Analytica Chimica Acta xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Tutorial

Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II

Anneli Kruve [a], Riin Rebane [a], Karin Kipper [a], Maarja-Liisa Oldekop [a], Hanno Evard [a], Koit Herodes [a], Pekka Ravio [b], Ivo Leito [a,*]

[a] University of Tartu, Institute of Chemistry, Ravila 14a, 50411 Tartu, Estonia
[b] Finnish Customs Laboratory, Tekniikantie 13, PL 53, FI-02151 Espoo, Finland

HIGHLIGHTS

- The status of validation of LC–MS methods is comprehensively reviewed.
- Clarity is brought into validation-related terminology.
- Recommendations on difficult validation-related issues in LC–MS are given.

GRAPHICAL ABSTRACT

ARTICLE INFO

Article history:
Received 4 February 2015
Accepted 9 February 2015
Available online xxx

Keywords:
Liquid chromatography–mass spectrometry
Validation
Electrospray
Ionization suppression
Limit of detection
Robustness

ABSTRACT

This is the part II of a tutorial review intending to give an overview of the state of the art of method validation in liquid chromatography mass spectrometry (LC–MS) and discuss specific issues that arise with MS (and MS–MS) detection in LC (as opposed to the "conventional" detectors). The Part II starts with briefly introducing the main quantitation methods and then addresses the performance related to quantification: linearity of signal, sensitivity, precision, trueness, accuracy, stability and measurement uncertainty. The last section is devoted to practical considerations in validation. With every performance characteristic its essence and terminology are addressed, the current status of treating it is reviewed and recommendations are given, how to handle it, specifically in the case of LC–MS methods.

Abbreviations: CRM, certified reference material; CV, coefficient of variation; EMA/EMEA, European Medicines Agency; ESI, electrospray ionization; FDA, United States Food and Drug Administration; ICH, International Conference on Harmonization; ILC, interlaboratory comparison; ILIS, isotopically labeled internal standard; IS, internal standard; IUPAC, International Union of Pure and Applied Chemistry; LC–MS, liquid chromatography–mass spectrometry; LLoQ, lower limit of quantitation; LoD, limit of detection; LoF, lack-of-fit; LoQ, limit of quantitation; ME, ionization suppression/enhancement (matrix effect); MU, measurement uncertainty; PE, process efficiency; *r*, correlation coefficient; *R*, recovery; r^2, coefficient of determination; RSD, relative standard deviation; *s*, standard deviation; SLV, single lab validation; ST%, stability; ULoQ, upper limit of quantitation; VIM, International Vocabulary of Metrology.

* Corresponding author.
E-mail address: ivo.leito@ut.ee (I. Leito).

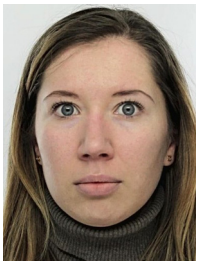
http://dx.doi.org/10.1016/j.aca.2015.02.016
0003-2670/

Contents

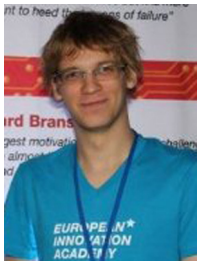
Anneli Kruve obtained her Ph.D. from the University of Tartu (UT) in 2011. Since 2005 she has been involved in HPLC and LC–MS method development and validation in various fields: bioanalysis, food analysis as well as industrial analysis. Starting from 2011 she works as a research fellow at the UT Institute of Chemistry. In 2008 and 2009 she has worked in the University of Helsinki in the field on miniaturization of MS ion sources. Her main research fields are method development for pesticide analysis, fundamental studies of ionization efficiency and design of MS ionization sources.

Riin Rebane obtained her Ph.D. in analytical chemistry from the University of Tartu in 2012 with a topic on optimization and validation of liquid chromatographic methods with mass spectrometric detection containing derivatization. For the past eight years her main research area has been LC–MS analysis, including method development and validation for various analytes and development of novel derivatization reagents for LC–MS. She also works as a quality assurance specialist in the Estonian Environmental Research Centre.

Karin Kipper obtained her Ph.D. from the University of Tartu (UT) in 2012. Since 2004 she has been involved in the bioanalytical method development and validation for HPLC-UV/VIS and LC–MS analysis, working at the UT Institute of Pharmacology and Institute of Chemistry. Starting from 2012 Karin Kipper works as a research fellow at the UT Institute of Chemistry. Her main research fields are pharmaceutical bioanalysis (pharmacokinetic/pharmacodynamic studies), pharmaceuticals' pathways in environment and development of novel eluent additives for LC–MS in order to improve separation and peak shape of basic compounds.



Maarja-Liisa Oldekop obtained her M.Sc. from the University of Tarty (UT) in 2013 and is now a Ph.D. student in the chair of analytical chemistry at UT. Her main field of research is development of LC–MS methods using derivatization. The focus is on matrix influences on this type of analysis, stressing the importance of trueness of the analysis results but also the sensitivity of the method. Since the beginning of 2013 Maarja-Liisa Oldekop works as the quality manager of the UT Testing Centre, which is an ISO/IEC 17025 accredited testing and calibration laboratory.



Hanno Evard graduated from the Applied Measurement Science master's programme at the University of Tartu in 2012. His master's thesis was about Paper spray Ionization. He is now a Ph.D. student at the University of Tartu. His research is focused on fundamental studies and developing new applications for different mass spectrometry ion sources.

Koit Herodes obtained his Ph.D. from the University of Tartu (UT) in 2002. Since 2008 he works as the head of the UT Testing centre – a unit providing testing and analysis services and accredited according to ISO 17025 by the Estonian Accreditation Centre. Since 2005 Koit Herodes works as associate professor of analytical chemistry at the UT Institute of Chemistry. He has been the principal investigator of numerous projects involving LC–MS analyses. Currently he is the principal investigator of the project "Development of software for validation of chromatographic methods", which aims at creating web-based software for validation of chromatographic methods.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

Pekka Ravio obtained his M.Sc. from the University of Helsinki in 1981. Since 1984 he has been working in the Finnish Customs Laboratory being responsible for various types of mass spectrometric analyses. His main area of expertise is pesticide residue analysis and quality assurance.

Ivo Leito obtained his Ph.D. from the University of Tartu (UT) in 1998. During 2002–2005 he reorganized the UT Testing centre – a unit providing testing, measurement, analysis and training services to laboratories and industry – and was its director. Starting from 2005 Ivo Leito works as a professor of analytical chemistry at the Institute of Chemistry, University of Tartu. His main research directions are on the borderline of analytical chemistry with other disciplines: chemistry of superacids and superbases; metrology in chemistry (MiC); liquid chromatography and mass spectrometry; sensors and their metrological characterization; applications of instrumental methods in analysis of historical objects. He teaches analytical chemistry and its metrological aspects at UT and has been involved in setting up several international MiC-related educational activities.

1. Introduction

Part I of this tutorial review [1] introduced the principles of operation of LC–MS, emphasizing the aspects important from the validation point of view, in particular the ionization process and ionization suppression/enhancement; reviewed the main validation guideline documents and discussed in detail selectivity/specificity/identity, ruggedness/robustness, limit of detection, limit of quantification, decision limit and detection capability. This second part of the tutorial review focuses on performance characteristics related to quantitation.

2. Quantitation methods

In addition to the different MS operation modes (discussed in Part I of this review [1]) the user can also select different quantitation methods. The most widespread are external calibration graph, calibration graph with internal standard and the method of standard additions. The calibration graph approaches can use either a calibration graph in solvent or a calibration graph in the extract of an appropriate matrix – the so-called matrix-matched calibration (the standard addition methods are intrinsically matrix-matched).

The calibration graph method in solvent is suitable for the simplest cases – simple matrices with high recovery and low ionization suppression. If significant ionization suppression occurs, then matrix-matched calibration offers an advantage over calibration in solvent. However, for using this approach, blank matrix sufficiently similar to the sample matrix, is needed. This similarity has to be demonstrated during validation (see Section 3.2 below). In practice the use of matrix-matched calibration can be impossible, if a suitable matrix is not available, or impractical, if the matrices of the routinely analyzed samples vary significantly [2,3].

A modified matrix-matched calibration approach has been proposed by the SANCO guide [4], where subsamples of blank sample matrix are spiked with the analyte at different levels. These spiked matrices are processed in a similar way to the real samples and the calibration graph is produced. If the matrix matches that of the real samples then this approach compensates for recovery loss and ionization suppression/enhancement effect. The drawbacks are that the blank matrix has to be available and it is also work-intensive and time-consuming.

Using an internal standard (IS) can help correcting for sample preparation recovery and to a lesser extent also for ionization suppression. Two types of IS can be distinguished: structural analogs of the analytes and isotopically labeled internal standards (ILIS). ILIS may either be isotopically labeled analyte or some other isotopically labeled compound. The IS is usually added to the samples at an as early as possible stage of sample preparation. It is also added to the calibration solutions. The calibration graph is presented as dependence of the ratio of analyte and IS signals on the ratio of analyte and IS concentrations. All types of IS help correcting results for sample preparation recovery. In order to account for ionization suppression the IS should co-elute with the analyte [5] and have the same "sensitivity" to matrix compounds. Therefore, ILIS corresponding to the analyte usually helps to correct for ionization suppression as well, though in some cases misleading results can be obtained if retention times of the analyte and the ILIS differ [6]. This problem occurs with deuterated ILIS, but generally not in the case of ^{13}C or ^{15}N labeled ILIS [6].

Our experience shows that if ILIS is added to the sample after sample preparation it is possible to use ILIS for ionization suppression/enhancement correction in order to reduce the costs (ILIS are expensive and a smaller amount is used in this case). Because this approach does not take into account the analyte loss during sample preparation, it is important to confirm beforehand that the sample preparation recovery is reproducible and reasonably close to 100%. Adding IS before injection also helps to account for injection repeatability.

The standard addition methods can potentially correct for recovery and ionization suppression (and other possible systematic effects), however, only if these effects are proportional [7,8]. The main drawback of the standard addition methods is that they are highly work-intensive and time consuming (both personnel and apparatus) as several additions have to be made and usually adjusting the added concentration is needed during conducting the analyses. Therefore standard addition is not cost-effective and is significantly less popular than the two previous quantitation approaches.

As all above described quantitation methods are somewhat prone to errors, one of the aims of validation is to verify the accuracy of the chosen quantitation method.

In the LC–MS literature all these methods have been used. An important issue in many of the works has been, whether or not the respective quantitation method is efficient against ionization suppression/enhancement. Few examples follow. Zamora et al. [9] observed during method optimization that classical calibration with standards in solvent cannot be used, as the matrix suppresses ionization in electrospray ionization (ESI) source. Even in 5 fold diluted orange and banana matrices the analyte signal was suppressed by 50%. While using matrix-matched calibration, no significant suppression/enhancement effect between standards and samples was observed. Gentili et al. [10] were able to overcome ionization suppression/enhancement by using appropriate quantitation method – addition of ILIS – for non-steroidal

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

anti-inflammatory drugs determination in bovine milk and muscle tissue. Margoum et al. [11] overcame ionization suppression, caused by co-extracted compounds from environmental water via combining usage and matrix-matched calibration. Kruve and Leito [12] observed that ionization suppression effect cannot be always accounted for via matrix-matched calibration. As a conclusion, none of the above described approaches are always accurate and therefore need to be validated.

3. Parameters of LC–MS methods

3.1. Linearity of signal, linear range and sensitivity

3.1.1. Linearity of signal

The term linearity of signal can in the context of LC–MS have two closely linked meanings: (1) linear relationship between analyte signals and analyte concentrations in calibration samples and (2) linear relationship between analyte signals and analyte concentrations in samples containing matrix components. The latter meaning is becoming increasingly more used [4,13a,14]. The reason is that if the analyte signal in samples is linear then almost certainly it is linear also in calibration solutions, while the opposite is not necessarily true.

The linearity of signal in a LC–MS method is influenced by several factors. Firstly, we can say that the ion source "behaves linearly" if the ionization efficiency of the analyte in the ion source [15] is independent of its concentration in the effluent, i.e., the amount of ions generated is proportional to the analyte concentration. For the ESI source it has been established that at low analyte concentrations (common in LC–MS analysis) this linear dependence holds [16], while at higher concentrations the excess charge or surface of droplets becomes limiting [17]. However, co-eluting compounds can markedly influence the ionization process in the ESI source (e.g., by charge competition) and lead to decrease or loss of linearity [18,19]. Therefore, investigation of linearity using sample matrix is encouraged.

Secondly, during the ion transport from the ion source to the mass analyzer the number of successfully transported ions must be proportional to the number of ions formed in the source. However, Page et al. have found that "the transmission efficiency is not constant, meaning that the amount of transmitted current is not proportional to the amount of current entering the capillary, but decreases as the amount of current increases" [20].

Thirdly, the linearity of ion signal depends on the mass analyzer design and on the linearity of the detector's signal. The contemporary ion detectors are highly linear [21], so that mass analyzer design is the deciding factor here. Mass analyzers are characterized by transmission, which is the ratio of ions that are finally detected and ions that entered the mass analyzer. The most widespread mass analyzer in routine LC–MS analysis is triple quadrupole, followed by ion trap. Triple quadrupole works in continuous mode and displays very good linearity. Time-of-flight and ion trap mass analyzers work in pulsed mode and linearity can be affected by the amount of charge collected in the mass analyzer [22–24].

In most validation guidelines linearity is defined as the method's "ability to obtain test results which are directly proportional to the concentration of analyte in the sample" [25–27]. However, NordVal does not mention linearity at all.

Regardless of the chosen quantitation method a calibration graph is constructed for almost all quantitative analysis done with LC–MS. In the majority of cases when LC–MS is used, the calibration graph is linear and analysts prefer this. Linearity of the calibration graph is closely related to the chosen calibration model and analytical range. The term "linear" in mathematics often does not refer to a straight-line relationship but rather to the fact that the statistical models considered are linear in terms of all parameters (coefficients), also including polynomial relationships. However, in this paper we limit the meaning of "linear" only to the case where a straight line can be used to describe the relationship between LC–MS signal and analyte concentration. Therefore, assessing linearity is essential before choosing the correct calibration model for reliable quantitative analysis. An overview of the recommendations for linearity evaluation given in validation guidelines is presented in Table 1.

It is evident from the table that the recommendations are not complete in any of the validation guidelines: all of them miss some aspects (how solutions should be prepared, in which order to measure them, etc.). The last row in Table 1 presents our recommendations which are discussed from now on.

3.1.1.1. Experiment planning for evaluation of linearity. In our opinion the following aspects are important in experiment planning for evaluation of linearity:

(1) The type of the calibration samples: either matrix-containing or matrix-free.
(2) The number of concentration levels in the series.
(3) The range of concentrations (the ratio between the highest and lowest analyte concentrations) and the distribution of the points along calibration line.
(4) Measurement protocol: the order of measuring the solutions in the series and the number of replicate measurements with every solution.

Our recommendations on these points are:

(1) If possible, matrix-containing solutions should be used (see Section 3.2.2 for the importance of matrix on ionization efficiency in LC–MS) and this is often applied [28–30]. Blank matrix extracts, preferably the same type as the sample, should be used. When the samples need to be diluted, the matrix concentration in the matrix-matched standards should be diluted proportionately [4]. The analyte standard should be spiked into the blank matrix at levels described in the next paragraphs. If solvent calibration is desired, a comparison of calibration graphs in matrix and solvent should be carried out, e.g., as in Refs. [31] and [32].
(2) The minimum number of 6 concentration levels is accepted by most validation guidelines and this also agrees with the recommendation by Hibbert [33] where 4 degrees of freedom are considered minimally acceptable. However, during validation we recommend a minimum of 10 concentrations because some points may fall out of linear range leaving too few points to draw any conclusions. This practice is followed in several literature sources: 9 [34], 10 [35] and 11 [36] calibration points.
(3) Choosing the concentrations of the calibration solutions, the ratio between the highest and lowest concentration should be appropriate for the method, keeping in mind the possible variation of analyte levels in the samples but should be at least 5. The points on the calibration graphs should be more or less evenly spaced, so that all parts of the calibration graph would contain data points with approximately the same density. If the range of the method spans several orders of magnitude then within each order the points should be approximately evenly spaced. The calibration solutions should not be prepared as consecutive dilutions (as this leads to accumulation of errors) but the solutions should be made by independent dilutions from the same stock solution.
(4) For linearity study the order of analysis of calibration standards should be random so that instrumental drift (which can be significant in the case of LC–MS, caused, e.g., by gradual ion

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

Table 1
Recommendations for evaluation of linearity in different validation guidelines and our recommendations.

Guideline	Experiment planning	Evaluation of data
ICH [26]	Min. 5 concentration levels. Dilutions of standard stock solution or weighing different amounts of analyte standard	First evaluated by visual inspection of the plot of signal–concentration relationship. In case of linearity, additional statistical method should be applied, e.g., regression line using least squares method. Correlation coefficient, *y*-intercept, slope of regression line and residual sum of squares should be used to evaluate linearity In addition: analysis of deviation of actual data points from the regression line (analysis of residuals)
AOAC [27]	6–8 points, approximately equally spaced, measured in duplicates at random. Calibration solutions obtained by dilution of stock solutions	Visual examination is usually sufficient and in addition use of plot of residuals is suggested. An acceptable fit produces random pattern of residuals with a 0 mean
Eurachem [25]	Blank and reference materials or spiked blanks at 6–10 concentrations evenly spaced exceeding ±10% or ±20% of the expected concentration range (2–3 measurements)	Visual inspection of the line and residuals may be sufficient, but statistical test, e.g., goodness-of-fit are recommended
IUPAC [48]	Min. 6 calibration standards at evenly spaced concentrations covering the range. Samples should be analyzed in random order at least in duplicate	Examination of plot of residuals or lack-of-fit test, but preferably both
EMA [14]	Min. 6 calibration standards at evenly spaced concentrations, a blank and a zero sample analyzed at least in duplicate. Calibration standard is defined as matrix spiked with analyte	Matrix-matched calibration. Back-calculated concentrations of the calibration standards should be within 15% of the nominal value (20% for lower limit of quantitation (LLoQ) and upper limit of quantitation (ULoQ)) for at least 75% of calibration standards
FDA [13a]	A blank sample,[a] a zero sample[b] and 6 min non-zero[c] samples analyzed in duplicate in 6 sequences over several days. Calibration samples should be prepared in the matrix as study samples	Matrix matched calibration. Standard curve is acceptable when 75% of non-zero standards are within 15% of the nominal concentration (20% for LLoQ)
SANCO [4]	Three or more concentrations. Use of two levels is appropriate if the difference of concentrations is below 10 times	Matrix matched calibration. Visual inspection or calculation of residuals. Relative residuals should be within ±20%. Weighted linear regression is preferred
Our recommendations	Solutions with at least 10 different concentrations, approximately equally spaced, with ratio of the highest concentration to the lowest at least 5, containing the matrix components (if possible), measured in random order and at least in duplicate	We recommend a two-stage approach: (1) Linearity is first assessed visually. In clear-cut situations (see Fig. 1 for explanations) visual assessment of linearity is sufficient. (2) In unclear situations (Fig. 1) statistical tests should be used

[a] Blank sample: matrix sample without the analyte and internal standard.
[b] Zero sample: matrix sample with internal standard, without analyte.
[c] Non-zero sample: matrix sample containing both analyte and internal standard [13a].

source contamination during a sequence) would not influence the results of linearity experiments [37]. It is useful to analyze all solutions in a manner as similar as possible to the real-life situation (i.e., calibration samples randomly ordered and placed between unknown samples in the run). For linearity evaluation these samples should be analyzed at least twice. In the literature also three [32] or six replicate measurements [38,39] have been used.

Because of the instability of signal in MS, calibration graphs usually cannot be used for longer than one sequence and, therefore, after validation the calibration samples should be involved in every sequence [28,30,38]. In some cases, usually involving internal standards, this might not be necessary if validation results confirm this. For example, using IS and linear regression with $1/x$ weighting Nilsson and Eklund found that the calibration graph slope was constant for 10 weeks (coefficient of variation CV 1,6%, $n = 6$) [40].

3.1.1.2. Recommendations for evaluating linearity test data. When evaluating the data of signal linearity experiments, the following aspects are important:

(1) Is visual evaluation of linearity sufficient or is there a need for using statistical tests.
(2) If visual inspection is sufficient, then what are the criteria of linearity.
(3) If statistical tests are used then which tests.

Fig. 1 represents three different cases when evaluating linearity which are now discussed.

(1) In our opinion, there is no need for statistical tests in clear-cut situations (i.e., if linearity is obvious by visual examination, Fig. 1a) but in unclear situations (Fig. 1c) visual inspection is not sufficient and statistical tests should be used.

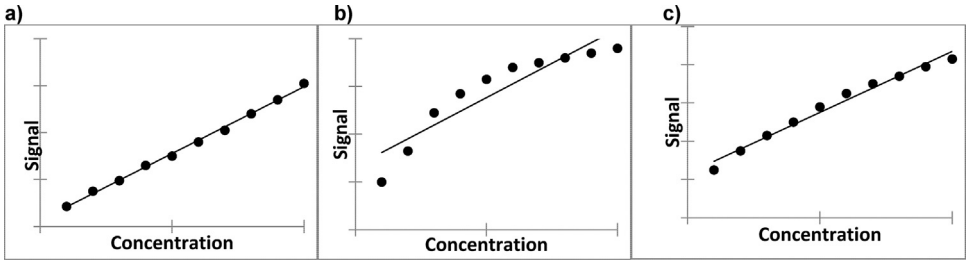


Fig. 1. Three cases while evaluating linearity: (a) visual inspection confirms linearity, (b) clearly non-linear graph and (c) graph where linearity needs to be checked with a statistical test.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

(2) When evaluating linearity visually, the criterion of linearity is random distribution of residuals. Fig. 1 presents two clear-cut situations. In situation (a) the residuals are clearly distributed randomly above and below the regression line and linearity can be confirmed. In situation (b) the residuals in the beginning and in the end of the graph have the same sign and in the middle of the graph have the opposite sign. This clearly indicates non-linearity.

(3) In situation (c) it is not obvious whether the graph is linear or not and a statistical test is required. Using statistical methods for linearity evaluation is still not widespread although recommended by accreditation bodies [41]. It is a common practice to evaluate calibration graph's linearity using correlation coefficient r or coefficient of determination r^2. The multitude of different limits used for r^2 in the literature demonstrates that it is not straightforward to evaluate linearity based on r^2 [29,34,42–46]. It has been demonstrated by several authors [33,47] that correlation coefficient as measure of linearity might be misleading [48,49].

In order to use simple linear regression, the standard deviation of the signal should be constant over the chosen concentration range (the homoscedasticity requirement). If the standard deviation of the signal increases with concentration, then the data is heteroscedastic and weighted calibration functions should be used. [50] The *F*-test is suggested for homoscedasticity testing, whereby the ratio of variance of the signal obtained at the lowest concentration and variation of signal obtained at the highest concentration is compared to the tabulated value [51]. In routine LC–MS analysis, when working in a narrow concentration range, the heteroscedasticity of data is usually not large enough to change markedly the calibration graph parameters or the concentrations found from calibration graphs.

The plot of residuals is suitable to confirm simple linear regression, both by evaluating the random distribution of residuals as well as the homoscedasticity of the data [51]. The lack-of-fit test, goodness-of-fit and Mandel's fitting test are used to evaluate the fitness of any kind of regression model for the data. While r^2 is unsuitable for establishing linearity of the calibration graph, it is actually suitable for deciding whether linear calibration model can be used with a certain set of calibration points, as long as the data points are evenly spaced [52].

3.1.1.2.1. Plot of residuals. A suitable approach to confirm linearity is the analysis of residuals from the linear regression analysis [25a,42,48], applied in Refs. [28,53]. Here, two approaches are available: using absolute or relative residuals. The absolute residuals are expressed as the difference between experimental (y_i) and calculated signal ($\hat{y}_i$) values [50] and are more suitable for the upper and middle part of the linear range. For the lower end of the range the relative residuals might be more suitable. SANCO guideline considers acceptable relative residuals lower than 20% [4]. The relative residual Y_i is expressed as follows:

$$Y_i = \frac{y_i - \hat{y}_i}{\hat{y}_i}. \quad (1)$$

Also, the standard deviation of relative residuals can be used and its value should be less than 0.1 for the suitable calibration model. The standard deviation of relative residuals is calculated as follows:

$$S_{Y_i} = \sqrt{\frac{\sum (Y_i - \overline{Y})^2}{n-2}} \quad (2)$$

Here, $\overline{Y}$ is the mean value of relative residuals Y_i and n is the total number of calibration data points, counting both different concentrations and replicates [54].

3.1.1.2.2. Lack-of-fit test. In addition to the plot of residuals we also recommend using the lack-of-fit (LoF) test that is based on the analysis of residual variance [37]. If a total number of n concentration levels are analyzed with p replicates at one concentration level, then the pure error can be estimated from replicates and compared to the error from lack of fit of the model. For the lack-of-fit test the ratio of the mean sum of squares of the lack of fit (MSS_{LoF}, $n-2$ degrees of freedom) and the random error (MSS_e, $n(p-1)$ degrees of freedom) is found:

$$F = \frac{MSS_{LoF}}{MSS_e} = \frac{\sum (\overline{y}_i - \hat{y}_i)^2 / n-2}{\sum (y_i - \overline{y}_i)^2} / n(p-1) \quad (3)$$

If the experimentally found F is higher than the tabulated value, the model cannot be considered fit for the data [37]. The lack-of-fit test is applied in Refs. [28,55,56].

3.1.1.2.3. Goodness-of-fit test. There is also a test for evaluation of the goodness of the model [57]. The fitness can be estimated using the ratio of mean sum of squares of the factors (MSS_{fact}) and the residuals (MSS_R) as follows:

$$F = \frac{MSS_{fact}}{MSS_R} = \frac{\sum (\hat{y}_i - \overline{y}_i)^2 / p-1}{\sum (y_i - \hat{y}_i)^2} / n-p \quad (4)$$

The experimentally found F being higher than the tabulated F means that the model differs systematically from the data. However, we find that this test can give excessively optimistic results and is not critical enough.

3.1.1.2.4. Mandel's fitting test.. Mandel's fitting test uses the difference of variances of residual standard deviation of linear (S_{y1}) and potential second-order (S_{y2}) calibration models. This is compared to the standard deviation of the potential second order calibration model using the *F*-test.

$$F = \frac{S_{y1}^2 \times (n-2) - S_{y2}^2 \times (n-3)}{S_{y2}^2} \quad (5)$$

The calibration function can be considered linear if the found value is lower than the tabulated value. If the test value is higher, narrowing the concentration range or use of the second order calibration function should be considered [58]. Mandel's fitting test is suggested by IUPAC and applied in Refs. [59–61].

3.1.1.3. Choosing the calibration model. Calibration model should be chosen based on the linearity experiments and on the shape of the concentration–signal relationship. The simplest model that adequately describes the concentration–signal relationship should be used and the use of more complex models should be justified [13a]. This is especially true in LC–MS where the scatter of data points around calibration graph can be quite high. In such cases small non-linearity can easily go unnoticed. On the other hand, if scatter of data points obscures non-linearity, then it is fair to say that non-linearity is statistically insignificant compared to random variation and linear calibration model can be used.

Best fit can be found by minimizing the sum of squared residuals. Although the linear regression is mostly desired also in LC–MS analysis, there are several possibilities once the non-linearity is confirmed. Weighted regression is often suggested as one option [25,48]. Weighted regression does not remove non-linearity, but is beneficial to decrease the influence of the higher concentration measurements if calibration graph spans a wide range of concentrations. Another tool to use would be to try to transform the data and use for example log y or log x values on the

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

plot [50]. Also quadratic [62,63] or cubic [64,65] curve might be suitable for characterizing the data. A combined method of weighted ($1/x$) quadratic model with IS method is often used [66,67].

Irrespective of whether the calibration model is a linear or a polynomial, its fit with the calibration data has to be demonstrated. This can be done with either by the lack-of-fit, goodness-of-fit or Mandel's test as discussed previously. When the second order (quadratic) model is chosen, the evaluation of significance of the quadratic effect needs to be justified. This can be done by carrying out a *t*-test [50]. However, if a non-linear model is chosen then more points compared to the linear model should be used for calibration [68] because the exact shape of the curve changes and depends on the instrument condition.

3.1.2. Linear range

Linear range can be found from linearity assessment experiments as the lower and upper concentrations of analyte in the sample for which precision, accuracy and response function has been established [14,26] (see Section 3.2). In addition to linear range, many other uses of the term range regarding method validation can be found in the literature: analytical range, calibration range, dynamic range, working range [41], suitability range [69] and validated range [48]. Linear (or linear dynamic or calibration) range is a more specific term than dynamic (or working or analytical) range, where the correlation between analyte concentration and signal may be non-linear [70].

Linear ranges may vary for different matrices because linearity itself can depend on matrix as was explained above. Therefore, during validation linearity should be confirmed for the expected working range. It is suggested that the concentration range of 0–150% or 50–150% of the expected analyte concentration should be covered [48]. The criteria for the linear range can be different. For example, a plot of response factor versus concentration, where the concentrations within linear range have the response factor lower than 2.5% for assay method and 5% or 10% for impurity method [64].

However, the linear range for LC–MS instruments is usually fairly narrow (and depends on the compound) [71]. The linear range can be widened using ILIS [40]. The signals of the compound and the ILIS may not be linear but their ratio is. However, even in this case, care has to be taken when working outside of the linear range of the signals. One way of increasing the linear range is by working at lower concentrations (and diluting samples), if the analyte signal is intense enough [72]. Another way to increase the linear range is to decrease charge competition by lowering flow rate in ESI source, e.g., by using nano-ESI [17].

3.1.3. Sensitivity

Sensitivity is, similarly to linearity, associated with calibration graph. Sensitivity is usually defined as "the change in the response of a measuring instrument divided by corresponding change in the stimulus" [25] or in the other words "the gradient of the calibration function" [48]. ICH and EMA do not mention sensitivity. Several other validation guidelines interpret sensitivity as related to the detection/determination ability. NordVal and AOAC define sensitivity as the ratio between the number of obtained positive results and the expected total number of positive results. In the recent FDA's Bioanalytical Method Validation guidance document [13b] sensitivity is defined as "the lowest analyte concentration that can be measured with acceptable accuracy and precision (i.e., LLoQ)". This definition effectively matches that universally (including this review) accepted for LoQ. In this review we use sensitivity as defined by Eurachem and IUPAC. We recognize that sensitivity is not an essential parameter in method validation, however it is very important in method optimization and routine monitoring of the instrument.

The numerical value (and unit) of sensitivity is arbitrary and depends on the instrument used and its settings [48]. This is the reason why sensitivity is not universally included among validation parameters [14,26,73,74]. On the other hand, sensitivity can be important to consider when an instrumental technique is used for the analysis [25] and it may be especially useful for method optimization and quality assurance procedures for testing the instrument performance [48]. Moreover, as the slope of the regression line, sensitivity is among parameters needed for evaluation of linearity [26].

Although sensitivity is not to be confused with limit of detection (LoD) and limit of quantitation (LoQ), these terms are interrelated: with a given signal to noise ratio, the higher the sensitivity, the lower the LoD and LoQ [26,75]. Sensitivity is also directly related to ionization suppression – in fact the essence of ionization suppression is decrease of sensitivity due to co-eluting compounds [76] (see below).

In our opinion the main use of sensitivity as a parameter is threefold: (1) optimization (for maximizing sensitivity) of the method parameters during method development, (2) daily optimization of the instrument parameters and (3) monitoring of the instrument performance.

3.2. Precision, trueness, accuracy

Precision, trueness and accuracy are related to the agreement between replicate measurement results or the agreement between the measured value and a reference value [77]. The definitions and usage of the terms, as well as the recommended experimental approaches related to precision, trueness and accuracy, vary considerably between different literature sources. The view accepted in this paper is expressed in Fig. 2.

An analyst is interested in the quality (closeness to the true value) of an individual analysis result. This result is affected by two groups of error sources – systematic and random [78]. These errors are caused by a range of reasons, such as imperfect analyte recovery during sample preparation, possible ionization suppression of the analyte, possible instability of the analyte and others. These errors put together give us the (total) error. This error would be a suitable characteristic for characterizing our result, but for a sample with unknown analyte content, this error is not experimentally accessible and thus cannot be used (see Section 1 of Ref. [79]).

A different approach is taken instead, whereby replicate measurements are used, leading to the average analysis result. If analysis is repeated on the same sample (not necessarily with known analyte content) on different days, then the standard deviation of the results, s_{RW}, is called intermediate precision and it characterizes the range within which the random error is expected to lie with ca. 68% probability (this probability is based on the properties of the normal distribution, see Section 2.2 of Ref. [50]). If analysis is repeated on the same sample with known analyte content C_{ref} (e.g., a certified reference material, CRM, is used) then the difference between the average result $C_{average}$ and C_{ref} is called bias. The intermediate precision and bias are influenced by different factors that cause errors. If suitable precision and bias data are available and are combined, then it is possible to obtain estimate of measurement uncertainty, which takes into account both random and systematic error sources and characterizes the accuracy of analysis results.

If everything is done correctly, then this accuracy, although obtained with samples different from the specific individual sample that was analyzed, can be used to characterize the accuracy of individual results obtained in the future: the exact magnitudes of the errors are not known but the true value is expected to lie within the uncertainty range with a predefined (high) probability.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

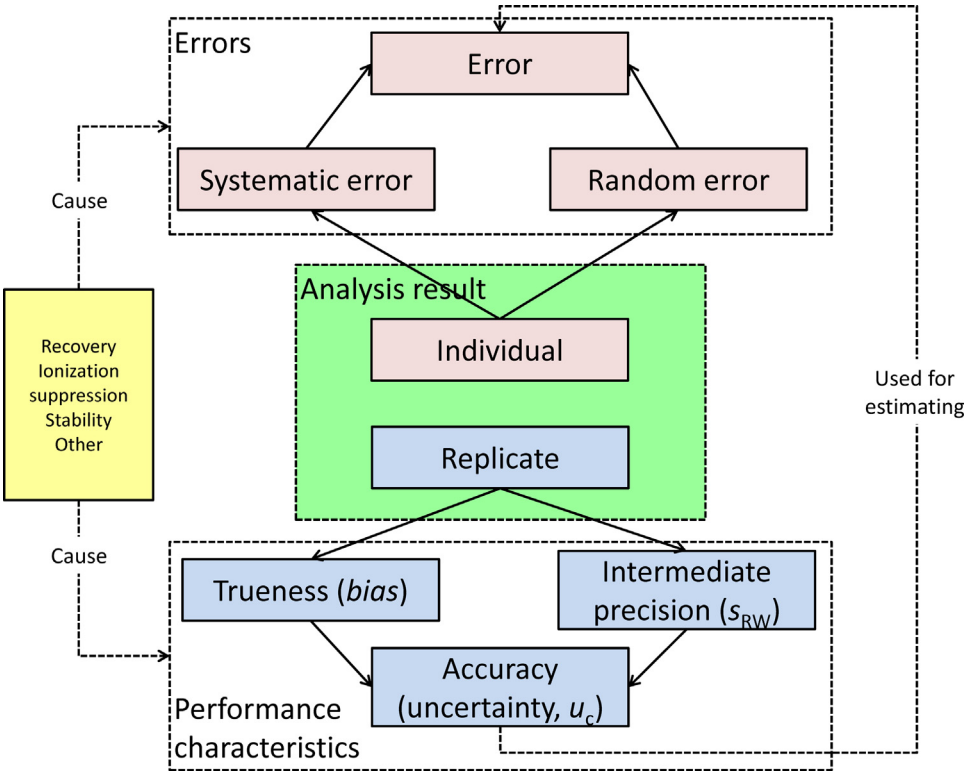


Fig. 2. Relations between different types of errors, precision, trueness, accuracy and measurement uncertainty.

3.2.1. Precision

Precision characterizes the closeness of agreement between the measured values obtained by replicate measurements on the same or similar objects under specified conditions [77]. Precision relates to the random error of a measurement system and is a component of measurement uncertainty [80].

By large, in all guidelines, two types of precision are mentioned that can be determined for an analytical method, namely repeatability [25–27,48,77] (also called within-run precision [13a,14]) and intermediate precision [26,27,77] (also called as between-run precision [13a,14], intra-lab reproducibility [25] within-laboratory reproducibility [78], precision under run-to-run conditions [48]). In addition to these two types of precision, AOAC and ICH also address the inter-laboratory reproducibility.

Repeatability is expected to give the smallest variation in results and is obtained by repeatedly analyzing independently prepared subsamples from a homogenous sample in the laboratory, by one operator, using one experimental setup and one set of reagents on one day. Intermediate precision, differently from repeatability, is the precision obtained within a single laboratory over a longer period of time (generally at least several months [78]) and takes into account changes such as different analysts, calibrants, batches of reagents, columns, spray needles etc. [41,73,81]. Not to be mistaken with the intermediate precision, reproducibility expresses the precision between measurement results obtained at different laboratories and is important to evaluate if the method is going to be used in several laboratories [44,75]. Some authors have used the term reproducibility for within-laboratory studies at the level of intermediate precision. However, this should be avoided in order to prevent confusion [81]. Reproducibility is not always needed for single-lab validation [26,27]. However, ICH suggests evaluating reproducibility when standardization of an analytical procedure is carried out [26].

Different guidelines each have their own specifics concerning how the experiments should be carried out (see below). It is important to keep in mind that ideally determination of precision should be carried out for all matrices that are within the scope of application of the method [27]. This general principle becomes even more important for LC–MS methods, where matrix influence can be quite significant (sample preparation, ionization suppression, etc.).

Precision is usually expressed as "imprecision": an absolute standard deviation (s), relative standard deviation (RSD), variance (s^2) or coefficient of variation (CV). Precision does not relate to the reference values [81,82]. For LC–MS, it is important to determine precision at different concentrations, as most guidelines suggest, since there is strong evidence that repeatability in liquid chromatography is concentration-dependent [83].

Specification of acceptance criteria for precision is quite difficult since it should be tailored to each specific application, or as it is called, methods should be "fit-for-purpose" [84]. ICH, Eurachem and IUPAC do not specify the acceptance criteria for precision. FDA and EMA say that within-run and between-run CV should be within 15% of the nominal value (20% at LoQ level). SANCO and EU directive 2002/657/EC state that CV should be smaller than 20%.

If there are no methods with which to compare the precision parameters, AOAC and NordVal refer to the Horwitz function [27] for the acceptance criteria. Horwitz has derived, by analyzing thousands of inter-laboratory comparison datasets, empirical functions, which relate expected repeatability or reproducibility RSD with the analyte concentration level in the sample [27]. According to the Horwitz function, higher relative variability is expected when lower concentrations are determined. LC–MS analysis is generally applied for analysis at low concentrations, leading to high Horwitz RSD values in most cases [84]. The criteria proposed by Horwitz state that in order to be in accordance with the Horwitz equation, the so-called Horwitz ratio (HORRAT), which is the ratio of found and calculated (by Horwitz function) RSD should be between 0.5 and 2 [27]. Although elegant and having

Table 2
Terms used for trueness in different guidance materials.

Organization	Term	Meaning according to VIM
Eurachem, AOAC, ISO	Accuracy	Accuracy
	Trueness	Trueness
ICH, FDA, EMA	Accuracy	Trueness
IUPAC, NordVal	Trueness	Trueness

high predictive ability while requiring very limited data, the Horwitz equation has two main drawbacks. Firstly, it completely neglects the specific properties of analytes and matrices. Secondly it is mainly based on data from half a century ago, while analytical instrumentation and quality assurance approaches have been intensely developed since then. The latter drawback means that the RSD values calculated by the Horwitz equation are often higher than those obtained nowadays. Many articles state using Horwitz function for acceptance value for LC–MS analysis [85,86]. There is also a Thomson modification of the Horwitz function for smaller concentrations [87] which has also been used [55]. However, we suggest that the acceptance criteria of precision should first of all be based on the requirements for a concrete application.

Eurachem [25] suggests using standard deviation of repeatability and reproducibility to calculate repeatability and reproducibility limits (precision limits). These limits help analyst to decide whether the difference between duplicate results (under repeatability or reproducibility conditions, respectively) of samples is significant [88].

It is possible to determine the repeatability and intermediate precision simultaneously. Subsamples of the sample of interest are analyzed in replicate under repeatability conditions across a number of different runs while varying conditions between the runs (days, analysts, equipment etc.). Via one-way ANOVA, repeatability can be calculated as the within-group precision, while the intermediate precision is obtained as the square-root of the sum of squares of the within-group and between-group precision. The advantage of this design is that it provides an efficient way of obtaining sufficient number of degrees of freedom for estimating repeatability and intermediate precision, and results in a balanced number of degrees of freedom for both estimates. Concrete examples and additional explanations are available from the revised Eurachem validation guide [25]. This approach has been applied for LC–MS analysis [39,89].

3.2.2. Trueness

Trueness relates to the systematic error of a measurement system [74,77,80] and if rigorously defined, refers to the agreement between the average of infinite number of replicate measured values and the true value of the measured quantity [77]. In practice, trueness is evaluated from a finite but reasonably large number of measurements and reference values are used instead of the true value. Different guidance materials use different terms for expressing trueness (Table 2). In this review we use the term trueness with the meaning given in the International Vocabulary of Metrology (VIM) [77]. The term accuracy has a different meaning and is discussed below.

Trueness of a method is usually quantitatively expressed as bias or relative bias. Bias is defined as estimate of the systematic error [77]. In practice bias is determined as difference between the mean obtained with a large number of replicate measurements and a reference value. The main reasons why LC–MS results can be biased can be termed as bias constituents:

1. Bias caused by analyte loss during sample preparation, expressed quantitatively by recovery (R).
2. Bias due to the limited analyte stability (see Section 3.3) of the analyte in sample solution (B_{stab}).

Table 3
Different terms related to trueness.[a]

Expression	Calculation	Comments
Bias	Absolute bias: $\text{bias} = X_{lab} - X_{ref}$	Bias takes into account the effects that are systematic over a long term, occurring at any stage of the analytical process Bias can be expressed in absolute or relative terms
	$\text{bias} = \text{ME}_{ionization} \times R \times B_{stab} \times B_{other}$	Absolute bias is useful when it is either constant over the used concentration range or it will be evaluated separately at different concentrations
	Relative bias[b]: $\text{bias}(\%) = \frac{X_{lab} - X_{ref}}{X_{lab}} \times 100\%$	Relative bias is useful when the absolute bias is proportional to the concentration
Process efficiency,[b] PE	$\text{PE} = \frac{X_{lab}}{X_{ref}}\text{PE} = \frac{m_{analyte\ detected}}{m_{analyte\ in\ sample}} = \text{ME}_{ionization} \times R$	Process efficiency refers to the joint effect of possible losses during sample preparation and ionization suppression/enhancement in the ion source. PE is a useful parameter for characterizing an analysis method when process efficiency is either required for characterization of the method or when it is intended to carry out correction with PE
Recovery,[b] R	$R = \frac{m_{analyte\ extracted}}{m_{analyte\ in\ sample}}$	Recovery[c] expresses the efficiency of the sample preparation step: the proportion of an analyte obtained from the sample during sample preparation
Ionization suppression/ enhancement (matrix effect), $ME_{ionization}$[b,d]	$\text{ME}_{ionization} = \frac{m_{analyte\ detected}}{m_{analyte\ extracted}}$	In LC–MS the term matrix effect refers to the suppression or enhancement of analyte ionization by co-eluting compounds originating from the sample matrix $m_{analyte\ detected}$: analyte amount detected in the sample $m_{analyte\ extracted}$: analyte amount actually extracted from the sample
Stability, B_{stab}	See Section 3.3 for discussion	This bias constituent takes into account losses due to analyte decomposition. Depending on which stage of sample preparation decomposition occurs there are different types of stability (see Section 3.3)
B_{other}		This bias constituent takes into account bias sources that are not connected to the above mentioned factors

[a] X_{lab}: average of results obtained by laboratory; X_{ref}: reference value.
[b] Can be expressed as percentage.
[c] In the case of most other analytical techniques recovery would include also the possible matrix effects, so that it would effectively be equal to PE as defined above.
[d] There are different ways of expressing the matrix effect. We use the way which is similar to expressing recovery and process efficiency.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

3. Bias due to the ionization suppression/enhancement ($ME_{ionization}$).
4. Bias due to other possible effects (B_{other}), i.e., purity of the standard substance, calibration bias of volumetric ware.

Table 3 presents the relations between these terms.

Fig. 3 illustrates interrelations between the different bias components. Process efficiency embraces both sample preparation recovery and possible ionization suppression/enhancement in the ion source. An additional important source of bias is possible instability of the analyte. The remaining area on the figure stands for all other (usually smaller) bias components, such as, e.g., calibration of glassware. In LC–MS literature process efficiency (PE) is often used as an LC–MS specific term for the overall trueness [5] (if B_{other} and B_{stab} are insignificant).

There is an important difference between precision and trueness. Although the repeatability and/or reproducibility standard deviation can be decreased, they cannot be fully eliminated. In contrast, elimination of bias by correction is in principle possible, although care must be exercised (see the excellent guidance in Ref. [90]). In practice bias correction (often called recovery correction) is frequently done or at least attempted. In the case of LC–MS method one of the most effective ways is the isotope labeled internal standard method described above (see Section 2). If such correction is successfully carried out then the remaining bias can be negligible, even though PE and/or R may be below 100%. In this case PE, R, etc. cannot be called bias constituents anymore. The discussion related to bias below is given from a wider point of view and does not assume that bias has been corrected for. Like precision, bias can also be determined in different time frames and, consequently, involves different extent of variations in analysis. The so-called run bias (within-day bias) refers to bias occurring during a given analytical run. It takes into account method bias and also the effects that are systematic (i.e., constant) within the given run, but can change between runs. The latter effects become random over a longer time period and are generally accounted for by the intermediate precision, because its determination is more convenient. Therefore, run bias is of limited practical interest and we do not address it in this section. If a longer time frame is considered, then many of the effects that are constant within day (and cause within-day bias) become random and are not included in bias any more, but instead are included in the intermediate precision. If the time period considered is in the order of one year, then most of the effects become random and bias takes into account only the method bias and laboratory bias. This is the so-called "method and laboratory bias" [78] and it is the most useful type of bias from the validation point of view. If bias is dependent on variations of matrix (which is not rare [78]) then the bias can be expressed either as the bias corresponding to a specific matrix or as an average bias corresponding to a group of matrices.

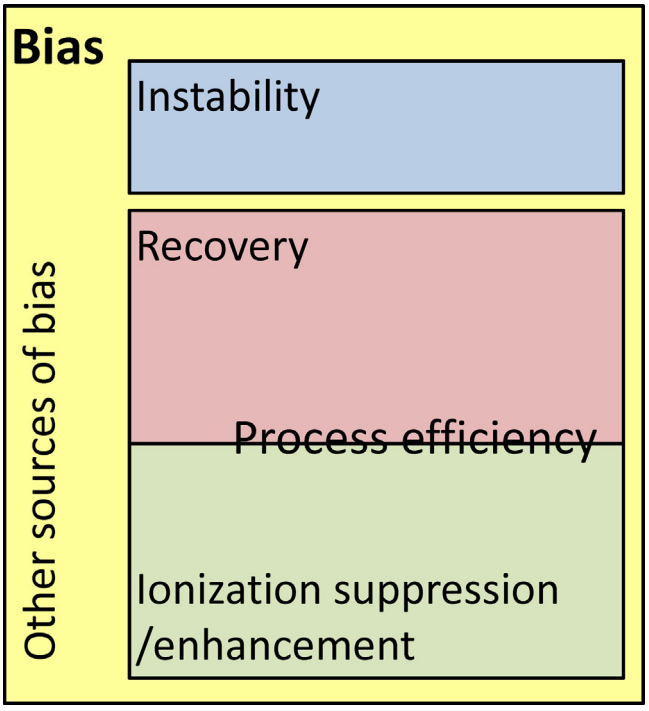


Fig. 3. Relations between the bias constituents.

In order to determine bias, a reference value of the measured quantity is required. By large, bias can be evaluated against a reference value that is either carried by a well-characterized material (reference material) (see Section 3.2.4 for practical suggestions) or provided by a second well-characterized reference method (also called cross-validation [13a,14]). Both methods – the method under investigation and the reference method – are used to analyze a number of typical samples, preferably covering a range of concentrations that is expected. Results can be compared with a suitable statistical method (for example, t-test, with due checks for homogeneity of variance and normality [48]). In addition, Deming regression has been used to compare LC–UV to LC–MS–MS [91] and LC–MS–MS to UPLC–MS–MS methods [92] as well as results obtained by different laboratories [93].

Whether or not to correct for bias is a difficult question and is often overlooked during validation. There are serious arguments both in favor of [94] and against correction [90]. Out of the guidelines discussed, SANCO suggests adjustment with mean recovery (to be understood here in the sense of bias taking into account all bias constituents) when it is not in the range of 70–120%. It is important that if corrections are used, it must be indicated when results are reported [4].

3.2.2.1. Recovery, ionization suppression and process efficiency. As presented in Table 3, recovery, ionization suppression and process efficiency (PE) are all related to trueness. In LC–MS, recovery is a more complex parameter than in case of most other analytical techniques. The quantity termed in this paper as process efficiency would in the context of most other analytical techniques be termed as recovery. In the case of LC–MS methods, it is useful to make a distinction between recovery (referring to the losses of analyte during sample preparation) and PE (referring to loss of analyte signal including the effects from sample preparation – recovery – and analyte ionization/detection) [3,5,81]. The three terms – recovery, ionization suppression/enhancement and process efficiency – can all be separately estimated. In LC–MS context the most important of them is ionization suppression/enhancement, which evaluation is described below.

Significant confusion of the calculation methods of these parameters is observed in the scientific literature. Different validation guides do refer to recovery from the sample preparation point of view [13a,25,95]. Recovery can be calculated as the percentage of the analyte signal in a solution obtained from sample preparation compared to a standard solution at the same concentration level. Nevertheless, guidance documents consider recovery and request its determination at high and low concentrations. Several guidance documents [27,95] give acceptable recovery values that are dependent on the concentration of the analyte and SANCO specifies acceptable recovery between 70 and 120% (which is in this case a requirement for PE) [4]. It is important to keep in mind that low PE also influences the LoD and LoQ of the method. Therefore, methods with poor PE may end up with unacceptably high LoD and LoQ values.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

The matrix effect in the broader sense is addressed in several validation guidelines [4,14,27], but ionization suppression/enhancement in LC–MS ion source, in spite of its large impact [5], has until recently not been in the focus of most validation guides. As one of the few, FDA is taking steps towards that with the most recent updated validation guide for bioanalytical method validation [13b] where the LC–MS ionization suppression/enhancement issue is addressed. This guide stresses the need to assure the absence of ionization suppression/enhancement and offers some technical guidance (see below). SANCO also discusses the issue of matrix effects for LC–MS analysis of pesticides [4]. In scientific analytical chemistry literature, on the other hand, ionization suppression/enhancement is nowadays almost routinely considered, sometimes as a separate validation parameter [96] but sometimes as part of robustness [97], trueness or recovery [59,98–100].

In the literature several methods, both qualitative and quantitative, have been proposed for evaluation of the ionization suppression/enhancement (matrix effect).

1. Qualitative estimation of ionization suppression/enhancement.
   i. The first method is suitable for detecting the presence of ionization suppression by recording the matrix effect profile [101] with post-column infusion method. For this the blank sample extract – not containing the analyte – is injected into the LC. At the same time a stream of analyte solution is mixed with the chromatographic effluent exiting the column and the mixture is delivered into the ion source. MS monitors the analyte signal. The same is done while injecting a blank solvent into the LC column. If no suppression/enhancement is present, then the analyte MS signal intensities are equal for sample and solvent injection within the precision of the MS signal. For a retention time region where ionization suppression occurs, the MS signal in the sample injection decreases and for ionization enhancement increases. For a method, not affected by ionization alteration effect, the analyte peak should elute away from the suppression and enhancement region. The matrix effect profile method has been used in several papers [102,103]. This approach does not enable quantitative evaluation of ionization suppression/enhancement but only its presence/absence. Still it can be very useful if chromatographic separation of the analyte and possible compounds causing ionization suppression is desired.
2. The new FDA validation guide proposes evaluation of ionization suppression by assessing the parallelism of dilution plots [13b]. Unfortunately no more specific guidance is given. We suggest building these plots in axes analyte signal versus dilution factor (defined as V_{sample}/V_{sum}). This approach is also applicable if blank matrices are not available although no quantitative conclusions can be made based on the comparison of dilution curves.

- Quantitative estimation of ionization suppression/enhancement is possible with post-extraction addition methods
  i. For this the analyte standard solution with known concentration is prepared in solvent and analyzed with LC-ESI-MS giving the peak area $A_{standard}$. Also a blank sample extract is prepared and spiked with the analyte at the same concentration level and thereafter analyzed giving peak area A_{sample}. The ionization suppression/enhancement effect can be calculated:

$$ME_{ionization} = \frac{\text{Analyte signal}_{\text{post extraction spiked matrix}}}{\text{Analyte signal}_{\text{solvent}}} \times 100\% \quad (6)$$

$ME_{ionization}$ value 100% indicates no effect, less than 100% indicates ionization suppression and $ME_{ionization}$ over 100% indicates ionization enhancement due to coeluting sample compounds. [3,5,104]. From this definition, most often used in LC–MS literature, some possible misunderstandings arise. The expression "reduce matrix effect" does not mean reduced value of %ME, but a $ME_{ionization}$ value becoming closer to 100%.

- Sometimes also positive/negative $ME_{ionization}$ scale is used, where 0% denotes no effect, values above 0% ionization enhancement and below 0% suppression [96,105]. The corresponding equation is:

$$ME_{ionization} = \frac{\text{Analyte signal}_{\text{post extraction spiked matrix}} - \text{Analyte signal}_{\text{solvent}}}{\text{Analyte signal}_{\text{solvent}}} \times 100\% \quad (7)$$

- Instead of comparing the peak areas, calibration graph slopes can be compared [106]. In this approach, two calibration graphs are constructed, one in the solvent and the other in the post-extraction spiked samples. Several aspects have to be kept in mind. First the intercepts of both calibration graphs have to be negligible so that the ionization suppression/enhancement would not depend on the analyte concentration. Unfortunately, the latter is not always true [19]. This approach is usable also in the case when blank matrix is unavailable [107] as the analyte already present is assumed to affect only the intercept of the graph and not its slope. A similar approach is described in the FDA upcoming validation guide [13b]. Still, before using the approach based on slope, linearity needs to be studied. In the literature this approach for ionization suppression/enhancement is often used and sometimes also combined with *F*- and *t*-test [55,108–110] or ANOVA [111] to evaluate statistical significance of the obtained matrix effect values.

All of the above described calculations can be done either in signal scale [96,112–114] (as used in Eqs. (1) and (2)) or concentration scale [115]. The obtained results are fairly similar if the samples used for ionization suppression/enhancement study are within the linear range and the intercept of the calibration graph is negligible. If these requirements are not fulfilled, it is more useful, from the method point of view, to use concentration-based calculations.

In the literature all the above described ways for ionization suppression/enhancement evaluation have been used. Often ionization suppression/enhancement is studied even if the corresponding validation guide does not require it [116,117] but sometimes vice versa is also observed – validation requires assessment of "matrix effect" but it is not evaluated during validation [118]. Sometimes qualitative and quantitative methods have been used hand-in-hand to study matrix effect [92].

If signal- or concentration-based calculations are used (not slope-based), the number of samples and replicates used for suppression/enhancement assessment during validation becomes an issue. Often several replicates are run at one [36] or more [31] concentration levels. For example Gupta et al. [36] studied ionization suppression at the lower limit of quantitation in 6 replicates for each matrix. On the other hand, Beaudry et al. [119] studied ionization suppression at 3 concentrations (covering the concentration range over 3 orders of magnitude) for propofol determination in human blood and plasma.

It has been often shown [2,120] that matrix effect depends on the sample source (e.g., different patient, different variety of fruit). It is, therefore, also recommended to use different matrices for suppression/enhancement evaluation. In the literature the number

of matrices used varies a lot. For example Dubreil-Chéneau et al. [121] used 24 different sources of honey to evaluate ionization suppression. Nirogi et al. used 5 different matrix sources of human plasma samples [122] for suppression studies during validation. Sometimes also the number of matrices is not specified.

In the literature it has been observed that ionization suppression/enhancement may strongly vary from day to day [123–125] and it cannot be estimated once during method optimization/validation and be used later for result correction. Pizzuti et al. [123] recommended using the ionization suppression/enhancement from validation data only as an indicative tool.

Also matrix effect can be studied together with other effects (e.g., recovery) influencing trueness. For example Careri et al. [59] studied matrix effect together with recovery and was still able to improve the method based on the obtained data.

3.2.3. Accuracy

Accuracy of a measurement result refers to the closeness of agreement between the measured value and the true value [40,77]. The main difference from trueness is that accuracy can be applied to an individual result, not just to the mean of a large number of results. Accuracy of a method refers to the ability of the method to deliver accurate results. As such, accuracy of a method is affected by systematic (bias) as well as random (precision) error components [25a,73,126] and is, therefore, studied as two components: trueness and precision [25]. Accuracy is quantitatively expressed as measurement uncertainty [40,126], which is discussed below. A number of regulatory bodies (ICH, FDA, EMA) define accuracy as the degree of agreement between the experimental value, obtained by replicate measurements, and the accepted reference value [13a,14,26], which is identical to the definition of trueness. For the evaluation of acceptability of accuracy, different evaluation criteria can be used: E_n-numbers, z-scores or ξ(zeta)-scores [127]. Recently, the accuracy profile approach [128] has been proposed as a useful tool to map the dependence of accuracy on analyte content in the sample.

3.2.4. Planning precision, trueness and accuracy experiments for LC–MS

The experiments for precision and trueness evaluation are often carried out in parallel. If this is not the case then before trueness experiments, precision of the method must be checked. This is because precision affects evaluation of trueness (or its constituents), but not vice versa.

Planning precision/bias experiments includes the following choices/decisions: (1) which sample types/matrices to use; (2) which concentration levels to use; (3) which time range to use; (4) how many replicates to make and (5) how to study ionization suppression. Overall, it is important that the obtained precision and bias values are representative of the likely analysis conditions, including sample matrices [48].

1. Most guidelines agree that the used matrices should as well as possible correspond to the matrices encountered in routine analysis. Each replicate determination should be made from a separate subsample and include all steps of the method (including sample preparation).
   i. When choosing samples for precision testing, the following are important considerations:
      a. The sample matrices should represent those routinely analyzed in the laboratory. The sample(s) can in principle be of any origin: leftovers from large routine samples, leftovers from proficiency testing samples, etc. Accurate reference values are not necessary and it is generally not advised to use certified reference materials (CRM) as samples for determining precision, since these are often better homogenized than real samples, leading to too optimistic estimates of precision [48].
   ii. The sample from which the subsamples are taken must be homogenous to the extent that the variability between the subsample results due to inhomogeneity is significantly lower than the variability due to the analysis.
2. Sufficient amount of sample must be available so that the required number of replicates can be carried out.
3. In the case of determining intermediate precision, the sample has to be stable during a longer period of time.

- For trueness evaluations reference values are needed and three types of samples can be used:
  a. Certified reference materials (CRM) with sufficiently small uncertainty of the reference value [48]. CRMs can very well address method and laboratory bias, as long as the CRM matrix matches the sample matrix. If samples with different matrix composition are analyzed, then additionally the matrix variation effect has to be taken into account. [80] CRM should match both the matrix routinely analyzed by the validated method as well as the range of the expected concentrations of the analyte in real samples.
- Reference materials that do not have to have a certified uncertainty estimate. These can be materials characterized by a reference material producer, but whose values are not accompanied by an uncertainty statement or are otherwise qualified; materials characterized by a manufacturer of the material; materials characterized in the laboratory for use as reference materials; and materials subjected to a restricted round–robin exercise, or distributed in a proficiency test [48]. If from background information there is a reason to assume high quality of the reference value, then they can be used the same way as CRMs. The ideal reference material is very similar to the sample of interest [25]. In addition, it is mandatory to use a particular reference material for one purpose during a validation study. As a most common example: for trueness evaluation a different reference material should be used than that is used for calibration [25b].
- When no reference materials are available, bias can be investigated by spiking studies. Sample is split into two aliquots – one is analyzed in its original state and the other is analyzed after a known amount of the analyte has been added. Calculation of bias is carried out by equations in Table 2. Importantly, in the case of many sample matrices, it may be difficult or impossible to spike the analyte into the sample in such a way that its molecular environment will be the same as that of the native analyte. Thus, the molecular interactions experienced by the spiked analyte may differ from those experienced by the native analyte. As a result, the spiked analyte may behave somewhat differently from the native analyte and the obtained bias value may not accurately reflect the bias operating on the native analyte. In most cases, the native analyte is more strongly bound by the matrix than the spiked analyte, resulting in somewhat optimistic bias estimates [25,48]. Therefore, spiking/recovery studies are accordingly very strongly subject to the observation that while good recovery is not a guarantee of trueness, poor recovery is certainly an indication of lack of trueness [25,48]. Strictly, this type of trueness studies only assesses bias due to effects operating on the added analyte [25,48]. The smaller the recovery, the larger the bias affecting the method and the lower the trueness of the method [80].

1. The precision as well as bias of LC–MS (as most instrumental techniques) depend on the concentration [12]. Therefore, it is important to investigate precision at different concentration levels. Validation guidelines differ concerning the recommended

number of concentration levels: 2 [48], 3 [13a,95], 4 [14,27]. The Eurachem guide does not give concrete guidance but suggests a number of concentration levels, stating that "a number of concentrations" across the working range should be used. The ICH guide recommends two approaches. If the analyte content in the sample is expected to vary significantly (e.g., in determination of contaminants or impurities) then altogether 9 measurements should be carried out, covering the expected concentration range, whereby the user can decide how to group the measurements between concentration levels (e.g., 3 concentration levels, 3 replicates each). If the analyte content in the sample is not expected to vary significantly (e.g., in assay determination) then 6 determinations should be carried out at the expected analyte concentration.

In our opinion the reasonable number of concentration levels is first of all determined by the purpose of the method. A number of concentration levels may be needed for determination of impurities (which can be present at vastly different levels), while just one concentration level may be sufficient in the case of assay. The actual working range of the method, as well as dependence of precision on analyte concentration, are also important.

If necessary and possible, bias can be expressed as a function of concentration [25,48] or at least the range where acceptable bias is obtained should be defined [129].

1. For determination of precision and bias, a sufficient number of replicate measurements over a suitable timespan is important. This requirement is universal to all analytical techniques, but in the case of LC–MS it is amplified by its rather poor precision and tendency of the instrument performance to change between days (due to, e.g., contamination). Repeatability study can be carried out during a single day, but s_{RW} should be determined over a longer period of time. Most guidelines do not specify a concrete timespan, except the Nordtest and EMA guides. In the Nordtest guide at least several months, preferably one year is suggested as suitable time span [78]. The EMA guide suggests for evaluation of the between-run precision (in principle corresponding to intermediate precision) that at least three runs per concentration level should be analyzed on at least two different days. This practice, however, cannot be widely recommended, because two days is too few, by any standards and the time interval between these days it is not specified. In addition, most published research works have intermediate precision experiments in the range of 3 days [30,85,116,117] showing that when validation is carried out according to the guidelines, very short time span is used. On rare occasions, periods as long as few weeks to two months, have been reported [40,89,130]. The time period should be chosen such that all variability factors (solvent and reagent batches, column, regular instrument maintenance, environmental conditions, ...) will have changed over this period of time. We recommend using the suggestion of the Nordtest guide [78], i.e., at least several months, preferably one year. It can be difficult under routine lab conditions to devote such long time to method validation. Therefore, a pragmatic approach would be that a preliminary estimate of s_{RW} is obtained during one month and if acceptable then the routine use can start. Simultaneously with routine use of the method more s_{RW} data are collected so that eventually a more reliable s_{RW} estimate is obtained for the method.
2. There is more agreement between guidelines considering the necessary number of replicate measurements. The most common recommendation is to make 5 replicates at each concentration level [13a,14,27,95]. ICH allows at least 3 (see above) and Eurachem suggests 6–15 replicate measurements in their current guide [[25b] (10 in the previous guide [25a]). It should be kept in mind that with too few replicate measurements, reliable precision estimate might not be obtained and bias cannot be successfully separated from precision. Our recommendation is to perform 5–10 replicate measurements or more.

Evaluation of trueness assessment is often combined with precision experiments. However, if separate experiments are used, a reduced design may be used for trueness, e.g., 2 replicates in five days. This is allowed because trueness estimates of comparable reliability can be obtained with fewer replicates than the respective precision estimates [84]. For intermediate precision, it is acceptable to have fewer replicates if at the same time measurements are performed over a longer period of time, because then a larger variety of effects will be accounted for [79].

Pooled standard deviation has also been recommended for intermediate precision evaluation since pooling different standard deviation enables obtaining more degrees of freedom than is usually possible when determining individual standard deviations, thereby improving the reliability of the estimated standard deviation [40,131].

5. The ionization suppression can be either determined individually or be included within the overall bias. Even though it is tempting to study all possible effects to the measurement accuracy one by one it is not always practical. Evaluation of the trueness allows to have a good quick estimation of the methods overall performance. Therefore, we suggest starting by studying ionization suppression within trueness during validation. If the trueness study yields unexpected or unsatisfactory results then a closer look should be taken to the method performance in order to improve and reoptimize the critical steps in the method. In this context separate evaluation of ionization suppression/enhancement is often needed and useful information about the method can be obtained.

When planning ionization suppression determination experiments some caution has to be taken:

(1) The effect should be studied in the whole range of analytical interest – either using replicates at high, medium and low concentration or using a calibration graph covering the whole linear range. Suppression assessment at one concentration is justified if the analytical range is narrow.
(2) As complete as possible coverage of the routinely analyzed matrices should be attempted. This can become very work-intensive if a wide variety of matrices are included in the analytical method. Also, even if formally only one type of samples is analyzed by the method, the possible variations of the sample matrix should be taken into account. For example, in different varieties of apples, ionization suppression can occur to different extent [3]. Analogous results have been obtained with pomegranates [86]. In order to have an acceptable method ionization suppression/enhancement should be acceptable for all matrices under study.
(3) If blank matrices are not available, then preferably addition method – e.g., comparing the slopes of the addition plots for sample and solvent – should be used (see above). These methods are applicable only in the linear range.
(4) if possible, all calculations should be made with the concentrations based equations in order to estimate the effects that will appear in the real samples.

3.3. Stability

Analyte stability is not universally included in validation guidelines as a validation parameter. For example, ICH [26],

NordVal [95], Eurachem [25] and IUPAC [48] guidelines do not specifically address analyte stability. The reason is that if the analyte is unstable its decomposition influences the trueness and precision of the procedure and is, thus, accounted for by these two parameters. Decomposition usually leads to lowering of the analyte content. In the specific case of analysis of decomposition products decomposition can actually lead to increase of the analyte content. In either case, decomposition is essentially a systematic effect. However, the rate of decomposition can be strongly dependent on fine experimental details (matrix, access of oxygen, temperature, light etc.). For this reason both systematic and random effects are usually involved and analyte instability affects both trueness and precision. Furthermore, besides the analyte in the samples analyte in the standards can also decompose. If both occur at the same rate then decomposition only affects precision. If not, then both trueness and precision are affected.

SANCO [4] specifies briefly that analyte stability in prepared sample extracts has to be evaluated. However, a very large share of LC–MS analysis is done in the bioanalytical field, where unstable analytes are rather a rule than an exception. In that field, possible analyte decomposition is of very high importance to quality of the results and, therefore, deserves special attention. For this reason, the EMA, FDA and AOAC validation guidelines specifically address analyte stability as a separate validation parameter. In addition, the EMA guide stresses that analyte stability cannot be proven by literature data [14], further outlining the importance of analyte stability testing.

The analyte or internal standard stability in test or reference solutions, ST%, expresses the part of the analyte or internal standard in a sample that does not decompose before the actual LC–MS analysis of the sample [69]. Stability can be evaluated either via chromatographic peak areas (1) or via concentrations (2).

(1) Stability can be evaluated via peak areas as follows:

$$\mathrm{ST}\% = \frac{S_t}{S_0} \times 100\% \tag{8}$$

where S_0 is the initial peak area, determined without introducing any extra pauses in the analysis process; S_t is the peak area obtained when analysis is carried out with making a pause with duration t in the analysis.

- Stability can be evaluated via concentrations as follows:

$$\mathrm{ST}\% = \frac{C_t}{C_0} \times 100\% \tag{9}$$

where C_0 is the initial concentration, determined without introducing any extra pauses in the analysis process; C_t is the concentration obtained when analysis is carried out with making a pause with duration t in the analysis.

The guidelines by EMA [14] and FDA [13b] as well as the AAPS/FDA white paper [132] from 2007 (the latter actually forms the basis for the FDA stability-related recommendations) specify a set of different conditions at which stability estimation is to be carried out.

FDA [13b] distinguishes the following types of stability: freeze and thaw stability, bench-top stability, long-term stability, stock solution stability, processed sample stability. This guide distinguishes between the analyte stability in calibration and stock solutions and stability in sample matrix and stresses the importance of storage conditions, matrix and container system on stability, besides the intrinsic properties of the analyte itself.

According to EMA [14] stability of the analyte is evaluated using both low- and high-level quality control samples. The investigation of stability should cover short-term stability at room temperature or sample processing temperature and freeze–thaw stability. In addition, long-term freezer stability should be studied at each temperature at which study samples will be stored. The emphasis in the EMA guide is not as much on the intrinsic stability of the analyte, as on its stability in the specific matrix.

AOAC [27] is less specific on the experiments that have to be carried out and recommends checking the stability of the stock and initial diluted solutions, stored at room or lower temperatures, by repeating their measurements several days or weeks later.

Stability should be studied at least at two concentration levels – low and high concentration level [13b,14] and with matrix, matching the "real-life" matrix. For this the native blank biological matrix should be used and the analyte should be spiked into the matrix at these concentration levels [13a,14,132,133]. Bench-top stability [13a] or short-term stability at room temperature or sample processing temperature [14] will indicate analyte stability under sample preparation conditions. Freeze–thaw stability is evaluated usually during three thawing cycles to predict the possible delays and glitches on sample handling [13b,14].

Stability should be evaluated in different time points and the samples should be analyzed in six replicates. Stability can be evaluated as the average percentage of analyte found in the sample under the specific conditions. The freshly prepared calibration standards (at the same concentration levels as the stability study samples) are considered as containing 100% of the initial analyte content.

Different criteria exist for assessing whether the observed analyte stability is sufficient. The FDA and EMA guides recommend using the simple acceptance criteria: at least 2/3 of samples have to fall into the range 80–120% [14] or 85–115% [13a] in order to be acceptable [13a,14,132,133]. More sophisticated stability evaluation has been recommended, based on the regression analysis of analyte content change in time [134]. According to Épshtein, analytes and internal standards can be considered stable if the difference (100%–ST%) does not exceed the relative error of determination of the main component (or impurity) according to the given analytical procedure [69]. By its essence this criterion is good, but it can be difficult to apply it in practice.

Épshtein has also proposed an alternative way of assessing stability, not involving explicit determination of ST(%). The change of analyte concentration in time can be statistically confirmed by the regression equation

$$A_t = A_0 + b \times t \tag{10}$$

where A_t is the area under the peak at time t, A_0 is the area at the initial moment and b is the slope of the regression line. The statistical significance of coefficient b can be tested using the t-test [69].

Our recommendation is the following:

(1) The ST% is calculated either via peak areas or via concentrations using at least three replicates at two concentration levels.
(2) If the difference $(100\% - \mathrm{ST}\%)$ in 2/3 of the samples is lower than 1/5 of the expected relative standard uncertainty of the method, then the stability can be considered sufficient and no further action is necessary. If not, one or more of the following should be done:
   a. Reduce analyte decomposition by changing the method parameters (e.g., sample preparation time, autosampler temperature, etc., see Fig. 4 for possibilities).
(3) Use appropriate ILIS.
(4) Increase the measurement uncertainty estimate of the method to account for the limited analyte stability.
(5) Evaluate the analyte decomposition and correct the results for it.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

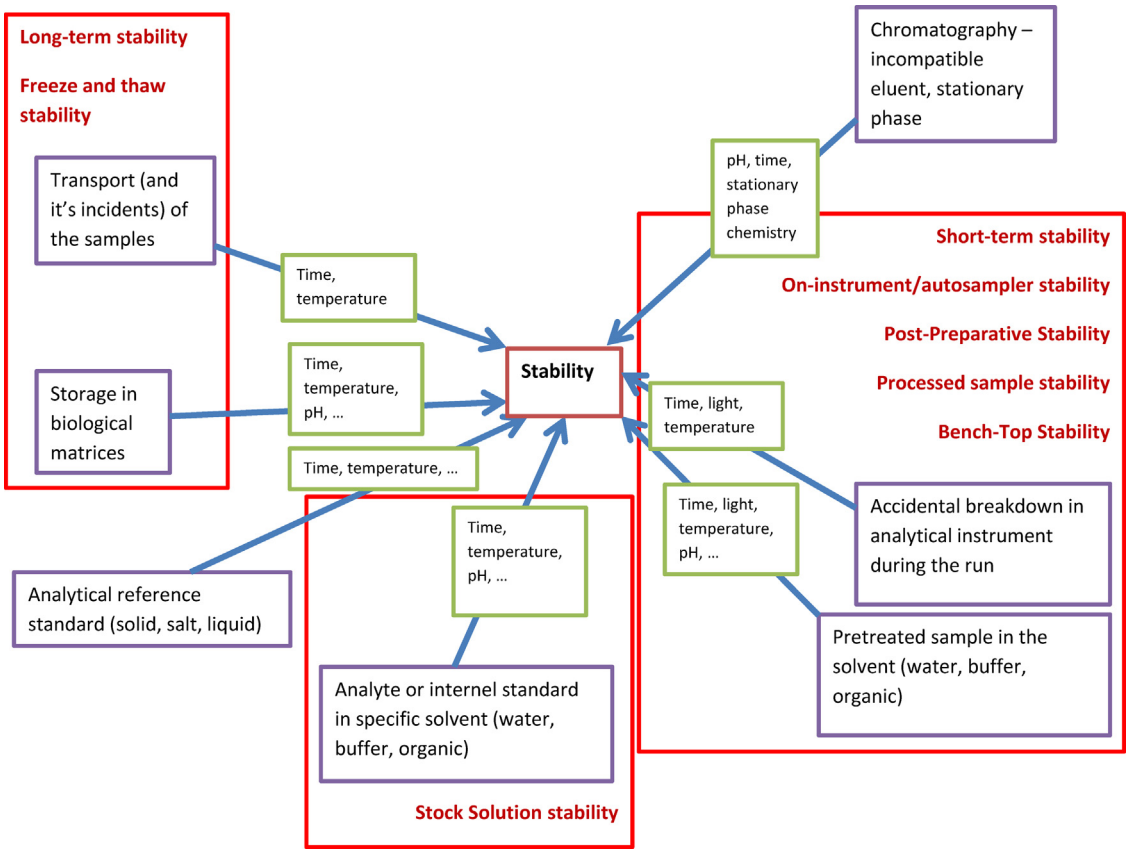


Fig. 4. Different types of stability in LC–MS analysis and the chemical/physical parameters influencing stability.

The time during which stability is tested, is important, but is usually not specified in the guidelines, only brief recommendations are given. The reason is that depending on the particular situation, the suitable length of the stability study can be very different and should be decided by the analyst. Usually when stability of the analyzed sample solutions or standard solutions is evaluated, the injections of the sample and standard solutions are carried out overnight [44,75,83]. Testing of (not yet prepared) samples should be carried out during at least a 48 h period, after which the components should be quantified and the terms for storage and conditions that improve the stability over the time, should be identified [44,83].

The experimental design of stability testing should take into account most important parameters – time and temperature of the analysis and the stability results should be used as the indicator to adjust one or the other in order to improve the stability of analyte or internal standard.

3.3.1. Stability report

The stability report should include the concentrations of analyzed samples and storage/analysis times, pH-s, mobile phases, elution parameters (gradient) and column used in the analyses.

3.3.2. Practical examples

The bench-top/short term stability of a beta-lactam antibiotic – meropenem – in processed blood plasma samples at the room temperature (+20 °C) was investigated during 24 h using overnight injections [135] and after 12 h the concentration of meropenem in the samples was only 17% of the original concentration at both concentration levels 1 ug mL^{-1} and 25 ug mL^{-1}. After keeping the autosampler at 4 °C and samples at −20 °C until loading into the autosampler, the ST% remained around 95% after 3 h and 92% after 12 h storage in autosampler (+4 °C). Thus, storage of samples and solutions at low temperature is a powerful means of increasing stability.

The well-known antibiotic penicillin G is also well known for its instability in solutions. Our unpublished data show that the bench-top/short term stability of processed blood plasma samples in an autosampler thermostated at +4 °C showed rapid degradation: ST% was 20% after 16 h for samples with pH 7.0. After pH adjustment to 6.2 processed blood plasma samples showed significant improvement on ST%, resulting in 78% after 16 h at the +4 °C. This example convincingly demonstrates how small changes in conditions can strongly influence analyte stability.

Degradation of the samples can be caused also by light. Photostability of the analytes has been investigated in several studies [136,137].

3.4. Measurement uncertainty

Many of the sources that address validation of LC–MS methods also touch measurement uncertainty (MU) [138,139]. This is fully justified, since MU is closely linked to several validation parameters (precision, trueness) and validation data form an important source of information for measurement uncertainty estimation [78,140].

In spite of this, in most validation-related sources MU is not included as a validation parameter but is kept separately, since MU traditionally characterizes rather an analysis result than an analysis method. We have nevertheless decided to briefly address MU in this review because of two reasons: (1) according to modern views [140] MU can also be understood as a parameter characterizing a method and (2) estimation of MU is nowadays often required from laboratories [141] and they often have problems with estimating MU of results of complex analyses. The ICH guide [26] does not address MU. The IUPAC guide [48]

includes some information and help for MU estimation. The Eurachem guide [25] briefly touches MU estimation.

In very broad terms there are two main approaches to MU estimation [140]: the modeling approach (sometimes also called the ISO GUM [142] Modeling approach) and the statistical approach based on validation data (sometimes also called single laboratory validation approach). Both of these have been used for MU estimation in LC–MS analysis. The Eurolab report [140] actually presents four approaches to MU estimation but only the above mentioned two of them can be recommended and are further discussed here.

The modeling approach [142] is considered the de facto standard approach for measurement uncertainty estimation. It is a rigorous and investigative approach, which, if carried out correctly, enables obtaining uncertainty budget – the list of uncertainty sources together with their contributions to the uncertainty of the result. This is a valuable tool for improving and optimizing the method. Yet in the modeling approach there are two difficult issues: (1) identifying all uncertainty sources and (2) estimation/quantification of their contributions to the MU of the results. Their difficulty is further amplified by the usual situation with LC–MS analysis methods: complex matrices, low analyte concentrations and the non-robust nature of the MS detector as well as the ion source. Quantification of uncertainty contributions is in such cases especially difficult and there are uncertainty sources for which rigorous and reliable quantification of uncertainty contributions is almost impossible: partial analyte decomposition during sample preparation, sample carryover, decrease of ionization efficiency due to ion source contamination, analyte ionization suppression by closely eluting matrix constituents.

Either neglecting some uncertainty sources or incomplete understanding of their magnitude will usually lead to underestimated measurement uncertainties if the modeling approach is used. Therefore, it is often necessary to use experience- or opinion-based uncertainty estimates in such cases. Although experience- or opinion-based estimates are not entirely satisfactory, the alternative – neglecting the uncertainty sources – would be still worse as it would lead to underestimated uncertainties.

In spite of these problems the modeling approach, if competently used, can lead to reliable MU estimates [124,143].

In the single lab validation (SLV) approach detailed investigation of uncertainty sources is not carried out. Instead, very general uncertainty contributions, which include a number of individual uncertainty sources, are considered and quantified. Perhaps the best-known formalization of the SLV approach is that published by Nordtest [78]. In the Nordtest guide the combined standard uncertainty (u_c) is viewed as composed of the intermediate precision $u(R_w)$ and uncertainty due to possible lab/method bias u(bias):

$$u_c = \sqrt{u(R_w)^2 + u(\text{bias})^2} \quad (21)$$

The $u(R_w)$ component is usually obtained either from control charts [78] or from pooled standard deviation [79] of routine sample results. The bias component is usually obtained from analysis of CRMs, from spiking studies or from results of participation in interlaboratory comparisons (ILCs) [78]. Detailed explanation of the SLV approach (the Nordtest implementation) is available in Ref. [78]. Explanations together with a worked LC–MS example are also available as part of a web course on MU estimation [79]. These sources also address the practical questions, such as at which concentration levels to carry out MU estimation, which time period to choose for $u(R_w)$ estimation, how often should the MU estimate be revised, etc.

The SLV approach does not have the above mentioned drawbacks of the modeling approach: careful investigation and individual quantification of all uncertainty sources is not needed, the danger of underestimating uncertainty is lower. Often the data needed for MU estimation with the SLV approach are automatically available in lab. The main difficulty with the SLV approach is finding samples with reliable reference values for estimating the u (bias) component. CRMs are a good choice, but often unavailable for a specific analyte–matrix combinations and consensus values of ILCs have low reliability, i.e., high uncertainty. If the reference value has high uncertainty then u(bias) and consequently also u_c will be inflated. In such cases the SLV approach tends to lead to overestimated uncertainties [79].

Comparing the pros and cons of the modeling and SLV approach it can be concluded that the modeling approach is suitable for expert labs, while for routine analysis labs the SLV is more suitable.

Other approaches have also been suggested. For example, SANCO recommends that whenever a laboratory has participated successfully in ILCs, the expanded measurement uncertainty at $k = 2$ level can be uniformly estimated as 50% relative. This is a very approximate approach but its undeniable virtue is its simplicity.

4. Practical considerations

The diverse possibilities but also the pitfalls inherent in LC–MS increase the importance of some aspects of LC–MS method validation (compared to LC methods with other detectors) as well as interpreting the validation results:

(1) The operation efficiency of the MS detector is strongly influenced by numerous parameters, both those that can be accurately preset (gas flows, different voltages) and those that are difficult to control (e.g., MS contamination level). Therefore, constant monitoring of the accuracy after the first stage of validation is often required, especially if complicated samples are analyzed. Also, validation of ruggedness and robustness can be more important than with simpler detectors.
(2) Precision- and trueness-related parameters should be determined from a larger number of replicates than is necessary in e.g., LC–UV. This is especially true for the trueness-related parameters (bias, ionization suppression, process efficiency).
(3) MS as a detector can be strongly influenced by matrix variations. In different matrices different matrix compounds can be incompletely chromatographically separated from the analyte. This can cause ionization suppression or enhancement and thereby influence the method's trueness. Therefore, trueness experiments should be carried out in all matrices belonging to the scope of the method.

4.1. Carrying out validation in practice

It is always of importance how to perform validation in the most effective way. As validation and optimization are closely related it is complicated to give strict instructions. Different sequences of operations have been suggested in the literature [25a,26,44,81]. Based on the literature data and our own experience we have tried to suggest a possible general sequence of validation in Fig. 5.

The validation should start with evaluating analyte stability and method selectivity as all the other parameters strongly depend on these. For example no linear relation can be achieved if analyte extensively decomposes in the autosampler. In that case, non-linear calibration models can be considered (see Section 3.1.1). Consequently we propose estimation of linearity as the next step, because for evaluation of trueness and precision we need to know the linear/working range of the method. We propose robustness studies as the last step of validation. It is sometimes suggested to test robustness as one of the first things in method validation or in

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

the end of the method development phase [26,144]. We find it important to have some insight, as to which are the most important performance characteristics (e.g., closest to the legal limits or requirements of the client) before deciding which of the method performance characteristics are varied during robustness studies.

It is often unnecessary to carry out determination of all possible method performance parameters. This kind of validation is often called partial validation. If a fully validated (e.g., using a collaborative trial) method is set up in the laboratory then it is necessary to carry out so-called verification: verify that it is able to achieve the published values of performance characteristics [25,26]. Partial validation is justified when a standard method (e.g., ISO, ASTM) is used [25], small changes are made to previously validated methods [14] or for methods with narrow application range [26]. The small changes can include transfer of the method to another laboratory, adding a new matrix, new reagent in sample preparation, etc. [14]. FDA [13a] states that in the case of

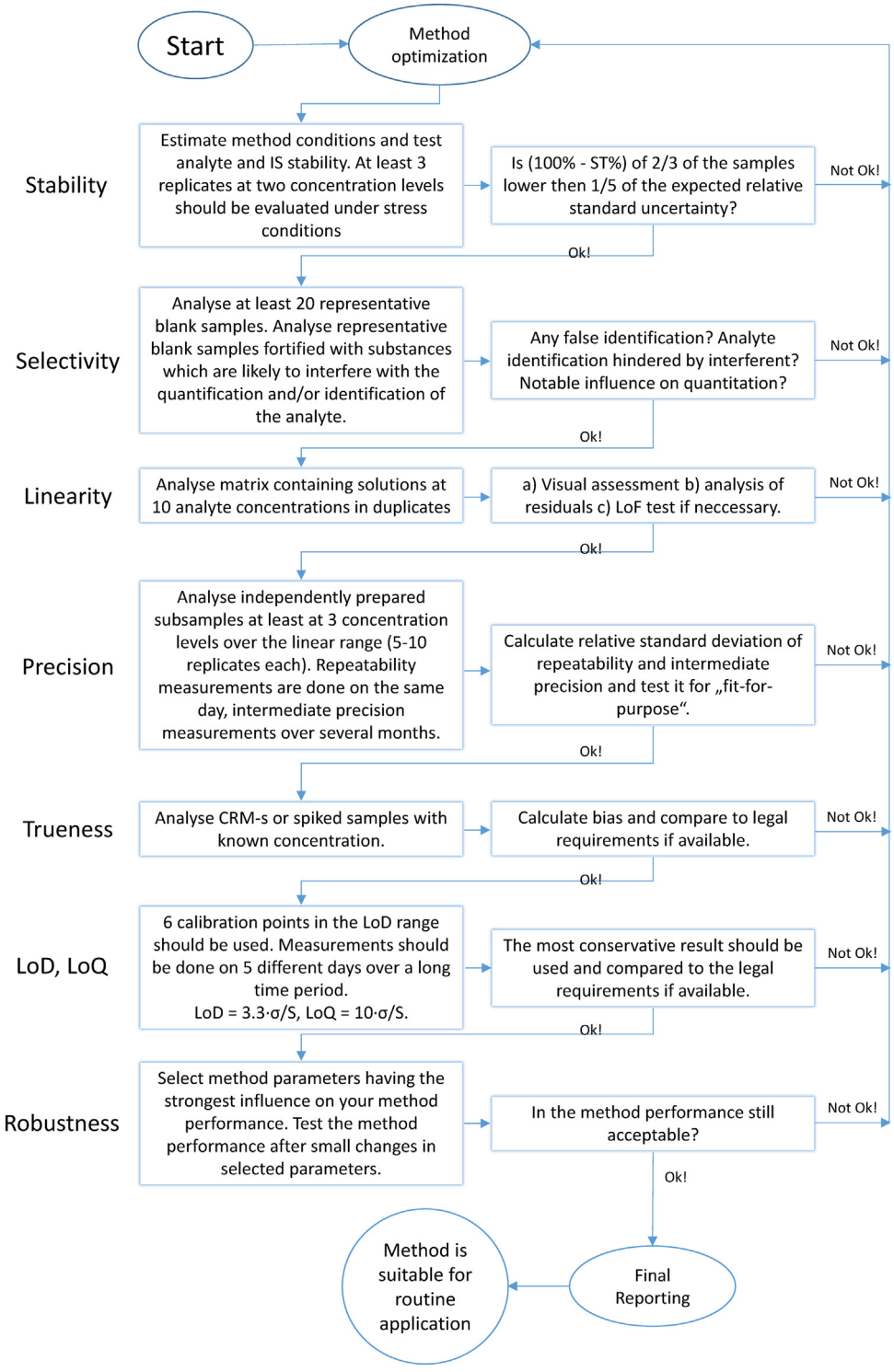


Fig. 5. A possible sequence of operations and decisions in LC–MS method validation.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

bioanalytical methods it is sometimes sufficient to determine only trueness and repeatability. ICH distinguishes between methods for identification, impurity analysis and assay analysis [26]. Different performance characteristics need to be determined for different types of methods.

Not only validation but also appropriate documentation of the validation is required for adequate interpretation as well as on the validity of the obtained results.

4.2. After the validation

As the last stage of validation an assessment of validity (fitness for the intended purpose) of the method should be given, based on the validation results. Validation guidelines almost lack advice concerning assessment of validity. Some general suggestions are given in the following paragraphs.

Before starting a validation a clear plan is compiled, which consists of the reason for validation, planned experiments as well as expected outcomes – requirements that need to be met by the method. The requirements often result from guidelines or from other regulatory documents. Making that plan depends on each different method under development and takes into account all specific aspects related to that method. After carrying out the necessary experiments, a decision must be made if the results are satisfying and consequently if the method is fit for purpose.

Guidelines generally give suggestions for evaluating separately each parameter and a suggestion for giving a decision for the whole method's validation is very general: validation has to prove that the values of all evaluated parameters are satisfactory. Few different cases arise.

(1) When methods are applied in the scope of standards, laws or directives, then these requirements must be followed and decision on validation should be based on these. When decision on validation suitability is based on the guidelines, then for each parameter a separate decision must be given according to the requirements.
(2) Sometimes the client can specify the requirements.
(3) If there are no external requirements then the analyst can set up requirements himself/herself based on his/her knowledge of the subject.

In addition to the estimation of performance characteristics during validation some LC–MS specific aspects have to be continuously evaluated during method usage. For example, the working condition of an MS system is strongly influenced by the contamination level of the samples, aging and other. It is, therefore, very important to monitor the performance of the LC–MS method during everyday application. Both internal as well as external quality control measures are useful for this type of monitoring [25].

Moreover, data about the method robustness should be monitored after the validation, while the method is in use and small changes in the method parameters are applied.

We find that one of the most convenient in-house approaches would be running quality control samples at different concentration levels and different matrices if applicable (including LoD and LoQ if the method is often used in the LoD or LoQ range) and monitoring the performance in the *X*-chart format. In case if *X*-chart is not applicable (e.g., the samples are instable) using an *R*-chart is recommended instead [25,78]. Sometimes also blind samples (in replicated test portions), prepared by another person than the analyst, are used in-house [25].

Finally, it is recommended to participate in proficiency testing schemes, especially for accredited laboratories [141]. If such schemes are not available, then analysis of reference materials other than the ones used in the validation, is also useful. Self-organized comparison measurements of the same sample with other laboratories are also a good way to monitor the performance of the method.

Acknowledgments

This work was supported by the institutional research funding IUT20-14 and personal research funding PUT34 from the Estonian Ministry of Education and Research, as well as the Estonian Science Foundation grant No 8572 and project "Development of software for validation of chromatographic methods" (with registration number 3.2.1201.13-0020) under the sub-measure "Supporting the development of R&D of info and communication technology" funded by the EU Regional Development Fund.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.aca.2015.02.016.

References

[1] A. Kruve, R. Rebane, K. Kipper, M.-L. Oldekop, H. Evard, K. Herodes, P. Ravio, I. Leito, Tutorial review on validation of liquid chromatography–mass spectrometry methods: part I, Anal. Chim. Acta (2015) , doi:http://dx.doi.org/10.1016/j.aca.2015.02.017.
[2] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, Anal. Chem. 75 (2003) 3019–3030.
[3] A. Kruve, A. Künnapas, K. Herodes, I. Leito, Matrix effects in pesticide multi-residue analysis by liquid chromatography–mass spectrometry, J. Chromatogr. A 1187 (2008) 58–66.
[4] SANCO/12571/2013, Guidance document on analytical quality control and validation procedures for pesticide residues analysis in food and feed. 2013.
[5] P.J. Taylor, Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry, Clin. Biochem. 38 (2005) 328–334.
[6] J. Wieling, LC–MS–MS experiences with internal standards, Chromatographia 55 (2002) S107–S113.
[7] S.L.R. Ellison, M. Thompson, Standard additions: myth and reality, Analyst 133 (2008) 992–997.
[8] M. Thompson, S.L. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis. IUPAC Technical report, Pure Appl. Chem. 74 (2002) 835–855.
[9] T. Zamora, O.J. Pozo, F.J. Lopez, F. Hernandez, Determination of tridemorph and other fungicide residues in fruit samples by liquid chromatography–electrospray tandem mass spectrometry, J. Chromatogr. A 1045 (2004) 137–143.
[10] A. Gentili, F. Caretti, S. Bellante, L.M. Rocca, R. Curini, A. Venditti, Development and validation of two multiresidue liquid chromatography tandem mass spectrometry methods based on a versatile extraction procedure for isolating non-steroidal anti-inflammatory drugs from bovine milk and muscle tissue, Anal. Bioanal. Chem. 404 (2012) 1375–1388.
[11] A. Margoum, C. Guillemain, X. Yang, M. Coquery, Stirbar sorptive extraction coupled to liquid chromatography–tandem mass spectrometry for the determination of pesticides in water samples: method validation and measurement uncertainty, Talanta 116 (2013) 1–7.
[12] A. Kruve, I. Leito, Comparison of different methods aiming to account for/overcome matrix effects in LC/ESI/MS on the example of pesticide analyses, Anal. Methods 5 (2013) 3035–3044.
[13] (a) U.S. Department of Health and Human Services Food and Drug Administration, Guidance for Industry: Bioanalytical Method Validation, U.S. Department of Health and Human Services Food and Drug Administration, 2001;
(b) U.S. Department of Health and Human Services Food and Drug Administration, Guidance for Industry: Bioanalytical Method Validation, Draft Guidance, U.S. Department of Health and Human Services Food and Drug Administration, 2013. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM368107.pdf.
[14] Guidance on bioanalytical method validation, European Medicines Agency, 2011.
[15] M. Oss, A. Kruve, K. Herodes, I. Leito, Electrospray ionization efficiency scale of organic compounds, Anal. Chem. 82 (2010) 2865–2872.
[16] C.G. Enke, A predictive model for matrix and analyte effects in electrospray ionization of singly-charged ionic analytes, Anal. Chem. 69 (1997) 4885–4893.
[17] K. Tang, J.S. Page, R.D. Smith, Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry, J. Am. Soc. Mass. Spectrom. 15 (2004) 1416–1423.
[18] N.B. Cech, C.G. Enke, Practical implications of some recent studies in electrospray ionization fundamentals, Mass Spectrom. Rev. 20 (2001) 362–387.

[19] A. Kruve, R. Auling, K. Herodes, I. Leito, Study of liquid chromatography/electrospray ionization mass spectrometry matrix effect on the example of glyphosate analysis from cereals, Rapid Commun. Mass Spectrom. 25 (2011) 3252–3258.
[20] J.S. Page, R.T. Kelly, K. Tang, R.D. Smith, Ionization and transmission efficiency in an electrospray ionization–mass spectrometry interface, J. Am. Soc. Mass Spectrom. 18 (2007) 1582–1590.
[21] D.W. Koppenaal, C.J. Barinaga, M.B. Denton, R.P. Sperline, G.M. Hieftje, G.D. Schilling, F.J. Andrade, J.H. Barnes, MS detectors, Anal. Chem. 77 (2005) 417–427.
[22] E. de Hoffmann, V. Stroobant, Mass Spectrometry: Principles and Applications, Wiley, Chichester, 2011.
[23] C. Soler, J. Mañes, Y. Picó, Comparison of liquid chromatography using triple quadrupole and quadrupole ion trap mass analyzers to determine pesticide residues in oranges, J. Chromatogr. A 1067 (2005) 115–125.
[24] M. Holčapek, R. Jirásko, M. Lísa, Recent developments in liquid chromatography mass spectrometry and related techniques, J. Chromatogr. A 1259 (2012) 3–15.
[25] (a) Eurachem, Eurachem Guide: The Fitness for Purpose of Analytical Methods, Eurachem, Teddington, 1998;
(b) B. Magnusson, U. Örnemark, Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, 2nd ed., Eurachem, Teddington, 2014 Available from www.eurachem.org. ISBN 978-91-87461-59-0.
[26] ICH Harmonized Tripartite Guideline: Validation of Analytical Procedures: Text and Methodology Q2(R1), International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2005.
[27] AOAC, AOAC Guidelines for Single-laboratory Validation of Chemical Methods for Dietary Supplements and Botanicals, Official Methods of Analysis, 19th ed., AOAC, INTERNATIONAL, Gaithersburg, MD, 2012 Appendix K.
[28] R. Zeleny, S. Harbeck, H. Schimmel, Validation of a liquid chromatography–tandem mass spectrometry method for the identification and quantification of 5-nitroimidazole drugs and their corresponding hydroxy metabolites in lyophilised pork meat, J. Chromatogr. A 1216 (2009) 249–256.
[29] M. Martínez-Huel'amo, S. Tulipani, X. Torrado, R. Estruch, R.M. Lamuela-Ravento's, Validation of a new LC–MS/MS method for the detection and quantification of phenolic metabolites from tomato sauce in biological samples, J. Agric. Food Chem. 60 (2012) 4542–4549.
[30] G. Satpathy, Y.K. Tyagi, R.K. Gupta, Validation of LC–MS/MS electrospray ionization method for estimation of oxytocin in fruits and vegetables; surveillance of residues of oxytocin in five different commodities, Electron. J. Environ. Food Chem. 10 (2011) 2722–2734.
[31] R.P. Lopes, D.V. Augusti, L.F. de Souza, F.A. Santos, J.A. Lima, E.A. Vargas, R. Augusti, Development and validation (according to the 2002/657/EC regulation) of a method to quantify sulfonamides in porcine liver by fast partition at very low temperature and LC–MS/MS, Anal. Methods 3 (2011) 606–613.
[32] H. Sun, F. Wang, L. Ai, C. Guo, R. Chen, Validated method for determination of eight banned nitroimidazole residues in natural casings by LC/MS/MS with solid-phase extraction, J. AOAC Int. 92 (2009) 612–621.
[33] D.B. Hibbert, Further comments on the (miss-) use of *r* for testing the linearity of calibration functions, Accredit. Qual. Assur. 10 (2005) 300–301.
[34] U. Koesukwiwat, S. Jayanta, N. Leepipatpiboon, Validation of a liquid chromatography–mass spectrometry multi-residue method for the simultaneous determination of sulfonamides tetracyclines, and pyrimethamine in milk, J. Chromatogr. A 1140 (2007) 147–156.
[35] D.V. Bharathi, K.K. Hotha, P.V. Sagar, S.S. Kumar, A. Naidu, R. Mullangi, Development and validation of a sensitive LCMS/MS method with electrospray ionization for quantitation of pramipexole in human plasma: application to a clinical pharmacokinetic study, Biomed. Chromatogr. 23 (2009) 212–218.
[36] A. Gupta, P. Singhal, P.S. Shrivastav, M. Sanyal, Application of a validated ultra performance liquid chromatography–tandem mass spectrometry method for the quantification of darunavir in human plasma for a bioequivalence study in Indian subjects, J. Chromatogr. B 879 (2011) 2443–2453.
[37] Analytical Methods Committee, Is my calibration linear? Analyst 119 (1994) 2363–2366.
[38] E. Saar, D. Gerostamoulos, O.H. Drummer, J. Beyer, Identification and quantification of 30 antipsychotics in blood using LC–MS/MS, J. Mass Spectrom. 45 (2010) 915–925.
[39] J. Beyer, F.T. Peters, T. Kraemer, H. Maurer, Detection and validated quantification of nine herbal phenalkylamines and methcathinone in human blood plasma by LC–MS/MS with electrospray ionization, J. Mass Spectrom. 42 (2007) 150–160.
[40] L.B. Nilsson, G. Eklund, Direct quantification in bioanalytical LC–MS/MS using internal calibration via analyte/stable isotope ratio, J. Pharm. Biomed. 43 (2007) 1094–1099.
[41] P. Araujo, Key aspects of analytical method validation and linearity evaluation, J. Chromatogr. B 877 (2009) 2224–2234.
[42] J. Van Loco, M. Elskens, C. Croux, H. Beernaert, Linearity of calibration curves: uses and misuses of the correlation coefficient, Accredit. Qual. Assur. 7 (2002) 281–285.
[43] G.A. Shabir, W.J. Lough, S.A. Arain, T.K. Bradshaw, Evaluation and application of best practice in analytical method validation, J. Liq. Chromatogr. Related Technol. 20 (2007) 311–333.
[44] S. Chandran, R.S.P. Singh, Comparison of various international guidelines for analytical method validation, Pharmazie 62 (2007) 4–14.
[45] P. Jedziniak, T. Szprengier-Juszkiewicz, M. Olejnik, Determination of benzimidazoles and levamisole residues in milk by liquid chromatography–mass spectrometry: screening method development and validation, J. Chromatogr. A 1216 (2009) 8165–8172.
[46] K.M. Kasiotis, C. Anagnostopoulos, P. Anastasiadou, K. Machera, Pesticide residues in honeybees honey and bee pollen by LC–MS/MS screening: reported death incidents in honeybees, Sci. Total Environ. 485–486 (2014) 633–642.
[47] W. Huber, On the use of the correlation coefficient *r* for testing the linearity of calibration functions, Accredit. Qual. Assur. 9 (2004) 726.
[48] IUPAC, Harmonized guidelines for single-laboratory validation of method of analyses (IUPAC technical report), Pure Appl. Chem. 74 (5) (2002) 835–855.
[49] Analytical Methods Committee, Uses (proper and improper) of correlation coefficients, Analyst 113 (1988) 1469–1471.
[50] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, 5th ed., Pearson/Prentice Hall, 2005.
[51] A.M. Almeida, M.M. Castel-Branco, A.C. Falcão, Linear regression for calibration lines revisited: weighting schemes for bioanalytical methods, J. Chromatogr. B 774 (2002) 215–222.
[52] S.L.R. Ellison, In defence of the correlation coefficient, Accredit. Qual. Assur. 11 (2006) 146–152.
[53] C.J. Anagnostopoulos, P. Aplada Sarli, K. Liapis, S.A. Haroutounian, G.E. Miliadis, Validation of two variations of the QuEChERS method for the determination of multiclass pesticide residues in cereal-based infant foods by LC–MS/MS, Food Anal. Methods 5 (2012) 664–683.
[54] O. Tiryaki, Method validation for the analysis of pesticide residues in grain by thin-layer chromatography, Accredit. Qual. Assur. 11 (2006) 506–513.
[55] S. Szilagyi, B. de la Calle, Development and validation of an analytical method for the determination of semicarbazide in fresh egg and in egg powder based on the use of liquid chromatography tandem mass spectrometry, Anal. Chim. Acta 572 (2006) 113–120.
[56] G.M. Hadad, S. Emara, W.M.M. Mahmoud, Optimization and validation of an LC method for the determination of cefdinir in dosage form and human urine, Chromatographia 70 (2009) 1593–1598.
[57] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, 2nd ed., Wiley-VCH, Weinheim, 2007.
[58] W. Funk, V. Dammann, G. Donnevert, Quality Assurance in Analytical Chemistry: Applications in Environmental, Food, and Materials Analysis, Biotechnology, and Medical Engineering, 2nd ed., Wiley-VCH, Weinheim, 2007.
[59] M. Careri, C. Corradini, L. Elviri, I. Nicoletti, I. Zagnoni, Liquid chromatography–electrospray tandem mass spectrometry of *cis*-resveratrol and *trans*-resveratrol: development, validation, and application of the method to red wine, grape, and winemaking byproducts, J. Agric. Food Chem. 52 (2004) 6868–6874.
[60] M. Mattarozzi, M. Milioli, C. Cavalieri, F. Bianchi, M. Careri, Rapid desorption electrospray ionization-high resolution mass spectrometry method for the analysis of melamine migration from melamine tableware, Talanta 10 (2012) 453–459.
[61] M. Suman, E. Bergamini, D. Catellani, A. Manzitti, Development and validation of a liquid chromatography/linear ion trap mass spectrometry method for the quantitative determination of deoxynivalenol-3-glucoside in processed cereal-derived products, Food Chem. 136 (2013) 1568–1576.
[62] L. Howells, M.J. Sauer, Multi-residue analysis of avermectins and moxidectin by ion-trap LC–MSn, Analyst 126 (2001) 155–160.
[63] A. Koehler, R. Oertel, W. Kirch, Simultaneous determination of bupivacaine mepivacain, prilocaine and ropivacain in human serum by liquid chromatography–tandem mass spectrometry, J. Chromatogr. A 1088 (2005) 126–130.
[64] H. Gréen, K. Vretenbrant, B. Norlander, C. Peterson, Measurement of paclitaxel and its metabolites in human plasma using liquid chromatography/ion trap mass spectrometry with a sonic spray ionization interface, Rapid Commun. Mass Spectrom. 20 (2006) 2183–2189.
[65] K.A. Mortier, V. Renard, A.G. Verstraete, A. Van Gussem, S. Van Belle, W.E. Lambert, Development and validation of a liquid chromatography–tandem mass spectrometry assay for the quantification of docetaxel and paclitaxel in human plasma and oral fluid, Anal. Chem. 77 (2005) 4677–4683.
[66] D.N. Heller, J.O. Peggins, C.B. Nochetto, M.L. Smith, O.A. Chiesa, K. Moulton, LC/MS/MS measurement of gentamicin in bovine plasma urine, milk, and biopsy samples taken from kidneys of standing animals, J. Chromatogr. B 821 (2005) 22–30.
[67] A. El-Beqqali, A. Kussak, M. Abdel-Rehim, Determination of dopamine and serotonin in human urine samples utilizing microextraction online with liquid chromatography/electrospray tandem mass spectrometry, J. Sep. Sci. 30 (2007) 421–424.
[68] V.P. Shah, K.K. Midha, J.W.A. 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, Bioanalytical method validation—a revisit with a decade of progress, Pharm. Res. 17 (2000) 1551–1557.
[69] N.A. Épshtein, Validation of HPLC techniques for pharmaceutical analysis, Pharm. Chem. J. 38 (2004) 212–228.
[70] IUPAC. Compendium of Chemical Terminology, 2nd ed. (the Gold Book). Compiled by A.D. McNaught, A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

(2006-) created by M. Nic, J. Jirat, B. Kosata, updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook Last update: 2014-02-24; version: 2.3.3.
[71] R. Rebane, M.-L. Oldekop, K. Herodes, Comparison of amino acid derivatization reagents for LC–ESI–MS analysis. Introducing a novel phoshpazene-based derivatization reagent, J. Chromatogr. B 904 (2012) 99–106.
[72] Z.J. Lin, S.-X. Qiu, A. Wufuer, L. Shum, Simultaneous determination of glycyrrhizin a marker component in radix glycyrrhizae, and its major metabolite glycyrrhetic acid in human plasma by LC–MS/MS, J. Chromatogr. B 814 (2005) 201–207.
[73] F.T. Peters, O.H. Drummer, F. Musshoff, Validation of new methods, Forensic Sci. Int. 165 (2007) 216–224.
[74] E. Rozet, R.D. Marini, E. Ziemons, B. Boulanger, P. Hubert, Advances in validation, risk and uncertainty assessment of bioanalytical methods, J. Pharm. Biomed. 55 (2011) 848–858.
[75] G.A. Shabir, Validation of high-performance liquid chromatography methods for pharmaceutical analysis: understanding the differences and similarities between validation requirements of the US Food and Drug Administration, the US Pharmacopeia and the International Conference on Harmonization, J. Chromatogr. A 987 (2003) 57–66.
[76] L. Tang, P. Kebarle, Effect of conductivity of the electrosprayed solution on the electrospray current. Factors Determining analyte sensitivity in ES mass spectrometry, Anal. Chem. 63 (1991) 2709–2715.
[77] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), JCGM 200:2012.
[78] Nordtest Report TR 537, Handbook for calculation of measurement uncertainty in environmental laboratories, 2nd ed., Nordtest, Espoo, 2004.
[79] Estimation of measurement uncertainty in chemical analysis, on-line course. https://sisu.ut.ee/measurement/uncertainty, University of Tartu, 2014.
[80] I. Taverniers, M. De Loose, E. Van Bockstaele, Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance, Trends Anal. Chem. 23 (2004) 535–552.
[81] F.T. Peters, Method validation using LC–MS, in: A. Polettini (Ed.), Applications of Liquid Chromatography–Mass Spectrometry in Toxicology, Pharmaceutical Press, London, 2006, pp. 71–96.
[82] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, Analysis of recent pharmaceutical regulatory documents on analytical method validation, J. Chromatogr. A 1158 (2007) 111–125.
[83] R. Bonfilio, E.C.L. Cazedey, M.B. de Araújo, H.R.N. Salgado, Analytical validation of quantitative high-performance liquid chromatographic methods in pharmaceutical analysis: a practical approach, Crit. Rev. Anal. Chem. 42 (2012) 87–100.
[84] D. Stöckl, H. D'Hondt, L.M. Thienpont, Methods validation across the disciplines – critical investigation of major validation criteria and associated experimental protocols, J. Chromatogr. B 877 (2009) 2180–2190.
[85] H.G.F. Smienk, D. Calvo, P. Razquin, E. Domínguez, L. Mata, Single laboratory validation of a ready-to-use phosphatase inhibition assay for detection of okadaic acid toxins, Toxins 4 (2012) 339–351.
[86] S.C. Utture, K. Banerjee, S. Dasgupta, S.H. Patil, M.R. Jadhav, S.S. Wagh, S.S. Kolekar, M.A. Anuse, P.G. Adsule, Dissipation and distribution behavior of azoxystrobin carbendazim, and difenoconazole in pomegranate fruits, J. Agric. Food Chem. 59 (2011) 7866–7873.
[87] M. Thompson, Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing, Analyst 125 (2000) 385–386.
[88] ISO 5725 series, Accuracy (trueness and precision) of measurement methods and results.
[89] B. Cordero, F. Canale, D. Del Rio, C. Bicchi, Identification quantitation, and method validation for flavan-3-ols in fermented ready-to-drink teas from the Italian market using HPLC–UV/DAD and LC–MS/MS, J. Sep. Sci. 32 (2009) 3643–3651.
[90] B. Magnusson, S.L.R. Ellison, Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements, Anal. Bioanal. Chem. 390 (2008) 201–213.
[91] O. Roth, O. Spreux-Varoquaux, S. Bouchet, P. Rousselot, S. Castaigne, S. Rigaudeau, V. Raggueneau, P. Therond, P. Devillier, M. Molimard, B. Maneglier, Imatinib assay by HPLC with photodiode-array UV detection in plasma from patients with chronic myeloid leukemia: comparison with LC–MS/MS, Clin. Chim. Acta 411 (2010) 140–146.
[92] S. Bouchet, E. Chauzit, D. Ducint, N. Castaing, M. Canal-Raffin, N. Moore, K. Titier, M. Molimard, Simultaneous determination of nine tyrosine kinase inhibitors by 96-well solid-phase extraction and ultra performance LC/MS–MS, Clin. Chim. Acta 412 (2011) 1060–1067.
[93] J.F. Emory, L.A. Seserko, M.A. Marzinke, Development and bioanalytical validation of a liquid chromatographic–tandem mass spectrometric (LC–MS/MS) method for the quantification of the CCR5 antagonist maraviroc in human plasma, Clin. Chim. Acta 431 (2014) 198–205.
[94] T.P.J. Linsinger, Use of recovery and bias information in analytical chemistry and estimation of its uncertainty contribution, TrAC, Trends Anal. Chem. 27 (10) (2008) 916–923.
[95] NordVal Protocol No. 2, Guide in validation of alternative propietary chemical methods, 2010.
[96] A.J. Giron, K. Deventer, P. Van Eenoo, Development and validation of an open screening method for diuretics, stimulants and selected compounds in human urine by UHPLC–HRMS for doping control, Anal. Chim. Acta 721 (2012) 137–146.
[97] D.N. Heller, Ruggedness testing of quantitative atmospheric pressure ionization mass spectrometry methods: the effect of co-injected matrix on matrix effects, Rapid Commun. Mass Spectrom. 21 (2007) 644–652.
[98] R. Bossi, K.V. Vejrup, B.B. Mogensen, W.A.H. Asman, Analysis of polar pesticides in rainwater in Denmark by liquid chromatography–tandem mass spectrometry, J. Chromatogr. A 957 (2002) 27–36.
[99] W. Naidong, P.R. Ring, C. Midtlien, X. Jiang, Development and validation of a sensitive and robust LC–tandem MS method for the analysis of warfarin enantiomers in human plasma, J. Pharm. Biomed. Anal. 25 (2001) 219–226.
[100] P.R. Kootstra, C.J.P.F. Kuijpers, K.L. Wubs, D. van Doorn, S.S. Sterk, L.A. van Ginkel, R.W. Stephany, The analysis of beta-agonists in bovine muscle using molecular imprinted polymers with ion trap LCMS screening, Anal. Chim. Acta 529 (2005) 75–81.
[101] R. Bonfiglio, R.C. King, T.V. Olah, K. Merkle, The effects of sample preparation methods on the variability of the electrospray ionization response for model drug compounds, Rapid Commun. Mass Spectrom. 13 (1999) 1175–1185.
[102] W. Jian, R.W. Edom, Y. Xu, J. Gallagher, N. Weng, Potential bias and mitigations when using stable isotope labeled parent drug as internal standard for LC–MS/MS quantitation of metabolites, J. Chromatogr. B 878 (2010) 3267–3276.
[103] M. Klötsel, B. Gutsche, U. Lauber, H.-U. Humpf, Determination of 12 type A and B trichothecenes in cereals by liquid chromatography-electrospray ionization tandem mass spectrometry, J. Agric. Food Chem. 53 (2005) 8904–8910.
[104] W.M.A. Niessen, P. Manini, R. Andreoli, Matrix effects in quantitative pesticide analysis using liquid chromatography–mass spectrometry, Mass Spectrom. Rev. 25 (2006) 881–899.
[105] A. Berlioz-Barbier, A. Vauchez, L. Wiest, R. Baudot, E. Vulliet, C. Cren-Olive, Multi-residue analysis of emerging pollutants in sediment using QuEChERS-based extraction followed by LC–MS/MS analysis, Anal. Bioanal. Chem. 406 (2014) 1259–1266.
[106] S.J. Lehotay, K. Mastovska, A.R. Lightfield, R.A. Gates, Comparison of QuEChERS sample preparation methods for the analysis of pesticide residues in fruits and vegetables, J. AOAC Int. 93 (2010) 355–367.
[107] M.C. Hurtado-Sanchez, R. Romero-Gonzalez, M.I. Rodriguez-Caceres, I. Duran-Meras, A.G. Frenich, Rapid and sensitive on-line solid phase extraction-ultra high performance liquid chromatography–electrospray–tandem mass spectrometry analysis of pesticides in surface water, J. Chromatogr. A 1305 (2013) 193–202.
[108] M. Rokka, K. Peltonen, Simultaneous determination of four coccidiostats in eggs and broiler meat: validation of an LC–MS/MS method, Food Addit. Contam. 23 (2006) 470–478.
[109] J. Ascari, S. Dracz, F.A. Santos, J.A. Lima, M.H.G. Diniz, E.A. Vargas, Validation of an LC–MS/MS method for malachite green (MG) leucomalachite green (LMG), crystal violet (CV) and leucocrystal violet (LCV) residues in fish and shrimp, Food Addit. Contam. 29 (2012) 602–608.
[110] C. Anagnostopoulos, G.E. Miliadis, Development and validation of an easy multiresidue method fort he determination of multiclass pesticide residues using GC–MS/MS and LC–MS/MS in olive oil and olives, Talanta 112 (2013) 1–10.
[111] C. Soler, B. Hamilton, A. Furey, K.J. James, J. Manes, Y. Pico, Optimization of LC–MS/MS using triple quadrupole mass analyzer for the simultaneous analysis of carbosulfan and its main metabolites in oranges, Anal. Chim. Acta 571 (2006) 1–11.
[112] R.K. Trivedi, R.R. Kallem, R. Mullangi, N.R. Srinivas, Simultaneous determination of rosuvastatin and fenofibric acid in human plasma by LC–MS/MS with electrospray ionization: assay development, validation and application to a clinical study, J. Pharm. Biomed. 39 (2005) 661–669.
[113] S. Esposito, K. Deventer, G. T'sjoen, A. Vantilborgh, P. Van Eenoo, Doping control analysis of desmopressin in human urine by LC–ESI–MS/MS after urine delipidation, Biomed. Chromatogr. 27 (2013) 240–245.
[114] B. Arora, A. Narayanasamy, J. Nirmal, N. Halder, S. Patnaik, A.K. Ravi, T. Velpandian, Development and validation of a LC–MS/MS method for homocysteinethiolactone in plasma and evaluation of its stability in plasma samples, J. Chromatogr. B 944 (2014) 49–54.
[115] W.Z. Shou, H.-Z. Bu, T. Addison, X. Jiang, W. Naidong, Development and validation of a liquid chromatography/tandem mass spectrometry (LC/MS/MS) method for the determination of ribavirin in human plasma and serum, J. Pharm. Biomed. 29 (2002) 83–94.
[116] M.J. Rumpler, Quantitative analysis of 11-nor-9-carboxy-tetrahydrocannbinol(THC–COOH) in urine by LC–MS/MS following a simple filtration, J. Chromatogr. B 957 (2014) 77–83.
[117] E.N. Ediage, J.D. Di Mavungu, S. Monbaliu, C. Van Peteghem, S. De Saeger, A validated multianalyte LC–MS/MS method for quantification of 25 mycotoxins in cassava flour, peanut cake and maize samples, J. Agric. Food Chem. 9 (2011) 5173–5180.
[118] F.D. Madureira, F.A. da Silva Oliveira, W.R. de Souza, A.P. Pontelo, M.L.G. de Oliveira, G. Silva, A multi-residue method for the determination of 90 pesticides in matrices with a high water content by LC–MS/MS without clean-up, Food Addit. Contam. B 29 (4) (2012) 665–678.
[119] F. Beaudry, S.A. Guenette, A. Winterborn, J.-F. Marier, P. Vachon, Development of a rapid and sensitive LC–ESI/MS/MS assay for the quantification of propofol using a simple off-line dansyl chloride derivatization reaction to enhance signal intensity, J. Pharm. Biomed. 39 (2005) 411–417.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016

[120] C. Kmellár, P. Fodor, L. Pareja, C. Ferrer, M.A. Martínez-Uroz, A. Valverde, A.R. Fernandez-Alba, Validation and uncertainty study of a comprehensive list of 160 pesticide residues in multi-class vegetables by liquid chromatography–tandem mass spectrometry, J. Chromatogr. A 1215 (2008) 37–50.
[121] E. Dubreil-Cheneau, Y. Pirotais, E. Verdon, D. Hurtaud-Pessel, Confirmation of 13 sulfonamides in honey by liquidchromatography–tandem mass spectrometry for monitoring plans: validation according to European Union Decision 2002/657/EC, J. Chromatogr. A 1339 (2014) 128–136.
[122] R.V.S. Nirogi, V.N. Kandikere, M. Shukla, K. Mudigonda, S. Maurya, R. Boosi, A. Yerramilli, Development and validation of a sensitive liquid chromatography/electrospray tandem mass spectrometry assay for the quantification of olanzapine in human plasma, J. Pharm. Biomed. 41 (2006) 935–942.
[123] I.R. Pizzutti, A. de Kok, M. Hiemstra, C. Wickert, O.D. Prestes, Method validation and comparison of acetonitrile and acetone extraction for the analysis of 169 pesticides in soya grain by liquid chromatography–tandem mass spectrometry, J. Chromatogr. A 1216 (2009) 4539–4552.
[124] A. Kruve, K. Herodes, I. Leito, Electrospray ionization matrix effect as an uncertainty source in HPLC/ESI-MS pesticide residue analysis, Rapid Commun. Mass Spectrom. 93 (1) (2010) 306–314.
[125] A. Kloepfer, J.B. Quintana, T. Reemtsma, Operational options to reduce matrix effects in liquid chromatography–electrospray ionization-mass spectrometry analysis of aqueous environmental samples, J. Chromatogr. A 1067 (2005) 153–160.
[126] A. Menditto, M. Patriarca, B. Magnusson, Understanding the meaning of accuracy, trueness and precision, Accredit. Qual. Assur. 12 (2007) 45–47.
[127] ISO13528: Statistical methods for use in proficiency testing by interlaboratory comparisons, ISO, 2005.
[128] A.G. González, M.Á. Herrador, A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles, Trends Anal. Chem. 26 (2007) 227–238.
[129] F. Lecomte, C. Hubert, S. Demarche, C. De Bleye, A. Dispas, M. Jost, F. Frankenne, A. Ceccato, E. Rozet, P.H. Hubert, Comparison of the quantitative performances and measurement uncertainty estimates obtained during method validation versus routine applications of a novel hydrophilic interaction chromatography method for the determination of cidofovir in human plasma, J. Pharm. Biomed. 57 (2012) 153–165.
[130] N. Venisse, C. Grignon, B. Brunet, S. Thévenot, A. Bacle, V. Migeot, A. Dupuis, Reliable quantification of bisphenol A and its chlorinated derivatives in human urine using UPLC–MS/MS method, Talanta 125 (2014) 284–292.
[131] T.M. Houts, Useful estimates of assay performance from small data sets, Clin. Chem. 50 (2004) 1958–1959.
[132] 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, Workshop/conference report—quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays, AAPS J. 9 (2007) E30–E42.
[133] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J. Smolec, B. Lee, M. Khan, R. Tacey, H. Hill, A. Celniker, Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules, Pharm. Res. 20 (2003) 1885–1900.
[134] D. Hoffman, R. Kringle, J. Singer, S. McDougall, Statistical methods for assessing long-term analyte stability in biological matrices, J. Chromatogr. B 877 (2009) 2262–2269.
[135] K. Kipper, K. Anier, I. Leito, J. Karjagin, K. Oselin, K. Herodes, Rapid determination of meropenem in biological fluids by LC: comparison of various methods for sample preparation and investigation of meropenem stability, Chromatographia 70 (2009) 1423–1427.
[136] C.C.G.O. Lopes, H.R.N. Salgado, Development of a validated stability-indicating LC assay and stress degradation studies of linezolid in tablets, Chromatographia 69 (2009) S129–S135.
[137] S. Thomas, S. Shandilya, A. Bharati, S.K. Paul, A. Agarwal, C.S. Mathela, Identification, characterization and quantification of new impurities by LC–ESI/MS/MS and LC–UV methods in rivastigmine tartrate active pharmaceutical ingredient, J. Pharm. Biomed. Anal. 57 (2012) 39–51.
[138] N. Salgueiro-Gonzalez, E. Concha-Grana, I. Turnes-Carou, S. Muniategui-Lorenzo, P. Lopez-Mahia, D. Prada-Rodriquez, Determination of alkylphenols and bisphenol A in seawater samples by dispersive liquid–liquid microextraction and liquid chromatography tandem mass spectrometry for compliance with environmental quality standards (Directive 2008/105/EC), J. Chromatogr. A 1223 (2012) 1–8.
[139] M. De Boevre, J.D. Di Mavungu, P. Maene, K. Audenaert, D. Deforce, G. Haesaert, M. Eeckhout, A. Callebaut, F. Berthiller, C. van Peterghem, S. De Saeger, Development and validation of an LC–MS/MS method for the simultaneous determination of deoxynivalenol, zearalenone, T-2-toxin and some masked metabolites in different cereals and cereal-derived food, Food Addit. Contam. A 29 (5) (2012) 819–835.
[140] Measurement Uncertainty Revisited. Eurolab Technical Report No 1/2007. Eurolab, 2007. Available on-line from http://www.eurolab.org/documents/1-2007.pdf.
[141] ISO/IEC 17,025:2005, General requirements for the competence of testing and calibration laboratories.
[142] JCGM 100:2008 Evaluation of measurement data—Guide to the expression of uncertainty in measurement. JCGM, 2008. Available on-line from http://www.bipm.org/en/publications/guides/gum.html.
[143] N. Majcen, P. Taylor (Eds.), Practical Examples on Traceability, Measurement Uncertainty and Validation in Chemistry, 1, EC-JRC IRMM, Geel, 2007.
[144] D. Dadgar, P.E. Burnett, M.G. Choc, K. Gallicano, J.W. Hooper, Application issues in bioanalytical method validation, sample analysis and data reporting, J. Pharm. Biomed. 13 (2) (1995) 89–97.

Please cite this article in press as: A. Kruve, et al., Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Anal. Chim. Acta (2015), http://dx.doi.org/10.1016/j.aca.2015.02.016