far as robustness and ruggedness are concerned.

Theoretically, no method of analysis should show important differences in its results when small changes are made to the experimental conditions. This property, known as robustness, is often confused or used indistinctly with ruggedness, which has a very similar meaning. Ruggedness refers to the changes that are observed in the response, but when external operation conditions are changed (i. e. operator, laboratory or equipment) [1-5].

Robustness and ruggedness are usually studied with quantitative methods of analysis. Practitioners have several helpful documents which discuss how to perform robustness and ruggedness studies. These documents are not only practical descriptions [6-12], but also references from institutions involved in method validation [13].

The AOAC INTERNATIONAL does not consider ruggedness as a formal part of the validation process, as is stated in the Methods Committee Guidelines for “Validation of Qualitative and Quantitative

Food Microbiological Official Methods of Analysis" [14]. Although it is not compulsory, submitting the method of analysis to the AOAC® *Official Methods*^{SM} Program (OMA) may provide valuable information.

An example of an indistinct use of the terms robustness and ruggedness is provided in the EURACHEM Guide "Fitness for Purpose of Analytical Methods" [15]. They are both used to refer to the performance characteristic related to the comparability of results within one method of analysis in different conditions.

The European Commission, when describing the performance of analytical methods and the interpretation of results [16], only uses the term ruggedness. It is defined as "the susceptibility of the method to changes in the experimental conditions" and it is evaluated in the validation procedure, which is presented as a two-stage process. In the first stage, a particular set of performance parameters should be unequivocally determined. These parameters are specificity, trueness, stability and calibration curves. Ruggedness is included in this first stage and it is also called applicability in this particular situation. It is evaluated by introducing reasonable minor variations, which should match usual deviations, in factors such as sample pre-treatment, clean-up and analysis. It seems that the term robustness, which is defined in the second paragraph, would fit this idea, i.e robustness to minor changes. Then, the results should be interpreted. The second stage of the validation procedure depemds on the intrinsic characteristics of the method of analysis. Other performance parameters such as recovery, repeatability or reproducibility, among others, should be determined. In this stage, ruggedness to the so-called major changes (i. e. different species, matrices or sampling conditions) should be evaluated. This concept is the term defined in the second paragraph as ruggedness. The Youden approach is the statistical tool frequently used to determine the corresponding effects.

In this sense, qualitative methods of analysis should also be evaluated for robustness and ruggedness. However, so far validation guidelines have neglected this parameter. The aim of this chapter is to present a procedure that assesses the robustness of a qualitative method of analysis that provides a YES/NO instrumental response.

As a starting point, studies of the robustness of quantitative methods of analysis are very valuable. However, the data will then be analysed differently because of the peculiar characteristics of the binary type result.

5.2 ROBUSTNESS IN QUALITATIVE ANALYSIS: A PRACTICAL APPROACH

Trends in Anal. Chem. In press

E. Trullols, I. Ruisánchez, E. Aguilera[a], R. Lucena[a], S. Cárdenas[a] and M. Valcárcel[a].

Departament de Química Analítica i Química Orgànica. Universitat Rovira i Virgili. C/ Marcel·lí Domingo s/n,. 43007 Tarragona (Spain)
[a]Department of Analytical Chemistry, Marie Curie Annex Building, Campus de Rabanales, University of Córdoba, 14071 Córdoba (Spain)

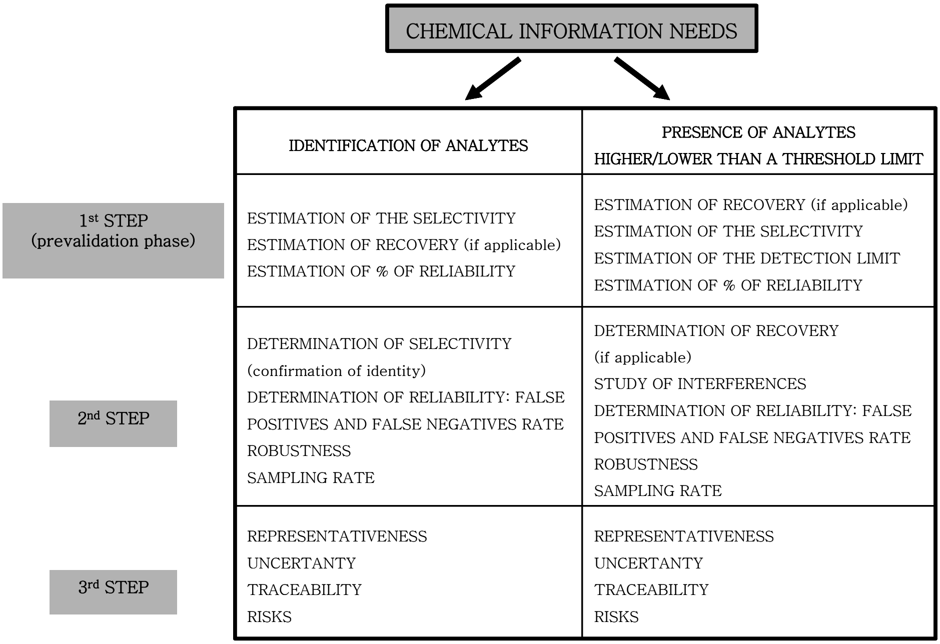
Abstract

The growing importance of qualitative information as output in nowadays analytical laboratories in response to client's demands is unquestionable. Therefore, the number of reliable, validated qualitative methods available for their implementation in routine laboratories is increasing in the same way. Unfortunately, no metrological support for this type of measurement process is yet available. In this paper, a practical approach about the assessment of the robustness of a qualitative method is presented. The proposed procedure is based on the selection of the critical variables and the estimation of the reliability and false positives and false negatives rates. The qualitative procedure selected is an automated configuration developed for monitoring the degree of exhaustion of alkaline degreasing baths based on the total oil/grease and surfactant contents. The study was carried out at two concentration levels for each family of compounds.

Keywords: Robustness; ruggedness; qualitative analysis

1. Analytical properties of qualitative analysis

The quality indicators of an analytical process are the so-called analytical properties. They have been mainly used to characterize a quantitative result and therefore the associate (bio) chemical measurement process. As it is the case with other facets of qualitative analysis, few approaches to define/adapt the analytical properties to qualitative test methods have been systematically carried out [1]. The analytical properties in qualitative analysis can also be ranked into three categories (capital, basic and productivity related) existing, as in quantitative analysis, basic, contradictory and complementary relationship among them. Some adaptations of the quantitative analytical features are required taking into account the peculiarity of the qualitative binary response and the test methods. Therefore, reliability defined as a combination of accuracy and precision, is used in qualitative analysis [2] as capital analytical property (together with representativeness) and characterized the yes/no binary response. Reliability depends on sensitivity, selectivity and robustness of the method. The dependence on sensitivity and specificity is not a mathematical function but a conceptual one. Reliability includes the information regarding the results which are proved to be true. Therefore, there are included both the results truly given as positive, i.e. sensitivity; and the results truly given as negative, i.e. specificity. Then, it is expressed as a rate. Also, the classical concept of uncertainty should be replaced by unreliability in this context. This analytical property defines an interval around the cut-off or threshold limits where qualitative errors (false positives and false negatives) are produced. Finally, the analytical properties are different depending on the type of qualitative analysis being under consideration, viz. analyte identification or sample qualification/classification. Fig. 1 shows a general procedure for the determination of the suitability of a qualitative method to the chemical information needs posed by the clients through the estimation/determination of the analytical features. In a first step it is necessary to establish the fitness-for-purpose of the qualitative method through a rough estimation of capital and basic properties. As can bee seen, one of the properties to be estimated regardless the type of qualitative analysis employed is the robustness.



CHEMICAL INFORMATION NEEDS

	IDENTIFICATION OF ANALYTES	PRESENCE OF ANALYTES HIGHER/LOWER THAN A THRESHOLD LIMIT
1st STEP (prevalidation phase)	ESTIMATION OF THE SELECTIVITY ESTIMATION OF RECOVERY (if applicable) ESTIMATION OF % OF RELIABILITY	ESTIMATION OF RECOVERY (if applicable) ESTIMATION OF THE SELECTIVITY ESTIMATION OF THE DETECTION LIMIT ESTIMATION OF % OF RELIABILITY
2nd STEP	DETERMINATION OF SELECTIVITY (confirmation of identity) DETERMINATION OF RELIABILITY: FALSE POSITIVES AND FALSE NEGATIVES RATE ROBUSTNESS SAMPLING RATE	DETERMINATION OF RECOVERY (if applicable) STUDY OF INTERFERENCES DETERMINATION OF RELIABILITY: FALSE POSITIVES AND FALSE NEGATIVES RATE ROBUSTNESS SAMPLING RATE
3rd STEP	REPRESENTATIVENESS UNCERTANTY TRACEABILITY RISKS	REPRESENTATIVENESS UNCERTANTY TRACEABILITY RISKS

Figure 1. Proposal of a general, flexible action list to determine the performance characteristics of a qualitative method

2. Robustness versus ruggedness

The robustness of an analytical method is an estimation of its capability to remain unaffected by small but deliberated changes in method variables. It provides a qualitative estimation of its reliability when it performs analyses in standard conditions [3]. Through an experimental design, it is possible to define allowable limits for critical parameters. There is another term with a similar meaning that is often used when referring to robustness. Ruggedness is defined as 'the degree of reproducibility when the procedure is subjected to changes in external conditions such as different laboratories, analysts, instruments [4]. Although the difference may be slight, both terms must be employed in the right situation. Then, if the variables considered belong to the method of analysis, the study will check robustness. On the contrary, if the variables studied are of environment nature (e.g., laboratory temperature, analyst, brand of the reagents), ruggedness will be examined [5-7].

It is necessary to use both terms correctly because they represent such different features of the method of analysis: robustness is related to the practicability and to the stability of the method of analysis using as a starting point the intrinsic variables; and ruggedness is related to the inter-laboratory method transferability [6, 7].

Robustness and ruggedness testing should be carried out during or nearly at the end of method development stage [8, 9]. The reason is that they can help in evaluating the precision of the analytical method [10]: they identify critical factors or variables, which may have influence in the performance of the analytical method. Then, they are crucial for the subsequent validation of the analytical method.

The distinction made between robustness and ruggedness hardly affects the design of a robustness and/or ruggedness study. In any case, it involves the selection of the suspected sources of variation, the experimentation, the estimation of the effects and the statement of the conclusions. The experimental domain should include the values of the variables when any change in the experimental conditions of the analysis (e.g., different equipment, different analyst, and different value of an inherent variable of the method of analysis) occurs. So the most common values for the variables under study are included in this experimental domain. Once the results are obtained, the adequate conclusions are inferred i.e. whether any change in the equipment, the analyst or any particular condition of the method of analysis will affect the final result or not.

Robustness in qualitative analysis is an analytical property of the qualitative test method, as in quantitative methods of analysis, rather than of the binary response, whose ultimate purpose is to define the experimental weakness of the qualitative method by defining what variables are critical to ensure the reliability of the responses. This property is very relevant as test methods are usually handled by unskilled and even different personnel, being therefore crucial to guarantee that the response obtained does not depend on external factors; but also on intrinsic ones, such as the stability of biochemical and immunoassay reagents, widely used for this purpose. Robustness also depends on the concentration of the analytes as experiments performed within the unreliability region will show higher influence of experimental factors and lower the robustness as result.

A practical approach of these theoretical considerations is the subject matter of this article.

3. Robustness studies in qualitative analysis. A case study

Robustness and ruggedness are analytical features which should be evaluated whenever it is necessary. However, these studies have been carried out commonly over quantitative methods of analysis due to their more extensive development. The growing importance of qualitative methods of analysis suggests the evaluation of these performance parameters in addition to the capital ones. Nonetheless, the main problem is that this process has not been systematized enough. As a starting point, the difference between the type of result provided by a quantitative method of analysis (i.e., a numerical value) and a qualitative method of analysis (i.e., binary outcome, YES/NO) should be considered. Then, this binary nature of the result is crucial in the subsequent data treatment and conclusions statement about the different factors or variables examined. The data evaluated are not the changes in the final numerical results but in two capital performance parameters such as reliability and false response rates [1]. These performance parameters, which are closely related, are very important because they reveal how good the method of analysis classifies the samples. Robustness and ruggedness move on the same direction as reliability: the closer to the decision value of the method the study is performed (i.e., lower reliability and higher false response rates), the lower the robustness and ruggedness are [1].

3.1. Robustness of an autoanalyzer for monitoring the exhaustion of alkaline degreasing baths

The qualitative method selected for this study has been previously described by our research group [11]. It consists of a manual liquid-liquid extraction of the total surfactant and oil/grease contents in n-hexane followed by the injection of a 250 µl aliquot of the extract in a continuous flow manifold. The sample passes through a silica column carried by a stream of n-hexane where the surfactant is retained while the oil/grease is driven to the evaporative light scattering detector. Surfactants are afterwards eluted by means of an ethanol stream. It was optimized and validated for monitoring the degree of exhaustion of industrial degreasing baths considering the global level of two families of compounds. If the amount of surfactant is high enough and/or the concentration of oil/grease is not so high, the degreasing bath can be keep on using whereas for a low surfactant concentration and/or high amount of oil, the bath should not be used any longer. The decision is made according to a threshold concentration fixed by the clients (in this case a surface technology industry): 0.39 g/l for the surfactant and 1 g/l for the mineral oil. The four possibilities derived from the combination of these two parameters are depicted in Fig. 2. The study will consider the change in the reliability and false response rates when different experimental conditions inherent to the method of analysis are varied. Therefore, it is a *robustness* study. Reliability and false response rates are checked for the two families of compounds individually at two different concentration levels for each index.

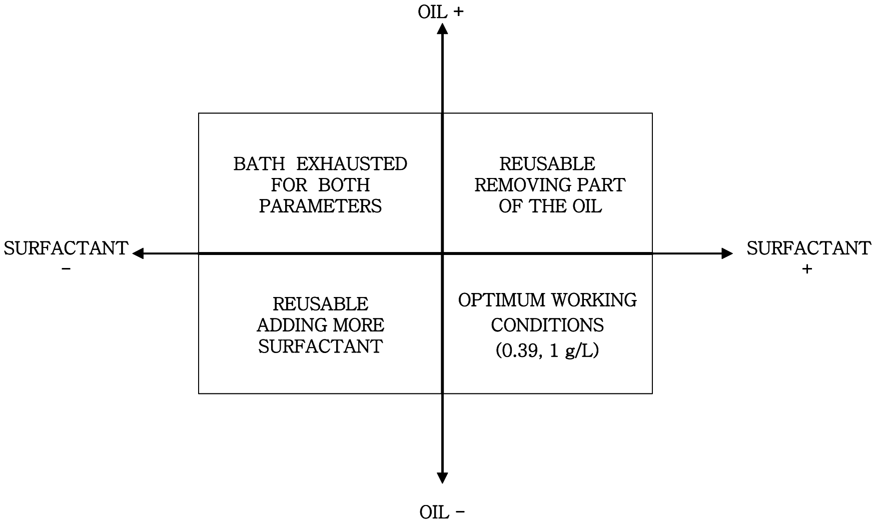

OIL +
BATH EXHAUSTED FOR BOTH PARAMETERS
REUSABLE REMOVING PART OF THE OIL
SURFACTANT -
SURFACTANT +
REUSABLE ADDING MORE SURFACTANT
OPTIMUM WORKING CONDITIONS (0.39, 1 g/L)
OIL -

Figure 2. Control plot for internal quality control of the ageing process of an alkaline degreasing bath as regards the surfactant and oil contents

In general terms, the study will include: a selection of the variables (or factors), a selection of the number of levels for each variable (usually 2), selection of the best experimental design, establishment of the response value that evaluate the changes in the factors and experimentation, calculation of the effects, i.e. individual and interaction between factors, and statement of the conclusions.

The automated configuration used for global indices determination consists of a high pressure pump, a six-port injection valve, a silica sorbent column and an evaporative light scattering detector. The signal measured is the light scattered by the analyte particles, after solvent evaporation which provides peak height (mV) as a response which depends on the analyte concentration [12]. Concerning sample preparation, it includes a liquid-liquid extraction, which entails several critical factors that affect the recoveries of the analytes. This effect is minimised by performing several sample preparations consecutively and by mixing extracts. Doing so, the probable difference in the values of every sample recovery is minimised. In addition to this, the automated system involves several intrinsic variables which can influence the final result, such as the flow rate of the eluents, the pressure of the nebulizing gas in the detector, the photomultiplier gain, the temperature of the nebulizing chamber or the post-time required for system conditioning between runs.

To achieve the objective proposed, it is necessary to consider the previous information gathered during the optimization [11] and during the validation [13] of the method of analysis. These two stages which involve a considerable experimental part allow the identification of the main variables which directly affect the performance of the method of analysis. The variables to be included are the detector gain, the temperature of the nebulizing chamber and the post-time between analyses. As it has been described elsewhere [11], the detector gain is different depending on the analyte measured, being 350 mV and 550 mV the optimum values for oil/grease and surfactant, respectively. This parameter is changed during analyses and thus, it may lead to photomultiplier gain values slightly different from the right ones. A variation of 5 mV above and below each value is considered as suitable. Although the temperature is constant during the analysis, some fluctuations have been observed. The analyses are performed at a nebulizing chamber temperature of 75 ºC and the variations observed usually do not exceed 3 ºC. Therefore this interval is considered for the robustness study. Similarly, some variations observed in the peak height are due to different post-times programmed and a variation of 4 min is considered for this variable in this study. The definitive levels (optimum and tolerated interval) for each factor are summarized in Table 1.

Table 1. Levels of the factors chosen for the robustness study

Variable	Optimum level value	Tolerated interval
Photomultiplier gain (mV)	350 mineral oil 550 surfactant	345-355 mineral oil 545-555 surfactant
Nebulizing chamber temperature (ºC)	75	72-78
POST-TIME (min)	5	3-7

3.2. Experimental design

In order to cover the experimental domain defined in Table 1, the experiments should be carried out following a pre-set experimental design. Due to the information exposed in the previous section, a full factorial design is the best option because there are not so many variables and the number of levels is two for each variable. Considering the 3 factors under study and two levels for each one, the total number of experiments is 8 and they are summarized in Table 2. The interactions between factors will be also evaluated.

Table 2. Matrix of experiments following a 2-factor fully factorial experimental design. A is the photomultiplier gain, B is the nebulizing chamber temperature and C is post-time.

Experiment	A	B	C	AB	AC	BC	ABC
1	−	−	−	+	+	+	−
2	+	−	−	−	−	+	+
3	−	+	−	−	+	−	+
4	+	+	−	+	−	−	−
5	−	−	+	+	−	−	+
6	+	−	+	−	+	−	−
7	−	+	+	−	−	+	−
8	+	+	+	+	+	+	+

The results obtained from these experiments will be decisive for the subsequent decision about their influence in the quality parameters chosen, i.e. reliability and false response rates. As they are rates, their calculation involves a considerable number of experiments and results. This feature must be also considered before performing the experimental part, so each experiment will be carried out 20 times. Doing so, any change of 5% in any rate will be noticed. The experiments are performed according to the experimental conditions specified in Table 2. The analyses are carried out during eight days. In order to avoid any possible systematic error, a fraction of 5 analyses for 4 different experiments is performed daily. Then, each day, 20 analyses are carried out but every 5 analyses the experimental conditions are changed.

The samples are chosen according to different degrees of exhaustion, so different results should be achieved. The two samples used contain different concentration levels of mineral oil and surfactant: the first sample contains: 0.099 g/l surfactant and 1.246 g/l mineral oil and the second sample contains 0.619 g/l surfactant and 0.707 g/l mineral oil.

3.3. Definition of the initial conditions

The method of analysis provides an instrumental response which is converted into a binary final result: YES the bath is exhausted and therefore, must be replaced; or NO the bath is not exhausted and replacement is not required yet. The decision of YES or NO is taken according to the instrumental response obtained for each analyte. A first approach, proposes a comparison of the response obtained with the response value corresponding to the threshold concentrations. However, in the present application the probabilities of error (type I and type II) are considered and then a new response value arises. This is the screening response or $r_{screening}$ [13]. According to this, the comparison is performed between the response of the sample and the screening response.

As it has been previously described, the two analytes measured correspond to two different families of compounds. Although it is more interesting to consider both families simultaneously, they will be examined separately on account of their different behaviour in the analytical (or detection) system. The oil fraction is not retained on the silica sorbent while the signal for the surfactant fraction appears between 4.75 and 5 min after the automated system starts as the likely result of its interaction with the sorbent column. Then, the variables of the method of analysis may affect in a different manner these two global indexes. Bearing this in mind, the experiments are performed with the two samples and the responses are examined for each analyte. The screening responses considered for each analyte are summarized in Table 3.

Table 3. Mean responses and screening responses for the surfactant and the mineral oil [13]

	Mean response at the specification limit, $\bar{r}_{SL}$	$r_{screening}$ (β=10%)
Surfactant	2.6	0.53
Mineral oil	1.9	0.67

The reliability is defined as the ‘proportion of right answers provided by the qualitative method of analysis carried out independently on aliquots of the same sample [1]. It is calculated according to Equation /1/:

Reliability (%) = 100 %−FP (%)−FN (%) /1/

Then, false responses should be defined previously:

1) A false positive result (FP) is to state that the degreasing bath is exhausted, provided it is not exhausted yet and it must be replaced. Then:

Surfactant: $r_i < r_{screening}$ (1.9 mV) and mineral oil: $r_i > r_{screening}$ (0.67 mV)

2) A false negative result (FN) is to state that the degreasing bath is not exhausted when indeed it is. Then:

Surfactant: $r_i > r_{screening}$ (1.9 mV) and mineral oil: $r_i < r_{screening}$ (0.67 mV)

3.4. Robustness study using a real sample obtained from an exhausted bath

For sample 1 due to the concentration levels of the analytes: a) surfactant (0.099 g/l) and b) mineral oil (1.246 g/l), the surfactant measurements should provide a response below its $r_{screening}$ (1.9 mV) and the mineral oil should provide a response value above its $r_{screening}$ (0.67 mV).

The results regarding the reliability of both the surfactant and the mineral oil are summarized in Table 4. The experiments performed with the optimal conditions show a reliability of 100% for both families of compounds. The false positive rate is 0% for the surfactant and the false negative rate is also 0% for the mineral oil.

Table 4. Plan of experimentation carried out with the first sample. The reliability and the false response rates are calculated for the surfactant and the mineral oil.

Experiment	A (mV)	B (ºC)	C (min)	[a]Rel. (%)	[b]Rel. (%)	[c]F. positive rate (%)	[d]F. negative rate (%)
1	345-545	72	3	100	80	0	20
2	355-555	72	3	100	100	0	0
3	345-545	78	3	100	55	0	45
4	355-555	78	3	100	100	0	0
5	345-545	72	7	100	80	0	20
6	355-555	72	7	100	100	0	0
7	345-545	78	7	100	65	0	35
8	355-555	78	7	100	100	0	0

[a, c] Reliability and false positive rate for the surfactant and [b, d] Reliability and false negative rate for the mineral oil.

As it can be seen, reliability is 100 % for the surfactant. Then the method is robust as far as this family of compounds is concerned.

According to the reliability obtained for the mineral oil, the factors studied have an effect (*D_F*) on the responses. This effect must be estimated (Equation /2/)

$$D_F = \frac{\sum Y(+) - \sum Y(-)}{n} \qquad /2/$$

$Y(+)$ is the reliability at the upper level of factor F, i.e. all the experiments with (+) in Table 2. $Y(-)$ is the lower level of factor F, i.e. all the experiments with (-) in Table 2, and *n* is the number of experiments performed at each level of the factor under study.

Then, it is calculated for the three factors under study, i.e. photomultiplier gain, temperature, post-time and for the interaction between two and three factors (Table 5).

Table 5. Effects on the reliability for the three factors studied. Interactions between two and three factors are also showed.

Factor or interaction between factors	Effect (D_F)
Photomultiplier gain	30% = 0.3
Nebulizing chamber temperature	10% = 0.1
Post-time	2.5% = 0.025
Photomultiplier gain and nebulizing chamber temperature	10% = 0.1
Photomultiplier gain and post-time	-2.5% = -0.025
Nebulizing chamber temperature and post-time	2.5% = 0.025
Photomultiplier gain nebulizing chamber temperature and post-time	-2.5% = -0.025

Once the effect (*D_F*) of each factor is calculated, a t-test (Equation /3/) will conclude if the factor significantly affects the responses and the results:

$$t = \frac{\left|D_F\right| \times \sqrt{n}}{s_I \times \sqrt{2}} \qquad /3/$$

where s_I is the standard deviation computed in intermediate precision conditions. This value is 0.0388 for the mineral oil. D_F should not be used as a percentage but as the corresponding value between 0 and 1.

The calculated t-value is compared with the corresponding tabulated value for a specific level of significance (e.g., 95%) and degrees of freedom. The degrees of freedom are these associated to s_I and are computed using Satterthwaite [14] approach. For the mineral oil, this value is 103. The comparison between both t-values is summarized in Table 6.

Regarding the surfactant as Table 4 shows, reliability is 100%. Then, the changes in the experimental conditions have not affected the final results. Therefore, the calculations performed with the mineral oil are not necessary.

Table 6. Calculated and tabulated t-values related to the effects of the three factors studied and their interactions for the mineral oil.

Factor or interaction between factors	Calculated t-value	Tabulated t-value
Photomultiplier gain	10.92	1.66
Nebulizing chamber temperature	3.64	1.66
Post-time	0.911	1.66
Photomultiplier gain and nebulizing chamber temperature	3.64	1.66
Photomultiplier gain and post-time	0.911	1.66
Nebulizing chamber temperature and post-time	0.911	1.66
Photomultiplier gain nebulizing chamber temperature and post-time	0.911	1.66

The main conclusion is that both the photomultiplier gain and the temperature of the nebulizing chamber affect the reliability at this concentration level of mineral oil. In addition to these effects, the one corresponding to the interaction between both factors is also relevant. The effect of the post-time as well as the effect of the interactions between post-time and the other two factors is not significant. As it was expected, the interaction between the three factors is also not relevant. On the contrary, the system is robust at this concentration level of surfactant.

The effect on false response rates can be easily inferred, once the reliability has been studied. The results of false response rates for the first sample are also summarized in Table 4.

The effects calculated for the false negative response rate are the same as for the reliability as far as the mineral oil is concerned. Therefore, the same factors are relevant at this concentration level for the mineral oil.

3.5. Robustness study using a real sample obtained from a non-exhausted bath

For sample 2 due to the concentration levels of the analytes: a) surfactant (0.619 g/l) and b) mineral oil (0.707 g/l), the surfactant measurements should be above its $r_{screening}$ (2.6 mV) because higher responses are provided by true negative samples. Similarly, the mineral oil should provide a response value below its $r_{screening}$ (0.53 mV) because lower responses are given by true positive samples (Table 3). Table 7 summarizes the results obtained. As reliability is 100 %, false response rates are 0% and therefore not listed in the table. As for the first sample, the experiments performed under the optimal conditions show a reliability of 100% for the surfactant and the mineral oil. The false positive rate is 0% for the surfactant and the false negative rate is also 0% but for the mineral oil.

Table 7. Plan of experimentation carried out with the second sample. The reliability is calculated for the surfactant and the mineral oil.

Experiment	A (mV)	B (ºC)	C (min)	[a]Reliability (%)	[b]Reliability (%)
1	345-545	72	3	100	100
2	355-555	72	3	100	100
3	345-545	78	3	100	100
4	355-555	78	3	100	100
5	345-545	72	7	100	100
6	355-555	72	7	100	100
7	345-545	78	7	100	100
8	355-555	78	7	100	100

[a] surfactant [b] mineral oil

If reliability is 100% at this concentration level for both analytes, the false response rates are 0%. Then, it is not necessary to study the effect of the factors. The conclusion is that the method is robust at this concentration level of surfactant and of mineral oil.

4. Conclusions

The need of reliable information provided by qualitative methods of analysis has been crucial in the apparition of validation guidelines. However, not all quality parameters are considered in the proposed guidelines. One of these parameters is the robustness which is not very often considered when validating this particular group of methods of analysis. In this paper, we have presented an approach of how robustness studies could be performed as far as methods of analysis providing binary type responses is concerned. An automated configuration that measures the degree of exhaustion of alkaline degreasing baths based on two families of compounds has been chosen as the case study.

The particular characteristics of the binary type response have defined how to design the robustness study. As for the case of quantitative methods of analysis, the robustness study proposed also involves an experimental design to evaluate the different factors or variables previously chosen. However, several concentration levels should be tested and the effects are calculated considering one or more performance parameters and not the numerical value of the response. The conclusions are also inferred by means of a t-test.

The robustness study is performed with the automated configuration at two different concentration levels of analytes. The effects of the variables photomultiplier gain, nebulizing chamber temperature and of their interaction are rather relevant at one concentration level of one family of compounds. Nonetheless, the method of analysis is robust for the other family of compounds at both concentration levels tested.

As robustness depends on the concentration level of the analyte studied, i.e. family of compounds, the two samples tested show significantly different results. The first sample comes from a bath containing oil/grease within the unreliability region. Then, the effect of the factors needs to be considered. However, the surfactant content is not within the unreliability region so the effect of the factors is not relevant. The second sample is from a clearly non-exhausted degreasing bath as long as the concentration of both analytes is far from the unreliability region.

Acknowledgements

This work was supported by grant CTQ2004-01220, BQU2003-500 and BQU2003-1142 from the MCyT. Esther Trullols would like also to thank the URV for providing a doctoral fellowship.

References

[1] S. Cárdenas, M. Valcárcel, Trends Anal. Chem. 24 (2005) 477.

[2] B. M. Simonet, A. Ríos, M. Valcárcel, Anal. Chim. Acta 516 (2004) 67.

[3] *ICH Q2A, CPMP/ICH/381/95*. ICH 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: Definitions and Terminology, 1994, (http//www.ifpma.org/ich1.html).

[4] The United States Pharmacopeia USP XXIII. Validation of Compendial Method, United States Pharmacopeia Convention, Rockville, 1995.

[5] D. R. Jenke, J. Liq. Chrom. & Rel. Technol. 19 (1996) 1873.

[6] L. Cuadros-Rodríguez, R. Romero, J. M. Bosque-Sendra, Crit. Rev. in Anal. Chem. 35 (2005) 57.

[7] R. Kellner, J.M. Mernet, M. Otto, M. Valcárcel, H.M. Widmer (eds), Analytical Chemistry, 2nd edition, Willey-VCH, 2004.

[8] D. Dadgar, P.E. Burnett, M.G. Choc, K. Gallicano, J.W. Hooper, Journal of Pharmaceutical and Biomedical Analysis 13 (1995) 89.

[9] H. Fabre, Journal of Pharmaceutical and Biomedical Analysis 14 (1996) 1125.

[10] J.A. Van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, P.J. Schoenmakers, M. Mulholland, Chemometrics and Intelligent Laboratory Systems 10 (1991) 337.

[11] E. Aguilera-Herrador, R. Lucena, E. Trullols, S. Cárdenas, M. Valcárcel, Anal. Chim. Acta Accepted for publication.

[12] J. M. Charlesworth., Anal. Chem. 50 (1978)1414.

[13] E. Trullols, I. Ruisanchez, E. Aguilera-Herrador, R. Lucena, S. Cárdenas, M. Valcárcel, Anal. Chim. Acta. Submitted.

[14] F.E. Satterthwaite, Psychometrika 6 (1941) 309.

5.3 REFERENCES

[1] The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for human use (ICH), Text on Validation of Analytical Procedures Q2A, Geneva, Switzerland, 1994. Available at http://www.ich.org/LOB/media/MEDIA417.pdf.

[2] The United States Pharmacopeia USP XXIII, Validation of Compendial Method, United States Pharmacopeia Convention, Rockville, 1995.

[3] D. R. Jenke, *J. Liq. Chromatogr. Relat. Technol. 19*, 1996, 1873.

[4] L. Cuadros-Rodríguez, R. Romero and J. M. Bosque-Sendra, *Crit. Rev. Anal. Chem. 35*, 2005, 57.

[5] R. Kellner, J.M. Mernet, M. Otto, M. Valcárcel, H.M. Widmer (eds), *Analytical Chemistry,* Wiley-VCH,2nd ed., Weinheim, 2004.

[6] Y. Vander Heyden, A. Nijhuis, J. Smeyers-Verbeke, B. G. M. Vandeginste and D. L. Massart, *J. Pharm. Biomed. Anal. 24*, 2001, 723.

[7] Y. Vander Heyden, K. De Braekeleer, Y. Zhu, E. Roets, J. Hoogmartens, J. De Beer and D.L. Massart. *J. Pharm. Biomed. Anal. 20,* 1999, 875.

[8] Y. Vander Heyden, M. Jimidar , E. Hund, N. Niemeijer, R. Peeters, J. Smeyers-Verbeke and D.L. Massart, J. Hoogmartens, *J. Chromatogr. B*, *845,* 1999, 145.

[9] A. Nijhuisa, H.C.M. van der Knaapa, S. de Jong and B.G.M. Vandeginste, *Anal. Chim. Acta 391,* 1999, 187.

[10] Y. Vander Heyden, F. Questier and D. L. Massart, *J. Pharm. Biomed. Anal. 18,* 1998, 43.

[12] L. Cuadros, R. Blanc, A. M. Garcıa Campana and J. M. Bosque Sendra, *Chemometrics Intell. Lab. Syst. 41,* **1998**, 57.

[13] L. M. B. C. Álvares-Ribeiro and A. A. S. C. Machado, *Anal. Chim. Acta 355,* **1997**, 195.

[14] W. J. Youden, E. H. Steiner, *Statistical Manual of the Association of Official Analytical Chemists,* Association of Official Analytical Chemists, Arlington, **1975**.

[15] P. Feldsine, C. Abeyta and W. H. Andrews, *J. AOAC Int. 85*, **2002**, 1187.

[16] EURACHEM, The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics, EURACHEM Secretariat, Teddington, Middlesex, UK, **1998**. Available at http://www.eurachem.ul.pt.

[17] Decision from the Commission. Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. CO (2002) 3044 final (12.08.02).

6. CONCLUSIONS

6.1 INTRODUCTION

The main contributions of this doctoral thesis are the theoretical approach to the validation of qualitative methods of analysis and the three practical applications. The conclusions drawn from these applications are presented in this chapter. Some suggestions for future research are also made.

6.2 CONCLUSIONS

1. The bibliographic review in the third chapter provided an overview of validation in qualitative analysis.

1.1 The values of the different operational performance characteristics (such as rapidity, ease of handling or economy), as well as the statistical values (such as the unreliability region, sensitivity or selectivity) favour the growing acceptance of qualitative methods of analysis.

1.2 However, the main problem is still the lack of harmony as far as terminology is concerned. Some terms are used rather imprecisely to refer to these methods and to designate their performance characteristics. Classifications are often made with different criteria.

1.3 Written standards are needed in the field of validation of qualitative methods: to help choose the best validation procedure, to standardise the nomenclature and to unambiguously define the performance parameters. Although some institutions participate in defining the main quality parameters, few of them have proposed validation schemes for qualitative methods of analysis. These schemes are usually addressed to manufacturers of commercial test kits, who require a quality assessment before marketing their products. However, the end user should also be able to carry out the validation processes proposed. In this respect, specific validation guidelines endorsed by renowned institutions would provide valuable support to this task.

1.4 The validation of a method of analysis should provide information about what the method must do for a particular analytical problem. In chemical analysis, in addition to the analytical problem, the intrinsic characteristics of the method of analysis chosen define the validation methodology to be followed. This situation is very common in qualitative methods of analysis; therefore, different protocols should be developed depending on the specific target.

2. Performance characteristic curves are useful for validating Aflacard B_1 because it provides a binary type result.

2.1 The plot of the probabilities of positive, negative or inconclusive results in the range of concentration levels of interest leads to the estimation of the performance parameters.

2.2 Although the qualitative method performs well in comparison with the confirmatory method, a bias towards false positive results is detected. This bias has been set by the manufacturer to avoid false negative results, and it can be either accepted or corrected. The correction, which means changing the experimental conditions, involves moving the unreliability region to higher concentration levels.

3. Statistical intervals are very useful tools for validating the methods of analysis that provide an instrumental response but a

binary type final result. However, statistical intervals can be used in different ways depending on the intrinsic characteristics of the method.

3.1 In those situations in which control samples are used, their probability distribution function must first be characterised. Two performance parameters (i. e. sensitivity and specificity rates) are estimated using the data obtained when measuring the control samples. The performance parameters related to false response rates are estimated using two particular samples because the response of the control samples is not near the unreliability region around the cut-off value.

3.2 The need to measure the responses of two analytes with a home-made autoanalyzer has also been overcome with statistical intervals. These intervals have been used because the method of analysis provides an instrumental response but the final result is binary. The statistical intervals are used differently from the case of the VZV IgG because no control samples are measured. The screening response value and the assessment of the quality parameters come after the response associated to the corresponding specification limit has been set and the one-sided prediction boundary has been defined.

4. As far as the intrinsic characteristics of the method of analysis and the information required by the end user are concerned, neither the contingency tables nor Bayes' Theorem are as good as the performance characteristic curves and the statistical intervals at validating the three qualitative methods studied.

4.1 Although the information extracted from the contingency tables is similar to the information gathered with the performance characteristic curves, the unreliability region around the chosen decision value cannot be estimated.

4.2 The information that can be obtained from Bayes' Theorem is not complete enough. Its well-consolidated theory provides outcomes for both the conditional probabilities and the likelihood ratio, but for only one sample at a time (e. g. the probability of a true positive being a false negative or vice-versa). To calculate these conditional probabilities the performance parameters of the method (e.g. sensitivity and specificity rates) need to be known.

5. Robustness can also be evaluated in qualitative methods of analysis. Although the procedures for evaluating robustness in qualitative and quantitative methods of analysis are apparently similar, the difference between them lies in how they treat data. In the field of qualitative methods of analysis, the different experimental conditions affect the performance parameters and not the experimental response. The procedure for evaluating robustness in qualitative methods of analysis presented in the paper could be applied to numerous methods.

6.3 FUTURE RESEARCH

The results presented in this doctoral thesis are just a small part of what needs to be done in the field of qualitative methods of analysis. Validation guidelines are a good starting point. Nonetheless, there are several issues that have not been dealt with here and which could be the subject of future research:

1) Laboratories often use methods of analysis based on simple chemical determination and the sensorial examination of the samples that provide the corresponding visual, olfactory or tactile outcomes is a rather extended practice. Validation guidelines should also be provided for these cases.

2) Methods of analysis that are part of routine quality control should have an internal quality system that includes control charts. In the particular case of qualitative methods that use control samples, measurements should also be used to determine if any undesirable variation occurs. Systematization is required in this area.

3) Performance characteristic curves also require further study. This model depends on two parameters (the so-called *a* and *b* parameters). How these two parameters affect the model should be examined. As for the linear model, it is known how the uncertainty varies, and a similar approximation could be done for the sigmoidal model. There is a proposal [1] that likens the central region of the performance characteristic curves to a straight line model. Although the authors have made the initial, important study, the issue should be studied in greater depth so that all the consequences of assuming one model or another can be determined.

6.4 PROFESSIONAL SKILLS ACQUIRED DURING THIS DOCTORAL THESIS

The process of attaining the degree of doctor has involved several stages in which I have acquired a variety of skills.

To start with, I have made an in-depth study of several areas of chemistry. Then, I carried out some bibliographic research which was presented orally and evaluated on the occasion of the Diploma of Advanced Studies This stage was a very important one as it was the first contact I had with the research topic of this thesis. As a result:

- I had a competitive advantage over other students in the first years of the doctorate because I had greater scientific knowledge.

However, although scientific knowledge was important I have also progressively acquired other skills:

- To position research in the most appropriate framework.
- To choose the best bibliographic source.
- To make the most of the information found.
- To regularly use bibliography in daily work.

During the final stages of the doctoral thesis which have led to the degree, my scientific knowledge of the research topic has increased considerably. However, I also acquired new skills which should be added to the ones I have mentioned in the paragraph above:

- I feel confident about my ability to design, participate in and use the scientific methodology of research projects in such different areas of analytical chemistry as environment and food contamination, drug compliance or quality control.

- I have learned to express myself and the research results correctly and clearly, orally and in writing, both in my own language and in English. Additionally, I have learnt to systematise what is to be said and to make it understandable.

The fact that other people have made contributions to this thesis has shown me how to:

- Collaborate with other scientists who have different opinions and to reach a consensus.
- Adapt to different groups of people with their own operational procedures. I am particularly grateful to the Laboratory of Public Health in Tarragona, to the Immunology Department at the Laboratorios de Análisis Dr. Echevarne and Prof. Valcárcel' s research group at the University of Córdoba.

6.5 REFERENCES

[1] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhardt, R.W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel, *Accred. Qual. Assur. 8*, 2003, 68.

APPENDIX

LIST OF PAPERS AND MEETING CONTRIBUTIONS

Papers presented by the author in chronological order:

1. Validation of qualitative analytical methods.
   E. Trullols, I. Ruisánchez and F. X. Rius.
   Trends in Analytical Chemistry 23, 2004, 137-145.
   Chapter 3

2. Qualitative method for determination of aflatoxin B_1 in nuts.
   E. Trullols, I. Ruisánchez, F. X. Rius, M. Òdena and M. T. Feliu.
   Journal of AOAC International 87, 2004, 417-423.
   Chapter 4

3. Validation of qualitative methods of analysis that use control samples.
   E. Trullols, I. Ruisánchez, F. X. Rius and J. Huguet.
   Trends in Analytical Chemistry 24, 2005, 516-524.
   Chapter 3

4. Validation of qualitative test kits with instrumental responses. Detection of Varicella-Zoster Virus IgG antibodies in human serum.
   E. Trullols, I. Ruisánchez, F. X. Rius and J. Huguet.
   Journal of Pharmaceutical and Biomedical Analysis. Submitted.
   Chapter 4

5. Statistical intervals to validate an autoanalyzer for monitoring the exhaustion of alkaline degreasing baths.
   E. Trullols, I. Ruisánchez, E. Aguilera, R. Lucena, S. Cárdenas and M. Valcárcel.
   Analytica Chimica Acta. In press.
   Chapter 4

6. Robustness in qualitative analysis: a practical approach.
   E. Trullols, I. Ruisánchez, E. Aguilera, R. Lucena, S. Cárdenas and M. Valcárcel.
   Trends in Analytical Chemistry. In press.
   Chapter 5

Meeting contributions presented by the author in chronological order:

1. Validation of Screening Test Kits for the Determination of aflatoxins in Nuts.
   E. Trullols, I. Ruisánchez, F. X. Rius, M. Òdena and M. T. Feliu.
   Eighth International Conference on Chemometrics in Analytical Chemistry. Seattle, Washington, USA, September 22-26, 2002
   Poster communication.

2. Estimation of the quality parameters of test kits that provide visual response.
   Esther Trullols, Itziar Ruisánchez y F. Xavier Rius
   Seminario: Análisis de Micotoxinas. Madrid, June 26th, 2003
   Oral communication.

3. Validation of analytical methods providing binary responses.
   E. Trullols, I. Ruisánchez, F. X. Rius, M. Òdena and M. T. Feliu.
   Workshop on the impact of qualitative chemical analysis in the VI Framework Program. Budapest, November 7-8, 2003
   Poster communication.

4. Validation of an ELISA qualitative kit providing instrumental response.
   E. Trullols, I. Ruisánchez, F. X. Rius and J. Huguet.
   Euroanalysis XIII. European Conference on Analytical Chemistry. Salamanca, Spain, September 5-10, 2004
   Poster communication.