JOURNAL OF SEPARATION SCIENCE

24|2021
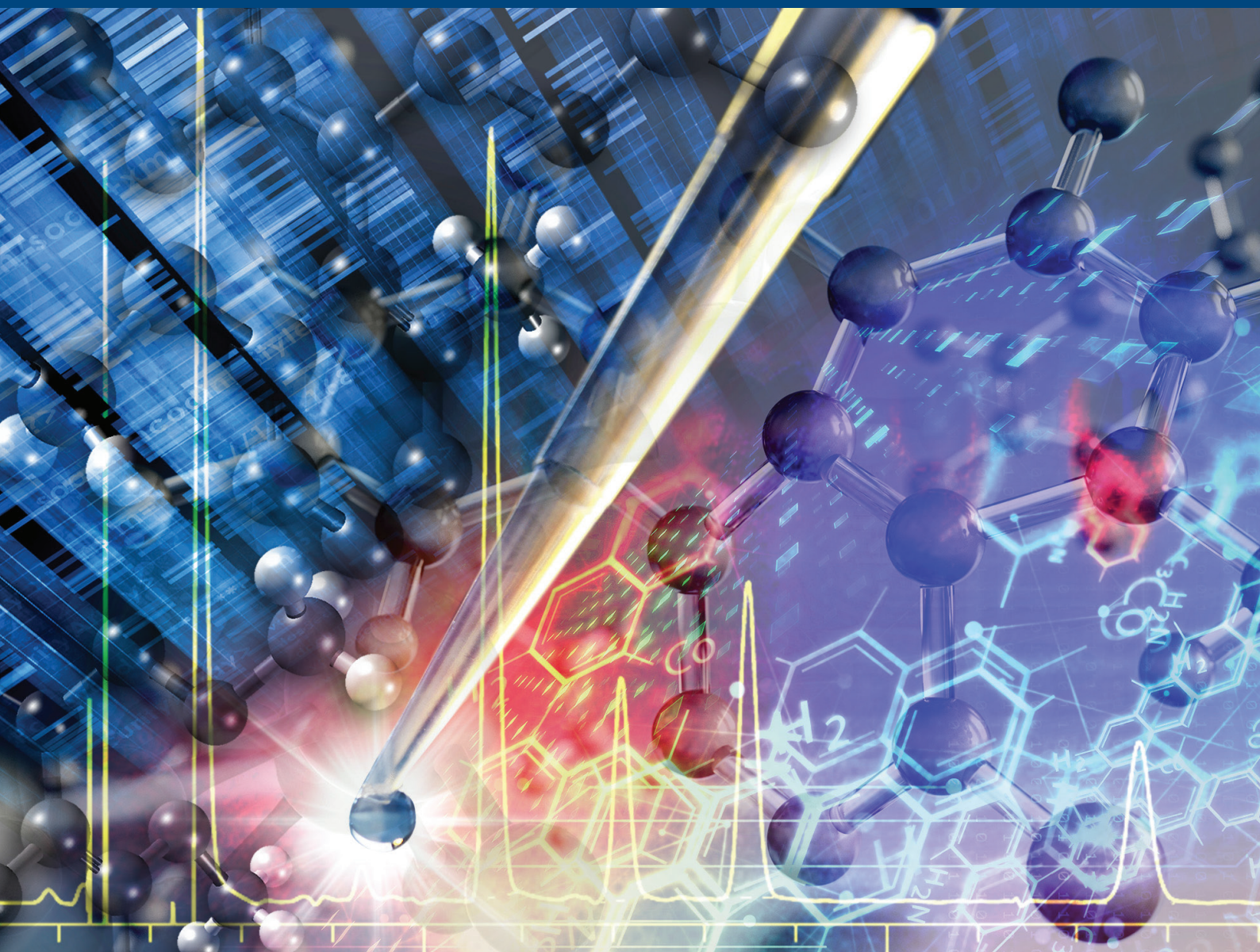
RESEARCH ARTICLE

JOURNAL OF
SEPARATION SCIENCE

The inadequate use of the determination coefficient in analytical calibrations: How other parameters can assess the goodness-of-fit more adequately

Juan M. Sanchez

Department of Chemistry, Faculty of Sciences, University of Girona, Girona, Spain

Correspondence
Juan M. Sanchez, Department of Chemistry, Faculty of Sciences, University of Girona, Aurèlia Capmany, 69, 17003-Girona, Spain.
Email: juanma.sanchez@udg.edu

Simple linear regression using ordinary least-squares is the most common function applied in laboratories for analytical calibrations. The determination and/or the correlation coefficients are usually the parameters applied for assessing the goodness-of-fit of a simple linear calibration. However, these parameters are unable to detect the highly biased results at low calibration levels that are obtained with ordinary least-squares. In this study, the use of other parameters based on the relative standard errors of the calculated contents is evaluated. It has been found that these alternative parameters can detect the biased results obtained at low calibration levels with ordinary least-squares, being the relative standard error the one that seems to provide the most adequate results. Ordinary least-squares should only be applied if the lower limit of quantification is set to at least five times above the conventional limit of quantification. For trace analysis, where the lowest possible limit of quantification is required, weighted least-squares should be applied to obtain accurate estimates, especially at low concentrations. One of the greatest advantages of the relative standard error is that this parameter can be determined for all types of regression functions and is not limited to calibrations with linear relationships between the variables.

KEYWORDS
analytical calibration, goodness-of-fit, lower limit of quantification, regression, relative standard error

1 | INTRODUCTION

It has been argued that calibration is a key and critical property of any analytical method [1]. The selection of the calibration model is one of the most important decisions to be made when performing quantitative analysis and the choice of an inappropriate regression function may easily lead to the determination of biased results, which will result in incorrect decisions being taken. For this reason, it is required to determine a parameter that can assess adequately the goodness-of-fit (GOF) of a calibration function.

The most common analytical methods used in laboratories are modeled using equations based on linear relationships between the independent and the dependent

---

Article Related Abbreviations: GOF, goodness-of-fit; LLOQ, lower limit of quantification; OLS, ordinary least-squares; QC, quality coefficient; RA, residual accuracy; RE, relative error; RSE, relative standard error; SSE, sum-of-squares error; SSR, sum-of-squares regression; SSTO, sum-of-squares total; WLS, weighted least-squares

variables, being ordinary least-squares (OLS) and weighted least-squares (WLS) the most used functions. Overwhelmingly, the GOF of these models is determined by the correlation coefficient (R) or the determination coefficient (R^2), which is the squared value of R only for linear relationships:

$$R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSTO}} = 1 - \frac{\sum_{i=1}^{n} \left(y_{i,exp} - y_{i,pred}\right)^2}{\sum_{i=1}^{n} \left(y_{i,exp} - y_{avg}\right)^2} \quad (1)$$

where SSR is the sum-of-squares regression, SSE is the sum-of-squares error, SSTO is the sum-of-squares total (SSTO = SSR+SSE), $y_{i,exp}$ is the experimentally measured value for the dependent variable, $y_{i,pred}$ is the predicted values from the linear model, and y_{avg} is the average experimental values. However, the disadvantages of using R and R^2 for assessing the GOF of linear analytical calibrations are well-known [2–7].

The most used linear regression function, OLS, is based on the minimization of the SSE term (also known as residuals) because this model was developed with the assumption that absolute errors of the dependent variable (measured as SD or variance, SD^2) are constant all along the range studied (homoscedasticity). However, in analytical and bioanalytical calibrations the most common situation is that absolute errors are not constant (heteroscedasticity) and the parameter that remains approximately constant is the relative error (RE; RSD) [6,8–13]. In this situation, OLS regression overestimates the effect of calibrators at high concentration ranges, and the higher variations at this level have a much greater influence on R^2 than small deviations present at low ranges [14]. Logue and Manandhar [7] showed, using a modeled geometric series, with ratio ½, of seven calibrators covering two orders of magnitude with a constant RSD of 10%, that 96% of the SSTO value (Equation (1)) is based on the SSE results of the largest two concentrations, and 99% on the largest three concentrations. Therefore, the first four calibrators contribute <1% to the calculated R^2 value. Another limitation of OLS with heteroscedastic data is that the values at high concentrations are given too much weight and estimations at this point work like a lever, which can easily lead to considerable bias at low concentrations. It has been reported that when the data have a proportional error, neglect of weighting can increase the uncertainty by a factor ≥10 at the lower concentration level [15,16]. For these reasons, it has been reported that OLS should not be used when samples are expected to be determined close to the LOQs [1,6–9,12,13,15,17–19], as is very common in trace analysis.

When WLS is used, a weighting factor based on the inverse of the variance of the measurement is applied for each calibrator ($w_i = 1/SD_i^2$), which forces the regression line to track closer to the points with the lowest variance (higher weight). In this situation, the lower concentration standards become more important. In all the studies that have compared OLS with WLS regression with heteroscedastic calibrations, it has been found that significantly biased results with OLS regression were obtained at low calibration levels [1,12,13,17,20], which are solved applying WLS. However, at a certain level above the LOQ of the method, both linear regression models tend to yield equivalent results. This is due to the fact that OLS can handle a certain level of heteroscedasticity, which allows OLS regression results to be unaffected unless a severe deviation from homoscedasticity is present [21].

Despite the great advantages when using WLS, especially when the analytical results are measured at the lower end of the calibration range, the R^2 values obtained by WLS and OLS regressions are very similar [7]. Therefore, the R^2 is unable to differentiate between a poor linear fitting, OLS, and a good one, WLS, and cannot be considered as a parameter to assess the GOF of the experimental calibrations.

Another disadvantage of R^2 is that it is based on absolute errors of the dependent variable (signal measured, y_i). However, the most important function of a calibration curve is to accurately predict the concentration of unknown samples. For this reason, GOF parameters based on differences between the nominal and calculated concentrations are more desirable. Moreover, due to the common heteroscedasticity usually present in analytical calibrations, parameters based on REs are also preferred.

Some parameters have been proposed for assessing GOF of calibration curves taking into account these considerations, like the quality coefficient (QC) [22]:

$$QC\,(\%) = 100 \times \sqrt{\frac{\sum_{i=1}^{n} \left(\frac{x_{i,true} - x_{i,calc}}{x_{i,true}}\right)^2}{n-1}} \quad (2)$$

the residual accuracy (RA) [7]:

$$RA\,(\%) = 100 \times \frac{\sum_{i=1}^{n} \left(1 - \left|\frac{x_{i,true} - x_{i,calc}}{x_{i,true}}\right|\right)}{n} \quad (3)$$

and the relative standard error (RSE) [17]:

$$RSE\,(\%) = 100 \times \sqrt{\frac{\sum \left(\frac{x_{i,true} - x_{i,calc}}{x_{i,true}}\right)^2}{(n-p)}} \quad (4)$$

where, $x_{i,calc}$ is the experimentally measured concentration with the assessed regression, $x_{i,true}$ is the true value of the concentration level, n is the number of calibration points,

and p is the number of terms in the fitting equation (linear $p = 2$ and quadratic $p = 3$).

These three parameters are equivalent and use the RE in the predicted concentration values. The only difference appears in the denominator of the equations. It has to be noted that the only parameter that takes into account the degrees of freedom of the equation function evaluated ($n - p$) is the RSE. For this reason, this parameter has been proposed for any type of regression model, linear or non-linear. The use of the RSE has been recommended by the National Environmental Laboratories Accreditation Conference Institute [23] for the evaluation of calibrations in environmental analysis and recent revisions of the US-Environmental Protection Agency methods for waters also accept this methodology for assessing the acceptability of the calibration curves [1,24].

It is clear that the use of the determination and correlation coefficients should be avoided for assessing the GOF of calibration and a different parameter should be taken into account. As Burrows and Parr wrote [1], "the combination of using unweighted linear regression function with evaluating the correlation coefficient is particularly pernicious" because this allows calibrations yielding large bias at the low end of the calibration range to pass typical method criteria.

In the present study, WLS regression, applying different procedures for calculating the weighting factor, has been compared against OLS with different experimental calibrations with linear relationships between the variables. In all the models evaluated, the GOF, using the R^2 and the other parameters explained in this introduction are evaluated in order to assess the efficiency of the proposed alternatives as a substitute of the R^2. The evaluated parameters are also applied for the evaluation of calibrations with non-linear relationships (i.e., curvature).

This work tries to reinforce what should be the current practice for analytical laboratories and to assess whether there is a parameter other than the R^2 that can evaluate the GOF of calibrations with linear relationships and any other type of calibration equations, solving the limitations and deficits of the R^2.

2 | METHODS

All the calibration data in this study have been obtained from experimental calibrations. The detailed information about the standards prepared, the individual signals obtained for the independent replicates measured at each concentration level, linearity assessment, and results of the calibrations evaluated are shown in the Supporting Information. The homogeneity/heterogeneity of the variances in each calibration was evaluated using the Levene test [25,26].

A total of six calibration models were assessed for each calibration with linear relationships, one using OLS and five with WLS regression (one using the inverse of the experimental variance, $1/SD_i^2$, as the weight and the others applying $1/x_i$, $1/y_i$, $1/x_i^2$, and $1/y_i^2$). In the case of calibrations with non-linear relationships between the variables, the quadratic regression was compared against OLS and WLS ($w_i = 1/SD_i^2$).

The SD at blank level (SD_{bl}) was determined from the analysis of at least seven independent replicates of spiked blanks. LOQ values were obtained taking into account the determined precision at the blank level and calculated as $10 \times SD_{bl}$, as proposed by different validation guidelines [27,28].

IBM SPSS Statistics 27 was used for the statistical and regression calculations. A difference was considered as significant when $p < 0.05$.

3 | RESULTS AND DISCUSSION

The results obtained for all the evaluated calibrations show that the variability associated with the dependent variable ranges between 1 and 13% (Supporting Information), whereas the variability associated with the independent variables was always <0.1% (determined from error propagation calculations taking into account the variabilities of the stock solutions and material used in the preparation of the standards). This confirms that the independent variable always had a significantly lower error than the dependent variable, which allows classical calibration to be used (regression of y on x) rather than inverse calibration (regression of x on y).

3.1 | Assessment of weighting factors with calibrations with linear relationships

One of the main limitations argued by many researchers against the use of WLS with routine calibrations is the requirement for a large number of replicates to be performed with each standard in order to obtain the weighting factors since they are defined as the inverse of the variance for each calibrator, which requires the measurement of independent replicates at each level.

As mentioned in the Introduction, the SD is not constant with the concentration in many analytical and bioanalytical methods and the parameter that tends to be so is the RE (i.e., heteroscedasticity). In these conditions, when classical calibration is applied, absolute errors tend to increase proportionally to both the concentration of the analyte in the sample and the measured response. The results obtained in this study confirm this fact (Figure 1).
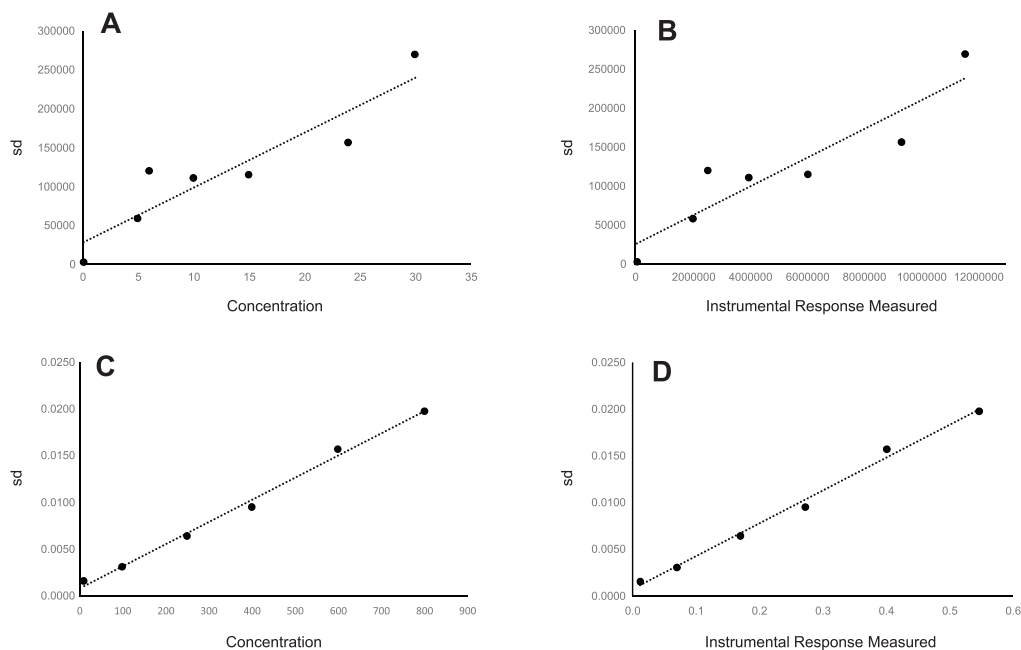
FIGURE 1 Increase in the experimental SD with the concentration (A,C) and with the instrumental response (B,D). (A) and (B) correspond to case study #1, whereas (C) and (D) correspond to case study #5. Seven independent replicates were measured at each calibration level

Further confirmation was obtained statistically by applying the Levene test to evaluate the homogeneity of the variances in each calibration (Table 1). All calibrations yield $p < 0.05$, confirming heteroscedasticity, except for case study #10, which only covers one order of magnitude.

Due to the proportional increase of the absolute errors to the two variables, different mathematical approaches have been proposed for a simpler and more practical way to determine the weighting factors without the need to analyze replicates, such as $1/x_i$, $1/y_i$, $1/x_i^2$, and $1/y_i^2$ [6,8,11,15,29,30]. The majority of these studies found that $1/x_i^2$ and $1/y_i^2$ usually give the best results. These approaches have been evaluated in the present study and the results obtained, despite not being novel, confirm that WLS is the most appropriate linear calibration for obtaining accurate estimates at low levels (see RSE values in Table 1). Moreover, the results obtained also confirm that the need to measure replicate standards is not an absolute requirement to perform WLS calculations and to obtain good estimates of the slope, the intercept, and their SDs. The five mathematical approaches evaluated for calculating the weighting factors have yielded equivalent results, and it was confirmed that applying $1/x_i^2$ and $1/y_i^2$ as weighting factors seems to give better estimates for the SD of the intercept.

It has been reported that WLS does not significantly alter the slope estimate obtained by OLS, whereas the intercept can be moderately affected [15,16,31,32]. In our case, these trends were confirmed. In the case of the slopes,

although in some calibrations statistical differences were found between the slopes calculated with the OLS and the WLS models (ANOVA test, case studies #1 $p < 0.001$, #2 $p = 0.006$, #3 $p = 0.004$, #4 $p = 0.007$, and #5 $p = 0.016$), the percentage of variation was always <5%, a value that from a practical point of view can be considered as acceptable and equivalent. In the case of the intercepts, no differences were found between the results obtained for all weighted regression models ($p > 0.15$). Interestingly, when the values obtained with OLS were assessed, two different behaviors were found. First, a significant difference between the intercepts determined with the unweighted model and the weighted models was usually found for those calibrations where the first standard was set at a value close to the LOQ, with 2–12 times higher values being obtained with the OLS model (#1, #2, #3, #4 $p < 0.001$, #5 $p = 0.027$, and #6 $p = 0.135$). Second, no significant differences were found when the first standard was set at a value well above the LOQ (#7 $p = 0.959$, #8 $p = 0.431$, #9 $p = 0.228$, and #10 $p = 0.132$). The OLS estimates were <2 times higher in these calibrations.

The SD values of the intercepts determined by OLS were always higher than those with the WLS functions. These differences were significantly larger in 70% of the calibrations, except in #7 $p = 0.152$, #8 $p = 0.120$, and #10 $p = 0.154$. In these three calibrations, the first standard was set at a level well above the LOQ of the method. F-tests were also applied to compare the values of the experimental variance determined at the blank level and the variance of the

TABLE 1 Relative standard error (RSE, %), quality coefficient (QC, %), residual accuracy (RA, %) and determination coefficient (R^2) calculated for the calibrations evaluated with linear relationship applying the six regressions evaluated

Study case	Calibration range	LOQ ($10 \times SD_{bl}$)	Model	RSE (%)	QC (%)	RA (%)	R^2
#1	0.1–30 mg/L	0.05 mg/L	OLS	147	134	52	0.9995
			WLS ($w_i = 1/SD_i^2$)	4	3	98	0.9995
			WLS ($w_i = 1/x_i^2$)	3	3	98	0.9991
Levene test	$p = 0.002$		WLS ($w_i = 1/y_i^2$)	3	3	98	0.9991
			WLS ($w_i = 1/x_i$)	5	5	97	0.9993
			WLS ($w_i = 1/y_i$)	6	6	96	0.9993
#2	0.05–20 mg/L	0.05 mg/L	OLS	96	89	59	0.9996
			WLS ($w_i = 1/SD_i^2$)	4	4	97	0.9992
			WLS ($w_i = 1/x_i^2$)	4	3	98	0.9995
Levene test	$p < 0.001$		WLS ($w_i = 1/y_i^2$)	4	3	98	0.9988
			WLS ($w_i = 1/x_i$)	5	5	97	0.9994
			WLS ($w_i = 1/y_i$)	6	1	97	0.9994
#3	1–50 mg/L	1 mg/L	OLS	47	42	83	0.9989
			WLS ($w_i = 1/SD_i^2$)	6	6	97	0.9982
			WLS ($w_i = 1/x_i^2$)	5	5	97	0.9975
Levene test	$p < 0.001$		WLS ($w_i = 1/y_i^2$)	6	5	97	0.9976
			WLS ($w_i = 1/x_i$)	8	7	95	0.9984
			WLS ($w_i = 1/y_i$)	8	7	95	0.9985
#4	0.4–21 ppbv	0.2 ppbv	OLS	52	46	77	0.9978
			WLS ($w_i = 1/SD_i^2$)	11	10	92	0.9944
			WLS ($w_i = 1/x_i^2$)	10	9	93	0.9896
Levene test	$p < 0.001$		WLS ($w_i = 1/y_i^2$)	10	9	93	0.9909
			WLS ($w_i = 1/x_i$)	15	13	91	0.9967
			WLS ($w_i = 1/y_i$)	13	12	91	0.9967
#5	10–800 mg/L	10 mg/L	OLS	19	17	93	0.9996
			WLS ($w_i = 1/SD_i^2$)	4	3	98	0.9997
			WLS ($w_i = 1/x_i^2$)	2	2	99	0.9997
Levene test	$p < 0.001$		WLS ($w_i = 1/y_i^2$)	2	2	98	0.9997
			WLS ($w_i = 1/x_i$)	3	3	98	0.9998
			WLS ($w_i = 1/y_i$)	4	2	98	0.9998
#6	10–800 mg/L	9 mg/L	OLS	34	31	86	0.9980
			WLS ($w_i = 1/SD_i^2$)	3	3	98	0.9993
			WLS ($w_i = 1/x_i^2$)	3	3	98	0.9993
Levene test	$p = 0.002$		WLS ($w_i = 1/y_i^2$)	3	3	98	0.9993
			WLS ($w_i = 1/x_i$)	3	3	98	0.9991
			WLS ($w_i = 1/y_i$)	4	3	97	0.9990
#7	25–300 mg/L	5 mg/L	OLS	3	2	98	0.9990
			WLS ($w_i = 1/SD_i^2$)	3	3	98	0.9990
			WLS ($w_i = 1/x_i^2$)	3	2	98	0.9989
Levene test	$p = 0.042$		WLS ($w_i = 1/y_i^2$)	3	2	98	0.9989
			WLS ($w_i = 1/x_i$)	3	3	98	0.9993
			WLS ($w_i = 1/y_i$)	3	3	98	0.9993
#8	50–500 μM	10 μM	OLS	4	3	97	0.9991
			WLS ($w_i = 1/SD_i^2$)	3	3	97	0.9980

(Continues)

TABLE 1 (Continued)

Study case	Calibration range	LOQ $(10 \times SD_{bl})$	Model	RSE (%)	QC (%)	RA (%)	R^2
			WLS ($w_i = 1/x_i^2$)	3	2	98	0.9988
Levene test	$p = 0.002$		WLS ($w_i = 1/y_i^2$)	3	2	98	0.9987
			WLS ($w_i = 1/x_i$)	3	3	98	0.9991
			WLS ($w_i = 1/y_i$)	3	3	98	0.9991
#9	50–500 μM	10 μM	OLS	3	3	98	0.9991
			WLS ($w_i = 1/SD_i^2$)	2	2	99	0.9997
			WLS ($w_i = 1/x_i^2$)	2	2	99	0.9995
Levene test	$p = 0.004$		WLS ($w_i = 1/y_i^2$)	2	2	99	0.9995
			WLS ($w_i = 1/x_i$)	2	2	99	0.9995
			WLS ($w_i = 1/y_i$)	2	2	99	0.9995
#10	11–92 mg/L	2 mg/L	OLS	6	5	96	0.9975
			WLS ($w_i = 1/SD_i^2$)	3	3	97	0.9984
			WLS ($w_i = 1/x_i^2$)	3	2	98	0.9990
Levene test	$p = 0.077$		WLS ($w_i = 1/y_i^2$)	3	2	98	0.9989
			WLS ($w_i = 1/x_i$)	3	3	98	0.9986
			WLS ($w_i = 1/y_i$)	3	3	98	0.9984

Abbreviation: OLS, ordinary least-squares; WLS, weighted least-squares.

ppbv = parts per billion by volume

intercept values for each calibration model evaluated. It was found that in 80% of the calibrations the SD of the intercept determined by OLS regression was significantly higher than the experimental variance of the blank. On the other hand, the variance of the intercept determined by WLS was equivalent to the blank variance in 80% of the cases. These results agree with those obtained by Vesterlund and Ramebäck [16] on analyzing certified reference materials at levels close to the LOQ and confirm that WLS regression is better at reproducing the variability at the low level of the calibration curves, near the LOQ.

3.2 | Assessment of the GOF with calibrations yielding linear relationships

In order to perform the evaluation of the linear calibrations, an approximate LOQ for each method was determined by a conventional validation procedure, based on $10 \times SD_{bl}$ [27,28]. Taking into account the results found in the previous section, the calibrations were divided into two groups. First, those calibrations where the first standard was set at a level equivalent to the calculated LOQ (case studies #1, #2, #3, #4, #5, and #6 in Table 1 and Supporting Information). Second, calibrations where the first standard was set clearly above the calculated LOQ (case studies #7, #8, #9, and #10).

As can be seen from the results obtained (Table 1), all linear calibrations evaluated were heteroscedastic (except

#10, $p = 0.077$), and gave appropriate values for R^2 (>0.99) for all the regression functions evaluated, also for OLS. However, when the other GOF parameters were taken into account, those calibrations where the first standard was set at a level equivalent to the LOQ only gave correct results with the WLS regressions (set in this study at <15% for RSE and QC, and >90% for RA). It was found in the calibrations evaluated that when the first standard was set at ≥5 × LOQ the results obtained with OLS and WLS regression did not yield any significant difference for any of the GOF parameters evaluated.

To confirm the lack-of-fit of the OLS regressions at low levels, the RE (%) for the back-calculated values for all standards was determined. In all cases, the WLS models always yielded RE < 10%, independently of the calibrator assessed, which were considered acceptable (US-FDA sets RE acceptance criteria at <15% of nominal concentration, except at LOQ level where should be <20% [33]). However, for those calibrations where the first standard was set at the LOQ level, the OLS model gave unsatisfactory results for the lowest level standards, with RE values ranging from 25 to 328%. When the first standard was set at a level ≥5 × LOQ RE values obtained with OLS regression were acceptable (RE < 10%).

In those cases where the OLS model yielded unsatisfactory results, calibrations were re-evaluated removing those standards that had concentrations <5 × LOQ (Table 2). It can be observed that once the lower standard is set at a level >5 × LOQ, the OLS function yielded satisfactory

TABLE 2 Goodness-of-fit (GOF) parameters (relative standard error [RSE], quality coefficient [QC], residual accuracy [RA]) calculated for the calibrations with linear relationship evaluated with the first standard (lower LOQ [LLOQ]) set at different levels in the case of ordinary least-squares (OLS) regression

Study case	LLOQ	LOQ ($10 \times SD_{bl}$)	Model	RSE (%)	QC (%)	RA (%)
#1	0.1 mg/L	0.05 mg/L	OLS	146.8	134.0	51.7
	5 mg/L		OLS	2.8	2.5	98.4
	0.1 mg/L		WLS ($w_i = 1/SD_i^2$)	3.6	3.3	97.8
#2	0.05 mg/L	0.05 mg/L	OLS	95.9	88.5	58.8
	0.1 mg/L		OLS	45.8	41.8	83.1
	1 mg/L		OLS	5.0	4.5	97.3
	0.05 mg/L		WLS ($w_i = 1/SD_i^2$)	3.7	3.8	97.4
#3	1 mg/L	1 mg/L	OLS	47.1	42.2	82.5
	10 mg/L		OLS	3.8	3.3	97.9
	1 mg/L		WLS ($w_i = 1/SD_i^2$)	6.2	5.5	96.5
#4	0.4 ppbv	0.2 ppbv	OLS	51.8	46.3	76.5
	0.8 ppbv		OLS	15.3	13.2	91.9
	1 ppbv		OLS	6.0	4.9	96.2
	0.4 ppbv		WLS ($w_i = 1/SD_i^2$)	11.4	10.2	92.3
#5	10 mg/L	10 mg/L	OLS	19.2	17.2	93.0
	100 mg/L		OLS	1.1	1.0	99.3
	10 mg/L		WLS ($w_i = 1/SD_i^2$)	3.6	3.2	97.9
#6	10 mg/L	9 mg/L	OLS	34.5	30.8	85.9
	100 mg/L		OLS	4.7	4.0	96.7
	10 mg/L		WLS ($w_i = 1/SD_i^2$)	3.0	2.7	97.9

Abbreviation: WLS, weighted least-squares

ppbv = parts per billion by volume

results for all the GOF parameters evaluated. In case studies #2 and #4, two standards had to be removed because the second standard was at $<5 \times$ LOQ, and at this level, the results measured with the OLS function were still unsatisfactory.

The results obtained clearly confirm that R and R^2 fail as GOF parameters. Using these coefficients, it is not possible to detect the highly significant bias resulting from OLS when analyzing compounds at low concentration levels. However, all the other parameters assessed were able to detect this problem. From the three proposed parameters, it seems that RSE gives better results. For example, in case study #5, the RA parameter (93%) was unable to detect the excessive bias for the first standard (38%) in this calibration. In case study #4, RA (92%) and QC (13%) also failed to detect the bias (25%) of the second standard used (set at 4 × LOQ). The only parameter that was able to detect all the excessive bias at low levels with OLS regression was the RSE, which might be associated with the fact that it is the only parameter that takes into account the degrees of freedom of the calibration model evaluated (n − p).

3.3 | Lower LOQ

Calibration results can be reported starting at the LOQ of the proposed method. However, the most common methodologies used in the experimental determination of LOQ values are only based on the precision at a single concentration, without taking into account the trueness at this level. The conventional concept of LOQ is based on a multiple of the experimental SD measured at blank level ($k \times SD_{bl}$, usually $k = 10$), which is simply an extension of the methodology applied for determining the LOD [28]. This procedure does not take into account the regression function applied and the variability introduced by this model. The International Union of Pure and Applied Chemistry has recommended modifying this concept [27]. In the new recommendations, the variability at the quantification level (σ_Q) should not be only based on an estimate of the SD at blank level, and the propagation of uncertainties in slope and intercept must be considered. Therefore, the regression model chosen has a significant effect on the correct determination of the LOQ for any analytical method, and its value can change for the same

analytical method depending on the regression function applied.

Some recent validation guidelines require the evaluation of trueness in the determination of the LOQ [33–35]. However, the proposed methodologies require the analysis of a large number of replicates at each calibration level, which is only possible during the validation of a method and is not practical in the daily routine of a laboratory. From a practical point of view, it is more appropriate to use the concept of lower LOQ (LLOQ), a criterion already required for calibrations in the environmental analysis [23]. It has been recommended that laboratories establish the LLOQ at concentrations where both quantitative and qualitative requirements can consistently be met [24].

The results obtained in this study show that probably the most important parameter to be taken into account to determine the linear calibration model that can be applied is the level set for the first standard (i.e., LLOQ), independently of whether the calibration is heteroscedastic or not. Once a correct LLOQ level is determined after the evaluation of the RSE values, the regression model chosen can be applied with adequate trueness and precision. As can be seen in Table 2, it is possible to specify appropriate LLOQ levels using the RSE parameter with any type of regression model for each calibration. The limits were found to change significantly for the same calibration results depending on the regression function applied. For example, in case study #1 the LLOQ can be set at 0.1 mg/L using WLS regression but must be increased to 5 mg/L with the used calibrators if OLS regression needs to be applied.

Other authors have also evaluated the heteroscedasticity effects on their analytical methods [10]. They found their results to be heteroscedastic, used the data obtained in recovery tests to assess the quality of both OLS and WLS regressions, and found no significant effects on recovery between the two calibration models. For this reason, it was concluded that, in general, heteroscedasticity should rarely affect accuracy. The calibration results presented in that study have been evaluated to calculate the RSE value of both regression methods. The results obtained are RSE = 6% for the OLS regression and RSE = 2% for the WLS model, which confirms that the LLOQ used in that study (0.5 mg/L) is adequate for applying the OLS regression. Therefore, no differences between OLS and WLS should be found if there are no measurements below this LLOQ. An approximation of the LOQ value was determined using the SD of the intercept determined by WLS, and LOQ = 0.08 mg/L was found, confirming that the first standard (LLOQ) was set at $>5 \times$ LOQ.

3.4 | Non-linear relationships between dependent and independent variables

The use of the RSE criterion has the advantage that it can be applied to all types of regression functions, not only those with linear relationships between the variables. Therefore, the ability of RSE to assess functions with non-linear relationships between the dependent and independent variables (i.e., curvature) has also been evaluated (Supporting Information, case studies #11, #12, and #13).

Case study #11 corresponds to an HPLC-UV method where the curvature is questionable. In this case, the lack-of-fit test suggests linearity ($p = 0.131$) but Mandel's test suggests non-linearity ($p < 0.001$). The lack-of-fit test is more robust with heteroscedastic calibrations because it takes into account the variability at the different levels evaluated, whereas Mandel's test only uses the mean value for each standard without taking into account the variabilities. However, the only test that can be applied with single measurements usually performed in the daily routine of laboratories is Mandel's test. In such situations, this calibration could be considered non-linear.

The results obtained with the RSE test confirmed that the lower value is obtained with the quadratic calibration (RSE = 1%). The acceptable results obtained with the WLS regression (RSE = 6%) may be associated with the questionable non-linearity of this calibration. The LOQ of this method is 1 mg/L, for this reason when the first standard (LLOQ) was set at 10 mg/L, the OLS regression also became satisfactory (RSE = 3%). These results also demonstrate that the RSE calculation can be used to analyze whether or not a linear model can be used with calibrations that seem to present curvature. As can be seen, despite the best RSE value being obtained with the quadratic regression, the curvature is not experimentally significant and the results obtained with the linear models, once the correct LLOQ value is set, do not introduce significant deviations in the REs calculated (RSE ≤ 5%).

The other two cases (case studies #12 and #13) correspond to methods using calibrations ranges that yield significant curvature in the response (Figure 2). In case study #12, the two statistical tests to assess linearity, lack-of-fit ($p < 0.001$) and Mandel's test ($p = 0.021$), confirmed the curvature in the calibration range evaluated. As expected in this situation, the two linear regression models evaluated give unsatisfactory results (OLS: RSE = 61%; WLS; RSE = 47%). The only model that yields a satisfactory RSE result is the quadratic regression (RSE = 6%). In case study #13, the two tests for linearity also confirmed non-linearity (lack-of-fit: $p < 0.001$, Mandel's test: $p = 0.002$). As in the previous example, the RSE values determined with the two linear models confirm the non-acceptability of the linear
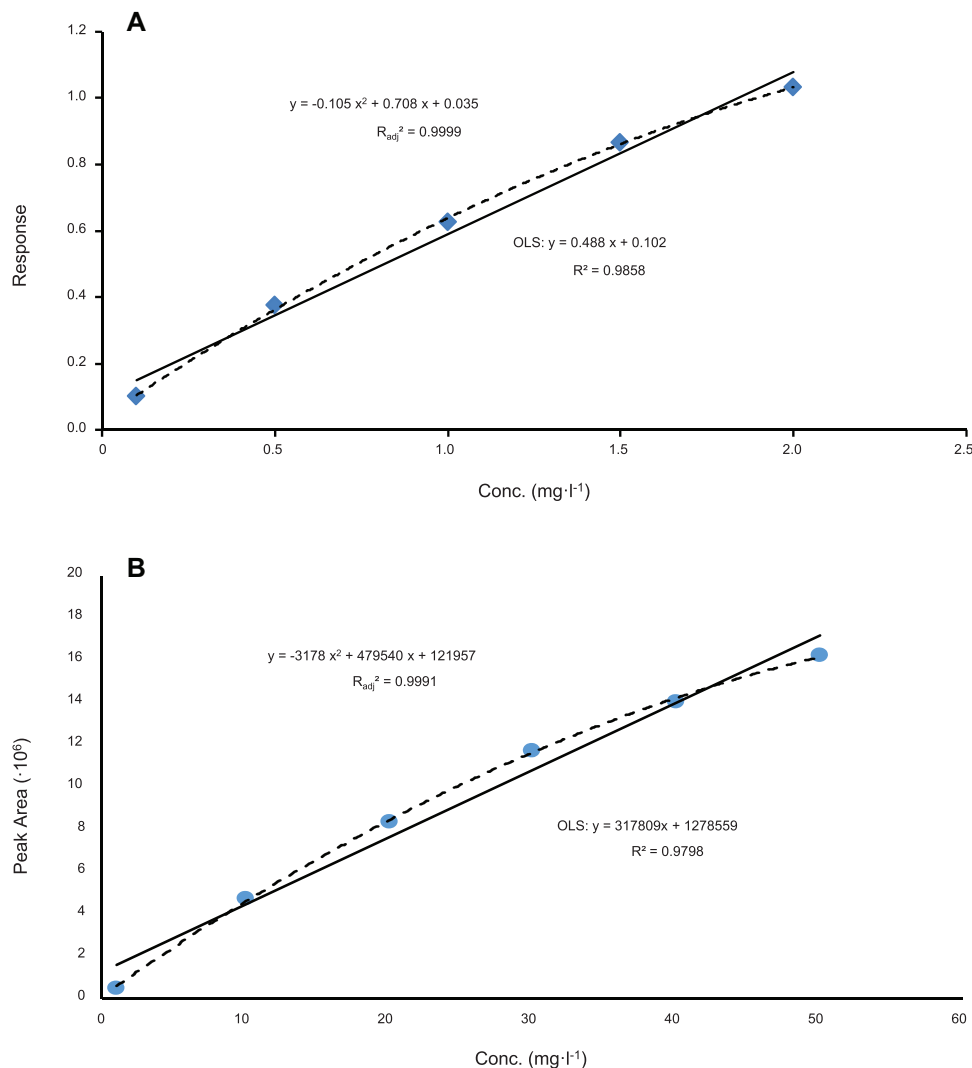
A

$y = -0.105 x^2 + 0.708 x + 0.035$

$R_{adj}^2 = 0.9999$

OLS: $y = 0.488 x + 0.102$

$R^2 = 0.9858$

B

$y = -3178 x^2 + 479540 x + 121957$

$R_{adj}^2 = 0.9991$

OLS: $y = 317809x + 1278559$

$R^2 = 0.9798$

FIGURE 2   Curves obtained for the non-linear calibrations evaluated. (A) Case study #12 and (B) case study #13

models (OLS: RSE = 211%; WLS: RSE = 53%). The quadratic regression gives an acceptable RSE value (11%).

4 | CONCLUDING REMARKS

Many people working in laboratories often perform linear regression based on OLS fitting without thinking about the limitations of how it really works. It has been demonstrated in many studies that WLS is the only linear model that can yield good results when analyzing samples where contents at low levels are expected. Despite this, it is possible to see comments in many internet forums dealing with calibration such as it is not clear (or not easy to see) what weights should be used and it is not clear if the WLS gives any advantage. Moreover, the most widely used parameter for assessing linear calibrations is still the correlation or the R^2s. However, it must be taken into account that when

R or R^2 are used as acceptance criteria for calibrations with linear relationships, they are irrelevant.

Here, the use of the RSE has been extensively evaluated with experimental calibrations obtained in our laboratory. First, it has been confirmed that it is not required to analyze replicates for each standard to perform WLS. Using approximations such as $w_i = 1/x_i^2$ or $w_i = 1/y_i^2$ allows correct estimates of the slopes and intercepts to be obtained, as well as of their SDs, equivalent to those obtained from the experimental variance after measuring a large number of independent replicates for each standard.

The use of the RSE calculation has proven to be a robust tool for assessing the quality of the calibration regression used without the need to perform multiple replicates for each standard. Another advantage of RSE is that it is not necessary to take into account the R or R^2 values to assess linear regression models. Moreover, this calculation also permits calibrations with non-linear and linear

relationships to be easily differentiated for those calibrations where the curvature is questionable.

Although it has been extensively demonstrated that OLS should not be applied as the regression model when samples are expected to be determined close to the LOQ [1,6–9,12,13,15,17–19], this function is still the most widely applied in all types of laboratories. This fact can mainly be associated with two factors. First, OLS is practically the only regression model that almost everyone learns during their training. A survey carried out at our university of more than 500 biotechnology and chemistry students over six academic years showed that none of the students had received training about any regression model other than OLS, only around 15% of the students answered that they had heard something about WLS in some analytical chemistry subjects. Given this situation, it is not surprising that OLS ends up being used practically for anything and everything. Second, OLS has great mathematical simplicity in comparison with other calibration functions. The widespread use of OLS is also almost certainly attributable to the fact that OLS calculations are included in popular software packages such as Microsoft Excel and in scientific calculators. Other regression functions are more complex mathematically and require the use of specialized statistical software to perform the calculations. It is true that some modern instruments already include controlling software that is able to perform other types of regression calculations but the generalized use of OLS has not decreased in recent years.

The most significant conclusion obtained in this study is that OLS can be applied with any type of calibration with linear relationships between the variables, but its LLOQ will always be significantly higher than the conventionally measured LOQ. A minimum LLOQ $\geq 5 \times$ LOQ seems to be required to reach acceptable RSE values when OLS regression is used. However, if we want to reach the smallest possible LLOQ level, equivalent to the conventional LOQ value, WLS regression must be applied, which is a common situation in trace analysis. The need to set the LLOQ $\geq 5 \times$ LOQ using OLS regression found in this study was also suggested in studies using more complex calculations, such as accuracy profile plots [13,34], which require many analyses to be performed, doing replicates at all calibration levels. The great advantage of RSE is that it does not require any extra measurement to be performed other than the conventional measurements required for a set of calibration standards, as is done in the daily routine of a laboratory.

Another significant characteristic of the application of the RSE value in calibrations is that the evaluation of R or R^2 is not required to assess the GOF of the regression function. As can be seen in Table 1, all the analyzed regression functions give $R^2 > 0.99$ (or $R > 0.995$), one of the most common criteria for acceptance of calibrations in laboratories. However, it is clear that this parameter is not adequate to assess the quality of the regression functions because it would lead to having to accept OLS regressions with >20% of RE at low concentration levels. As was also suggested by Burrows and Parr [1], necessary changes to remove the use of the correlation coefficient and the coefficient of determination must be considered for methods and their use as a criterion for the acceptance of a calibration equation should be avoided.

FUNDING INFORMATION

CONFLICT OF INTEREST
The author has declared no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available in the supplementary material of this article.

ORCID
Juan M. Sanchez https://orcid.org/0000-0002-4139-7273

REFERENCES

1. Burrows R, Parr J. Evaluating the goodness of instrument calibration for chromatographic procedures. LC-GC North Am. 2020;38:35–8.
2. Van Arendonk MD, Skogerboe RK, Grant CL. Correlation coefficients for evaluation of analytical calibration curves. Anal Chem. 1981;53:2349–50.
3. Analytical Methods Committee. Uses (proper and improper) of correlation coefficients. Analyst 1988;113:1469.
4. Mulholland M, Hibbert DB. Linearity and the limitations of least squares calibration. J Chromatogr A. 1997;762:73–82.
5. Van Loco J, Elskens M, Croux C, Beernaert H. Linearity of calibration curves: use and misuse of the correlation coefficient. Accredit Qual Assur. 2002;7:281–5.
6. Raposo F. Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: a tutorial review. TrAC - Trends Anal Chem. 2016;77:167–85.
7. Logue BA, Manandhar E. Percent residual accuracy for quantifying goodness-of-fit of linear calibration curves. Talanta 2018;189:527–33.
8. Almeida AM, Castel-Branco MM, Falcão AC. Linear regression for calibration lines revisited: weighting schemes for bioanalytical methods. J Chromatogr B Anal Technol Biomed Life Sci. 2002;774:215–22.
9. Baumann K, Wätzig H. Appropriate calibration functions for capillary electrophoresis II: heteroscedasticity and its consequences. J Chromatogr A. 1995;700:9–20.

10. Donati GL, Amais RS. Fundamentals and new approaches to calibration in atomic spectrometry. J Anal At Spectrom. 2019;34:2353–69.

11. Kiser MM, Dolan JW. Selecting the best curve fit. LC-GC Eur. 2004;17:138–43.

12. Sanchez JM. Ordinary least squares with laboratory calibrations: a practical way to show students that this fitting model may easily yield biased results when used indiscriminately. World J Anal Chem. 2017;5:1–8.

13. Sanchez JM. Linear calibrations in chromatography: the incorrect use of ordinary least squares for determinations at low levels, and the need to redefine the limit of quantification with this regression model. J Sep Sci. 2020;43:2708–17.

14. Nagaraja NV, Paliwal JK, Gupta RC. Choosing the calibration model in assay validation. J Pharm Biomed Anal. 1999;20:433–8.

15. Tellinghuisen J. Weighted least-squares in calibration: What difference does it make? Analyst 2007;132:536–43.

16. Vesterlund A, Ramebäck H. Achieving confidence in trace element analysis for nuclear forensic purposes: ICP-MS measurements using external calibration. J Radioanal Nucl Chem. 2019;322:941–8.

17. Edgerley DA Tecniques for improving the accuracy of calibration in the environmental laboratory. In: 14th Annual Waste Testing & Quality Assurance Symposium; 1998:181–7.

18. Yolci Omeroglu P, Ambrus A, Boyacioglu D. Uncertainty of pesticide residue concentration determined from ordinary and weighted linear regression curve. Food Addit Contam - Part A Chem Anal Control Expo Risk Assess. 2018;35:1324–39.

19. Szabo GK, Browne HK, Ajami A, Josephs EG. Alternatives to least squares linear regression analysis for computation of standard curves for quantitation by high performance liquid chromatography: applications to clinical pharmacology. J Clin Pharmacol. 1994;34:242–9.

20. Zabell APR, Lytle FE, Julian RK. A proposal to improve calibration and outlier detection in high-throughput mass spectrometry. Clin Mass Spectrom. 2016;2:25–33.

21. Bohrnstedt GW, Carter TM. Robustness in regression analysis. Sociol Methodol. 1971;3:118–46.

22. Knecht J, Stork G. Prozentuales und logarithmisches verfahren zur berechnung von eichkurven. Fresenius' Zeitschrift Für Anal. Chemie. 1974;270:97–9.

23. The NELAC Institute. TNI standard: EL-V1M4 sections 1.7.1 and 1.7.2. 2015. https://nelac-institute.org/docs/standards/2015/chemistry/TNI_Standard_ELV1M4_1.7.1_and_1.7.2_0415.pdf Last Accessed September 12, 2021.

24. US-EPA. SW-846 test method 8000D: determinative chromatographic separations. 2018. https://www.epa.gov/hw-sw846/sw-846-test-method-8000d-determinative-chromatographic-separations Last Accessed September 12, 2021.

25. Levene H. Robust tests for equality of variances. In: Olkin I, editor. Contributions to probability and statistics: essays in honor of harold hotelling.. Palo Alto: Stanford University Press; 1960. p. 278–92.

26. Brown MB, Forsythe AB. Robust tests for the equality of variances. J Am Stat Assoc. 1974;69:364.

27. Currie LA. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). Anal Chim Acta. 1999;391:105–26.

28. Magnusson B, Örnemark U. The fitness for purpose of analytical methods: a laboratory guide to method validation and related topics. 2nd ed. Teddington: Eurachem; 2014.

29. Gu H, Liu G, Wang J, Aubry AF, Arnold ME. 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. Anal Chem. 2014;86:8959–66.

30. Zeng QC, Zhang E, Dong H, Tellinghuisen J. Weighted least squares in calibration: estimating data variance functions in high-performance liquid chromatography. J Chromatogr A. 2008;1206:147–52.

31. Oppenheimer L, Capizzi TP, Weppelman RM, Mehta H. Determining the lowest limit of reliable assay measurement. Anal Chem. 1983;55:638–43.

32. Zorn ME, Gibbons RD, Sonzogni WC. Weighted least-squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. Anal Chem. 1997;69:3069–75.

33. US-FDA. Bioanalytical method validation guidance for industry, FDA 2018. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry Last Accessed March 8, 2021.

34. Hubert P, Chiap P, Crommen J, Boulanger B, Chapuzet E, Mercier N. The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: from the Washington conference to the laboratory. Anal Chim Acta. 1999;391:135–48.

35. ICH. Validation of analytical procedures: text and methodology Q2(R1). 2005. https://www.ich.org/page/quality-guidelines Last Accessed March 8, 2021.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.