

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Linearity of Calibration Curves for Analytical Methods: A Review of Criteria for Assessment of Method Reliability

Seyed Mojtaba Moosavi and Sussan Ghassabian

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.72932>

Abstract

Calibration curve is a regression model used to predict the unknown concentrations of analytes of interest based on the response of the instrument to the known standards. Some statistical analyses are required to choose the best model fitting to the experimental data and also evaluate the linearity and homoscedasticity of the calibration curve. Using an internal standard corrects for the loss of analyte during sample preparation and analysis provided that it is selected appropriately. After the best regression model is selected, the analytical method needs to be validated using quality control (QC) samples prepared and stored in the same temperature as intended for the study samples. Most of the international guidelines require that the parameters, including linearity, specificity, selectivity, accuracy, precision, lower limit of quantification (LLOQ), matrix effect and stability, be assessed during validation. Despite the highly regulated area, some challenges still exist regarding the validation of some analytical methods including methods when no analyte-free matrix is available.

Keywords: analytical method, calibration, linearity, regression analysis, validation

1. Introduction

Calibration curve in bioanalytical method is a linear relationship between concentration (independent variable) and response (dependent variable) using a least squares method. This relationship is built to predict the unknown concentrations of the analyte in a complicated matrix. The unknown samples can be from a wide range of sources: food and agricultural, pharmaceutical formulations, forensic and the clinical pharmacology studies. This chapter is more focused on the bioanalytical methods in which an analyte is measured in blood, plasma, urine or other biological matrices. However, the main concepts are applicable to the other analytical approaches.

The quality of a bioanalytical method is highly dependent on the linearity of the calibration curve [1]. A linear calibration curve is a positive indication of assay performance in a validated analytical range. Other characteristics of the calibration curve, including regression model, slope of the line, weighting and correlation coefficient, need to be carefully evaluated. In the following sections, each of those parameters is explained, and few practical examples have been used to further discuss the concepts.

After the calibration model is chosen, it is required to demonstrate that all future measurements will be close to the true values of the content of the analyte in the sample. This will be achieved during validation of the analytical method. There are international guidelines for the validation of the analytical methods, which need to be followed closely in order to have more consistent data throughout different laboratories and increase the chance of their acceptability by the regulatory authorities.

2. Aims

The aim of this chapter is to discuss different aspects of linearity and relevant assumption as a practical guide to develop a robust analytical method in order to predict true concentrations of the analytes in samples.

3. Calibration curve: definitions and characteristics

3.1. Regression analysis

Regression analysis is a deterministic model, which allows predicting of the values for a dependent variable (Y) when an independent variable (X) is known. The model determines the kind of relationship between X and Y. The experimental values rarely fit the mathematical model, and there are differences between the observed and the predicted values provided by the model, which are called residuals (**Figure 1**). The sum of squared residuals needs to be minimised to have the best estimate of the model parameters, and it can be done using the “method of least squares.” The simplest regression model is the linear one in which the relationship between X (known without error) and Y (known with error) is a straight line, $Y = a + bX$, where a is the y-intercept and b is the slope of the line [1].

The relationship between an instrument response and the known concentrations of an analyte (standards), which is used as the calibration curve can be explained by a similar regression model. To have a robust calibration line (or curve), a series of replicates of each standard (at least three replicates of 6–8 expected range of concentration values) are recommended. The assumption for this model is that the measurement error is the same and normally distributed for each sample. If this assumption is not applicable, an extended or weighted least squares analysis will be required. The assumption regarding the measurement error must be verified to validate the results found. The distribution properties of the residuals are expected to be normal and centred on zero (Kolmogorov–Smirnov test). If the results found cannot support

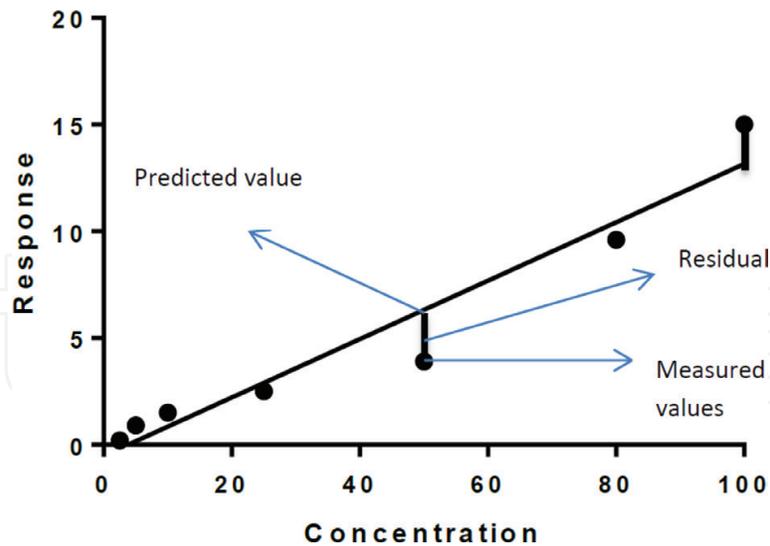


Figure 1. Linear regression model in which the differences between observed and predicted values (residuals) have been shown.

this assumption, the estimated parameters using the model cannot be used, and the model needs to be modified, e.g. using a non-linear model which requires more standard concentrations compared with having a linear relationship between concentrations and instrument response. A linear regression model between calculated standard points and the nominal ones used to evaluate the quality of the fit should have a unit slope and a zero intercept. In case of linear calibration method, the slope should be statistically different from 0, the intercept should not be statistically different from 0 and the regression coefficient should not be statistically different from 1. In case of having a significant non-zero intercept, the accuracy of the method must be demonstrated [2].

A standard 0 must be included in the calibration curve because the instrumental signal is subjected to the same kind of error for all points. The signal for the standard zero should not be subtracted from the response values for other standards before calculating the equation of the regression line because it can cause imprecision during the determination of the concentration values for unknown samples [3].

If one of the standard points deviates greatly from the calibration curve (outlier), it can be removed from the equation provided that six non-zero standards remain after removing the outlier and inclusion of that point can cause the loss of sensitivity or it clearly biases the quality control (QC) results, and the back-calculated standard concentrations deviate from its nominal value. The poor chromatography can also be considered as a justification for removing the outlier standard [4].

In order to verify the accuracy and precision of the analytical method during the period of sample analysis, quality control (QC) samples are prepared and stored frozen at the same temperature as is intended for the storage of the study samples. The calibration curve standards are prepared by spiking the reference standard solutions to the matrix (e.g. plasma or urine) either freshly or by freezing and storage with QC samples [4].

3.2. Weighting in linear regression

When the range in x-values is large, e.g. more than one order of magnitude, the variance of each data point might be quite different. However, the simple least squares method considers that all the y-values have equal variances. Larger deviations at larger concentrations tend to influence the regression line more than smaller deviations associated with smaller concentrations (heteroscedasticity) leading to the inaccuracy in the lower end of the calibration range (see the practical example 1). A simple and effective way to counteract this situation is to use weighted least squares linear regression (WLSLR) [1]. WLSLR is able to reduce the lower limit of quantification (LLOQ) and enables a broader linear calibration range with higher accuracy and precision especially for bioanalytical methods.

Two most commonly used regression models, particularly for liquid chromatography tandem mass spectrometry (LC-MS/MS) calibration curves, are linear and quadratic regression models using non-weighted or weighted least squares regression algorithm. To select the type of calibration curve and weighting, "Test and Fit" strategy is widely used due to its simplicity and lack of statistical analysis and causes inaccuracy in the regression model based on the limited set of test results. The Food and Drug Administration (FDA) guideline suggests that "the simplest model that adequately describes the concentration-response relationship should be used and selection of weighting and use of a complex regression equation should be justified" [5]. However, other experts suggested that a weighting should be used if homoscedasticity was not met for the analytical data. By neglecting the weighting for analysing data with heteroscedastic distribution, a precision loss as big as one order of magnitude in the low concentration region of the calibration curve could happen [4].

For most immunoassay methods, the response is a non-linear function of the analyte concentration, and the standard deviations (SD) of the calculated concentrations are not a constant function of the mean response; therefore, a weighted, non-linear least squares method is generally recommended for fitting dose-response data. The nonevidence-based weights (e.g. $1/Y$ or $1/X$) are not recommended without assessment of the response-error relationship. A reference model for immunoassay data employs the four-parameter logistic (4PL) equation to fit the concentration-response relationship and a power-of-the-mean (POM) equation to fit the response-error relationship [6].

3.3. Correlation coefficient

Linearity of the calibration curve is usually expressed through the coefficient of correlation, r , or coefficient of determination, r^2 . A correlation coefficient close to unity ($r = 1$) is considered by some authors' sufficient evidence to conclude that the calibration curve is linear. However, r is not an appropriate measure for the linearity. The FDA guidance for validation of analytical procedures [5] recommends that the r should be submitted when evaluating a linear relationship and that the linearity should be evaluated by appropriate statistical methods, e.g. analysis of variance (ANOVA). This guidance does not suggest that the numerical value of r can be used as a degree of deviation from linearity.

Other mathematical measures, including slope standard relative deviation or goodness of fit, can be used to evaluate the linearity [3]. Using residual plots is a simple way to check the

linearity. The residuals are expected to be normally distributed for a linear model, so a plot of them on a normal probability graph may be useful. Any curvature suggests a lack of fit (LOF) due to a non-linear effect. A segmented pattern indicates heteroscedasticity in data, so weighted regression model should be used to find the straight line for calibration [7].

A clear curved relationship between concentration and response may also have an r value close to one. Two statistical tests, including the lack-of-fit and Mandel's fitting tests, are suitable for the validation of the linear calibration model (practical example 2 [8]). A straight-line model with r close to 1, but with a lack of fit, can produce significantly less accurate results than its curvilinear alternative. A straight-line calibration curve should always be preferred over curvilinear or non-linear calibration models if equivalent results can be obtained and is easier to implement [8].

3.4. Slope of the curve and application in matrix effect and detection limit

Slope of the calibration curve can be used to estimate the detection limit of the assay [9]. Three times the standard deviation value of the response corresponding to the blank according to Eq. (1), obtained for seven determinations, divided by the slope of the calibration line (note that we are calculating the standard deviation of the concentration corresponding to the blank equation, and again the imprecision of the value of the slope is not taken into account) [3]:

$$\text{LOD} = 3.3 \times (S_y/a). \quad (1)$$

S_y denotes the SD of responses, Y , for blanks or around expected LOD (limit of detection) and "a" for the slope of a linear calibration line. If the calibration curve is linear, "a" is constant, and the estimation of LOD is easy to calculate. However, when the calibration curve is not linear, e.g. in enzyme-linked immunosorbent assay (ELISA), the definition needs to be modified. In the case of ELISA, when there is a semilogarithmic calibration curve over a wide range of concentrations, the detection limit is calculated using a differential coefficient which is obtained using a computer programme [9].

It is assumed that a validated analytical method should have constant slope over the period of sample analysis. Variation in the slope might be due to the laboratory errors during sample preparations, change in the internal standard (IS) of working solution concentrations between preparations, instrument variations such as changes in mass spectrum (MS) calibrations, MS signal cross contributions between analyte and IS and matrix effect (ME) [10]. Although there is no criteria in the international guidelines to report the slope, monitoring the slope can provide valuable information regarding the quality of the sample analysis.

ME can also affect the slope of the calibration curve. Coeluting of the matrix components escaped during extraction may reduce the signal intensity and affect the accuracy and precision of the MS-based assays. The phenomenon is called ion suppression, and it has been shown that the electrospray ionisation responses of organic bases decrease with an increase in concentrations of other organic bases present in the matrix. The ME is especially dependent on the degree of sample clean-up and chromatographic separation of the analyte. When developing high-throughput assays using a short run time, a careful assessment of the ME and ion suppression is necessary [11].

3.5. Internal standard (IS)

IS is a chemical substance that is added in equal amounts to all samples, and it changes the way that calibration curve is prepared. Instead of analyte response, the ratio of the analyte to the IS signal versus the analyte concentration is plotted. The benefits of adding the IS are to correct or compensate analyte losses during sample preparation including transfer loss, adsorption loss, evaporation loss and variation in injection volume and in MS response due to ion suppression or enhancement (ME).

The IS must have similar physicochemical properties and show similar behaviour to the analyte when extracted or run through the analytical column or detection in the analytical system. An external standard also behaves similarly with the analyte, but it is run alone at different concentrations, so a standard curve can be generated. External standards do not correct for losses that may occur during preparation of the sample. Using IS is usually more effective due to lower measurement uncertainty and therefore is more common in analytical chemistry [12].

Two common types of ISs are used: structural analogues and stable isotope-labelled (SIL) ISs or isotope dilution mass spectrometry (IDMS). SIL ISs are more effective. To reduce the interferences between IS and analyte, SIL IS molecular weight is preferred to be ideally 4 or 5 Da higher than that of the analyte. Labelled SIL ISs with ^{13}C and/or ^{15}N are usually superior to those labelled with deuterium (^2H , D or d) in terms of performance; however, the synthesis of deuterated ISs is easier and cheaper. The location of stable isotope atoms should be in a way that deuterium-hydrogen exchange is minimised during sample preparation.

A structural analogue of the analyte can be used if SIL ISs are not available or expensive. In this case, the IS should preferably have key structure and functionalities (e.g. $-\text{COOH}$, $-\text{SO}_2$, NH_2 , halogen and heteroatoms) of the analyte with difference only being C-H moieties (length and/or position). Modifications in key chemical structure and/or functionalities cause significant differences in ionisation pattern and even extraction recovery. The IS should not be similar or converted to any *in vivo* biotransformed products of the analyte (e.g. hydroxylated or N-dealkylation metabolites). An appropriate structural analogue IS can be selected from the same therapeutic class as the analyte or by key chemical structure and preferably a compound that is not very commonly prescribed because those compounds may be present in pooled blank plasmas used for preparation of the calibrators and QCs. Other parameters for choosing a right structure analogue IS are physicochemical properties, such as log D (hydrophobicity), pKa and water solubility. For selection of the IS, it may be difficult to have a compound to track the analyte of interest in all the three distinctive stages of LC-MS bioanalysis, sample preparation (extraction), chromatographic separation and mass spectrometric detection. The IS should be chosen depending on which step is more critical. For example, when the extracts of samples contain coeluting matrix components that cause ion suppression, then tracking the analyte during MS detection to avoid or minimise ME becomes more important. The choice of IS is also depending on the extraction method. Tracking an analyte during a simple protein precipitation procedure would be less stringent than that for liquid-liquid extraction (LLE) or solid-phase extraction (SPE) method [13].

It is possible to develop an assay without using any IS, for example, in early drug discovery stage or when clean extracts are used. In this case, ECHO peak technique can be used where

the analyte is used as its own IS. In this method, after the injection of the sample containing the analyte of interest, a standard solution is also injected, which result in two peaks for the analyte, one from the sample and the other from the standard solution with constant concentration (an echo peak). By using their response ratio for quantitation, the ME might be compensated for because the two peaks are affected by the coeluted matrix components similarly [13].

There is no general rule for choosing the IS concentrations. However, the accuracy and precision of the method may be affected if an inappropriate IS concentration is used. As shown in practical example 3, reducing the concentration of IS can lead to the increasingly non-linear calibration curve due to chemical impurity in the reference standard or because of isotope interferences.

When choosing the IS and its concentration, the magnitude of the cross signal contribution between the analyte and IS should be considered. The IS interference signal due to its impurity or isotope interferences should be equal or less than 20% of the LLOQ response and 5% of the IS response for IS-to-analyte and analyte-to-IS contributions, respectively [14]. The minimum IS concentration required (CIS-Min) and the maximum IS concentration allowed (CIS-Max) can be calculated using Eqs. (2 and 3):

$$\text{CIS-Min} = m \times \text{ULOQ}/5. \quad (2)$$

$$\text{CIS-Max} = 20 \times \text{LLOQ}/n. \quad (3)$$

where m and n represent the % of cross signal contributions from analyte to IS and IS to analyte, respectively. As an example, if the cross signal contribution from analyte to IS is 2.5%, the minimum IS concentration calculated accordingly is 50% of the ULOQ. A high IS concentration might be useful in reducing a systemic error in the analysis of unknown samples. If the IS coelutes more closely to the analyte, it will be more effective in minimising ME.

In some cases, the analyte signal might be suppressed by the coeluting IS signal, and therefore the IS concentration must be kept low to maintain a low detection limit. However, it might be required to increase the IS concentration when the analyte suppresses the IS signal.

IS should be added as early as possible to compensate for the variabilities during sample preparation and analysis; however, if the IS structure is not very close to the analyte, it can be used to reduce the variabilities due to the ion suppression or enhancement only and not sample extraction [13].

3.6. Linearity when no analyte-free matrix exists

For making calibrators and QCs, an analyte-free matrix is required. The presence of unknown amount of the analyte in the matrix makes the quantification difficult, and different approaches have been used to overcome the problem including using stripped matrices (filtration on activated charcoal-dextran or dialysis), substitute matrices (e.g. neat solutions, artificial matrices, human serum albumin or 0.9% sodium chloride) or diluted matrices. If the actual matrix is used, various methods are followed including, background subtraction, or the standard addition method [3, 15].

One of the approaches for validation of the assay is to determine the accuracy throughout the validation step, using the biological matrix containing the endogenous compound to prepare the standard curves and all pools of six or more assays of each QC sample [3]. The amount of the analyte in the matrix (C_{basal}) can be computed using a calibration curve in the substitute matrix, and the concentration of the analyte in the QC can be calculated by subtracting the C_{basal} from the calculated one as follows, $C_{\text{real}} = C_{\text{found}} - C_{\text{basal}}$, in which C_{found} is the concentration of the analyte in the QCs calculated against a calibration curve in the substitute matrix and C_{real} is the corrected concentration [3]. When using this approach, the LLOQ of the method cannot be smaller than the endogenous concentrations of the analyte in the matrix, and therefore a lot of blank matrices need to be screened to find the suitable one.

Alternatively, the endogenous concentration of the analyte in the matrix can be subtracted from the added concentrations and uses the subtracted concentrations to build the calibration curve. Using the actual biological matrix for making the calibrators and QCs reduces the recovery and matrix effects between samples and calibrators. Again, the limitation of this method is that the increase in background peak area after spiking with standards has to be at least 15–20% of the background peak area, and the LLOQ is limited by the endogenous background concentration even if much lower concentrations can be detected by the method. Another difficulty is when multiple analytes with different endogenous compounds need to be quantified [15].

Alternatively, the background concentration in the blank matrices can be lowered by dilution of the blank matrices before spiking with standards. However, by diluting the matrix, the composition of the matrices in the study samples versus calibration curve is different leading to different recoveries of the analytes. Therefore, the extraction recoveries of analytes between the matrix and diluted matrix should be determined before using this method [15].

3.6.1. Surrogate matrices

Surrogate matrices can vary widely from a simplest form, mobile-phase solvents (neat) or pure water to a synthetic polymer-based solution. Some biological matrices, e.g. cerebrospinal fluid or tears, are difficult to obtain. The surrogate matrix should simulate the authentic matrix in terms of composition, salt content, analyte solubility, recovery and ME. For example, phosphate-buffered saline (PBS) or bovine serum albumin (BSA) in PBS (20–80 g/L) has the similar protein and ionic strength as human plasma.

To use neat solutions as surrogate matrices, extraction recovery and ME are required to be comparable with the original matrix. For example, thromboxane B2 and 12(S)-hydroxyeicosatetraenoic acid were quantified in human serum using mixture of water/methanol/acetonitrile (80:10:10, v/v/v) as a surrogate matrix, and the ME and recoveries of the analytes were demonstrated to be comparable.

3.6.2. Stripped matrices

Biological matrices can be stripped from particular endogenous components to generate analyte-free surrogate matrices. Adding activated charcoal, for example, can adsorb and remove the analyte from the matrix, but the charcoal must effectively remove from the matrix before

spiking the analyte. Some analytes, e.g. homocysteine, cannot be removed by the charcoal and also the composition of the matrix may change or cause batch-to-batch variation after adding the charcoal leading to the altered analyte recovery and ME. Some light-sensitive analytes can be decomposed by heat or exposing to the light and therefore removed from the matrix.

3.6.3. Method of standard addition

In the standard addition method, every study sample is divided into aliquots of equal volumes, and the aliquots are spiked with known and varying amounts of the analyte to build the calibration curve. The sample concentration is then calculated as the negative x-intercept of the calibration line. This method is very accurate because it allows direct quantitation of endogenous analytes without manual subtraction of background peak areas. The disadvantage of the method is that it requires a large amount of sample and is very time-consuming and labour intensive. Examples of using this method when the analyte-free matrices are not available include measuring abscisic acid, a phytohormone from plant leaves and the emission of polycyclic aromatic hydrocarbons from petroleum refineries. Standard addition can also be used when some matrix components produce MS signals that interfere with the analytes of interest.

We have used this method by some modifications to measure homocysteine and pyridoxal 5-phosphate in samples of human serum and whole blood, respectively [16, 17]. The matrix was first spiked with different concentrations of the analytes, and the endogenous concentrations of the analytes were estimated using the negative x-intercept of the calibration line. Then, the endogenous concentrations were added to the spiked concentrations, and new calibration curves with real concentrations were constructed (practical example 4). QCs were prepared in both actual and surrogate matrices, and the sample volume reduced to only 20 μL to minimise the matrix effect.

3.7. Validation

All the developed analytical methods need to be validated to make sure that each measurement of the content of the analyte in the sample in routine analysis is close to the true values [7]. There are international guidelines for validation of the analytical methods including FDA [6], European Medicines Agency (EMA) [14], International Union of the Pure and Applied Chemistry (IUPAC) [18] and Association of Official Analytical Chemists (AOAC) International. The major parameters need to be validated including linearity, accuracy, precision, specificity, selectivity, sensitivity, ME and stability testing.

3.7.1. Selectivity and specificity

Selectivity is the ability of a method to determine a particular analyte in a complex matrix without interference from other ingredients of the matrix. Specificity, however, is the ultimate in selectivity, and it means that no interference is expected to occur, but these two terms are used interchangeably in the literature. If a method has specificity for an analyte, it means that either you have it or you do not. Selectivity can be graded as low, high, partial, good or bad, but the selectivity refers to 100% selectivity (or 0% interference) [19].

Selectivity can be calculated by comparing the chromatograms obtained after injection of a blank sample with and without the analyte or analytical solutions and with and without the matrix components.

3.7.2. Accuracy

Accuracy (or trueness or bias) is the most important aspect of validation and should be addressed in any analytical method. Accuracy shows the extent of agreement between the experimental value (calculated from replicate measurements) and the nominal (reference) values. Accuracy is a measurement of the systematic errors affecting the method. To estimate the accuracy of a method, the analyte is measured in comparison with a reference material or by spiking known amount of analyte in the blank matrix (QC samples) and calculating the percentage of recovery from the matrix. It can also be estimated using the comparison of the results from the method by a reference method [19].

The guideline for validation of analytical methods by the EMA [14] recommends checking the accuracy within run and between runs by analysing a minimum of five samples per four QC levels (LLOQ, low, medium and high) as a representative of the whole analytical range in at least two different days. The accuracy needs to be reported as the percentage of the nominal concentrations and the mean concentration should be within 15% of the nominal values for all QC levels, except LLOQ, which should be within 20% of the nominal values [14].

3.7.3. Precision

The term precision is defined as the closeness of repeated individual measurements of an analyte under specified conditions. This term is demonstrating the repeatability and reproducibility of the method and expressed as the coefficient of variation (CV). Precision should be measured for LLOQ, low, medium and high QC samples in the same run that accuracy is testing. The acceptance criteria are also similar to the accuracy evaluation [14, 19].

3.7.4. Uncertainty

To make sure that a method is correctly fit for the purpose of measurement, “uncertainty” of the method is required to be evaluated [7].

A detailed list of all possible sources of uncertainty needs to be prepared. A preliminary study may identify the most significant sources of uncertainty. Typically, the two sources of uncertainties are Type A or random error and Type B or systematic error. Random error is caused by unpredictable variations and gives rise to variations in repeated observations. The random error can generally be minimised by increasing the number of observations. Systematic error, however, is a type of errors, which remain constant, or its variation is predictable and therefore independent of the number of observations. The result should be corrected for all recognised significant systematic errors. The steps involved in uncertainty estimation are identification of uncertainty sources, quantification of uncertainty components and calculation of combined and expanded uncertainty. The main sources of uncertainty are sampling, environmental conditions, method validation, instruments, weighting and dilutions, reference materials,

chemicals and in high-performance liquid chromatography (HPLC) are repeatability of peak area, dilutions factors, reference materials and sampling. Sampling, calibration and repeatability were the most significant sources, which affect combined uncertainty [20].

3.7.5. LOD and LLOQ

The LOD is generally defined as the lowest amount of an analyte in a sample that can be detected by a particular analytical method. LOD is usually evaluated using the calculation of the signal/noise relationship considering the assumption that data normality, homoscedasticity and independency of residuals are met. The signal-to-noise ratio is determined by comparing the analytical signals at known low concentrations compared with those of blank sample up to a concentration that produces a signal equivalent to three times the standard deviation of the blank sample [19]. Determination of the LOD is not necessary during the validation, because the assay may have high variability in that level.

On the other hand, the lowest concentration of an analyte in a sample, which can be reliably quantified is defined as the LLOQ. The analyte signal at the LLOQ level should be at least five times the signal of blank sample and the accuracy and precision within 20% of the nominal concentrations. The LLOQ should be selected based on the expected concentrations in the study. For example, for bioequivalence studies the LLOQ should not be higher than 5% of the maximum concentration of the analyte in the samples (C_{\max}) [14].

3.7.6. Matrix effect (ME)

ME measurement is necessary for validation when the analytical method uses mass spectrometry as the detector due to the ion suppression or induction caused by the matrix components. The ME evaluation required spiking the analyte (at low and high concentrations) in six lots of matrix obtained from individual donors. First, the ratio of the peak area in the presence of matrix to the peak area in the absence of the matrix is calculated to achieve the matrix factor (MF), followed by the calculation of the IS normalised MF by dividing the MF of the analyte of interests by the MF of the IS. The CV of the IS-normalised MF is calculated from the six lots of the matrix and should be $\leq 15\%$ (practical example 5). In some cases that this method is not practical (e.g. online sample preparation), the variability of the response should be assessed by analysing at least six lots of matrix spiked at low and high levels. The overall CV should not be greater than 15%. The ME is also recommended to be tested in haemolysed, hyperlipidaemic matrices or plasma collected from renally or hepatically impaired patients depending on the target population of the study [14].

3.7.7. Stability

Stability testing must be planned based on the conditions applied to the samples during processing. The stability is tested using spiked concentrations of the analyte to the matrix at low and high QC levels (six replicates at two levels are generally sufficient). Short-term stability at room temperature (2–8 h depending on the latest period of time required for sample processing), long-term stability at storage temperature (e.g. at -20°C or -80°C), freeze and

thaw and stock solution stabilities are the most common tests. The stability of QC samples are analysed against a freshly prepared calibration curve, and the calculated concentrations should be within 15% of the nominal concentrations. The stability of processed samples in the autosampler temperature also determines how long samples can be stored in the autosampler without the analyte being degraded [14]. Any other variation during sample processing which can potentially affect the stability of the analyte of interest needs to be tested during validation.

4. Practical examples

4.1. Practical example 1: impact of weighting

See **Table 1**.

4.2. Practical example 2: linearity assessment

In **Table 2**, it shows that the linear regression model (LRM) must systemically be rejected at the 95% confidence level ($F_{\text{crit},95\%} = 4.53$) for lack-of-fit test and at 99% confidence level ($F_{\text{crit},99\%} = 10.56$) for Mandel's fitting test. Thus, despite the fact that r and quality coefficient (QC) are greater than 0.997 and lower than 5%, respectively, the linearity of the calibration lines was rejected based on the F-tests. So, the r is not a good measure of the linearity assessment. Even with a QC value less than 3%, the LRM is rejected at the 95% confidence level (**Table 2**). Alternatively, the residual plots give useful information to validate the chosen regression model.

The residual plot can be used to check if the principle assumptions, i.e. normality of the residuals and homoscedasticity, are met when evaluating the goodness of fit of the regression model. The U-shaped residual plot usually shows that a curvilinear regression model is a better fit than an LRM. In order to correct the non-linearity, a quadratic curvilinear function ($f(x) = a + bx + cx^2$) can be chosen. The "lack of fit" tests for the quadratic regression model

Concentration (ng/mL)	Area ratio	Accuracy (no weighting)	Accuracy (1/x weighting)
0	0.002	0	0
6	0.006	125	92.3
18	0.019	104	94.1
37.5	0.400	98.5	94.1
75	0.836	101	99.3
300	3.320	98.5	98.9
480	5.290	97.8	98.5
600	6.890	102	103

Table 1. Increasing the accuracy of the lower end of the calibration curve by applying the weighing.

Linear regression model				Quadratic regression model	
LOF	Mandel's test value	QC (%)	r	LOF	P-value on second-order coefficient
11.08	51.46	3.93	0.9982	0.63	0.0000
19.42	56.84	4.23	0.9978	1.58	0.0000
7.13	26.29	3.67	0.9985	0.94	0.0006
6.99	37.73	3.79	0.9984	0.18	0.0002
11.43	58.21	4.03	0.9981	0.31	0.0000
29.91	53.02	3.53	0.9986	4.08	0.0000
49.80	71.07	3.76	0.9984	5.69	0.0000
23.77	73.86	3.19	0.9989	1.66	0.0000
31.95	63.37	3.24	0.9988	3.55	0.0000
7.49	33.50	2.92	0.9991	0.54	0.0003
9.99	55.19	3.95	0.9983	0.15	0.0000
10.71	28.65	4.70	0.9975	1.89	0.0005
25.21	79.60	3.34	0.9987	1.62	0.0000
13.16	35.74	3.37	0.9987	1.93	0.0002

For the quadratic regression model, the F-value of the lack-of-fit test and the P-value for testing significance of the second-order coefficient for the quadratic regression model are represented. The significance value at the 95% confidence level is underlined (reproduced from Van Loco et al. [8] with permission from Springer-Verlag).

Table 2. The F-value of the lack-of-fit (LOF) test and Mandel's fitting test is compared with the quality coefficient for several linear calibration lines of Cd.

(QRM) are summarised in **Table 2**. The test for lack of fit indicates that this QRM fits the calibration data at 99% confidence level in all cases except one. To check the suitability of the order of polynomial regression model, the significance of the second-order coefficient needs to be estimated. The P-value on the second-order coefficient, shown in **Table 2**, is systemically smaller than 1%, and therefore a lower order model should not be considered. Moreover, residual plots (**Figure 2**) were constructed for the QRM, and the residuals were randomly scattered within a horizontal band around the centre line. Therefore, the QRM was selected as the reference model. It is noted that an increase of the variance is observed at higher concentrations [8].

As a summary, in this example, a linear model with $r > 0.997$ and $QC < 5\%$ but with lack of fit (LOF) yielded predicted values for a mid-scale calibration standard that significantly differ from the nominal ones. The accuracy was overestimated, while the precision on the results was comparable in both LRM and QRM [8].

4.3. Practical example 3: IS concentration and the linearity

The role of IS concentration on the linearity of the calibration curve has been demonstrated by Tan et al. [13]. They presented a case in which decreasing concentration of the IS from 100%

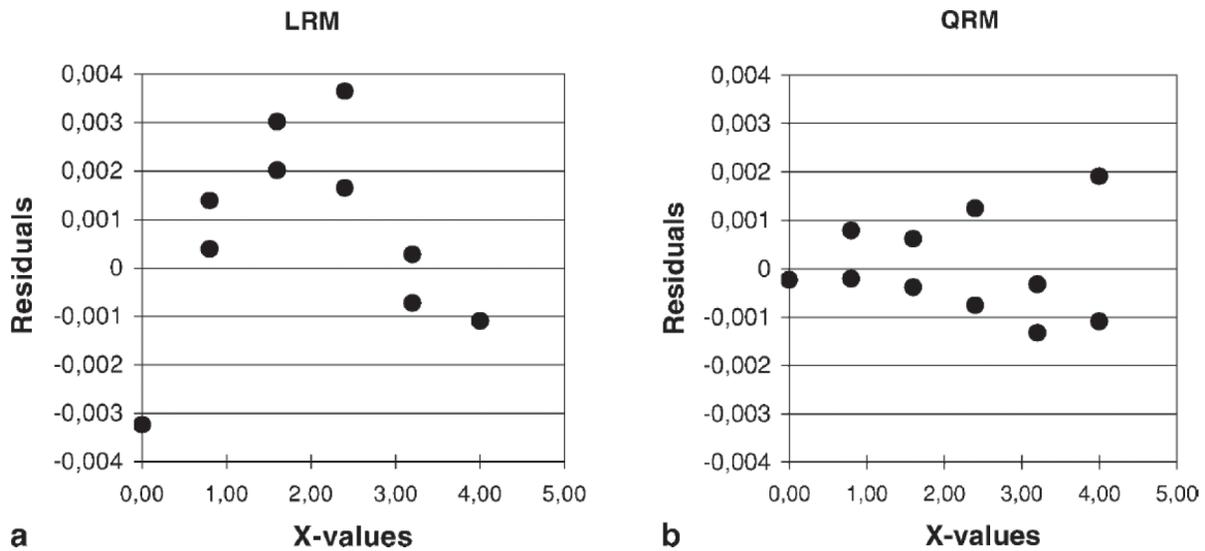


Figure 2. Plots of residuals for (a) the linear regression model (LRM) and (b) the quadratic regression model (QRM) versus predicted values (adopted from Van Loco et al. [8] with permission from Springer-Verlag).

to 5% ULOQ made the calibration curve non-linear. In that case, the cross-contribution from the analyte to the IS is equivalent to 5% of the concentration of the analyte. The cross-signal contribution from the analyte to the IS is either due to the isotope interference or chemical impurity in reference standard [13].

4.4. Practical example 4: method of standard addition for homocysteine calibration curve

Table 3 shows the calculated calibration curve data for homocysteine standard solutions spiked into a pooled human serum.

To estimate the endogenous concentrations of homocysteine in the sample of pooled human serum, the negative x-intercept of the curve is calculated:

$$x = \frac{-0.262}{0.000598} = 438 \text{ ng/mL.} \quad (4)$$

Spiked concentrations (ng/mL)	Calculated concentration (ng/mL)
0	N/A
50	30.87
600	632.76
1100	1107.05
1600	1652.48
2100	2167.02
2600	2584.02
3100	3031.75

Regression equation: $y = 0.000598x + 0.262$.

Table 3. Homocysteine calibration curve: the x-axis is representing the spiked concentrations.

Then, the nominated concentration is changed to the spiked + endogenous concentrations, and a new calibration is constructed (**Table 4**).

Now, by comparing the detector response for the unknown samples with the second calibration curve, the unknown sample concentrations can be calculated.

4.5. Practical example 5: ME calculations

Table 5 is representing the analyte peak area spiked in six different lots of human plasma. The MF has been calculated by dividing the area of analyte (or IS) in each matrix to the average peak area of the analyte (or IS) in the pure solutions. The IS-normalised MF is the ratio of the MF for the

Spiked + endogenous concentration (ng/mL)	Calculated concentration (ng/mL)
0 + 438 = 438	381.51
50 + 438 = 488	469.28
600 + 438 = 1038	1071.34
1100 + 438 = 1538	1545.77
1600 + 438 = 2038	2091.35
2100 + 438 = 2538	2606.98
2600 + 438 = 3038	3023.15
3100 + 438 = 3538	3471.01

Regression equation: $y = 0.000598x - 0.000762$.

Table 4. Homocysteine calibration curve: The x-axis is representing the spiked + endogenous concentrations.

Analyte of interest		Matrix factor	IS		Matrix factor	IS-normalised MF
Spiked	Peak area Pure (mean of three replicates)		Spiked	Peak area Pure (mean of three replicates)		
1,095,000	1,210,000	0.905	4,320,000	6,343,333	0.681	1.33
1,050,000		0.868	6,240,000		0.984	0.882
1,110,000		0.917	5,780,000		0.911	1.01
1,120,000		0.926	5,660,000		0.892	1.04
1,100,000		0.909	5,770,000		0.910	0.999
1,130,000		0.934	5,170,000		0.815	1.15
Mean						1.07
SD						0.154
CV%						14.4

Table 5. Calculation of ME.

analyte to the MF for the IS. The CV% in this example was 14.4%, which is within the acceptance limit for the matrix effect by the EMA guideline for validation of bioanalytical methods [14].

5. Key results

- Calibration curve is a regression model between a known concentration of an analyte and the response from an instrument enabling the estimation of the concentration of the analyte in an unknown sample.
- Weighted least squares linear regression (WLSLR) is necessary when the standard deviations across the standard range are not consistent. Weighting improves the sensitivity and accuracy of the lower end of the calibration range.
- Coefficient of correlation is not a suitable measure for the linearity of the calibration curve, and the linearity should be evaluated using an appropriate statistical analysis.
- Stable isotope-labelled compounds are the most preferable internal standards. However, carefully chosen structural analogues with similar functional groups and physicochemical properties can contribute to generation of comparable analytical methods.
- The concentration of the internal standard may affect the linearity of the calibration curve due to the cross signal contribution between the analyte and the internal standards.
- When an analyte-free matrix does not exist, the amount of endogenous analyte in the matrix can be estimated using the negative x-intercept of the regression equation and adding this value to the spiked concentrations of the analyte to calculate the actual concentrations of each standard.
- During validation of an analytical method, selectivity, specificity, accuracy, precision, uncertainty, LLOQ, matrix effect and stability are the minimum criteria to be evaluated.

Abbreviations

4PL	Four-parameter logistic
ANOVA	Analysis of variance
BSA	Bovine serum albumin
CIS-Max	Maximum IS concentration
CIS-Min	Minimum IS concentration
CV	Coefficient of variation
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
FDA	Food and Drug Administration

HPLC	High-performance liquid chromatography
IDMS	Isotope dilution mass spectrometry
IS	Internal standard
IUPAC	International Union of Pure and Applied Chemistry
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LLE	Liquid-liquid extraction
LLOQ	Lower limit of quantification
LOD	Limit of detection
LOF	Lack of fit
LRM	Linear regression model
ME	Matrix effect
MF	Matrix factor
MS	Mass spectrometry
PBS	Phosphate buffer saline
POM	Power of the mean
QC	Quality control
QRM	Quadratic regression model
SD	Standard deviation
SD	Standard deviation
SIL	Stable isotope labelled
SPE	Solid-phase extraction
ULOQ	Upper limit of quantification
WLSLR	Weighted least squares linear regression

Author details

Seyed Mojtaba Moosavi¹ and Sussan Ghassabian^{2*}

*Address all correspondence to: s.ghassabian@uq.edu.au

¹ The University of Technology Sydney, Sydney, Australia

² The University of Queensland, Brisbane, Australia

References

- [1] Almeida AM, Castel-Branco MM, Falcao AC. Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*. 2002;**774**(2):215-222
- [2] Bressolle F, Bromet-Petit M, Audran M. Validation of liquid chromatographic and gas chromatographic methods. Applications to pharmacokinetics. *Journal of Chromatography. B, Biomedical Applications*. 1996;**686**(1):3-10
- [3] Rodríguez LC et al. Estimation of performance characteristics of an analytical method using the data set of the calibration experiment. *Analytical Letters*. 1993;**26**(6):16
- [4] Gu H et al. Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay performance. *Analytical Chemistry*. 2014;**86**(18):8959-8966
- [5] FDA, editor. *Analytical Procedures and Methods Validation U.S.D.o.H.a.H. Services*. Rockville, Maryland: Food and Drug Administration (FDA); 2000
- [6] Findlay JW et al. Validation of immunoassays for bioanalysis: A pharmaceutical industry perspective. *Journal of Pharmaceutical and Biomedical Analysis*. 2000;**21**(6):1249-1273
- [7] Gonzalez AG, Herrador MA. A practical guide to analytical method validation including measurement uncertainty and accuracy profiles. *Trends in Analytical Chemistry*. 2007;**26**(3):11
- [8] Van Loco J, Elskens M, Croux C, Beernaert H. Linearity of calibration curves: Use and misuse of the correlation coefficient. *Accreditation and Quality Assurance*. 2002;**7**
- [9] Hayashi Y et al. Detection limit estimated from slope of calibration curve: An application to competitive ELISA. *Analytical Sciences*. 2005;**21**(2):167-169
- [10] Zhao Y et al. Reasons for calibration standard curve slope variation in LC-MS assays and how to address it. *Bioanalysis*. 2014;**6**(11):1439-1443
- [11] Matuszewski BK, Constanzer ML, Chavez-Eng CM. Matrix effect in quantitative LC/MS/MS analyses of biological fluids: A method for determination of finasteride in human plasma at picogram per milliliter concentrations. *Analytical Chemistry*. 1998;**70**(5):882-889
- [12] Oliveira EC. Internal standard versus external standard calibration: An uncertainty Case study of a liquid chromatography analysis. *Quimica Nova*. 2010;**33**(4):19
- [13] Tan A, Awaiye K. Use of internal standards in LC-MS bioanalysis. In: Li W, Zhang J, Tse FLS, editors. *Handbook of LC-MS Bioanalysis: Best Practices, Experimental Protocols, and Regulations*. John Wiley & Sons Inc.; 2013. pp. 217-227
- [14] EMEA, editor. *E.M.A., Guideline on Bioanalytical Method Validation*. London, England: E.M.A. (EMA); 2011. p. 23

- [15] Thakare R et al. Quantitative analysis of endogenous compounds. *Journal of Pharmaceutical and Biomedical Analysis*. 2016;**128**:426-437
- [16] Ghassabian S, Griffiths L, Smith MT. A novel fully validated LC-MS/MS method for quantification of pyridoxal-5'-phosphate concentrations in samples of human whole blood. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*. 2015;**1000**:77-83
- [17] Ghassabian S et al. Fully validated LC-MS/MS method for quantification of homocysteine concentrations in samples of human serum: A new approach. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*. 2014;**972**:14-21
- [18] Danzer K, Currie LA, Guidelines for Calibration In Analytical Chemistry. *Pure and Applied Chemistry*. 1998;**70**(4)
- [19] Araujo P. Key aspects of analytical method validation and linearity evaluation. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*. 2009;**877**(23):2224-2234
- [20] Traple MAL et al. Measurement uncertainty in pharmaceutical analysis and its application. *Journal of Pharmaceutical Analysis*. 2014;**4**(1):5

