This is the accept manuscript of the following article that appeared in final form in,Bioanalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect, Journal of Chromatography A,1353 : 10-27 (2014) which has been published in final form at https://doi.org/10.1016/j.chroma.2014.03.077

Bionalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect

Oskar González[a,b], María Encarnación Blanco[a], Gorka Iriarte[c], Luis Bartolomé[d], Miren Itxaso Maguregui[a], Rosa Mª Alonso[a]

[a] Analytical Chemistry Department, Science and Technology Faculty, the Basque Country University/EHU, P.O. Box 644, 48080 Bilbao, Basque Country (Spain)

[b] Analytical Bioscience Division, LACDR, Leiden University Einsteinsweg 55, 2333 CC Leiden (Netherlands)

[c] Laboratory of Public Health of Alava, (Public Health and Addictions Directorate, Basque Government), Santiago 11, 01002 Vitoria-Gasteiz, Basque Country

[d] Central Analytical Service (SGIker), Science and Technology Faculty, the Basque Country University/EHU, P.O. Box 644, 48080 Bilbao, Basque Country (Spain)

Corresponding author: Rosa Mª Alonso Rojas,

Email: rosamaria.alonso@ehu.es

Permanent address: Analytical Chemistry Department, Science and Technology Faculty, Basque Country University/EHU, P.O. Box 644, 48080 Bilbao, Basque Country (Spain)

Telephone: 34 94 601 2686 Fax: 34 94 601 35 00

Abstract

Method validation is a mandatory step in bioanalysis, to evaluate the ability of developed methods in providing reliable results for their routine application. Even if some organisations have developed guidelines to define the different parameters to be included in method validation (FDA, EMA); there are still some ambiguous concepts in validation criteria and methodology that need to be clarified. The methodology to calculate fundamental parameters such as the limit of quantification has been defined in several ways without reaching a harmonised definition, which can lead to very different values depending on the applied criterion. Other parameters such as robustness or ruggedness are usually omitted and when defined there is not an established approach to evaluate them. Especially significant is the case of the matrix effect evaluation which is one of the most critical points to be studied in LC-MS methods but has been traditionally overlooked. Due to the increasing importance of bioanalysis this scenario is no longer acceptable and harmonised criteria involving all the concerned parties should be arisen.

The objective of this review is thus to discuss and highlight several essential aspects of method validation, focused in bioanalysis. The overall validation process including common validation parameters (selectivity, linearity range, precision, accuracy, stability...) will be reviewed. Furthermore, the most controversial parameters (limit of quantification, robustness and matrix effect) will be carefully studied and the definitions and methodology proposed by the different regulatory bodies will be compared. This review aims to clarify the methodology to be followed in bioanalytical method validation, facilitating this time consuming step.

Keywords: Method validation, bioanalysis, limit of quantification, robustness, matrix effect

1. Introduction

Method validation is a necessary process to demonstrate that an analytical method is suitable for its intended use, thus, that it can offer accurate, precise and reproducible results. These reliable results are essential for bioavailability, bioequivalence, pharmacokinetic, pharmacodynamics or toxicological studies where analytes must be quantified in biological matrices such as urine or plasma. Consequently, method validation is a crucial step in bioanalysis and essential for laboratories to adhere to current Good Manufacturing Processes (GMP), Good Laboratory Practices (GLP) or International Organisation for Standardisation (ISO) regulations, such as ISO17025 [1] and ISO15189 [2]. Nowadays several regulatory bodies deal with bioanalytical validation and even though there are still some divergences, a relative consensus has been reached by the scientific community. Nevertheless, the continuous advances in instrumentation and the emergence of more demanding analytical challenges make bioanalysis a field in permanent evolution. Therefore, method validation guidelines should keep up with this progression in order to cope with real requirements, and to this end the participation of all the concerned parties (pharmaceutical industry, statisticians, analytical chemists, academicians…) is imperative.

The first attempt to harmonise bioanalytical validation dates back to 1990, when the United States Food and Drug Administration (FDA) and the American Association of Pharmaceutical Scientists (AAPS) sponsored the first bioanalytical method validation workshop in Crystal City (Arlington, VA). The aim of this workshop was to reach a consensus on the requirements in validation for analytical methods focused on bioavailability, bioequivalence and pharmacokinetic studies. Parameters considered essential were set (stability, accuracy, precision, sensitivity, specificity, response and reproducibility) and the outcome of this workshop was well received by the scientific community eager of a harmonised policy to work with. In consequence, the report of this workshop published in 1992 [3] became the basis of nowadays bioanalytical validation and the starting point for a stimulating discussion. Soon, due to the importance of method validation and the great

advances made in this field during the following years, the need for an official document became apparent, and finally, in 1999 the FDA issued the *"Draft Guidance for Industry: Bioanalytical Method Validation"* [4]. Shortly after, the FDA and AAPS organised the second Crystal City workshop in order to discuss the draft and give the scientific community a chance of sharing the experience accumulated over the 10 years elapse since the first workshop took place. New topics such as partial, full and cross-validation were discussed; stability experiments were studied more in depth and flourishing hyphenated mass spectrometry techniques (MS) were addressed. The summary of this workshop was published in 2000 [5] and it became the backbone for the official *"Guidance for Industry: Bioanalytical method validation"* issued in 2001 [6].

This document can be considered the cornerstone of bioanalytical validation since many laboratories standardised their validation procedures following this guideline and the later proposals for new guidelines used it as reference. Nevertheless, analytical chemistry and drug analysis are in an unceasing development and therefore new worries and problems are continuously emerging. This is so that after the publication of the guidance two new Crystal City meetings were organised in 2006 and 2008 and their respective white papers published in order to complete the initial guideline [7, 8]. The first one dealt with some questions that remained unresolved such as metabolite and stability studies, carryover and especially, matrix effect (ME) [9]. In this workshop incurred sample reanalysis (ISR) was discussed for the first time [10] but it was not until 2008 workshop that this topic was carefully studied. The white papers arising from these workshops are widely accepted by the scientific community despite not being official documents. Nevertheless and after more than 10 years the FDA needed to release an update of the validation guideline. Therefore, the draft of the guideline update was published in September 2013 [11]. The AAPS called the scientific community for a new workshop held in Baltimore in December 2013 where the FDA draft was carefully studied. As expected, a vivid discussion about hot topics such as validation of biomarkers,

endogenous compounds or anticoagulant change took place. Once the outcome of this conference is considered, the final guide will be released.

Besides FDA other regulatory agencies and organisations have dealt with bioanalytical method validation. In Europe, this field was to some extent covered by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) "*Validation of Analytical Procedures Q2 (R1)*" [12], developed between 1994 and 1996 by experts on method validation from European Union, Japan and USA. Nevertheless, this guideline is addressed to identification tests, control of impurities or active ingredient quantification. Obviously the European Union needed an official document to establish the regulation for bioanalytical validation and consequently, in December 2008 European Medicines Agency (EMA) released the "*Concept paper/recommendations on the need for a guideline on the validation of bioanalytical methods*" [13]. Immediately the European Bioanalysis Forum (EBF), an organisation comprised of bioanalytical scientists working within the pharmaceutical industry, expressed their concern about the possibility of contradiction between this new guideline and the world-widely recognised FDA guidance [14]. When the draft guidance was released in November 2009 [15] a lively discussion arose again and several meetings were organised in order to achieve a more harmonised document [16-20]. In this framework the Global Bioanalysis Consortium (GBC) was founded with the idea of merging all the existing and emerging bioanalytical guidances in a unified consensus document [19]. The GBC is a worldwide organisation of representatives of scientific associations with a well-coordinated organisation involving different harmonisation teams that are connected to the Steering Committee via the Scientific Leadership team [21]. Besides all the aforementioned there are other organisations which deal with bioanalytical validation, either organising frequent workshops and conferences such as the Canadian Calibration and Validation group (CVG) [22-27] and the Global CRO Council (GCC) [28] or developing domestic guidelines such as Brazilian ANVISA (*Agência*

Nacional de Vigilância Sanitária) [29], Japanese MHLW (*Ministry of Health, Labour and Welfare*) [30] or Spanish AEFI (*Spanish Association of the Industry Pharmaceutics*) [31].

The numerous meetings organised yearly around bioanalytical validation evidence the importance of the topic as well as the lack of consensus in some important issues which need to be further discussed and agreed. This review aims to summarise the different approaches around validation parameters in a clarifying way, providing clear information about the whole validation process but with special focus in the most controversial validation parameters: robustness, lower limit of quantification (LLOQ) and matrix effect, since according to our experience, these are the most ambiguous or overlooked parameters.

To our knowledge this is the first work that includes relevant aspects from the worldwide recognised FDA, EMA and ICH official guidelines plus ANVISA, MHLW and AEFI domestic guides. In addition it includes first comments on FDA guideline update draft.

2. Bioanalytical method validation

As aforementioned, method validation is the process through which the reliability of a method to quantify an analyte in a certain matrix is demonstrated. Depending on the previous degree of development of the method to be validated this process is more or less exhaustive leading to three different types of method validation: full, partial and cross-validation.

When a method is implemented for the first time or a new drug or metabolite is included a *full validation* is mandatory. According to FDA, a full validation should study all the fundamental parameters including accuracy, precision, selectivity, calibration curve, sensitivity, reproducibility and stability. In addition, EMA adds matrix effect to these fundamental parameters. In general, this kind of validation is necessary for each matrix and species under study and it should be carried out in the same matrix and using the same anticoagulant (if the matrix should be blood/plasma) as the study samples. Only an alternative matrix can be used if justified, as is the case of the rare matrices.

Usually, before a full validation is carried out a pre-validation process is necessary in order to characterise some parameters. The most important step during pre-validation is the definition of the lower and upper limits of quantification (quantification range) and the response function to be used during the validation. Additionally the number of calibration standards needed to build the calibration curve, recovery (REC) and selectivity can be evaluated [32, 33].

Sometimes minor changes are made to a previously validated method such as transfer between laboratories, change in concentration range, change in sample processing procedure, change in storage conditions… In these cases a full validation may not be necessary and a *partial validation* may be enough. According to the guidelines, the validation process to be carried out during partial validation can range from as little as the determination of the within-run precision and accuracy, to an almost full validation. The absence of an established protocol means that the validation process will rely on the criterion of the analyst.

Some of the minor changes that lead to a partial validation can be controversial. For example, even if EMA states that *"generally a full validation should be performed for each species and matrix concerned"* partial validation may be enough for a change in animal species or matrix. Anyway, when an extrapolation between animal species is carried out species-specific metabolism should be taken into consideration and when a change in matrix within species is performed it must not affect the outcome of the study [20]. There is still a lack of consensus regarding the anticoagulant change (for example from heparin to EDTA), even if it is clear that at least a partial validation must be carried out there is not full agreement regarding the use of the same anticoagulant with different counter-ion (for example sodium heparin and lithium heparin) [7, 24]. A survey carried out among 15 EBF companies showed that a change in counter-ion has not impact on assay performance and therefore they concluded that plasma samples containing same anticoagulant with different counter-ion should be regarded as equal matrices [34, 35]. This point was also a matter of discussion during the 5th Workshop on Recent Issues in Bioanalysis organised by the CVG [26] where they

recommended a benchtop stability study on matrix despite the fact that it has been demonstrated that the change in the counter-ion does not affect the sample beyond a slight change in the pH.

Finally, the last type of validation is *cross-validation* [36]. It consists in a comparison of the validation parameters of two or more different analytical procedures that are used in the same study. It can consist of a comparison between methods using different analysis techniques (LC-MS/MS vs ELISA) or between laboratories using the same method. According to EMA for the cross-validation the same quality control (QC) samples should be analysed by the different methods (laboratories) and the results cannot differ more than a 15%.

3. Validation parameters

Along this section the different parameters that should be evaluated in bioanalytical method validation for chromatographic assays will be studied. This review will be mainly focused on the guidelines published by the two main regulatory agencies in bioanalysis: *"Guideline on bioanalytical method validation"* by EMA [37] and "*Guidance for Industry: Bioanalytical Method Validation"* by FDA. For the latter, the document presented in 2001 [6] and the white papers issued in 2007 [7] and 2009 [8] will be discussed. The guideline update, which draft has been recently published [11], will be also taken into account but always bearing in mind that it is still in the review phase. Other documents regarding method validation will also be taken into consideration such as *"Validation of analytical procedures: text and methodology Q2(R1)"* [12] by ICH, *"Guide for validation of analytical and bioanalytical methods (RE 899/2003)"* [38] and "*Dispõe sobre os requisitos mínimos para a validação de métodos bioanalíticos empregados em estudos com fins de registro e pós-registro de medicamentos (RDC 27/2012)"* by Brazilian ANVISA [29], *"Draft Guideline on Bioanalytical Method Validation in Pharmaceutical Development"* [30] by Japanese MHLW, *"Validation of Analytical Methods"* [31] by Spanish AEFI and *"The Fitness for Purpose for Analytical Methods. Laboratory Guide to Method Validation and Related Topics"* by EURACHEM [39].

One of the main drawbacks of all these guidelines is that they are rather conceptual and even if they define validation criteria the methodology to carry out the validation procedure is not precise enough. In order to fill this gap, and help the professionals from the pharmaceutical industry the *"Société Française des Sciences et Techniques Pharmaceutiques"* (SFSTP) elaborated a validation procedure proposal with a smart and careful experimental design which has also be taken into consideration in this work [33, 40-42].

3.1. Selectivity and Specificity

These two parameters that are closely related are usually inappropriately interchanged [43]. Both FDA and EMA agree in the definition of selectivity as the ability of a bioanalytical method to measure and differentiate the analyte(s) of interest and internal standard (IS) in the presence of components which may be expected to be present in the sample. On the other hand, EMA defines specificity as the ability to measure the analyte unequivocally in the presence of other compounds, either exogenous or endogenous, in the matrix. Even if the difference between both terms is apparently faint, a clear distinction should be made: selectivity is a parameter that can be graded whereas specificity is absolute [44-48]. Selectivity is the measure of the extent to which an analytical method can determine an analyte without interference from other compounds. Only when a method is perfectly selective for the analyte it is considered specific. Very few methods are specific and therefore selectivity is considered to be the best term to be used in validation.

Main guidelines agree in the methodology for selectivity studies. At least six independent sources of the same matrix should be used and the response of the interferences compared to the signal of the analyte at the LLOQ. ANVISA RDC 27/2012 is more exhaustive and specifies that from the 6 different sources one should be hyperlipidaemic if the matrix is whole blood and one hyperlipidaemic and another one hemolysed if the matrix is plasma. The response of the interference should be lower than 20% the response of the analyte at the LLOQ and 5% the response of the IS. EMA allows the use of a smaller number of different sources if the matrix is

213 considered rare, and FDA does it if LC-MS methods are used. In this case, checking the matrix
214 effect is encouraged in order to demonstrate that precision and sensitivity are not affected.
215 Regardless the detection method used, the authors recommend checking the selectivity in at least 6
216 different sources especially when low molecular mass analytes or complex matrices are to be
217 analysed.

218 The regulatory agencies also propose a second approach for the study of selectivity based on the
219 investigation of potential interfering substances, such as metabolites, degradation compounds or
220 concomitant medication. This is a common practice in industry, where effect of co-medication is
221 widely studied according to a survey carried out by EBF [49]. On the other hand, very few scientific
222 papers follow this approach [50-53]. Ansermot *et al.* studied the selectivity of a method for the
223 quantification of selective serotonin reuptake inhibitors by spiking plasma samples with 63 common
224 medications in the psychiatric population [52]. They found out a potential interference with the
225 same retention time and m/z as the analyte and optimised the fragmentation of the confirmation ion
226 in order to avoid false positives in routine analysis. Hu *et al*. found out that the hydroxylated
227 metabolite of the brominated hypoglycemic agent G004 interfered with the parent drug after in-
228 source dehydration by analysing incurred samples [53]. It is evident that this infrequent
229 methodology offers complementary information about method selectivity and therefore it should be
230 considered especially when metabolites or degradation products are available and common co-
231 prescribed drugs known.

232 ***3.2. Accuracy***

233 According to the Joint Committee for Guides in Metrology [54] accuracy of an analytical method is
234 defined as the closeness of agreement between a measured quantity value and a true quantity value
235 of a measurand. In the FDA guideline text accuracy is defined as the closeness of *mean* test results
236 obtained by the method to the true value (concentration) of the analyte. This definition, which
237 involves the mean of some replicates, is misleading and corresponds to trueness, as it has been

discussed in depth by Rozet *et al.* [43, 55]. According to ISO definition, accuracy involves both random error and bias while trueness only reflects the bias of the analytical method [56]. Therefore, by using the methodology described in the guidelines, trueness of the analytical method is calculated instead of the accuracy. This inadequate definition of accuracy is thoroughly accepted in bioanalysis as it can be observed in different guidelines and publications. Nevertheless, in order to achieve a harmonised definition in the frame of the analytical validation this term should be reviewed in the close future.

The regulatory organisations recommend to measure accuracy using a minimum of 5 replicates at least at three different concentrations (QC samples) covering the whole calibration range: within 3 times the LLOQ, near the centre, and close to the upper calibration point. Taking into account the allowed error, the authors suggest setting the upper QC sample concentration between 75% and 85% of the upper calibration point in order to avoid extrapolation. Besides this three concentration levels, EMA and ANVISA ask to measure accuracy at LLOQ. Within-run (intra-batch) accuracy and between-run (inter-batch) accuracy should be evaluated and expressed as Relative Error (%RE). The deviation (in percentage) of the mean from the true value should be less than 15% except for LLOQ where up to 20% of error is accepted. For between-run accuracy at least 3 runs analysed on at least two days should be used.

3.3. Precision

Precision is defined as the closeness of agreement between a series of measurements obtained from multiple sampling of the same homogenous sample under the prescribed conditions and it is expressed as relative standard deviation or absolute coefficient of variation (%CV) (ratio between standard deviation and mean, multiplied by 100). The %CV should not exceed 20% for the LLOQ or 15% for the rest of the QC samples. Precision can be evaluated with the same experiments used for the determination of the accuracy. In this way, within-run (intra-batch) precision assesses precision during a single analytical run and between-run (inter-batch) precision measures the

precision with time. Despite FDA suggesting that the between-run precision may involve different analysts, equipment, reagents and laboratories, the inter-laboratory study of the precision fits better with the term reproducibility used in the glossary of the guideline.

In the guidelines for analytical method validation, ICH and ANVISA classify precision in a more unambiguous way by defining repeatability, intermediate precision and reproducibility. In this way, repeatability assesses precision under the same operating conditions within a short period of time and would be equivalent to within-run precision. Intermediate precision expresses within laboratory variations (different days, analysts or instrumentation) and would be equivalent to between-run precision without considering different laboratories. Finally, reproducibility is the precision inter-laboratory and involves collaborative studies. These organisations are less demanding in terms of number of replicates and request a minimum of three determinations per concentration level. This classification is unequivocal and avoids the confusion in inter-batch precision that can derive from FDA and EMA definitions. Reproducibility and intermediate precision are closely related to robustness and ruggedness, in fact ICH defines reproducibility as ruggedness, but this point will be thoroughly discussed in the following section.

3.4. Robustness and Ruggedness

Robustness and ruggedness are terms related to the ability of a given method to provide reliable results despite variations in the analysis conditions. Nevertheless, there is considerable confusion in the scientific literature with regard to the use of the terms robust and rugged and the associated characteristics robustness and ruggedness as applied to the description of analytical methods. This confusion does not only take place in the pharmaceutical or bioanalytical field but also in others such as environmental or food analysis. Differences between these terms have not been properly defined by IUPAC, ISO or similar bodies. Indeed, within IUPAC documents, confusion exists within the use of ruggedness as implying an inter-laboratory use, or a single-laboratory situation [57]. The most accepted and used definitions for robustness and ruggedness are given by ICH and

288 the U.S. Pharmacopeial Convention (USP). ICH defines robustness as follows: *"The robustness of*
289 *an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate*
290 *variations in method parameters and provides an indication of its reliability during normal usage"*
291 [12]. Therefore a study of robustness will involve an investigation into the effect of these small, but
292 deliberate variations. The USP has accepted the ICH definition for robustness and defines
293 ruggedness as: *"The ruggedness of an analytical method is the degree of reproducibility of test*
294 *results obtained by the analysis of the same samples under a variety of conditions such as different*
295 *laboratories, different analysts, different instruments, different lots of reagents, different elapsed*
296 *assay times, different assay temperatures, different days, etc. Ruggedness is normally expressed as*
297 *the lack of influence of operational and environmental factors of the analytical method. Ruggedness*
298 *is a measure of reproducibility of test results under the variation in conditions normally expected*
299 *from laboratory to laboratory and analyst to analyst"* [58]. As can be seen in the definition, the
300 ruggedness test could be performed without deliberately changing the method parameters but only
301 under different test conditions. Afterwards, their influences can be evaluated using a nested design
302 or nested Analysis of Variance (ANOVA). In fact, this definition is equivalent to that for
303 intermediate precision or reproducibility already mentioned in section 3.3. In this sense, is
304 moderately considered by some authors [59] that ruggedness is a property hierarchically above
305 robustness as put forward by the Canadian Drugs Directorate in their three level testing system. In
306 this system, Level I refers to the ICH definition of robustness and should include verification of
307 repeatability by using a second analyst. In Level II testing, the effects of more severe changes in
308 conditions are examined when the method is intended to be applied in a different laboratory with
309 different equipment. Level III considers "a full collaborative testing", which is rarely done. On the
310 other hand, other researchers such as Youden and Steiner [60] used the term ruggedness test for an
311 experimental set-up that examines influences of minor but deliberate and controlled changes in the
312 method parameters on the response, in order to detect those factors with a large influence (non-
313 rugged factors). In fact, Youden and Steiner's ruggedness test application could be more related

with robustness as described by ICH. As can be seen, several definitions for robustness or ruggedness exist which are, however, closely related and, in some cases, exchanged.

As it has been already mentioned the main objective of validation of bioanalytical methods is to demonstrate that the procedure is suitable for its intended purpose. For this reason, if a bioanalytical method is to be used routinely it needs to be robust so that the results obtained are reliable regardless small methodological variations. Although the robustness/ruggedness could be a significant parameter to be studied during the optimisation or validation step in bioanalysis, neither of these parameters are taken into account in the most important bioanalytical validation guidelines. According to FDA, bioanalytical method validation should include all the procedures/parameters that could make it reliable and reproducible for the intended use [6]. Nevertheless, neither robustness nor ruggedness is mentioned along the guide and no recommendation has been included regarding the study of the influence of small variations or different conditions. EMA guideline also omits robustness and ruggedness. On the other hand, both guidelines ask for a partial validation when methods are transferred between laboratories, which could be an unstated alternative to check the ruggedness (as defined by USP) of an initial bioanalytical method.

In contrast, ANVISA guideline [38] provides detailed information about robustness, which is defined according to ICH and is supposed to be evaluated before the validation step by studying the main parameters susceptibility to variations. Depending on the analytical methodology used (spectrometry, liquid chromatography or gas chromatography) the guide recommends checking different parameters. For example, in the case of LC, parameters that could affect to robustness are the pH variation of the mobile phase, variation in the composition of the mobile phase, different batches or column manufacturers, temperature and flow of mobile phase. It should be noted that these recommendations do not apply to identification or performance purpose methods.

Although the literature cited above refers almost exclusively to drug analysis and bioanalysis, the clear distinction between the descriptors robust and rugged is of wider application to other fields of

analysis under regulation, such as those of human foods, animal feed, environmental samples and of articles subject to tariff/customs control. In this sense, the council directive 96/23/ CE [61] lays down measures to monitor drugs and pharmaceuticals, among others, in live animals and animal products. Although strictly speaking it is not a guide in the bioanalytical field, it could be a great example of discussion about how robustness should be considered during the validation of the analytical procedures. This directive defines ruggedness (in the same way as ICH robustness) as *"the susceptibility of an analytical method to changes in experimental conditions which can be expressed as a list of the sample materials, analytes, storage conditions, environmental and/or sample preparation conditions under which the method can be applied as presented or with specified minor modifications. For all experimental conditions which could in practice be subject to fluctuation (e.g. stability of reagents, composition of the sample, pH, temperature) any variations which could affect the analytical result should be indicated."*. Furthermore, not only is given the definition but also some recommendations about how the analytical method should be tested. This directive suggests the inclusion of relevant changes in the robustness study such as stability of reagents, sample composition, pH or temperature. Besides that, unlike with all previous guides, according to this directive responses of these changes should be analysed by means of, for example, Youden focus [60]. Briefly, according to this methodology, the average differences between measurements under normal conditions and measurements under slight metrological changes are compared. The standard deviation of the sum of the quadratic differences should be less than the reproducibility standard deviation of the validated method. If this assumption is satisfied, the robustness of the method is assured [61]. In summary, in this particular case there are not only some suggestions to check the parameters relative to robustness but also, as well as in the case of other parameters such as reproducibility or accuracy, minimum tabulated levels to assure the applicability of analytical methods.

3.4.1. Methodological approach

If a partial literature revision of the last 5 years articles related to bioanalytical validation methods is carried out, only in approximately 20% of the total cases, a robustness study was performed by the authors [62-67]. Some of the authors explain that the lack of these studies is due to the fact that according to the guides they followed, robustness test was not required. In other works, robustness (following USP ruggedness concept) has been studied as a part of a reproducibility study but not deeply. In the articles in which a proper robustness study is performed it is carried out by means of One Variable At a Time (OVAT) procedures [68-70]. In this cases method overall precision, chromatographic resolution or peak symmetry are evaluated for changes in factors such as flow rate, wavelength, column temperature, pH of buffer and organic content in mobile phase as a measure of robustness. Briefly, levels of a given factor are varied while keeping the other factors at method working (nominal) levels to evaluate the effect of the former factor on the method response(s). This classical approach can be fast (only if few factors are studied), easy and applicable for bioanalytical methods [71, 72], but is not recommended for robustness testing [73] since interactions among factors are not taken into account. Moreover, when robustness tests are performed following this procedure, they are usually carried out at a late stage in the method validation probably due to the relationship between robustness/ruggedness concepts and reproducibility of the method, for which inter-laboratory studies are commonly performed in the final stage of the validation process. However, performing a robustness test at this point involves the risk of finding out a lack of robustness of the method too late, which involves that it should be redeveloped and reoptimised. At this stage much effort and money have already been spent in the optimisation and validation, and therefore this is an incident to be avoided.

A different approach to study robustness which is spreading lately and overcomes OVAT methodology drawbacks is based on the use of experimental designs. This option has large advantages compared to univariate treatment. When applying an experimental design, the effect of a given factor is calculated at several level combinations of the other factors, while with the OVAT

389 approach it is performed only at one level. Thus, in an experimental design, a reported factor effect
390 is an average value for the whole domain, and it represents more globally what is happening around
391 the nominal situation. Moreover, the univariate approach requires more experiments, especially
392 when the number of examined factors becomes larger, and the importance of factor interactions
393 cannot be taken into account. Design of Experiments (DoE) with predictive probability is a very
394 innovative framework to simultaneously optimise the separation and estimate the method
395 robustness over the experimental domain. Vander Heyden and co-workers explain thoroughly the
396 use of DoE for robustness testing during validation step in a brilliant guidance [73]. In Figure 1, the
397 steps they propose for a robustness/ruggedness test can be observed.

398 According to this author, regardless robustness is investigated as part of the development of the
399 method, at the end of method development during optimisation, or after method development
400 during validation, the method parameters which will be investigated remain the same, e.g., %
401 organic modifier in the mobile phase, pH of the mobile phase…[73, 74]. These are referred to as the
402 robustness factors and the first step of robustness testing is to choose them. The number of factors
403 selected determines how much information is gathered about how the changes in method
404 parameters affect the results, so, the higher number of factors studied the more information but also
405 the higher number of experiments to be performed. Once the factors have been chosen, the levels at
406 which they will be studied need to be defined. The amount of variation is referred to as the factor
407 level, typically, limits around a nominal value are investigated and the magnitude of these limits has
408 to be defined. It must be kept in mind that the aim is to understand the effect of changing a method
409 parameter, whereas when determining the robustness of a method during validation the aim is to
410 examine the variation which might be expected in routine use of the method. Thus the limits for the
411 former may be expected to be wider than in the latter scenario. The factors and factor levels define
412 the way in which the method will be set up for robustness experiments, but a way of measuring the
413 effect of the method variations is also required. This needs to take into account quantitative aspects

of the method such as an assay result (e.g., %w/w for main components or impurities), but also chromatographic criteria such as resolution, retention factor, asymmetry factor or number of theoretical plates. These are referred to as the method responses. Once factors, levels and responses have been selected, an experimental design is chosen and executed. The recommended designs are fractional factorial and Plackett-Burman [73, 74] although two-level full fractional designs also can be used, provided that the number of examined factors does not exceed four, since otherwise too many experiments are required. Statistical and graphical methods may be used to aid interpretation [74]. Although not addressed to bioanalysis, some international guides such as the FDA and ICH request a Quality-by-Design (QbD) approach based on DoE. QbD is defined as *"a systematic approach to development that begins with predefined objectives and emphasises product and process understanding and process control, based on sound science and quality risk management"*. Although the ICH guideline Q8 (R2) either does not explicitly discuss analytical method development, the QbD concept can be extended and results in a systematic approach that includes defining methods goal, risk assessment, developing of a Design Space, implementing a control strategy and continual improvement to increase method robustness and knowledge.

The emergence of the use of QbD principles in pharmaceutical manufacturing has led to the application of QbD to analytical methods in other fields such as bioanalysis. This in turn has highlighted the importance of the study of robustness during method development as the design space concept of QbD translates into knowledge about the effect of each method parameter on the final analytical result. A very smart and computer-assisted way of developing a chromatographic method is by using different kind of software modelling packages (R 2.13 free-ware software, DryLab®4 chromatography modelling) [75, 76]. Based on a small number of experiments, these software applications can predict the movement of peaks in reversed-phase liquid chromatography separations when changing the mobile phase composition or pH, temperature, flow rate and the column dimensions and particle size [77]. The DoE is not usually used as a test of robustness in

bioanalytical methods but it is possible to find it in some validations of drugs in other matrices than bulks or pharmaceutical formulations [50, 78-80].

3.4.2. Discussion

Robustness and ruggedness are overlooked in the main bioanalysis validation guides taking into account the importance of these parameters to guarantee the good performance of the analytical method during routine analysis. Despite the intensive method validation procedures required in bioanalysis in order to meet the strict regulations set by the regulatory authorities, only ANVISA, in the RE 899/2003 requires the study of robustness as a method validation parameter.

In order to cover these parameters, non bioanalysis specific guidelines have to be consulted and in these documents there is also significant confusion in defining both terms (robustness/ruggedness) and making a difference between them. For that reason, it would be highly recommended that organisations such as IUPAC would define these terms unambiguously. Ideally, robustness should be investigated as part of method validation because a method is not complete without an evaluation of its reliability in routine use, but unfortunately, as has been observed, it is often deferred, or completely overlooked, because of the time-consuming nature of the study. In the ICH validation guidelines, robustness is not included in the tabular summary of required characteristics to be tested during validation, which could lead to the mistaken belief that a study of robustness is not required. Robustness should be included in the guides in which is not currently present as one of the parameters for the validation and optimisation process such as accuracy, precision, selectivity, sensitivity, reproducibility or stability. Given the different ways to ensure robustness found in the literature in recent years, also would be advisable to include methodological recommendations. The use of design of experiments (DoE) seems to be a good alternative as a tool to evaluate robustness better than OVAT procedure. Moreover, the performance of a robustness test has been shifting to earlier stages in method development. Some guidelines, including ICH, as well as some authors working in bioanalysis consider robustness a method validation topic performed during the

development and optimisation phase of a method, while others consider it as belonging to the development of the analytical procedure. Therefore, the robustness test could be viewed as a part of method validation that is performed at the end of method development or at the beginning of the validation procedure. In any case, the exact position has relatively little influence on how it is performed.

3.5. Calibration curve: Range and linearity

The calibration curve is the mathematical equation that relates the instrument response to the calibration standards for a defined calibration range. After the pre-validation step this range should be fixed between the LLOQ and the upper limit of quantification (ULOQ). The former depends highly on the sensitivity of the analysis technique but it does not need to be the limit value which fulfils the specifications of the regulatory authorities. In fact, the LLOQ should be adapted to expected concentrations and to the aim of the study (fit-for-purpose). For example, EMA establishes the limit for the LLOQ for bioequivalence studies to be 5% of the C_{max} of the analyte. This means that even if the analysis method is able to quantify 0.01% of the C_{max} such a low LLOQ may not be necessary since it will not provide any additional information to the study. Regarding the ULOQ, it should be high enough to ensure that all the study samples will fall within the calibration range. If a sample concentration is higher than the ULOQ the concentration should not be calculated by extrapolation but by means of a reassay after dilution with the matrix as will be explained later (point 3.10).

According to the guidelines a calibration curve should consist of a blank sample (blank matrix), a zero standard (blank matrix spiked with the IS) and at least six calibration standards covering the whole calibration range. The use of replicates is not necessary but it can be helpful when labile compounds are to be analysed. The simplest mathematical model that can explain adequately the relationship between concentration and signal should be used. It is important to mention that forcing the calibration curve to have zero as intercept can introduce a significant bias, so this approach

should not be followed. Ideally a non-weighted univariate equation would be utilised, but taking into account the wide concentration ranges of bioanalytical methods, weighted linear regression or complex quadratic/logarithmic equations may be necessary [81-83]. Even if the guidelines do not mention it, homoscedasticity along the calibration curve should be studied and if the distribution of the variance is not homogeneous, statistical weights should be applied [84, 85]. When this step is not performed, the influence of the low concentration standards is usually underestimated due to the overestimation of the high concentration standards. That is the reason why many bioanalytical methods use $1/X$ or $1/X^2$ weights for the regression models [86-91].

Regarding the acceptance criteria for the calibration curve a slight incongruity can be found in the FDA official guideline. In the *Calibration Curve* section the requirement for acceptance is that 4 out of six non-zero standard have less than 15% of deviation from the nominal concentration (20% for LLOQ standard) whereas in the *Acceptance Criteria for the Run* section same cut-off is required for 75% of the standards. The latter criteria seems to be more adequate and fits with EMA's one. Furthermore, both organisations agree that at least 6 of the non-zero standards must satisfy the accuracy criteria including LLOQ and ULOQ standards. When replicate standards are used at least 50% of the calibration standards tested per concentration level should fulfill the accuracy criteria. In addition to standards accuracy, FDA also requires a goodness of fit test for the calibration curve and in the new update they ask to perform at least six calibration curves over the whole validation process with at least 4 concentrations analysed in duplicate (LLOQ, low, medium and high). This last requirement was only applied to microbiological and ligand-binding assays in previous guidelines and if it is finally accepted for chromatographic methods it would lengthen the validation procedure considerably.

3.6. Limit of quantification

The LLOQ is the lowest concentration of analyte which can be determined with acceptable precision and accuracy. The bioanalytical validation guidelines establish well harmonised

514 acceptance criteria. %RE and %CV values of five replicates should be less than 20% and the
515 analyte response at the LLOQ should be at least 5 times the response of the blank. In other words,
516 the signal of the blank at the retention time of the analyte must have an area no greater than 20% of
517 the area corresponding to the LLOQ [31].

518 *3.6.1. Methodological approach*

519 Even if all guidelines agree in the definition of the LLOQ, there is not such agreement in the
520 calculation of this parameter. In this way, different approaches can be found in the literature and the
521 guidelines to determine the LLOQ in the bioanalytical field. Taking into account the importance and
522 ambiguity on the calculation of this parameter many other approaches have been proposed in other
523 analytical fields which could be applied to bioanalyisis [92]. Nevertheless due to the scope of this
524 review only the approaches dealing with LLOQ have been considered.

525 *3.6.1.1 Based on the practical examination*

526 Based on a trial and error concept, the LLOQ can be determined by using at least five QC samples
527 spiked at a concentration close to the estimated LLOQ, from a single pool of matrix different from
528 the calibration standards [3, 93]. The mean values should be within pre-defined acceptance criteria
529 [94]. The advantage of this approach is the fact that the estimation of LLOQ is based on the same
530 quantification procedure used for real samples [95]. On the other hand it can be time consuming
531 since many different QC samples may be needed until reaching the concentration that meets the
532 criteria.

533 *3.6.1.2 Based on the "Standard Deviation at the LLOQ" (IUPAC approach)*

534 In practice, the quantification limit can be expressed as the concentration that can be determined
535 with a specified relative standard deviation [96]. Thus:

536 $$LLOQ = k \cdot \sigma$$ Equation 1

where k is the multiplier whose reciprocal equals the selected quantifying %CV and σ is the standard deviation of the concentration at the level of the LLOQ [97]. In bioanalysis, the required precision at the LLOQ is of 20%, thus, k=5. To estimate σ a number of independent determinations (n > 10) must be carried out using samples spiked at a concentration close to LLOQ. Since LLOQ is not known, some general analytical validation guides recommend to analyse a sample with a concentration between 2 and 5 times the estimated detection limit (LOD) [57]. However it is important to point out that in contrast to analytical method validation, LOD is not a term used in regulatory bioanalysis [16], thus, the analyst may need to perform some extra experiments to estimate it [12, 98-101]. Other guidelines suggest to perform the determinations on a blank sample [39]. Such a procedure, however, is discouraged, unless there is a strong evidence that the precision is constant between C = 0 and C = LLOQ.

3.6.1.3 Based on the well-known signal-to-noise (S/N) ratio approach

This approach is only applicable to instrumental analysis procedures providing a blank signal, background noise or baseline signal, such us UV-visible spectrophotometry or chromatographic methods [12, 31, 38, 102].

The value of the LLOQ is given by the concentration of analyte providing a signal at least 5 times higher than that of background noise, since a 5:1 S/N ratio is considered to be sufficient to discriminate the analyte from the background noise. Nevertheless, the required ratio can vary between 5 and 20 depending on the guideline [31] and the procedure followed to measure the noise. In this aspect, S/N can be defined as the ratio between the height of the analyte peak (signal) and the amplitude between the highest and lowest point of the baseline (noise) in a certain area around the analyte peak [103]. However, very often the noise is defined as half of the amplitude and consequently the limit for S/N becomes twice bigger.

560 Even if this is one of the most used approaches in chromatographic methods, this procedure
561 presents some disadvantages, starting with the fact that the calculation of the signal of the blank is
562 not straight forward which turns the estimation of the S/N ratio into a difficult task. Nevertheless
563 some rules have been detailed to facilitate it [104]. According to Hartmann [93], in order to
564 calculate the signal of the blank, it is advisable to perform from 6 to 10 consecutive analyses of a
565 blank sample recording the maximum fluctuation of the background noise measured over a certain
566 distance, in an area corresponding to 20 times the half peak width. Another inconvenience of this
567 approach is that it is dependent on the manner the noise is measured; thus, the S/N ratio will vary
568 from one instrument to another depending on the internal operational set-up such as signal data
569 acquisition rate or the detector time constant. Furthermore, thresholding and smoothing can have a
570 dramatic effect on the apparent S/N without necessarily reflecting an instrument improvement in
571 terms of precision or accuracy of the results, making the S/N ratio measure an extremely subjective
572 estimation [55, 105, 106]. In addition, the estimation of baseline noise can become quite difficult in
573 bioanalysis, if matrix peaks elute close to the analyte peak [95]. These peaks can interfere with the
574 background noise and lead to an overestimation of the noise amplitude and consequently also of the
575 LLOQ. From the point of view of experimental verification, it has been observed in many cases,
576 that it is possible to obtain equally precise and accurate results at concentration levels below the
577 theoretical LLOQ [31]. Another problem of this approach is associated to the magnitude used to
578 measure the S/N ratio and to build the calibration curve. For example, in chromatography with
579 photometric detection, the signal height is measured in absorption units, but calibration curve is
580 generally built using areas. Therefore, the quantification limit is not expressing the lowest level of
581 the analyte, but the lowest quantified absorbance [55].

582 Despite all the drawbacks, this is an interesting approach since the LLOQ is related to the ability of
583 the method to distinguish analyte signal from the background signal or noise. However, some
584 authors [107] identify LLOQ directly with sensitivity when these terms are conceptually different,

being the sensitivity the capability of the method to discriminate small differences in concentration of the analyte [108]. Therefore in practical terms, sensitivity is the slope of the calibration curve obtained when representing the response versus concentration. In this regard, a high sensitivity of the analytical method does not always imply lower LLOQ and it could be preferable a method with lower background noise at the expense of sensitivity [31, 109].

3.6.1.4 Based on the "Standard Deviation and the Slope of the Calibration Curve".

LLOQ can be calculated using the calibration curve and the signal of the blank using Equation 2 [31]:

$$LLOQ = \frac{y_{blank} + 10 \cdot \sigma}{b}$$ Equation 2

Where y_{blank} is the background signal or the signal of the blank, b is the slope estimated from a calibration curve built close to the concentrations levels of the LLOQ and σ is the standard deviation of the response. Depending on the approximation used to obtain σ this function will have different ways.

Using the standard Deviation of the Blank: $\sigma = s_{blank}$

This approach is only applicable for methods where s_{blank} can be estimated from replicate analyses of blank samples. In this case σ is the standard deviation of the response obtained from the analysis of the blank (i.e., the background noise of the system, y_{blank}). The application of Equation 2 is only valid if the major source of variation is due to the standard deviation of the blank (thus, s of the blank is much bigger than the s of the y-intercept and the s of the slope). Otherwise, the term corresponding to the standard deviation of the blank s_{blank} would be more complex to calculate [31].

As happens with the S/N approach the estimation of the y_{blank} and the s_{blank} is certainly the weak point of this approach. [110]. In order to avoid the sometimes challenging calculation of these

607 parameters, Equations 2 can be modified using some approximations. On the one hand, y_{blank} is
608 replaced by the y-intercept with the calibration curve (extrapolating the equation to zero
609 concentration). On the other hand, s_{blank} is replaced by the standard deviation, obtained from the
610 calibration curve, of this hypothetical blank [31].

611 *Residual Standard Deviation of the Calibration Curve: $\sigma = s_{x/y}$*

612 In this case σ is obtained from the residual standard deviation of the calibration curve ($s_{x/y}$) [12, 38,
613 102]. It is advisable to remember that this approach can only be used under homoscedasticity
614 conditions. If the precision along the calibration curve is significantly different a weighted
615 calibration curve should be used and the LLOQ calculated using the corrected $s_{x/y}$ [105].

616 *Standard Deviation of the Intercept with the y-Axis of the Calibration Curve: $\sigma = s_a$*

617 In this case σ is the standard deviation of y-intercept of the calibration curve (s_a). As explained in
618 the first option for this approach, it can be obtained directly from the standard deviation of the zero
619 concentration, but ANVISA states that s_a should be calculated as the standard deviation of the y-
620 intercept of at least 3 curves [38].

621 This approach has some problems that are worth mentioning. On the one hand, ICH Q2(R1) Guide
622 assumes that the calibration is linear, what is not always true [55]. On the other hand no guidance
623 has established any requirement related to calibration curves at low concentration ranges (number
624 of standards, correlation coefficient, number of calibration curves…). Obviously, the obtained
625 LLOQ will depend on all these parameters. Finally, it is important to take into account that all these
626 approaches based on the calibration curve should be used under homoscedasticity conditions, but
627 the use of a weighting procedure should be considered [110].

628 *3.6.1.5. Based on the "Graphical Examination of the %CV versus Concentrations close to the*
629 *Expected LLOQ" (EURACHEM approach)* [39].

The LLOQ can be calculated based on the %CV following the EURACHEM approach (also known as "Target %CV approach"). This approach is different in essence from those related to the estimation of the standard deviations since it is based on the direct measure of precision, in other words, it is based on the establishment of the minimum concentration of analyte that can be analysed with a precision equal to a %CV of 20%. For this aim, six replicates of a series of samples with decreasing concentration levels are measured. Then, the obtained %CV for each level is plotted against analyte concentration, as shown in Figure 2. Using this chart the LLOQ is calculated as the concentration for which the %CV is equal to 20% [39].

The strong point of the EURACHEM approach is that the precision of the measurement, i.e. one of the defining criteria for the LLOQ, is directly measured. This approach overcomes some of the problems stressed in the previous approaches, since it is no longer equipment and operator dependant [105]. However, only the precision of the analytical procedure is assessed without taking into account the trueness or the accuracy.

3.6.1.6. Graphical Examination of Relative Error versus Concentration Level" (Accuracy Profile approach)

Another procedure which allows the estimation of the LLOQ is the Accuracy Profile approach (also named Total Error Profile approach) which is part of a complete systematic validation strategy, based on the construction of the so called accuracy profiles that reflect the total error of the method [32, 33, 40-42, 111-114].

This approach re-focuses some validation criteria and proposes harmonised protocols by distinguishing, in particular, diagnosis rules and decision rules. These decision rules are based on the use of the Accuracy Profile, which integrates essential elements for the validation in a single graph (or table), i.e. the bias, the precision, and the quantification limits [41].

This powerful tool is based on tolerance intervals for the measurement's error [40]. As it can be seen in Equations 3 and 4, two terms are contained in the tolerance interval: one is trueness (bias or systematic error) and the other one, is the intermediate precision coefficient variation (random error). In this way the tolerance interval expresses the result's accuracy (understood as total error). Furthermore, the tolerance interval integrates an additional dimension, the chance (or risk), for future results, conditionally to past results, to fall within (outside) the acceptance limits [41].

For this approach two parameters must be set by the analyst: the acceptance limits (established by the regulatory Guidelines) and the risk of having future measurements falling outside those acceptance limits (1−β) [115]. If the analyst is ready to assume, for example a risk of a 5% (i.e. a $\pm$ 5% acceptance limit or 95% of tolerance level, β= 95%), (s)he will be able to guarantee that 95 times out of 100 the future measures given by his procedure will be included within the acceptance limits fixed according to the regulatory requirements [40].

In practice, the tolerance interval at each concentration level for the validation standards is computed as follows [33] (in order to achieve a better understanding of this approach the reader is addressed to Hubert *et al*. 's work [41]):

$$\boldsymbol{Low\ Tolerance\ Interval}_{j}\left(LTI_j\right) = Bias_j(\%) - Q_t\left(v;\frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{p\cdot n\cdot B_j^2}}\cdot CV_{IPj}$$

Equation 3

$$\boldsymbol{Upper\ Tolerance\ Interval}_{j}\left(UTI_j\right) = Bias_j(\%) + Q_t\left(v;\frac{1+\beta}{2}\right)\sqrt{1+\frac{1}{p\cdot n\cdot B_j^2}}\cdot CV_{IPj}$$

Equation 4

Where:

- $Bias_j(\%)$ is the trueness of the back-calculated concentration of the j-concentration level

$$RE\% = \frac{x - \mu_T}{\mu_T} \times 100$$

- $Q_t(\upsilon; (1+\beta)/2)$ is the β-quantile of the Student t distribution with υ degrees of freedom

$$\upsilon = \frac{(R+1)^2}{\frac{\left(R + \left(1/n\right)\right)^2}{p-1} + \frac{1-(1/n)}{p \cdot n}}$$

(where β is the chance level, p is the number of series analysed and n is the number of independent replicates per series).

- $B_j = \sqrt{\frac{R_j + 1}{n \cdot R_j + 1}}$
- $R_j = \frac{\hat{\sigma}^2_{B,j}}{\hat{\sigma}^2_{W,j}}$.
- $CV_{IP,j} = \frac{\hat{\sigma}^2_{IP,j}}{\hat{\mu}_j} \times 100$ $CV_{IP,j} = \frac{\hat{\sigma}^2_{IP,j}}{\hat{\mu}_j} \times 100$ is the intermediate precision (in terms of coefficient of variation%) for a j-concentration level, ($\hat{\sigma}^2_{IP,j} = \hat{\sigma}^2_{W,j} + \hat{\sigma}^2_{B,j}$ is the intermediate-precision variance for a j-concentration level, $\hat{\sigma}^2_{W,j}$ the intra-series variance and $\hat{\sigma}^2_{B,j}$ the inter-series variance; and $\hat{\mu}_j$ is mean of the back calculated concentrations of the j-concentration level).

To obtain an Accuracy Profile, the relative error is plotted versus the validation standards concentration levels and then, the lower tolerance limits LTl_j are joined together ($LTl_1 \rightarrow LTl_2 \rightarrow \ldots \rightarrow LTl_m$), as well as the upper tolerance limits UTl_j ($UTl_1 \rightarrow UTl_2 \rightarrow \ldots \rightarrow UTl_m$). Finally, the previously specified acceptance limits are reported in the graph as represented in Figure 3. If the whole Accuracy Profile is within in the acceptance limits, the analytical method is expected to provide

accurate results for its intended purpose and therefore declared valid over the whole range studied. If some point of the profile steps outside these limits then the method cannot be considered valid for that concentration level. In order to define the LLOQ the concentration value in which the 20% acceptability limit line crosses either the lower or higher tolerance limit should be graphically found.

Summarising, this approach not only simplifies the validation process of an analytical procedure, but, also constitutes a visual tool allowing the analyst to evaluate the capability of its method and even its future performance [41]. The analytical interpretation is easy and all the useful required statistics, such as trueness, precision, quantification limits, risk, linearity, are integrated. The use of the Accuracy Profile as single decision tool allows not only to reconcile the objectives of the procedure with those of the validation, but also to visually asses the capacity of the analytical procedure to fit its purpose [40, 111].

3.6.2. Discussion

Despite the consensus existing around the definition of the LLOQ concept, there is a lack of agreement regarding the way this parameter should be calculated. This is highly problematic since it can mean that the LLOQ values given by different laboratories using the same analytical procedure are not necessarily comparable. Consequently, when reporting the quantification limit for a given analytical method, it should be necessary to specify exactly how the limits have been determined.

The S/N approach is one of the most widely used, despite the difficulties in calculating the noise and the significant equipment dependence. Thus, all the approaches based on the measure of the background noise or its standard deviation do not appear to be robust. In fact, Mermet *et al.* confirmed that the estimation of the standard deviation of the blank signal is certainly the weak point of these procedures [110]. The approaches based on the deviations of the slope or the Y-

714 intercept could be an alternative but none of the guides set specific requirements in relation to the
715 calibration curves that should be used in this approach.

716 The EURACHEM approach could provide a good compromise for evaluating the LLOQ. However,
717 it is only able to discriminate based on the obtained precision values. It does not reflect any
718 information about accuracy (trueness), which obviously must be also considered. A possible
719 improvement would include the graphical presentation of %RE obtained from the same
720 experiments, but in that case a calibration curve would be necessary, making the procedure more
721 time-consuming.

722 The Accuracy Profile approach seems to be the best way to take into consideration both the bias and
723 the precision. In this case, the LLOQ is the concentration that fulfils the acceptability limits
724 established for the total error. This approach turns out to be revolutionary, since it involves a
725 substantial change in the concept of validation itself. The resources and efforts are optimised in an
726 obvious way, while the analysis of the results, in addition to be simple and complete, is visual and
727 easy to interpret.

728 ***3.7. Matrix effect***

729 Matrix effect (ME) is a parameter specifically related to mass spectrometry (MS) methods and is
730 defined as the effect of other compounds in the matrix, different than the analyte, on the
731 quantification of the analyte: signal suppression or enhancement, elevated baseline, impact on the
732 extraction or the retention time...[116]. With the spread of MS methods in bionalysis this
733 phenomenon has gained importance in the last years due to the impact that coeluting substances can
734 have in the ionisation efficiency, to the extent of becoming a fundamental parameter in method
735 validation. The high sensitivity and selectivity of the MS technique, especially when coupled to
736 liquid chromatography and used in tandem (LC-MS/MS), has led to a growing trend in methods
737 with little to no sample preparation and minimal chromatographic separation [117]. However, the

high selectivity of LC-MS/MS methods does not eliminate completely the problems caused by coeluting compounds. With less selective techniques, such as photometric detection, these compounds can usually be visually detected as they appear as peaks causing an interference, whereas in LC-MS they may not be visible in the m/z ratio or transition monitorised but still modify the signal of the analyte of interest by suppressing or enhancing it.

Selectivity and ME are terms closely related although they name different events. As it has been said previously, selectivity deals with interferences whose signal overlap with the signal of the analyte of interest. On the contrary, compounds producing matrix effect may not be visible, so that method appears to be selective and yet be affected by signal suppression or enhancement. The use of MS detectors instead of other less selective techniques such as photometric or electrochemical detection, especially when used in multiple reaction monitoring (MRM) mode, reduces the selectivity problem, but it is susceptible to ME complications. Even though ME is characteristic of the MS detection independently of the technique used for separation, it is largely more common in LC than in other techniques such as gas chromatography (GC) analysis, being usually neglected when GC is used. Some studies explore the effect of matrix in the signal intensity when GC/MS is used, but are mainly from environmental or alimentary fields [118-120]. For this reason this section will be focused on the observed ME when LC-MS is used in bioanalytical studies.

Although many theories have been proposed to explain ion suppression [121-123], the exact mechanism of the competition between analyte and coeluting components remains yet unknown. It is well known that ME is generated in the ion source, the interphase between LC and MS, where analytes are desolvated and charged. The most common ion sources used in bioanalysis are electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI). Since ionisation of molecules takes place differently in the different ion sources [124], the mechanisms by which ME is generated are specific for each type of source. Also, ion suppression has been demonstrated to be affected by the ionisation source design [125].

In the ESI source a process of charging and desolvation transforms the analytes in the liquid phase into gas ions that are introduced in MS analyser. Different theories have been proposed to explain this complicated and yet not fully understood process [126-128], but it seems clear that the coeluting compounds interfering with either the desolvation or the charging step alter the ionisation of the analyte. Although they are generally the principal cause, not only endogenous components in the biological matrix (salts, amines, triglycerides…) cause ME, also some exogenous compounds (plasticizers from sample containers or anticoagulants in case of plasma) are susceptible to alter the ionisation process [123]. Furthermore, there are other substances that can be present in the mobile phase and can alter the signal of the analyte by causing ion suppression or enhancement. Nevertheless this is not considered a ME source since it is not sample specific. Usual cases are some mobile phase additives such as non-volatile phosphate or sulphate salts which provoke poor desolvation of droplets [129-131] and components such as surfactants or ion-pairing reagents which make difficult the migration of an analyte to the surface of the droplet.

APCI source is less prone to suffer from ME. Since the analytes are not charged in the liquid phase but in the gas phase [132], droplet generation or desolvation problems are eliminated. Nonetheless, it is not completely free of ME [133]. This ionisation source is also vulnerable to non-volatile compounds and to competition between analyte and coeluting compounds for available charge.

Over the years, several researchers have proposed diverse methods to control and reduce matrix effects in bioanalysis [134-138]. Commonly the attempts to reduce matrix effect go through extensive clean-up, sample dilution, improved chromatography separation or the use of isotope labelled standards. When complex biological matrixes are studied it is recommendable a sample treatment step before the analysis [139, 140]. Therefore, in the choice of the most adequate sample treatment method, its impact in ME should be taken into account. For example, when plasma samples are studied, a major source of suppression comes from the residual phospholipids from the matrix that can be minimised by an extraction step previous to the analysis [141-143]. ME could be

788 considered one of the most important disadvantages of LC-MS methods in bioanalysis. It can
789 diminish or enlarge the signal of the analyte and alter the S/N ratio modifying therefore the LLOQ.
790 Furthermore, ME can vary dramatically among samples affecting the accuracy and precision of the
791 method, and causing lack of linearity, which can lead to deceitful results [144]. In spite of its
792 importance, ME has been traditionally overlooked during method validation and even though
793 current guidelines include the quantification of this parameter as a validation requirement, there are
794 still important voids in this topic especially regarding methodology.

795 The matrix effect phenomenon was first reported by Kebarle and Tang [145] in 1993, who showed
796 that the electrospray response of an analyte decreased as the concentration of other electrolytes was
797 increased. Notwithstanding, it was not until the 2nd Crystal City meeting held in 2000 that matrix
798 effect was considered by the FDA [5], later defined in its guideline as *"the direct or indirect*
799 *alteration or interference in response due to the presence of unintended analytes (for analysis) or*
800 *other interfering substances in the sample"* [6]. However, no methodology was suggested by the
801 FDA to evaluate the existence and magnitude of ME until the white paper published in 2007,
802 product of Crystal City III meeting. In this paper, the matrix factor (MF) is defined as a ratio of the
803 analyte peak response in the presence of matrix ions to the analyte peak response in the absence of
804 matrix ions [7]. According to this last definition, a MF value of one implies that the analysis is not
805 affected by any ME, a MF lower than one indicates ion suppression, and a MF value higher than
806 one indicates ion enhancement. The IS normalised MF (NMF) was also defined as the MF of the
807 analyte divided by the MF of the IS.

808 EMA gave the same definition for ME and NMF on the *Guideline on bioanalytical method*
809 *validation* of 2011 [37]. The definition of MF is expanded saying that the peak area in presence of
810 matrix should be measured by analysing blank matrix spiked after extraction and in absence of
811 matrix by analysing a pure solution of analyte. Both guides agree that the variability in the MF,
812 which would cause lack of reproducibility in the method, must be studied using six lots of blank

matrix from individual donors. In the case of matrices that are difficult to obtain, FDA as well as EMA accept the use of less than six individual blank matrices. It is also admissible according to the FDA guideline to use less than six matrices when the NMF with stable isotope IS is used. EMA emphasises that pooled matrices should not be used. The variability of MF should be determined in terms of %CV. A %CV of 15% is considered by both agencies as the maximum acceptable value.

Although FDA does not include any indication about the concentration at which ME should be studied, EMA suggests that the determination of the MF should be done at two levels of concentration, a lower level of maximum 3 times the LLOQ and a higher level close to the ULOQ, but no indication is given to compare the consistency of the matrix effect at the different levels.

ANVISA did not consider matrix effect among the validation parameters in RE 899/2003 document [38], however, a section focused on matrix effect assessment was included in RDC 27/2012 [29] recommending similar procedure as EMA.

In the case of plasma and serum, EMA suggests that hemolysed and hyperlipidaemic samples should be also studied. ANVISA RDC 27/2012 is more specific and states that for plasma eight samples must be studied, being two of them hemolysed, two hyperlipidaemic and four normal; for total blood six samples are necessary, being two of them hyperlipidemic and four normal. These kinds of samples are rather common and are considered to be particularly affected by ME. Even though ideally all type of populations should be included in a validation, this is unfeasible on a routine basis, but should be considered when those populations are targeted, for example in clinical trials of renally or hepatically impaired populations.

Sometimes, excipients used in formulations remain unaltered in-vivo and can be a source of ME. In these cases, EMA proposes that ME should be determined not only for blank matrix but also for matrix containing these excipients by spiking them directly into the blank matrix. In the case that those excipients are metabolised or somehow altered, matrix should be collected from subjects

837 administered only with the excipients.

838 Finally, Japanese MHLW [30] also briefly considers matrix effect as a parameter of validation when
839 mass spectrometric methods are used. This guideline, similarly to the FDA, recommends to
840 determine the MF by comparing the analyte response in presence and in absence of matrix from 6
841 different sources and that it may be normalised with IS so that the precision should not exceed 15%.

842 *3.7.1. Methodological approach*

843 Matrix effect is mainly assessed by two strategies, a qualitative assessment based on post column
844 infusion described by Bonfiglio *et al.* [129] and a quantitative evaluation as the absolute matrix
845 effect by post extraction addition firstly defined by Matuszewski *et al.* [146].

846 In the first protocol a constant flow of a solution of the analyte is added after the chromatographic
847 column using a "Tee" union while an injection of an extracted blank sample is carried out (Figure
848 4). Same procedure is followed with a blank solvent injection and afterwards both profiles are
849 compared. Ideally, if there is no ME, the profiles of the analyte for the blank sample and the blank
850 solvent should be the same. Ion suppression or enhancement is evidenced by variations in this
851 profile as shown in Figure 5. If more than one analyte is analysed with the same method each
852 analyte must be infused individually to evaluate its ME in order to avoid self-induced ME.

853 This methodology has been vastly used to study ME in bioanalysis [88, 134, 147] and provides a
854 qualitative way to examine the areas of the chromatogram affected by ME. Estimations of ME in
855 terms of percentage of signal intensity must be considered as approximations [136]. Kaufmann and
856 Butcher [148] developed a smart variation of this strategy in order to correct the ME. Segments of
857 analyte solution and analyte-free solvent are alternatively infused post column so that a pulsating
858 signal is registered. In this way the ME is quantitatively evaluated in real time and after data-
859 processing the signal of the analyte can be corrected. Tudela *et al.* [149] proposed the use of a ME
860 marker to identify samples whose ME differ from the accepted values. They found that isotopically

861 labelled budesonide may be a suitable marker to indicate samples with severe ME in the analysis of
862 diuretics in urine.

863 For a quantitative assessment, Matuszewski *et al.* [146] proposed a methodology based on the
864 calculation of ME by comparing the signals from the analysis of three sets of samples: a matrix free
865 standard (A), a sample spiked after the extraction of a blank matrix (B) and the same matrices
866 spiked before extraction (C), all at seven levels of concentration. They define ME as the ratio of the
867 response of B over the response of A, which is equivalent to the MF defined in the guidelines.

868 With these three sets the recovery (REC) and the process efficiency (PE) can be calculated as
869 follows: REC(%) = (C/B)x100 and PE(%) = (C/A)x100 = (RECxME)/100.

870 The advantage of this procedure is that recovery and matrix effect are assessed together [150-153]
871 and can be combined with accuracy and precision studies. Nevertheless it is not a common practice
872 to study them at seven levels. Usually recovery and ME are studied at three levels of concentration
873 (low, medium and high QC) or at two levels as it is recommended by EMA and ANVISA guidelines
874 (close to the LLOQ and the ULOQ). The authors of this approach also proposed assessing ME using
875 five different lots of matrices in order to demonstrate the absence of "relative" ME, referring to the
876 variability of the ME among different sources of the same matrix. However, currently it is more
877 common to use six different sources as recommended by the guidelines.

878 An alternative approach to determine the presence or absence of ME was also proposed by
879 Matuszewski [154]. By this methodology the slope of a calibration curve prepared in matrix is
880 compared to the slope of the same calibration curve prepared in clean solvent. Superposed lines
881 indicate absence of ME whereas differences in the slopes reveal the presence of ion suppression or
882 enhancement. Thus, the relative matrix effect would be assessed by comparison of the slopes
883 obtained in different matrix sources.

884 *3.7.2. Discussion*

885 ME could be considered as one of the major drawbacks of the LC-MS technique, especially when
886 such complex matrices as the bioanalytical ones are studied. As it is not visible unless it is
887 specifically addressed it could go unnoticed and appear as lack of reproducibility, poor precision or
888 deficient accuracy among other problems. Despite its potential deleterious effects, ME was initially
889 overlooked in the international guidelines of validation where it is vaguely defined or not even
890 mentioned. Due to the wide expansion of LC-MS use in bioanalysis laboratories, the assessment of
891 ME has become imperative for the favourable outcome of many bioanalytical methods.
892 Consequently, it is included in the most recent versions of almost all the guidelines. However, the
893 methodology to be followed remains sometimes unclear. In our opinion clear and concise
894 instructions about the methodology to evaluate ME should be given.

895 The quantitative methodology proposed by Matuszewski *et al.* [146] has proven to be adequate to
896 determine the presence or absence of ME interfering with the analyte of interest, and determining
897 the recovery of the method. Contrarily to bioanalytical validation guidelines, the authors
898 recommend applying this methodology at three levels of concentration (low, medium and high QC)
899 simultaneously with the accuracy and precision studies. In this way the number of experiments to
900 perform is minimised and at the same time information of ME along the whole calibration range is
901 obtained.

902 As it happens with the recovery, none of the guidelines gives a value admissible for the MF.
903 According to their indications, as long as ME is reproducible it does not necessarily need to be
904 eliminated, but identified and quantified. Nevertheless, Kollipara *et al.* [107] indicate that most of
905 the bioanalytical laboratories accept 0.80-1.20 as the limit for acceptable IS normalised MF value,
906 and consider ME negligible if it is in ±15% of nominal value with a %CV no more than 15%.

907 ***3.8. Recovery***

908 Recovery measures the ability of a method to extract an analyte from a biological matrix and it is
909 expressed as the percentage of the known amount of the analyte carried through the sample

extraction and processing steps of the method. It is important to point out that there is not a minimum established value for recovery since a bionalytical method with a low recovery could be suitable for a certain analyte if the sensitivity of the detection technique is high enough. Nevertheless, recovery of the analyte and the IS should be precise, reproducible and consistent over the calibration range. Indeed, as happens with ME, the guidelines do not focus on the recovery value (even if 100% is desirable) but in demonstrating that the obtained values are consistent.

According to FDA and ANVISA, recovery is calculated by comparing the analytical response for extracted samples at three concentrations with unextracted standards. Even if it is not explicitly indicated, by this approach the absolute recovery or PE is obtained (see section 3.7), which includes relative recovery and matrix effect. In our opinion both terms should be separately calculated and therefore the methodology for the calculation of recovery defined in the MHLW would be more suitable, where they compare the analyte response in a biological sample spiked before the extraction and processed with the response in a biological blank processed and then spiked with the analyte. This guideline recommends performing the analysis at each concentration at least in triplicate while FDA and ANVISA do not set the number of replicates. To our knowledge, the most efficient way to determine recovery is described in *matrix effect* section, where matrix effect, recovery and process efficiency are calculated simultaneously. The three required concentration levels can be the same ones used for accuracy and precision study.

It is noteworthy that recovery term is mentioned not even once in EMA guideline. This fact did not go unnoticed by the scientific community as it can be observed in the large amount of comments received by EMA during the consultation period regarding this topic [155]. The organisation claims that recovery is an issue to be investigated during the analytical method development and as such is not considered to be included in a validation guideline. The authors agree with EMA in this point and more taking into consideration that extraction recovery reproducibility is implicitly demonstrated when the accuracy, precision and linearity of the method fulfil the acceptance criteria.

935 Nevertheless, from a practical point of view the authors encourage to include relative recovery
936 assay during method validation since the number of experiments to be added is minimal if the
937 procedure is properly designed, and it can provide very useful information to understand the whole
938 analytical procedure.

939 ***3.9. Stability***

940 Studying the stability of the analyte in stock solutions and matrix is vital to ensure the reliability of
941 the results provided by the analytical method [156]. After the first Crystal City meeting only
942 stability during the collection process, sample storage period and two freeze-thaw cycles was
943 demanded. Nowadays instead, the different guidelines ask for more exhaustive studies. These
944 include assesses that cover all the situations that can be encountered during the whole analytical
945 procedure such as freeze-thaw stability, short and long term stability, stock stability and post
946 preparative stability. Even if there is not a complete consensus among the guidelines regarding the
947 last point, it has been already thoroughly discussed by Kollipara *et al*. who distinguished four
948 assays for processed sample stability: wet state stability, processed sample integrity, autosampler
949 stability and reinjection reproducibility [107]. It is noteworthy that FDA guideline does not specify
950 acceptance criteria for the stability exercises whereas the rest of the regulation bodies deemed that
951 the deviation between the calculated and the nominal values should be within ±15%.

952 Additionally to the aforementioned stability assays, during the last bioanalysis meetings, several
953 points regarding stability have been widely discussed such as the stability of the drug in blood
954 before plasma has been separated, stability in hemolysed or hyperlipidaemic samples, influence of
955 coanalysed compounds in analyte stability, the assumption of the stability at -70 ºC if it has been
956 demonstrated at 20 ºC [27, 28, 157]. All these aspects are still matter of debate and should be
957 eventually integrated in the official guidelines.

958 ***3.10. Dilution Integrity***

Sometimes, study samples' concentration may exceed the ULOQ, which means that a dilution and a reassay of the sample are necessary. During the validation procedure it should be demonstrated that this step do not affect the quality of the final results. It is important to emphasise that the dilution should be carried out before the sample treatment using the same matrix than the study samples (e.g., human to human urine). Otherwise, dilution of the processed sample would alter the amount of matrix compared to the calibration standards (unless they are diluted too) leading to unreliable results.

Both FDA and EMA include the demonstration of dilution integrity but only the latter proposes a methodology to study it. At least five replicate samples should be spiked at the highest expected concentration above the ULOQ and properly diluted to fit in the calibration range. The precision and accuracy should be within the set criteria (%CV, %RE 15%) and it should be assessed on at least one of the validation days [158, 159]. ANVISA RDC 27/2012 includes dilution integrity test with the study of precision and accuracy by adding one more QC level named CQD (Dilution Quality Control) in each batch.

3.11. Carryover effect

Carryover is a phenomenon caused by the presence of a residual amount of the analyte in the analytical instrument after an injection, which can affect the accuracy and precision of the results. Carryover can be a great obstacle when developing LC-MS/MS methods in bioanalysis due to its intrinsic high sensitivity and the broad calibration range necessary for some studies. This effect affects to a greater extent the low concentration samples due to the higher percentage influence of the remaining analyte in the measured response. This is especially troubling when the analysis of a low concentration sample is preceded by the analysis of a high concentration one, which can often happen when analysing study samples. But carryover does not affect only to subsequent sample(s), it can be random too, for example, if late eluting compounds are accumulated and eluted several injections later.

Despite the importance of carryover, FDA does not mention it in its official guideline and it is not until the White Paper published in 2007 that this effect is addressed [7]. They recommend injecting one or more blank samples after a high concentration sample or standard and compare the response of the analyte in the blanks with the response at limit of quantification. Even if there is not a cut-off value for the carryover (it should be minimised as much as possible), taking into account that the response at the LLOQ should be at least 5 times higher than the blank response, the carry-over cannot be higher than the 20% the response of the analyte at the LLOQ or 5% the response of the IS. This is indeed the procedure recommended by EMA , MHLW and ANVISA..

In some cases carryover cannot be avoided, hence special measures should be taken such as the use of a fast cleaning gradient or the injection of blank samples between study samples and after high concentration standards or QC samples [160, 161]. In these cases randomisation of the standards or study samples is not recommended in order to minimise the impact of the carryover effect in the accuracy and precision

3.12. Incurred Sample Reanalysis

Incurred Sample Reanalysis (ISR) has been a hot topic in bioanalysis during last years [162]. This topic was brought out in Crystal City III workshop when it was shown that even if the inter-day precision and accuracy was usually less than 6% when analysing incurred samples repeatedly the obtained concentration values could differ by 30-80% [7, 10]. It was obvious that in some cases the QC samples were not able to mimic the study samples mainly due to metabolites conversion to the precursor drug, differences in protein binding, recovery issues, sample inhomogeneity or matrix effects. Due to the importance of this topic a new workshop was organised in 2008 in order to set some recommendations to be followed by bioanalysis laboratories [8].

According to this workshop the number of samples to reanalyse depends on the size of the study, with a minimum of 5% of the samples for the large studies (5-10% for the rest). These samples

should be representative of the study and therefore it is recommended to choose samples around the t_{max} and around the elimination time. The results obtained for the reanalysed samples are then compared with the original values, so the number of replicates used in the analysis should be the same. In order to pass the acceptance criteria for the ISR the concentration value of at least 4 out of 6 samples should agree within the 20%. When EMA released the guideline ISR was included following the same criteria with the only exception that they set the number of reanalysed samples to 10% if the total sample number in the study is less than 1000 and 5% if the number of samples exceeds this number, which is the same procedure that MHLW draft guideline recommends. In the last FDA draft the number of ISR samples is set in the 7% of the study sample size, regardless the size. Both guidelines agree that ISR is at least expected for in vivo human bioequivalence studies, all pivotal pharmacokinetic/pharmacodynamic studies and toxicokinetic studies.

3.13. System suitability in routine drug analysis

Although this review is focused on method validation, system suitability and acceptance criteria for the run when applied to routine analysis will be shortly discussed due to the connection with the topic and the fact that the guidelines also address these points. System suitability is a test to confirm that the instrument performance is adequate and is carried out before running the batch, usually using a standard solution [108]. The parameters to be studied during system suitability test are selected before the validation procedure and depend on the analytical instrument to be used: retention time or peak width for chromatographic methods; analyte response or fragmentation ratios for MS detectors. Once the proper instrument performance has been demonstrated by verifying that the values of those parameters are inside the tolerance range the analytical run can be carried out.

An analytical run consists of a calibration curve (blank sample, zero sample and calibration standards), QC samples at least at three different concentrations in duplicate (low, medium and high concentrations) and the study samples. The QC samples and the calibration standards should be spiked using different stock solutions and the whole batch should be processed together without

interruption. In the case of large studies the number of QC samples should be at least the 5% of the number of studied samples.

During method validation the criteria explained in point 3.5 should be applied to the calibration standards and the regression model should be the one used during method validation. According to EMA, even if the calibration standards defined as LLOQ or ULOQ do not fulfil the accuracy criteria the calibration curve can be used, as long as six calibration standards meet the criteria and all the QC samples remain inside the calibration range. In those cases, the LLOQ for that run is the next lowest concentration standard and the ULOQ the next highest concentration standard. FDA does not give specific information about this fact but states that the values falling out the accuracy limits can be discarded, provided they do not change the model, so the same approach can be followed. In case the sample concentration falls below the LLOQ, it should not be reported as zero but as "BQL" (below the quantification limit) and it should not be used for any pharmacokinetic calculations [163].

Concerning the QC samples at least 4 out of 6 samples should meet the accuracy criteria, always taking into account that at least one of the replicates for each concentration level should be within the 15% of the nominal values (4-6-15 rule). The authors recommend designing the calibration curve and the QC samples in such a way that the concentration of the lowest QC sample is slightly higher than the second calibration standard and the concentration of the highest QC sample is lower than the second to last calibration standard. This allows the rejection of calibration curves when one of the extreme points does not fulfil the acceptance criteria. Another approach to minimise the risk of discarding a whole analytical run due to a single point out of the accepted range is to run duplicate calibration standards only for the LLOQ and the ULOQ. In that way, even if only one of the replicates would meet the acceptance criteria the run could be accepted.

4. Discussion

Current bioanalytical validation guidelines have demonstrated to be essential documents to orientate bioanalytical laboratories in method validation and to set the criteria that the developed methods should fulfil. Nevertheless, as it has been thoroughly discussed along this review, these guidelines are too ambiguous when defining some parameters, the methodology to be followed or when setting some of the acceptance criteria.

Among the established criteria for validation all guidelines agree that the accuracy and precision, expressed as %RE and %CV values respectively, should be lower than 15% (20% for LLOQ). This acceptance criterion is easy to implement but it can lead to unreliable results. On the one hand, % RE is calculated using the mean of the measures and therefore it is possible to arrive to the paradoxical situation where the accuracy and precision are acceptable but the bias of more than half of the measures is higher than 15% (e.g. a sample with a reference value of 100 a.u. for which the measure of 5 replicates are 84, 86, 84, 100 and 116). On the other hand both acceptance criteria are independent from each other and consequently there is not a combined measure of the bias and the precision. This means that a method could be validated with both %CV and %RE values close to 15%. Once again in this extreme situation the bias of most of the measures could be unacceptable. Hubert *el al*. [40] already stated that for a normal distribution with a %CV of 15% and no relative error only 66% of the measures would have a bias less than 15%. Obviously when increasing the relative error until being close to 15% this number would decrease (to around only 45% of the measures with a bias lower than 15%). It means that this hypothetical method that could be successfully validated according to current regulations, would hardly meet the 4-6-15 rule set for routine analysis. This fact is really contradictory taking into account that the main aim of validation is to demonstrate that the validated method offers reliable results during routine analysis. Obviously this scenario should be avoided in bioanalysis and an approach that implies the total error (accuracy + precision) should be considered. An interesting option is to use the Accuracy Profile approach explained in *limit of quantification* section in which the construction of the accuracy profiles and

1082 the use of the acceptability limits lead to more reliable results besides providing more complete
1083 information regarding the regression mode and the working range.

1084 This approach does not only offer a different way to process the data obtained from validation
1085 experiments but a new methodology to face the whole validation procedure about which some other
1086 authors have discussed earlier [32, 40, 43, 55].

1087 While bioanalytical validation guidelines define the majority of the validation parameters'
1088 acceptance criteria (with the exception of robustness), they leave a wide margin for interpretation in
1089 terms of methodology. The required content is clear, but not the overall process for obtaining it, i.e.,
1090 a validation program is not defined and the consecution of steps which the analyst must accomplish
1091 is not clearly expressed. Obviously the validation guidelines cannot just consist of a rigorous
1092 schedule detailing all the steps to follow because there should be room for modifications depending
1093 on each particular analysis but nevertheless, reaching some standardisation degree in the
1094 methodology would also help to achieve a more harmonised validation procedure in bioanalysis.

1095 The lack of agreement in methodological approach to calculate the LLOQ and the little attention
1096 that is paid on ME have been already carefully discussed together with the omission of robustness
1097 assays. Another issue not covered by the current regulations relates to the quantification of
1098 biomarkers. In the last years the advances in the field of personalised medicine and metabolomics
1099 have led to the need of quantifying these endogenous compounds that can indicate a certain
1100 biological condition such as an illness. Unfortunately, the bioanalytical guidelines are more
1101 addressed to exogenous compounds such as drugs and their metabolites and therefore no indications
1102 can be found related to the quantification of endogenous compounds. Aware of this situation, FDA
1103 included a section in the last draft of the guideline update called *"special issues"* where information
1104 about biomarkers and endogenous compound analysis is shown. They propose the validation
1105 procedure to be carried out using analyte free matrix, what is usually complicated in these cases.
1106 The use of alternative matrices (e.g. use a buffer instead of urine) is justified provided there is not

matrix effect difference whereas the possibility of using alternative methods such standard additions is not yet considered. The growing interest in quantification of biomarkers will probably result in a rise of the relevance of this topic in the bioanalysis field.

5. Conclusions

Bioanalytical method validation is a fundamental process to confirm that the results obtained for a certain analyte in a biological matrix are reliable. Despite the need of trustworthy results in all the fields involving drug or metabolite analysis (toxicology, pharmacokinetic studies, bioavailability studies…) there are still some vague aspects that have not been properly addressed by the different regulatory bodies.

After carefully reviewing the regulatory guidances and scientific works dealing with bioanalytical method validation, the authors concluded that the most controversial validation parameters are robustness, lower limit of quantification (LLOQ) and matrix effect.

Robustness assay is not yet mandatory according to the regulatory bodies, taking into consideration the importance of this parameter when routine analysis is carried out, a deep discussion about the different approaches to evaluate it was included. Experimental design emerges to be a very powerful tool for this aim, allowing a fast, systematic and easy to implement methodology [73].

Regarding to LLOQ, main bioanalytical validation guidelines agree in the definition of the LLOQ concept, but do not give information about the procedure to follow for its calculation, probably because it is considered to be evaluated in a pre-validation step. This turns problematic since the LLOQ values given by different laboratories using the same analytical procedure are not necessarily comparable. The Accuracy Profile approach seems to be the best way to estimate LLOQ, integrating trueness, precision, risk and linearity. The use of the Accuracy Profile provides a tool to assess the capacity of the analytical procedure to fit its purpose.

Even though in the last versions of the main validation guidelines ME has been included as an essential validation parameter, the specific methodology to be used remains imprecise. As ME could be considered one of the major drawbacks of the LC-MS technique, especially when complex matrices are studied, more precise directions should be given in order to avoid mismanagement of such an important phenomenon. The methodology proposed by Matuszewski appears to be adequate to determine the presence or absence of ME.

2001 FDA guideline is still followed in many laboratories although taking into account nowadays technology and requirements some of its aspects are out of date. The appearance of EMA guideline in 2011 solved some of the issues that were not covered by its predecessor such as ME or ISR, but there are still some aspects that should be included or improved in the following updates: robustness studies, validation of biomarkers and endogenous compounds, stability issues, acceptance criteria for IS repeatability, conditions under which partial validation is allowed... FDA 2013 guideline will deal with some of these points but due to the unceasing development in bioanalysis it is necessary that all the concerned parties keep involved in order to reach harmonised criteria for bioanalytical method validation.

Acknowledgements

Authors thank the Ministry of Science and Education, the Basque Country Government and the University of Basque Country/EHU for financial support (CTQ09-1169, IT89-13, IT789-13, PPM12/06 and UFI 11/23) and the SGIker support (UPV/EHU, MICINN, GV/EJ, ERDF and ESF). O. González and M.E. Blanco also thank the Basque Country Government for their postdoctoral and predoctoral grants respectively.

REFERENCES

[1] International Organisation for Standardisation, ISO/IEC17025, *General requirements for the competence of testing and calibration laboratories*, 2005.

[2] International Organisation for Standardisation, ISO/IEC15819, *Medical laboratories — Particular requirements for quality and competence*, 2012.

[3] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman, S. Spector, K.S. Albert, S. Bolton, M. Dobrinska, W. Doub, M. Eichelbaum, J.W.A. Findlay, K. Gallicano, W. Garland, D.J. Hardy, J.D. Hulse, H.T. Karnes, R. Lange, W.D. Mason, G. McKay, E. Ormsby, J. Overpeck, H.D. Plattenberg, G. Shiu, D. Sitar, F. Sorgel, J.T. Stewart, L. Yuh, Pharmaceutical research, 9 (1992) 588-592.

[4] U.S Department of Health and Human Services, Food and Drug Administration, *Draft Guidance for Industry: Bioanalytical Method Validation*, 1999.

[5] V.P. Shah, K.K. Midha, J.W. Findlay, H.M. Hill, J.D. Hulse, I.J. McGilveray, G. McKay, K.J. Miller, R.N. Patnaik, M.L. Powell, A. Tonelli, C.T. Viswanathan, A. Yacobi, Pharmaceutical research, 17 (2000) 1551-1557.

[6] U.S Department of Health and Human Services, Food and Drug Administration, *Guidance for Industry: Bioanalytical Method Validation*, 2001.

[7] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, M.J. Rose, J. Sailstad, V.P. Shah, J.P. Skelly, P.G. Swann, R. Weiner, Pharmaceutical research, 24 (2007) 1962-1973.

[8] D.M. Fast, M. Kelley, C.T. Viswanathan, J. O'Shaughnessy, S.P. King, A. Chaudhary, R. Weiner, A.J. DeStefano, D. Tang, The AAPS journal, 11 (2009) 238-241.

[9] S. Bansal, A. DeStefano, The AAPS journal, 9 (2007) E109-114.

[10] M.L. Rocci, Jr., V. Devanarayan, D.B. Haughey, P. Jardieu, The AAPS journal, 9 (2007) E336-E343.

[11] U.S Department of Health and Human Services, Food and Drug Administration, *Guidance for Industry: Bioanalytical Method Validation. Draft Guidance*, 2013.

[12] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), *Validation of Analytical Procedures: Text and Methodology Q2(R1)*, 2005.

[13] European Medicines Agency, EMEA/CHMP/EWP/531305/2008, *Concept paper/recommendations on the need for a guideline on the validation of bioanalytical methods*, 2008.

[14] B. Lausecker, P. van Amsterdam, M. Brudny-Kloeppel, S. Luedtke, P. Timmerman, Bioanalysis, 1 (2009) 873-875.

[15] European Medicines Agency, EMEA/CHMP/EWP/192217/2009, *Draft Guidance on Validation of Bioanalytical Methods*, 2009.

[16] R.W. Abbott, Bioanalysis, 2 (2010) 703-708.

[17] P. Timmerman, S. Lowes, D.M. Fast, F. Garofolo, Bioanalysis, 2 (2010) 683.

[18] G. Smith, Bioanalysis, 2 (2010) 929-935.

[19] P. van Amsterdam, M. Arnold, S. Bansal, D. Fast, F. Garofolo, S. Lowes, P. Timmerman, E. Woolf, Bioanalysis, 2 (2010) 1801-1803.

[20] H. Blume, E. Brendel, M. Brudny-Kloeppel, S. Grebe, B. Lausecker, G. Rohde, C. Siethoff, European Journal of Pharmaceutical Sciences, 42 (2011) 300-305.

[21] M. Imazato, Bioanalysis, 5 (2013) 281-283.

[22] F. Garofolo, The Canadian Workshop on the Crystal City AAPS/FDA White Paper, 2007.

[23] N. Savoie, B.P. Booth, T. Bradley, F. Garofolo, N.C. Hughes, S. Hussain, S.P. King, M. Lindsay, S. Lowes, E. Ormsby, R. Phull, M.L. Rocci, Jr., P.T. Vallano, A. Viau, Z. Zhu, Bioanalysis, 1 (2009) 19-30.

[24] N. Savoie, F. Garofolo, P. van Amsterdam, B.P. Booth, D.M. Fast, M. Lindsay, S. Lowes, R. Masse, L. Mawee, E. Ormsby, R. Phull, M.L. Rocci, Jr., P.T. Vallanou, X. Yin, Bioanalysis, 2 (2010) 53-68.

[25] N. Savoie, F. Garofolo, P. van Amsterdam, S. Bansal, C. Beaver, P. Bedford, B.P. Booth, C. Evans, M. Jemal, M. Lefebvre, A.L. Lopes de Silva, S. Lowes, J.C. Marini, R. Masse, L. Mawer, E. Ormsby, M.L. Rocci, Jr., C.T. Viswanathan, J. Wakelin-Smith, J. Welink, J.T. White, E. Woolf, Bioanalysis, 2 (2010) 1945-1960.

[26] F. Garofolo, M.L. Rocci, Jr., I. Dumont, S. Martinez, S. Lowes, E. Woolf, P. van Amsterdam, S. Bansal, A.C. Gomes Barra, R. Bauer, B.P. Booth, M. Carrasco-Triguero, B. DeSilva, J. Dunn, K. Gallicano, D. Gouty, S. Ho, R. Hucker, M. Jemal, N. Katori, O. Le Blaye, J. Lee, W. Li, S. Michael, C. Nehls, R. Nicholson, E. Ormsby, D. Tang, C.T. Viswanathan, R. Weiner, G. Young, Bioanalysis, 3 (2011) 2081-2096.

[27] B. DeSilva, F. Garofolo, M. Rocci, S. Martinez, I. Dumont, F. Landry, C. Dicaire, G. Szekely-Klepser, R. Weiner, M. Arnold, S. Bansal, K. Bateman, R. Bauer, B. Booth, S. Davis, S. Dudal, D. Gouty, J. Grundy, S. Haidar, R. Hayes, M. Jemal, S. Kaur, M. Kelley, M. Knutsson, O. Le Blaye, J. Lee, S. Lowes, M. Ma, T. Mitsuoka, J.T. Neto, R. Nicholson, E. Ormsby, J. Sailstad, L. Stevenson, D. Tang, J. Welink, C.T. Viswanathan, L. Wang, E. Woolf, E. Yang, Bioanalysis, 4 (2012) 2213-2226.

[28] M. Boterman, M. Doig, M. Breda, S. Lowes, J. Jersey, R. Shoup, F. Garofolo, I. Dumont, S. Martinez, S. Needham, M. Cruz Caturla, P. Couerbe, J. Guittard, J. Maltas, T. Lansing, M. Bhatti, C. Schiebl, P. Struwe, C. Sheldon, R. Hayes, T. Sangster, C. Pattison, J. Bouchard, L. Goodwin, R. Islam, R. Segers, Z. Lin, J. Hillier, W. Garofolo, D. Zimmer, L. Folguera, T. Zimmermann, M. Pawula, M. Moussallie, L.d.S. Teixeira, T. Rocha, D. Tang, P. Jardieu, J. Truog, J. Lin, R. Lundberg, C. Cox, A. Breau, C. Bigogno, D. Schoutsen, C. Dilger, M. Bouhajib, A. Levesque, S. Gagnon-Carignan, R. Nicholson, R. Jenkins, M.H. Lin, S. Karnik, T. De Boer, R. Houghton, R. Green, W. DeMaio, R. Sable, K. Smith, C. Siethoff, L. Cojocaru, M. Allen, T. Harter, S. Fatmi, F. Sayyarpour, M. Malone, S. Best, X. Fang, Bioanalysis, 4 (2012) 651-660.

[29] Agência Nacional de Vigilancia Sanitária, RDC 27/2012*, Dispõe sobre os requisitos mínimos para a validação de métodos bioanalíticos empregados em estudos com fins de registro e pós-registro de medicamentos*, 2012.

[30] Ministry of Health, Labour and Welfare, *Draft Guideline on Bioanalytical Method Validation in Pharmaceutical Development*, 2013.

[31] Asociación Española de Farmacéuticos de la Industria, *Validación de Métodos Analíticos*, 2001.

[32] P. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandejan, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, Analytica chimica acta, 391 (1999) 135-148.

[33] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, Journal of pharmaceutical and biomedical analysis, 45 (2007) 70-81.

[34] C.J. Sennbro, M. Knutsson, P. Timmerman, P. van Amsterdam, Bioanalysis, 3 (2011) 2389-2391.

[35] C.J. Sennbro, M. Knutsson, P. van Amsterdam, P. Timmerman, Bioanalysis, 3 (2011) 2393-2399.

[36] D.C. Muirhead, T.S. Smart, Chromatographia, 52 (2000) S72-S75.

[37] European Medicines Agency, EMEA/CHMP/EWP/192217/2009, *Guideline on Bioanalytical Method Validation*, 2011.

[38] Agência Nacional de Vigilância Sanitária, RE 899/2003, *Guide for Validation of Analytical and Bioanalytical Methods*, 2003.

[39] Eurachem, *The Fitness for Purpose of Analytical Methods. Guide to Method Validation and Related Topics*, 1998.

[40] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, Journal of pharmaceutical and biomedical analysis, 36 (2004) 579-586.
[41] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet, Journal of pharmaceutical and biomedical analysis, 45 (2007) 82-96.
[42] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, E. Rozet, Journal of pharmaceutical and biomedical analysis, 48 (2008) 760-771.
[43] E. Rozet, R.D. Marini, E. Ziemons, B. Boulanger, P. Hubert, Journal of pharmaceutical and biomedical analysis, 55 (2011) 848-858.
[44] B.A. Persson, J. Vessman, Trac-Trends in Analytical Chemistry, 20 (2001) 526-532.
[45] J. Vessman, Accreditation and Quality Assurance, 6 (2001) 522-527.
[46] J. Vessman, R.I. Stefan, J.F. Van Staden, K. Danzer, W. Lindner, D.T. Burns, A. Fajgelj, H. Muller, Pure & Appl. Chem., 73 (2001) 1381-1386.
[47] J. Vessman, Journal of pharmaceutical and biomedical analysis, 14 (1996) 867-869.
[48] H.Y. Aboul-Enein, Accreditation and Quality Assurance, 5 (2000) 180-181.
[49] B. Lausecker, in: *EBF Recommendation on testing of co-medications and interaction compounds*, EBF Open symposium, Barcelona, Spain, 2013, http://bcn2013.europeanbioanalysisforum.eu/ (last access March 2014).
[50] L. Kristoffersen, E.L. Oiestad, M.S. Opdal, M. Krogh, E. Lundanes, A.S. Christophersen, J Chromatogr B, 850 (2007) 147-160.
[51] L. Zufia, A. Aldaz, N. Ibanez, C. Viteri, Journal of pharmaceutical and biomedical analysis, 49 (2009) 547-553.
[52] N. Ansermot, M. Brawand-Amey, C.B. Eap, J Chromatogr B, 885 (2012) 117-130.
[53] L. Hu, L. Ding, X. Li, N. Zhou, S. Zhong, G. Wu, H. Zhang, J Chromatogr B, 899 (2012) 8-13.
[54] Joint Commitee for Guides in Metrology, *Evaluation of Measurement Data. Guide to the Expression of Uncertainty in Measurements*, 2008.
[55] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, Journal of chromatography. A, 1158 (2007) 111-125.
[56] International Organisation for Stardardisation, ISO5725-6, *Accuracy (trueness and precision) of measurement methods and results. Use in practice of accuracy values*, 1994.
[57] M. Thompson, S.L.R. Ellison, R. Wood, Pure & Appl. Chem., 74 (2002) 835-855.
[58] U.S. Pharmacopeial Convention, USP<1225>, *Validation of Compendial Methods*.
[59] M. Zeaiter, J.M. Roger, V. Bellon-Maurel, Trends in Analytical Chemistry, 23 (2004) 157-170.
[60] W.J. Youden, E.H. Steiner, Statistical Manual of the Association of Official Analytical Chemists, The Association of Official Analytical Chemists, Washington, DC, 1975.
[61] 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*, 2002.
[62] M. Aouri, A. Calmy, B. Hirschel, A. Telenti, T. Buclin, M. Cavassini, A. Rauch, L.A. Decosterd, Journal of mass spectrometry : JMS, 48 (2013) 616-625.
[63] M. Nieddu, G. Boatto, M.A. Pirisi, E. Azara, M. Marchetti, Journal of chromatography. B, Analytical technologies in the biomedical and life sciences, 867 (2008) 126-130.
[64] M. Andersson, E. Gustavsson, N. Stephanson, O. Beck, Journal of chromatography. B, Analytical technologies in the biomedical and life sciences, 861 (2008) 22-28.
[65] S. Ayciriex, A. Regazzetti, M. Gaudin, E. Prost, D. Dargere, F. Massicot, N. Auzeil, O. Laprevote, Analytical and bioanalytical chemistry, 404 (2012) 3049-3059.

[66] D.V. Bharathi, A. Naidu, B. Jagadeesh, K.N. Laxmi, P.R. Laxmi, P.R. Reddy, R. Mullangi, Biomedical chromatography : BMC, 22 (2008) 645-653.
[67] S.F. Teunissen, H. Rosing, R.H. Koornstra, S.C. Linn, J.H. Schellens, A.H. Schinkel, J.H. Beijnen, J Chromatogr B, 877 (2009) 2519-2529.
[68] R.J. Dhawan, R. Ravi, T. Subburaju, H. Revathi, C. Arul, N. Ganapathi, Int J Pharm Sci Res, 4 (2013) 1227-1232.
[69] G. Kher, V. Ram, M. Kher, H. Joshi, Res J Pharm, Biol Chem Sci, 4 (2013) 365-377.
[70] R. Yanamandra, C.S. Vadla, U. Puppala, B. Patro, Y.L. Murthy, P.A. Ramaiah, Scientia pharmaceutica, 80 (2012) 101-114.
[71] R. Brunetto Mdel, Y. Contreras, Y. Delgado, M. Gallignani, J.M. Estela, V.C. Martin, Journal of chromatographic science, 47 (2009) 478-484.
[72] G.B. Kasawar, M. Farooqui, Journal of pharmaceutical and biomedical analysis, 52 (2010) 19-29.
[73] Y. Vander Heyden, A. Nijhuis, J. Smeyers-Verbeke, B.G. Vandeginste, D.L. Massart, Journal of pharmaceutical and biomedical analysis, 24 (2001) 723-753.
[74] B. Dejaegher, Y.V. Heyden, Journal of chromatography. A, 1158 (2007) 138-157.
[75] D. Awotwe-Otoo, C. Agarabi, P.J. Faustino, M.J. Habib, S. Lee, M.A. Khan, R.B. Shah, Journal of pharmaceutical and biomedical analysis, 62 (2012) 61-67.
[76] L. Schmidt, J. Muller, T. Goen, Analytical and bioanalytical chemistry, 405 (2013) 2019-2029.
[77] K. Monks, I. Molnar, H.J. Rieger, B. Bogati, E. Szabo, Journal of chromatography. A, 1232 (2012) 218-230.
[78] M. Enrique, E. Garcia-Montoya, M. Minarro, A. Orriols, J.R. Tico, J.M. Sune-Negre, P. Perez-Lozano, Journal of chromatographic science, 46 (2008) 828-834.
[79] M. Giorgi, G. Ye, M. Sgorbini, M. Corazza, Journal of chromatographic science, 49 (2011) 79-83.
[80] R. Zeleny, S. Harbeck, H. Schimmel, Journal of chromatography. A, 1216 (2009) 249-256.
[81] B. Chou, R.S. Adler, M. Meng, S. Percey, B. Dean, C.E. Hop, Y.G. Shin, Journal of pharmaceutical and biomedical analysis, 70 (2012) 354-361.
[82] H.T. Karnes, C. March, Journal of pharmaceutical and biomedical analysis, 9 (1991) 911-918.
[83] L. Kirkup, M. Mulholland, Journal of chromatography. A, 1029 (2004) 1-11.
[84] S. Kollipara, G. Bende, N. Agarwal, B. Varshney, J. Paliwal, Chromatographia, 73 (2011) 201-217.
[85] A.M. Almeida, M.M. Castel-Branco, A.C. Falcao, J Chromatogr B, 774 (2002) 215-222.
[86] T. Alebic-Kolbah, M.S. Modesitt, Analytical and bioanalytical chemistry, 404 (2012) 2043-2055.
[87] O. Gonzalez, G. Iriarte, E. Rico, N. Ferreiros, M. Itxaso Maguregui, R. Maria Alonso, R. Maria Jimenez, J Chromatogr B, 878 (2010) 2685-2692.
[88] O. Gonzalez, R.M. Alonso, N. Ferreiros, W. Weinmann, R. Zimmermann, S. Dresen, J Chromatogr B, 879 (2011) 243-252.
[89] R.W. Sparidans, S. Durmus, A.H. Schinkel, J.H.M. Schellens, J.H. Beijnen, J Chromatogr B, 925 (2013) 124-128.
[90] N. Roth, B. Moosmann, V. Auwaerter, Journal of Mass Spectrometry, 48 (2013) 227-233.
[91] N.G. Knebel, R. Herzog, F. Reutter, K. Zech, J Chromatogr B, 893 (2012) 82-91.
[92] J. Corley, Best practices in establishing detection and quantification limits for pesticide residues in foods, in: P.W. Lee (Ed.) Handbook of Residue Analytical Methods for Agrochemicals, John Wiley & Sons Ltd., West Sussex, England, 2003, pp. 59-74.
[93] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, Journal of pharmaceutical and biomedical analysis, 17 (1998) 193-218.
[94] F. Bressolle, M. Bromet-Petit, M. Audran, J Chromatogr B, 686 (1996) 3-10.

[95] F.T. Peters, H.H. Maurer, Accreditation and Quality Assurance, 7 (2002) 441-449.
[96] R. Boque, A. Maroto, J. Riu, F.X. Rius, Grasas y Aceites, 53 (2002) 128-143.
[97] L.A. Currie, Pure & Appl. Chem., 67 (1995) 1699-1723.
[98] G.L. Long, J.D. Winefordner, Analytical chemistry, 55 (1983) 712A-724A.
[99] A. Hubaux, G. Vos, Analytical chemistry, 42 (1970) 849-855.
[100] C.A. Clayton, J.W. Hines, P.D. Elkins, Analytical chemistry, 59 (1987) 2506-2514.
[101] U.S. Environmental Protection Agency, *Assigning values to non-detected/non-quantified pesticide residues in human health food exposure assessments*, 2000.
[102] European Pharmacopoeia, *Technical Guide for the elaboration of monographs*, 2011.
[103] F.T. Peters, O.H. Drummer, F. Musshoff, Forensic science international, 165 (2007) 216-224.
[104] J.P. Foley, J.G. Dorsey, Chromatographia, 18 (1984) 503-511.
[105] J. Vial, A. Jardy, Analytical chemistry, 71 (1999) 2672-2677.
[106] J. Vial, K. Le Mapihan, A. Jardy, Chromatographia, 57 (2003) s303-s306.
[107] S. Kollipara, G. Bende, N. Agarwal, V. Brijesh, J. Paliwal, Cromatographia, 73 (2011) 201-217.
[108] L. Huber, Validation of analytical methods, Agilent Technologies, 2010.
[109] A.C. Mehta, Journal of clinical pharmacy and therapeutics, 14 (1989) 465-473.
[110] J.-M. Mermet, Spectrochimica Acta Part B: Atomic Spectroscopy, 63 (2008) 166-182.
[111] B. Boulanger, P. Chiap, W. Dewe, J. Crommen, P. Hubert, Journal of pharmaceutical and biomedical analysis, 32 (2003) 753-765.
[112] M. Feinberg, B. Boulanger, W. Dewe, P. Hubert, Analytical and bioanalytical chemistry, 380 (2004) 502-514.
[113] B. Boulanger, W. Dewe, A. Gilbert, B. Govaerts, M. Maumy-Bertrand, Chemometrics and Intelligent Laboratory Systems, 86 (2007) 198-207.
[114] M. Feinberg, Journal of chromatography. A, 1158 (2007) 174-183.
[115] E. Rozet, C. Hubert, A. Ceccato, W. Dewe, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, P. Hubert, Journal of chromatography. A, 1158 (2007) 126-137.
[116] C. Dicaire, J.-N. Mess, M. Furtado, A. Bergeron, F. Garofolo, Bioanalysis, 5 (2013) 2359-2370.
[117] O. Nunez, H. Gallart-Ayala, C.P. Martins, P. Lucci, R. Busquets, J Chromatogr B, 927 (2013) 3-21.
[118] C.F. Poole, Journal of chromatography. A, 1158 (2007) 241-250.
[119] E. Villaverde-de-Saa, C. Valls-Cantenys, J.B. Quintana, R. Rodil, R. Cela, Journal of chromatography. A, 1300 (2013) 85-94.
[120] S. Yu, X.M. Xu, Rapid communications in mass spectrometry : RCM, 26 (2012) 963-977.
[121] F. Gosetti, E. Mazzucco, D. Zampieri, M.C. Gennaro, Journal of chromatography. A, 1217 (2010) 3929-3937.
[122] H. Trufelli, P. Palma, G. Famiglini, A. Cappiello, Mass spectrometry reviews, 30 (2011) 491-509.
[123] H. Mei, Y. Hsieh, C. Nardo, X. Xu, S. Wang, K. Ng, W.A. Korfmacher, Rapid communications in mass spectrometry : RCM, 17 (2003) 97-103.
[124] N.B. Cech, C.G. Enke, Mass spectrometry reviews, 20 (2001) 362-387.
[125] C. Ghosh, C.P. Shinde, B.S. Chakraborty, J Chromatogr B, 893-894 (2012) 193-200.
[126] P. Kebarle, Journal of mass spectrometry : JMS, 35 (2000) 804-817.
[127] J.V. Iribarne, B.A. Thomson, The Journal of chemical physics, 64 (1976) 2887-2894.
[128] M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice, The Journal of chemical physics, 49 (1968) 2240-2249.
[129] R. Bonfiglio, R.C. King, T.V. Olah, K. Merkle, Rapid communications in mass spectrometry : RCM, 13 (1999) 1175-1185.

[130] Revista de Saude Publica, 40 (2006) 361-363.
[131] R. King, R. Bonfiglio, C. Fernandez-Metzler, C. Miller-Stein, T. Olah, Journal of the American Society for Mass Spectrometry, 11 (2000) 942-950.
[132] J. Sunner, G. Nicol, P. Kebarle, Analytical chemistry, 60 (1988) 1300-1307.
[133] B.M. Fong, S. Tam, K.S. Leung, Talanta, 116 (2013) 115-121.
[134] S. Souverain, S. Rudaz, J.L. Veuthey, Journal of chromatography. A, 1058 (2004) 61-66.
[135] E. Chambers, D.M. Wagrowski-Diehl, Z. Lu, J.R. Mazzeo, J Chromatogr B, 852 (2007) 22-34.
[136] R. Dams, M.A. Huestis, W.E. Lambert, C.M. Murphy, Journal of the American Society for Mass Spectrometry, 14 (2003) 1290-1294.
[137] A. Gonzalez-Antuna, J.C. Dominguez-Romero, J.F. Garcia-Reyes, P. Rodriguez-Gonzalez, G. Centineo, J.I. Garcia Alonso, A. Molina-Diaz, Journal of chromatography. A, 1288 (2013) 40-47.
[138] I. Marchi, S. Rudaz, J.L. Veuthey, Journal of pharmaceutical and biomedical analysis, 49 (2009) 459-467.
[139] C. Muller, P. Schafer, M. Stortzel, S. Vogt, W. Weinmann, J Chromatogr B, 773 (2002) 47-52.
[140] M. Yadav, V. Trivedi, V. Upadhyay, G. Shah, G.A. Baxi, S. Goswami, P.S. Shrivastav, J Chromatogr B, 885-886 (2012) 138-149.
[141] S. Ahmad, H. Kalra, A. Gupta, B. Raut, A. Hussain, M.A. Rahman, Journal of pharmacy & bioallied sciences, 4 (2012) 267-275.
[142] V. Pucci, S. Di Palma, A. Alfieri, F. Bonelli, E. Monteagudo, Journal of pharmaceutical and biomedical analysis, 50 (2009) 867-871.
[143] O.A. Ismaiel, T. Zhang, R.G. Jenkins, H.T. Karnes, J Chromatogr B, 878 (2010) 3303-3316.
[144] J. Schuhmacher, D. Zimmer, F. Tesche, V. Pickard, Rapid communications in mass spectrometry : RCM, 17 (2003) 1950-1957.
[145] P. Kebarle, L. Tang, Analytical chemistry, 65 (1993) 972A-986A.
[146] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Analytical chemistry, 75 (2003) 3019-3030.
[147] M. Pickering, S. Brown, Biomedical chromatography : BMC, 27 (2013) 568-574.
[148] A. Kaufmann, P. Butcher, Rapid communications in mass spectrometry : RCM, 19 (2005) 611-617.
[149] E. Tudela, G. Munoz, J.A. Munoz-Guerra, J Chromatogr B, 901 (2012) 98-106.
[150] L. Zhang, L. Teng, C. Gong, W. Liu, X. Cheng, S. Gu, Z. Deng, Z. Wang, C. Wang, Journal of pharmaceutical and biomedical analysis, 85 (2013) 162-168.
[151] Y. Zhou, M. Liu, J. Jiang, H. Wang, P. Hu, J Chromatogr B, 939 (2013) 59-66.
[152] X.Y. Wang, X.H. Ma, W. Li, Y. Chu, J.H. Guo, S.M. Li, J.M. Wang, H.C. Zhang, S.P. Zhou, Y.H. Zhu, Journal of pharmaceutical and biomedical analysis, 86 (2013) 82-91.
[153] B. Wang, S. Ji, H. Zhang, L. Zhao, L. Lv, Y. Li, G. Zhou, G. Zhang, Steroids, 78 (2013) 1164-1170.
[154] B.K. Matuszewski, J Chromatogr B, 830 (2006) 293-300.
[155] European Medicines Agency, EMA/434565/2010, *Overview of comments received on 'Guideline on validation of bioanalytical methods' (EMEA/CHMP/EWP/192217/2009)*, 2011.
[156] M. Pawula, G. Hawthorne, G.T. Smith, H.M. Hill, Best Practice in Biological Sample Collection, Processing and Storage for LC-MS Bioanalysis of Drugs, in: W. Li, J. Zhang, F.L.S. Tse (Eds.) Handbook of LC-MS Bioanalysis: Best Practices, Experimental Protocols, and Regulations, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2013, pp. 47-64.
[157] P. van Amsterdam, A. Companjen, M. Brudny-Kloeppel, M. Golob, S. Luedtke, P. Timmerman, Bioanalysis, 5 (2013) 645-659.
[158] T.W. Chow, A. Szeitz, D.W. Rurak, K.W. Riggs, J Chromatogr B, 879 (2011) 349-358.

[159] C.M. Nijenhuis, H. Rosing, J.H.M. Schellens, J.H. Beijnen, 88 (2014) 630–635.
[160] S. Ghassabian, S.M. Moosavi, Y.G. Valero, K. Shekar, J.F. Fraser, M.T. Smith, J Chromatogr B, 903 (2012) 126-133.
[161] D.H. Vu, R.A. Koster, A.M.A. Wessels, B. Greijdanus, J.W.C. Alffenaar, D.R.A. Uges, J Chromatogr B, 917 (2013) 1-4.
[162] M. Yadav, P.S. Shrivastav, T. de Boer, J. Wieling, P. Singhal, Current Understanding of Bioanalytical Assay Reproducibility – Incurred Sample Reanalysis (ISR), Incurred Sample Stability (ISS) and Incurred Sample Accuracy (ISA), in: W. Li, J. Zhang, F.L.S. Tse (Eds.) Handbook of LC-MS Bioanalysis: Best Practices, Experimental Protocols, and Regulations, John Wiley & Sons, Inc. , Hoboken, NJ, USA, 2013.
[163] S. Braggio, R.J. Barnaby, P. Grossi, M. Cugola, Journal of pharmaceutical and biomedical analysis, 14 (1996) 375-388.

Figure Captions.

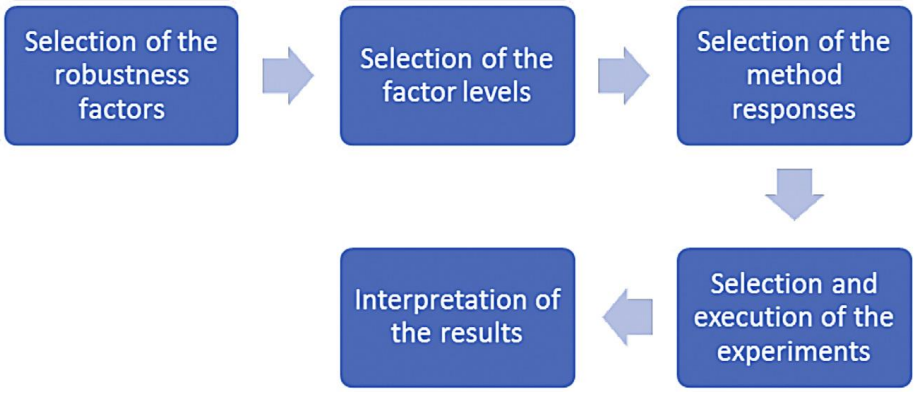


Figure 1: Summary of the procedure for robustness testing

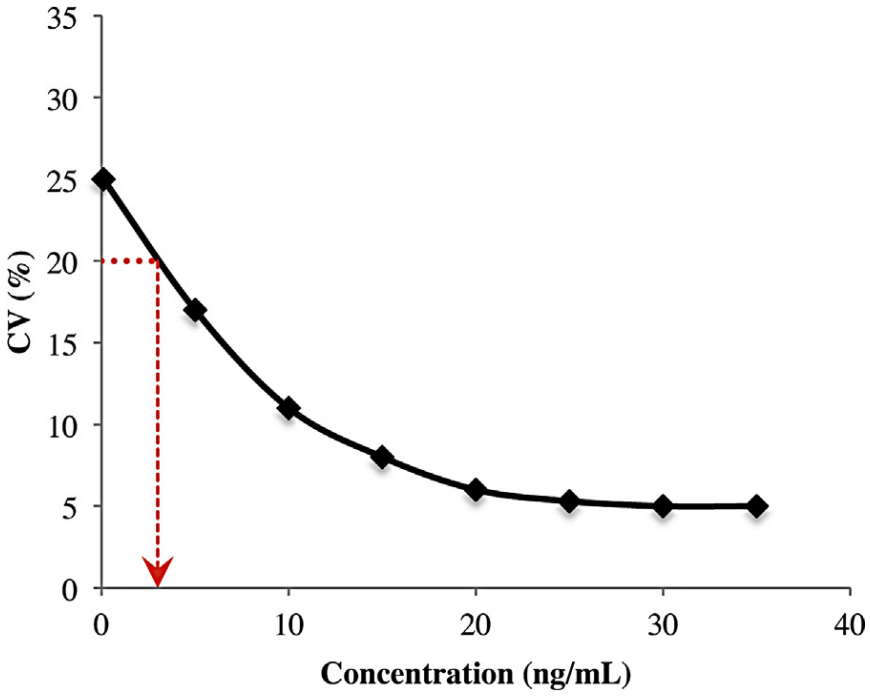
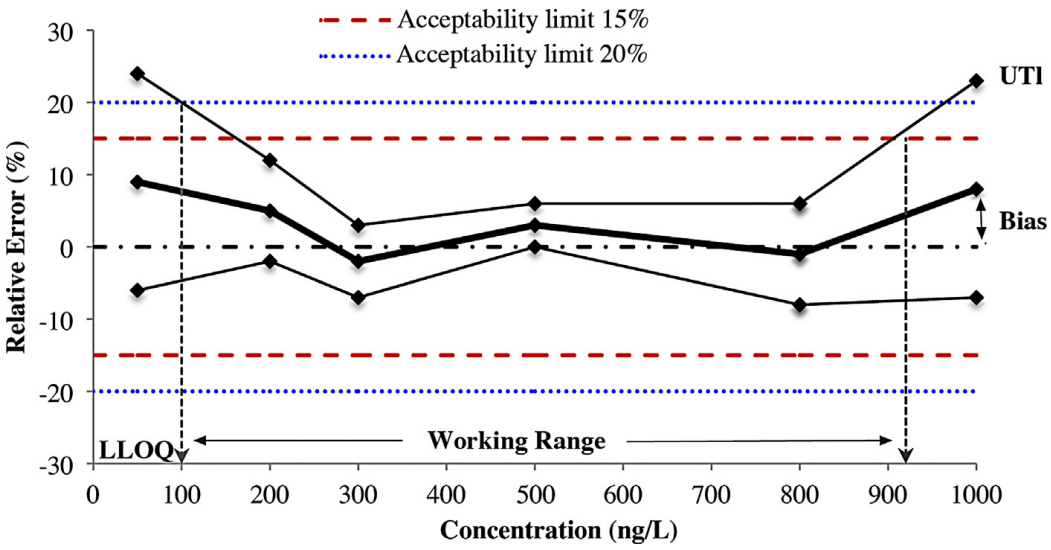
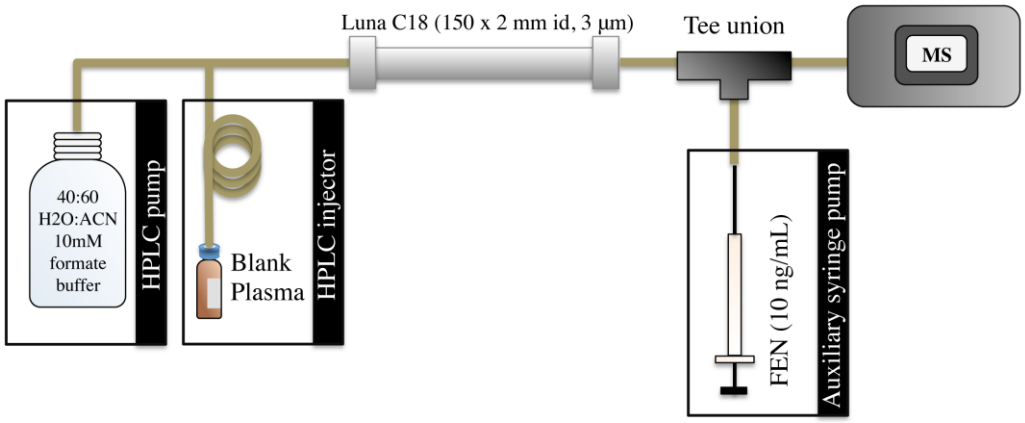


Figure 2: Example of precision profile, showing the evolution of the bioanalytical method precision with respect to analyte concentration. The LLOQ corresponds to the smallest concentration with a %CV not exceeding 20%. Based on the "Graphical Examination of the Relative Error versus Concentration Level" (Accuracy Profile approach).



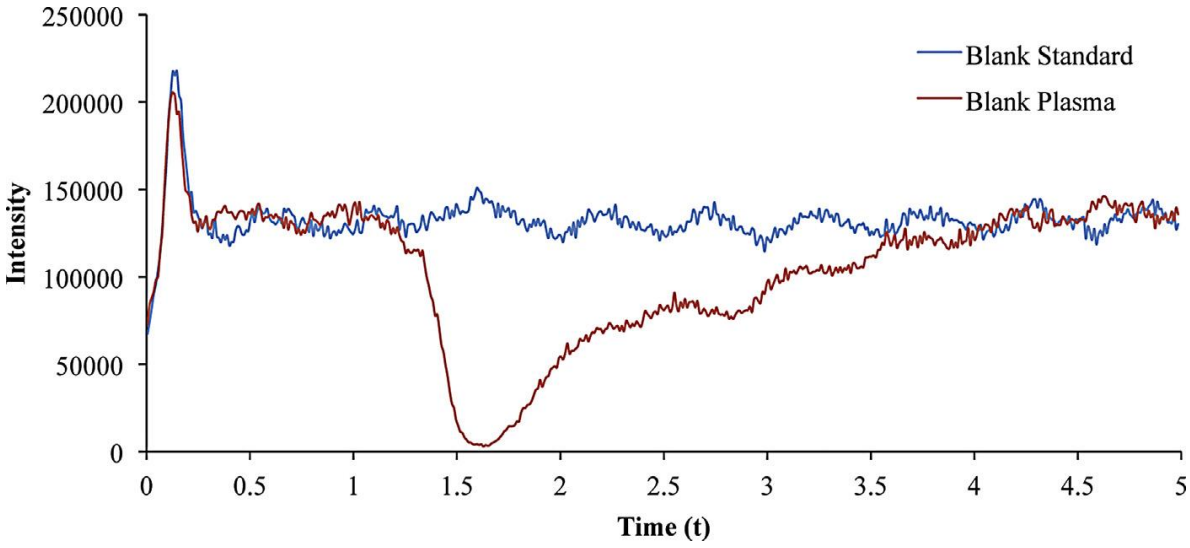
1477

1478 **Figure 3:** Schematic representation of an Accuracy Profile based on a 6 concentration levels. Two
1479 acceptance limits have been represented: in red, upper and lower acceptance limits (ER=15%) for
1480 the calibration range, except the LLOQ. In blue, upper and lower acceptance limits for the LLOQ
1481 (ER=20%). β-expectation tolerance limits (intervals) for each concentration level are represented in
1482 fine black. The observed relative bias (trueness) is reproduced in bold-black. While the intersection
1483 between the tolerance limit and the acceptance limit of 20% defines the LLOQ, the intersection
1484 between the tolerance limit and the acceptance limit of 15% defines the ULOQ.



1485

1486 **Figure 4:** Scheme of the post column infusion system



1487

1488 **Figure 5:** Post column infusion of fentanyl while injecting through the chromatographic system a

1489 blank standard (blue) or a blank plasma treated with protein precipitation using acetonitrile (red).

1490 The ion suppression between 1.5 and 3.5 minutes can be clearly observed.