Journal of Chromatography B, 911 (2012) 192–202

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

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

Comparison of different linear calibration approaches for LC–MS bioanalysis

Aimin Tan*, Kayode Awaiye, Besy Jose, Paresh Joshi, Fethi Trabelsi

BioPharma Services Inc., Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 29 August 2012
Accepted 5 November 2012
Available online 15 November 2012

Keywords:
Linear calibration
Linear regression
Calibration standard
Bioanalysis
Quantitation
LC–MS

ABSTRACT

Many different calibration approaches are used for linear calibration in LC–MS bioanalysis, such as different numbers of concentration levels and replicates. However, direct comparison of these approaches is rare, particularly using experimental results. The purpose of this research is to compare different linear calibration approaches (existing and new ones) through simulations and experiments. Both simulation and experimental results demonstrate that linear calibration using two concentrations (two true concentrations, not forced through zero) is as good as or even better than that using multiple concentrations (e.g. 8 or 10) in terms of accuracy. Additionally, two-concentration calibration not only significantly saves time and cost, but is also more robust. Furthermore, it has been demonstrated that the extrapolation of a linear curve at the high concentration end to a linearity-known region is acceptable. When multi-concentration calibration is used, the difference between the two commonly used approaches, i.e. singlet (one curve) or duplicate (two curves) standards per concentration level is small when a method is very precise. Otherwise, one curve approach can result in larger variation at the low concentration end and higher batch failure rate. To reduce the variation and unnecessary reassays due to batch failure or possible rejection of the lowest and/or highest calibration standards, a partially duplicate-standard approach is proposed, which has duplicate-standard-like performance but still saves time and cost as singlet-standard approach does. Finally, the maximum allowable degrees of quadratic (non-linear) response in linear calibration are determined for different scenarios. Because of its multiple advantages and potential application in regulated bioanalysis, recommendations as how to implement two-concentration linear calibration in practice are given and some typical "concerns" regarding linear calibration using only two concentrations are addressed, e.g. how does one know if the response is truly linear over a given range when only two concentrations are used?.

1. Introduction

Linear regression is the most widely employed calibration model in LC–MS based bioanalysis [1–7]. Typically, calibration standards (CSs) at several different concentration levels (e.g. 8) are analyzed in duplicate to construct a linear calibration curve [1–4], particularly in regulated bioanalysis because a minimum of six concentration levels are specified or implied for linear calibration in most regulatory guidelines on bioanalytical method validation [8–11]. Though it appears simple, in reality, considerable amounts of time and cost are required for the preparation of various working standard solutions and calibration standards as well as the actual analysis of calibration standards in each batch (run) during method validation and sample analysis. For example, Nilsson and Eklund [12] reported that roughly 20% of the total time for analysis is spent on the preparation and analysis of calibration standards. Hence, it is very much desirable to reduce the aforementioned time and cost while without sacrificing the accuracy of calibration.

Evidently, two approaches can be used to achieve this goal, i.e. reducing the number of concentration levels or the number of replicates per concentration level with the first one being most effective. While it is well known that two points define a straight line, i.e. two concentration levels being sufficient to define a linear calibration curve, multiple concentration levels are still commonly used for linear calibration in LC–MS bioanalysis [1–7]. The main reasons behind this might be due to the need to satisfy regulatory requirements, the tradition, the demonstration of linearity, and a perception of "more CS levels leading to better accuracy". Unfortunately, this perception as well as the requirement of a minimum of six standard concentration levels for a linear regression are rarely supported by published scientific research. On the contrary, they have been proven otherwise by many [13–16]. For examples, Renman and Jagner [13] have demonstrated through simulations that optimum accuracy and precision are obtained by using minimum calibration concentration levels while performing multiple measurements of these concentrations, instead of using multiple standard concentration

* Corresponding author at: BioPharma Services Inc., 4000 Weston Road, Toronto, ON M9L 3A2, Canada. Tel.: +1 416 747 8484; fax: +1 416 747 8480.
E-mail addresses: aimintan@hotmail.com, atan@biopharmaservices.ca (A. Tan).

1570-0232/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.11.008

levels. Boulanger et al. [14,15] reached a similar conclusion through simulation and theoretical consideration, i.e. best analytical results being obtained by using two extreme concentrations for linear calibration. Based on simulated data for atomic absorption spectrometric methods and from the perspectives of outlier detection and evaluation of goodness-of-fit, Penninckx et al. found that, among the various numbers of concentration levels studied (3, 4, 6, 9, and 12), three (evenly spaced) concentration levels gave the best results for linear calibration [16]. As these conclusions were drawn only from simulations and theoretical consideration and the focus was not on LC–MS bioanalysis, it is therefore desirable to perform simulations with the focus on LC–MS bioanalysis. It would be particularly interesting to compare two-concentration linear calibration with multiple-concentration linear calibration using experimental results and to further develop a workable plan for the implementation of two-concentration calibration if the experimental results corroborate those of simulations.

Peters and Maurer [17] performed an interesting comparison between one-point (one-concentration, linear through zero) calibration and multiple-point (multiple-concentration) full calibration using the experimental data of six mass spectrometry-based multianalyte bioanalytical assays. It was concluded that one-point calibration with a calibrator close to the centre of the full calibration range can be a feasible alternative to full calibration. However, some high bias and precision data were obtained for some analytes in the low concentration range. It should be pointed out that a true zero intercept is rare in LC–MS based bioanalysis due to analyte and internal standard cross signal contributions, matrix interferences, endogenous level or incomplete stripping of endogenous compounds [18–20]. In addition, the response-concentration relationship might be linear starting only from the lower limit quantitation (LLOQ), not necessarily from the zero concentration. For these reasons, single-point calibration was very successful in some situations but not so successful in others [12,17].

Since the required number of replicates for each concentration level is not specified in the regulatory guidelines, different approaches exist in the bioanalytical community [1–7,21]. Some use two replicates [1–4] while others use singlet [4–7,22] per calibration concentration level, which are sometimes referred to as "two calibration curves" and "one calibration curve", respectively; however, it should be clarified that it is always one regression equation, one curve, being produced at the end of regression. Obviously, each has its own real or perceived advantages and disadvantages. For instance, duplicate standard is usually perceived as being more accurate and reliable. In addition, more calibration standard replicates can be excluded from the regression while still meeting the regulatory requirement (a minimum of 75% of the total calibration standards accepted [8–10]), e.g. up to 4 out 16 CS replicates. On the other hand, those who choose singlet-standard approach may believe that it is equally accurate but with less cost and time, e.g. half of that for duplicate-standard approach. Unfortunately, to the best of our knowledge, no side by side comparison of their differences can be found in the literature.

Based on the above, it is important to compare different linear calibration approaches through both simulation and experiment for LC–MS based bioanalysis, especially when exists an ever stronger desire in the global bioanalytical community for scientifically sound practices and guidelines where rationale is given for each requirement [23,24]. Specifically, based on simulations for LC–MS bioanalysis and experimental results, is two-concentration linear calibration accurate and robust enough for regulated bioanalysis? How should it be best implemented? Provided that two concentrations are used, how can the linearity over the entire calibration range be demonstrated? When linearity has been demonstrated over a given concentration range (i.e. as a straight line), should extrapolation within this range be acceptable? Is there any difference between duplicate-standard and singlet-standard approaches? If yes, how much is it? Is there any other alternative approach that would have the benefits of these two common approaches (e.g. duplicate- or singlet-standard) yet avoid their respective disadvantages? All these questions will be addressed in this article. It is our hope that the results obtained herein will be useful to the establishment of scientifically sound bioanalytical guidelines and practices in the future.

2. Experimental

2.1. Chemicals and reagents

Triamcinolone acetonide, rosuvastatin calcium salt, rosuvastatin-d_6 sodium salt, valsartan, and valsartan-d_3 were all purchased from Toronto Research Chemicals (Toronto, Ontario, Canada). Triamcinolone-6-d_1 acetonide-d_6 was purchased from C/D/N Isotopes (Pointe-Claire, Quebec, Canada). Acetonitrile (HPLC grade), methanol (HPLC grade), methyl tertiary butyl ether (MTBE, OmniSolv), and water (HPLC grade) were obtained from EMD (Toronto, Ontario, Canada). Ammonium acetate (Ultra Pure) and formic acid (reagent) were bought from Caledon (Georgetown, Ontario, Canada). Human EDTA K_2 plasma was obtained from BioChemed Services (Winchester, Virginia, USA). High purity liquid nitrogen was supplied by Linde (Toronto, Ontario, Canada).

2.2. Stock solutions, calibration standard and quality control samples

The stock solutions were all prepared in methanol at various concentrations (triamcinolone acetonide, triamcinolone-6-d_1 acetonide-d_6, and rosuvastatin: 100 μg/ml; rosuvastatin-d_6: 50 μg/ml; valsartan: 1.6 mg/ml; valsartan-d_3: 72 μg/ml). All intermediate and working solutions were prepared by the successive dilutions of the corresponding stock solutions in methanol. The calibration standards and quality controls (QCs) were all prepared in control blank human EDTA K_2 plasma, except those of rosuvastatin, which were prepared in buffered human EDTA K_2 plasma (1 M ammonium acetate buffer (pH 4.0) mixed to plasma in the ratio of 1/10, v/v).

2.3. Sample processing

2.3.1. Triamcinolone acetonide

Two hundred microliters (200 μl) of human EDTA K_2 plasma sample was mixed with 150 μl of internal standard working solution (H_2O-based). Then, the mixture was loaded on an ISOLUTE SLE+ 400 μl supported liquid extraction plate (Biotage, Charlotte, NC, USA) and extracted with 800 μl of MTBE. After evaporation, the sample residual was reconstituted in 200 μl of mobile phase, acetonitrile/2 mM ammonium acetate buffer (pH 3.2) (55/45, v/v).

A large run was extracted, which included six replicates each of the lower limit of quantitation QC (LLQC), low QC (QC1), medium QC (QC2), high QC (QC3), and the upper limit of quantitation QC (ULQC), eight replicates each of double blank and zero standard (ZS), as well as eight (CS2–CS7) or 12 replicates (CS1 and CS8) of calibration standards. The concentrations of quality controls were 20 pg/ml, 60 pg/ml, 7 ng/ml, 15 ng/ml, and 20 ng/ml for LLQC, QC1, QC2, QC3, and ULQC, respectively. The concentrations of calibration standards were 20 pg/ml, 40 pg/ml, 0.4 ng/ml, 2 ng/ml, 4 ng/ml, 8 ng/ml, 16 ng/ml, and 20 ng/ml for CS1–CS8, respectively. To mimic a potential ten percent error of response in calibration standards, one replicate each of the eight CS levels was extracted with 180 μl (90% of 200 μl) of the corresponding prepared CS plus 20 μl of control blank plasma (to make the same total plasma volume, i.e. 200 μl).

During LC–MS/MS analysis, QC samples were injected in the middle of the batch while calibration standards were injected at the beginning and the end of the batch (equally distributed). Since multiple replicates were injected for each CS, the selection of which CS replicates to be used in the regression was based on the following pre-defined guideline. When an even number of replicates are required for a CS concentration level, e.g. 4, half (e.g. 2) will be chosen from those injected at the beginning of the batch and the other half from the end of the batch. For an odd number of replicates needed in the regression, e.g. 3, which cannot be equally distributed, the extra one replicate will be chosen from those injected at the beginning of the batch if the CS level number is also odd, e.g. CS1, CS3, CS5, and CS7. Otherwise, the extra one replicate will be selected from those injected at the end of the batch, i.e. for CS2, CS4, CS6, and CS8. In addition, the selection of replicates for the same CS level will start from the two extremities of the batch and gradually move to the centre.

The linear calibration schemes evaluated for tramcinolone acetonide were 8×7 (eight CS levels, seven replicates each), 8×2 (eight CS levels, two replicates each), 8×1 (eight CS levels, singlet for each), $8 \times 1 + 2$ (same as 8×1 except duplicate for CS1 and CS8), 5 + 5 + 5 (CS1, CS6, and CS8, five replicates each), 7 + 7 (CS1 and CS8, seven replicates each), 4 + 4 (CS1 and CS8, four replicates each), 2 + 2 (CS1 and CS8, two replicates each), and 4 + 4, T7 (CS1 and CS7, four replicates each, truncated range).

2.3.2. Rosuvastatin

Two hundred microliters (200 μl) of human EDTA K_2 plasma sample was mixed with 200 μl of internal standard working solution (0.2 M ammonium acetate, pH 4.5). Then, the mixture was loaded on an ISOLUTE SLE+ 400 μl supported liquid extraction plate and extracted with 900 μl of MTBE. After evaporation, the sample residual was reconstituted in 200 μl of acetonitrile/H_2O (50/50, v/v) with 5 mM ammonium acetate (pH 4.6).

The method of rosuvastatin was fully validated as per US FDA and EMA guidelines on bioanalytical method validation [8,10]. Each between-run (inter-run) used for the evaluation of accuracy and precision consisted of six replicates of each QC type, i.e. LLQC, low QC (QC1), medium QC (QC2), and high QC (QC3). For the evaluation of within-run (intra-run) accuracy and precision, additional quality controls, namely intermediate QC (QC4) and ULQC, were processed in six replicates each. The concentrations of quality controls were 0.10 ng/ml, 0.30 ng/ml, 4.02 ng/ml, 50.30 ng/ml, 80.47 ng/ml, and 100.59 ng/ml for LLQC, QC1, QC4, QC2, QC3, and ULQC, respectively. The calibration curve included ten CS levels, all extracted in singlet except those at the lowest and highest concentrations (CS1 and CS10) that were extracted in duplicate. The concentrations of calibration standards were 0.10 ng/ml, 0.20 ng/ml, 1.01 ng/ml, 2.01 ng/ml, 5.03 ng/ml, 10.06 ng/ml, 20.12 ng/ml, 60.35 ng/ml, 90.53 ng/ml, and 100.59 ng/ml for CS1–CS10, respectively. The odd level calibration standards (e.g. CS3, CS5) were injected at the beginning of the batch while the even level calibration standards (e.g. CS4, CS6) were injected at the end of the batch. For CS1 and CS10, one replicate each were injected at the beginning and the second replicates were injected at the end. Moreover, two replicates each of double blank and zero standard were also processed together each calibration curve to evaluate contamination and interference.

A total of six runs were extracted to evaluate accuracy without the introduction of deliberate errors for calibration standards. Four linear calibration schemes were compared, i.e. 10×1 (ten CS levels, singlet for each), $10 \times 1 + 2$ (same as 10×1 except duplicate for CS1 and CS10), 2 + 2 (duplicate for CS1 and CS10), and 2 + 1, T (duplicate for CS1 and singlet for CS8, truncated range).

2.3.3. Valsartan

One hundred microliters (100 μl) of human EDTA K_2 plasma sample was mixed with 300 μl of internal standard working solution (0.5 M ammonium acetate, pH 4.5). Then, the mixture was loaded on an ISOLUTE SLE+ 400 μl supported liquid extraction plate and extracted with 800 μl of MTBE. After evaporation, the sample residual was reconstituted in 400 μl of acetonitrile/H_2O (60/40, v/v) with 4 mM ammonium acetate.

A batch consisting of eight (CS2–CS9) or ten (CS1, CS10) replicates of calibration standards and two replicates each of double blank and zero standard was extracted. Six replicates each of the ten CS levels were processed as quality controls. The concentrations of CS1–CS10 were 15 ng/ml, 30 ng/ml, 150 ng/ml, 300 ng/ml, 750 ng/ml, 1.5 μg/ml, 3 μg/ml, 9 μg/ml, 13.5 μg/ml, and 15 μg/ml, respectively. Similar procedures as those mentioned in Section 2.3.1 were followed for sample injection and the selection of CS replicates to be used in regressions.

The linear calibration schemes tested for valsartan included 10×2 (ten CS levels, duplicate for each), 10×1, $10 \times 1 + 2$, 2 + 2 (CS1 and CS10, each in duplicate), and 2 + 2, T (CS1 and CS7, each in duplicate, truncated range).

2.4. LC–MS/MS conditions

The LC system consisted of an LC-20AD pump, a SIL-20AC HT autosampler, a CTO-20AC column oven, a DGU-20A3 degasser, and a CBM-20A controller from Shimadzu (via Mandel, Guelph, Ontario, Canada). A Sciex API 4000 triple quadrupole mass spectrometer equipped with a TurboIonSpray interface (MDS Sciex, Toronto, Ontario, Canada) was used for the detection and data analysis (mainly chromatogram integration) with the Analyst software (versions 1.5.1 and 1.5.2, MDS Sciex).

The important method-specific LC–MS/MS conditions are summarized in Table 1 for all the three analytes, i.e. triamcinolone acetonide, rosuvastatin, and valsartan.

2.5. Simulations

2.5.1. Generation of response data

Theoretical responses for CS and QC samples were first calculated using the following linear equation, $y = 0.0002 + 0.002 \times C$, where y is the response and C represents the concentration. The CS concentrations used were 1, 2, 20, 100, 200, 400, 800, and 1000 (arbitrary unit) for CS1 to CS8, respectively; while those of QCs were 1, 3, 350, 750, and 1000 (arbitrary unit) for LLQC, QC1, QC2, QC3, and ULQC, respectively. In one simulated batch, there were six replicates for each QC type and eight (CS2 to CS7) or 16 (CS1 and CS8) replicates for the calibration standards. Then, random errors were added to the calculated responses of CS and QC replicates using the "RAND()" function in Excel programme (version 2007). The maximum magnitudes of random error set for poor precision scenario were ±20% at the LLOQ level and ±15% at others. In the good precision scenario, the corresponding percentages of error were reduced to ±12% and ±7%, respectively. For the very good precision scenario, the magnitudes of error were further reduced to ±6% and ±4%, respectively. A total of 30 batches were simulated for each precision scenario.

2.5.2. Linear regression of simulated batches using different calibration schemes

Once the response data sets were generated, they were linearly regressed using different linear calibration schemes with the weighting factor of $1/C^2$. The linear calibration approaches evaluated included 8×2, 8×1, $8 \times 1 + 2$, 8 + 8 (eight replicates for both CS1 and CS8), 4 + 4, 2 + 2, and 6 + 5 + 5 (six replicates of CS1 plus five replicates for both CS6 and CS8). Since the numbers of CS

Table 1
LC–MS/MS conditions for triamcinolone acetonide, rosuvastatin, and valsartan.

	Triamcinolone acetonide	Rosuvastatin	Valsartan
LC column	Agilent Zorbax SB-Aq column (150 mm × 4.6 mm, 5 μm)	Shimadzu C18 column (50 mm × 4.6 mm, 5 μm)	Agilent Zorbax SB-Aq column (150 mm × 4.6 mm, 5 μm)
Mobile phase	ACN/2 mM AA (pH 3.2) (55/45, v/v)	ACN/H_2O (55:45, v/v), 2 mM AA, 0.05% (v/v) FA	ACN/H_2O (60:40, v/v), 4 mM AA
Flow rate (ml/min)	1.2	1	1
Injection volume (μl)	50	20	2
Ionization	Turbo Ion Spray, positive	Turbo Ion Spray, positive	Turbo Ion Spray, negative
Collision gas (CAD, psi)	10	12	12
Curtain gas (CUR, psi)	30	30	25
Nebulizer gas (GS1, psi)	65	40	60
Heater gas (GS2, psi)	65	70	50
Ion Spray Voltage (ISV, V)	4000	3000	-3000
Temperature (TEM, °C)	700	700	550
MRM transitions (*m*/*z*)	435.2 → 415.2 (442.2 → 422.2)	482.1 → 258.2 (488.1 → 264.0)	434.1 → 350.0 (437.0 → 349.7)
Declustering potential (DP, V)	45	115 (111)	−95 (−80)
Collision energy (CE, V)	15	45 (50)	−26
Collision cell exit potential (CXP, V)	11	18	−9 (−19)

Notes: ACN, acetonitrile; AA, ammonium acetate; FA, formic acid.
The values in the brackets are for the internal standards (if applicable).

replicates generated were more than what were required for some calibration schemes, the same procedure outlined in Section 2.3.1 was followed as to which replicates should be included in the regression.

2.5.3. Allowable maximum degree of quadratic response in linear regression

To mimic quadratic response, $y = A \times C^2 + C$ was used to generate responses without the addition of random errors. Four dynamic concentration ranges (1000-fold, 500-fold, 250-, and 100-fold) were involved. The concentrations used for linear regressions are presented in Table 3. In all cases, zero concentration was never used, i.e. never forced through zero. To determine the allowable maximum degree of quadratic response in linear calibration in a given concentration range by a calibration scheme, the parameter "A" was continuously changed (progressively more quadratic) until the bias for any sample (any concentration) in the given concentration range by the calibration scheme reached 15% or −15%.

2.6. Regression calculations

Both weighted ($1/C$ and $1/C^2$) and non-weighted least-squares linear regressions were performed using an in-house built Excel programme. Prior to its application to the various simulated and experimental data sets, the accuracy of the in-house built programme was successfully verified against that of the Analyst software.

3. Results and discussion

3.1. Comparison of different linear calibration schemes based on simulated data

3.1.1. Some considerations

The linear equation, $y = 0.0002 + 0.002 \times C$, was chosen because it represents a typical LC–MS bioanalytical method over the concentration range of 1–1000 (1000-fold dynamic range) with the concentration of internal standard set at half of the ULOQ [25]. Though the generally accepted limit for interference in a zero standard is 20% of the LLOQ response, efforts are commonly made to reduce this amount [26]. Therefore, the intercept of 0.0002, which corresponds to 10% of the LLOQ, is quite reasonable.

The weighted ($1/C^2$) linear regression was selected because $1/C^2$ is by far the most often used weighting factor in LC–MS bioanalysis [1–7] and it has also often been proven to be the most appropriate one at the end of statistical test if it is ever performed [22,27]. In addition, the utilization of the weighting factor of $1/C^2$ is helpful in reducing the potential impact of cross signal contribution between an analyte and its internal standard [18].

As to the magnitude of random errors, it deemed necessary to consider three different situations, i.e. poor precision, good precision, and very good precision, which represent typical precisions of different types of LC–MS bioanalytical methods. For example, a bioanalytical method utilizing stable-isotope labelled (SIL) internal standard (IS) or a bioanalytical method using structural analogue IS but being well-developed usually falls into the category of good precision [4,28]. Examples of very good precision are the bioanalytical methods that utilize SIL internal standards and are very well developed [3,29]. On the other hand, if a bioanalytical method uses structural analogue IS and simple protein precipitation extraction to quantify an analyte in a complex matrix like whole blood, the precision would most likely be poor [28]. For each precision scenario, 30 batches were selected as a balance between the work load of simulations and statistical significance of the results. Furthermore, 30 batches would be a reasonable number of batches for a typical bioanalytical study [30].

To adequately compare the accuracy of different linear calibration schemes, a different indicator other than mean bias is necessary because of the large number of batches (30). Otherwise, the chance of large positive biases being cancelled out by large negative biases (which still leads to small mean bias) would be very high. For this reason, a new indicator that adequately combines the magnitudes of bias (either negative or positive) and variation (CV) is proposed. This new indicator, termed mean apparent bias (MAB), is defined as: $\mathrm{MAB} = \sqrt{1/n(\sum_{i=1}^{n} b_i^2)}$, where b_i is the bias of individual replicate and n is the total number of replicates.

Table 2
Comparison of different linear calibration schemes using the same simulated raw data sets.

	CV (%) of responses for different batches		Calibration scheme	8 × 2	8 × 1	8 × 1 + 2	6 + 5 + 5	8 + 8	4 + 4	2 + 2
Poor precision[a]	LLQC	8.4–16.9	MAB[d] of LLQC (%)	15.8	16.9	16.3	15.2	14.7	15.0	16.6
	QC1	3.9–10.5	MAB of QC1 (%)	9.8	10.6	10.1	10.0	9.7	10.4	10.6
	QC2	4.3–12.5	MAB of QC2 (%)	9.7	10.4	9.9	10.3	9.6	11.0	10.6
	QC3	3.5–12.3	MAB of QC3 (%)	10.1	10.4	10.2	10.5	10.2	10.7	10.8
	ULQC	3.4–12.9	MAB of ULQC (%)	9.2	10.0	9.6	10.1	9.6	10.3	10.3
			Overall MAB[d] (%)	11.2	11.9	11.5	11.4	10.9	11.6	12.0
			No. of failed batches[e]	5	15	8	8	4	11	12
Good precision[b]	LLQC	3.3–9.9	MAB of LLQC (%)	9.0	10.7	9.4	8.3	8.1	8.7	9.4
	QC1	1.2–5.5	MAB of QC1 (%)	4.2	4.3	4.3	4.2	4.0	4.4	4.6
	QC2	1.7–5.6	MAB of QC2 (%)	4.0	4.4	4.2	4.2	4.2	4.5	4.7
	QC3	2.1–5.9	MAB of QC3 (%)	4.5	4.7	4.6	4.4	4.6	4.9	5.1
	ULQC	1.5–5.9	MAB of ULQC (%)	4.5	4.7	4.6	4.3	4.5	4.8	5.1
			Overall MAB (%)	5.6	6.2	5.8	5.3	5.3	5.7	6.1
			No. of failed batches	–	1	–	–	–	–	–
Very good precision[c]	LLQC	2.5–4.8	MAB of LLQC (%)	4.6	4.9	4.8	4.5	4.5	4.7	4.9
	QC1	0.8–3.2	MAB of QC1 (%)	2.6	2.6	2.6	2.6	2.7	2.8	2.8
	QC2	1.0–2.9	MAB of QC2 (%)	2.3	2.3	2.3	2.4	2.4	2.5	2.4
	QC3	1.2–3.1	MAB of QC3 (%)	2.4	2.4	2.4	2.4	2.3	2.5	2.7
	ULQC	1.1–3.2	MAB of ULQC (%)	2.4	2.4	2.3	2.4	2.4	2.5	2.5
			Overall MAB (%)	3.0	3.1	3.0	3.0	3.0	3.1	3.2
			No. of failed batches	–	–	–	–	–	–	–

Notes:
[a] Random errors within ±20% at the lower limit of quantitation (LLOQ) and within ±15% for all others.
[b] Random errors within ±12% at the LLOQ and within ±7% for all others.
[c] Random errors within ±6% at the LLOQ and within ±4% for all other levels.
[d] MAB: mean apparent bias, which is defined as the square root of the mean squared individual bias values ($n = 6 \times 30$ for MAB at individual QC levels and $n = 6 \times 30 \times 5$ for overall MAB).
[e] Out of 30 batches. Failure reasons: ≤2/3 individual QC replicates accepted and/or unacceptable CV/mean bias at one or more QC levels.

3.1.2. Two vs. eight or three concentration levels

As shown in Table 2, when the total number of standard replicates is the same, two-concentration linear calibration (8 + 8) is either equivalent to (very good precision scenario) or better than (poor or good precision scenarios) eight-concentration linear calibration (8 × 2) in terms of overall accuracy and batch failure rate. The improvement in accuracy by two-concentration linear calibration is mainly at low concentration end. For example, in the case of poor precision, the overall mean apparent bias of calibration scheme (8 + 8) is 10.9% while that of calibration scheme (8 × 2) is 11.2%. Out of the same 30 batches of simulated response data, calibration scheme (8 × 2) has 5 failed batches whereas calibration scheme (8 + 8) has 4 failed batches. While the differences in MAB between the two schemes at other QC levels are usually 0.4% or less, the difference at the LLOQ level is 1.1%. Similar conclusions can be made for the comparison between calibration schemes (8 × 1) and (4 + 4). Furthermore, it is interesting to notice that the performance of calibration scheme (2 + 2), which uses as few as only four standard replicates, is either very close to (very good precision scenario) or better than (good and poor precision scenarios) that of calibration scheme (8 × 1), a total of eight standard replicates.

Since the comparison between two- and three-concentration levels was not included in the research work of Penninckx et al. [16], two-concentration linear calibration (8 + 8) was compared with three-concentration calibration (6 + 5 + 5). The results (Table 2) demonstrate that the addition of a third concentration level near the middle of the calibration range at best did not improve either the accuracy or reliability. In fact, for the scenario of poor precision, the two-concentration calibration scheme is evidently better than the three-concentration one, especially in terms of batch failure rate. Specifically, 8 batches failed with three-concentration linear calibration while only 4 failed for two-concentration calibration.

3.1.3. Duplicate- vs. singlet- vs. partially duplicate-standard for eight-concentration calibration

As expected, when a method is very precise (very good precision scenario), the difference between duplicate-standard approach (8 × 2) and singlet-standard approach (8 × 1) is almost negligible, such as the overall mean apparent biases being 3.0% vs. 3.1% (Table 2). In the good precision scenario, the overall difference is also small except at the LLOQ level. For example, the difference at the LLOQ level is 1.7% while those at other levels are 0.4% or less. However, one batch failed for singlet-standard approach. For the poor precision scenario, the difference between the two approaches become more significant, particularly in terms of batch failure, i.e. 5 (two curves) vs. 15 (one curve).

However, by duplicating only the LLOQ and ULOQ standards (8 × 1 + 2), i.e. adding just two more samples, the accuracy is significantly improved, particularly at the low concentration end. The number of failed batches can also be reduced, such as from 15 to 8 for the poor precision scenario. A similar performance to that of the duplicate-standard approach is achieved. Furthermore, duplicating the LLOQ and ULOQ standards is very much desirable for a practical reason as well. When either the LLOQ or ULOQ standard is rejected in the singlet-standard approach such as due to error in sample processing, bad chromatography or unacceptable accuracy, all the samples whose concentrations are below that of CS2 or higher than that of CS7 have to be re-analyzed because calibration

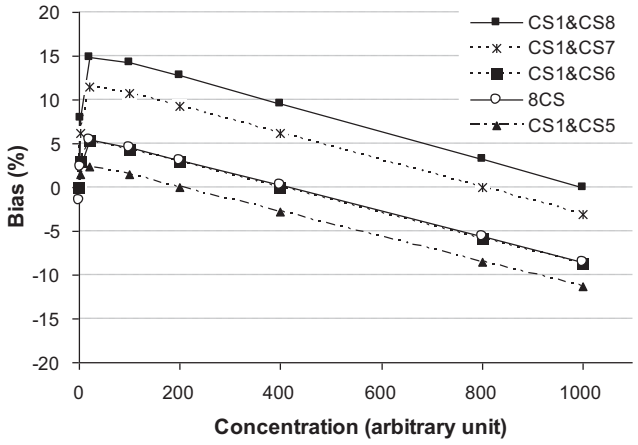

Fig. 1. Bias *vs.* concentration pattern when a given quadratic response ($y = C - 0.0001378C^2$) is regressed linearly using eight and two concentrations. The concentrations of CS1–CS8 are 1, 2, 20, 100, 200, 400, 800, and 1000 (arbitrary unit), respectively.

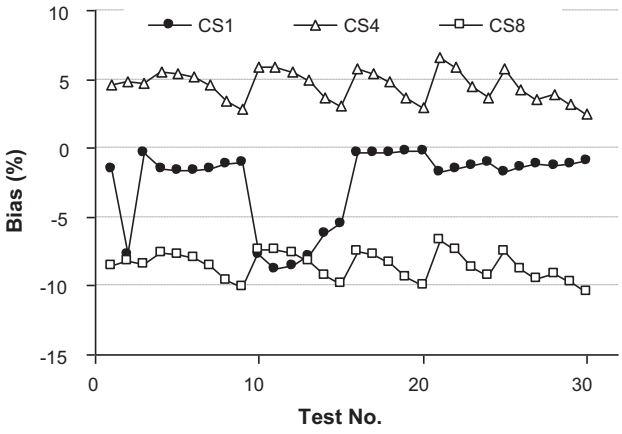

Fig. 2. Demonstration of bias variation at different concentration levels just due to different calibration standards (CSs) excluded from the linear regression of a quadratic response ($y = C - 0.0001378C^2$). Calibration scheme: 8×1 (eight CS concentrations, 1, 2, 20, 100, 200, 400, 800, and 1000 in arbitrary unit, singlet); Test 1: All CS included; Tests 2–9: one CS excluded; Tests 10–30: two non-adjacent CSs excluded.

range is truncated and the extrapolation is not recommended or allowed as per current regulatory guidelines [8,10].

3.1.4. *What happen when quadratic response is regressed as linear (simulation results)*

Non-linearity (quadratic) response often exists in LC–MS bioanalysis because of saturation, formation of dimer/multimer/cluster ions, and cross signal contribution from an analyte to its internal standard, particularly when the concentration range is wide [18,31]. Since linear regression is usually favoured due to its robustness or owing to the preference of regulatory agencies (at least generally perceived as such) [8,31–33], many would try to fit non-linear data linearly. To mimic this scenario, the following simulations were performed.

First, an effort was made to determine the highest degree of curvature that can be linearly fitted with acceptable accuracy (±15%) using two concentrations (LLOQ and ULOQ) for a typical dynamic concentration range of 1000-fold (1–1000). This maximum degree of quadratic can be expressed as: $y = C - 0.0001378 \times C^2$. By using the Q-factor proposed by Liu et al. [34], which was defined as ULOQ × (−*A*/*B*), this maximum degree of quadratic response corresponds to a Q-factor of 0.138.

When a quadratic curve is fitted linearly with two concentrations (LLOQ and ULOQ), the highest positive % bias usually occurs at the low concentration end while the biases at the two calibration concentrations equal zero (Fig. 1). For the exemplary range of 1–1000, the concentration of 32 has the highest positive bias (14.99%). This highest positive bias can be reduced to only 5.4% when eight concentration levels are used to regress the same quadratic response data set, which appears to be a significant improvement in calibration accuracy. However, the overall bias-concentration pattern remains almost the same except for being shifted down (Fig. 1). This kind of "improvement" can be easily matched by adjusting the higher standard concentration used in two-concentration calibration. As demonstrated in Fig. 1, when the higher concentration standard used in two-concentration calibration is changed from CS8 to CS7, CS6, and CS5, the bias-concentration curve progressively moves downwards. The one using CS1 and CS6 for two-concentration calibration has almost identical bias-concentration pattern as that of eight-concentration calibration.

Further simulations have determined the maximum degrees of quadratic response that can be linearly regressed for different concentration ranges and different calibration schemes, including the optimized 2nd CS concentrations and other relevant useful information (Table 3). Still using the range of 1–1000 as an example, the maximum tolerable quadratic response for eight-concentration calibration corresponds to an "*A*/*B*" ratio of −0.0002319. With this degree of non-linearity, the unit responses at the LLOQ and ULOQ are 1 and 0.768, respectively, which is equivalent to 23.2% decrease in unit response from the LLOQ to ULOQ. The highest positive bias and the lowest negative bias are reached at the concentrations of 23 and 1000, respectively. For this reason, the selection of QC concentrations should not be decided by some general criteria, such as middle of a calibration curve. These two worst scenario concentrations should be included to guarantee acceptable accuracy over the entire concentration range. On the other hand, for two-concentration calibration, the maximum tolerable degree of non-linearity with the optimized 2nd CS concentration of 522.2 corresponds to a Q-factor of 0.270, which represents a change of −27.0% in unit response from the LLOQ to the ULOQ. Also shown in the table, the optimized 2nd CS concentrations for different concentration ranges are between 52% and 57%, which agree well with the finding of Peters and Maurer [17] mentioned earlier in the introduction.

Moreover, linear regression of a quadratic data set using eight CS levels, such as 8×1 scheme, is intrinsically unstable. As in reality, there may be a need to exclude one or two calibration standards from a regression due to unaccepted back-calculated concentration(s), e.g. not within (100 ± 15) % of the nominal concentration, sample processing error, or bad chromatography. In these situations, depending on which calibration standard(s) is/are to be excluded, the predicated concentrations of other samples, i.e. quality controls and unknown samples can be variable even though their responses are accurate (Fig. 2), especially for low concentration samples. For example, when CS3 and CS5 are excluded from the regression, the bias for a QC at the concentration of CS4 level (10% of the ULOQ) is 6.63%. While CS6 and CS8 are removed from the regression, the bias of the same QC becomes 2.47% even though there is no error with this QC. In case it is necessary to compare the results from two different runs, such as stability evaluation, cross-validation, and ISR (incurred sample reanalysis), the interpretation of test results could be biased if different calibration standards are excluded from the regression for the two runs. Apparently, this type of variation does not exist in the two-concentration calibration, which can be translated as being more robust.

Table 3
Allowable maximum degree of quadratic response in linear calibration.

Conc. range (ULOQ/LLOQ)	1–1000 (1000-fold)		1–500 (500-fold)		1–250 (250-fold)		1–100 (100-fold)		
No. of conc. levels	2	8	2	8	2	8	2	8	8
Conc. of calibration standards	1, 1000	1, 522.2	1, 500	1, 265.6	1, 250	1, 136	1, 100	1, 56.8	1, 2, 5, 10, 20, 40, 80, 100
		1, 2, 20, 100, 200, 400, 800, 1000		1, 2, 15, 50, 100, 200, 400, 500		1, 2, 10, 25, 50, 100, 200, 250			
% of ULOQ for 2nd CS conc.	100	52.2	100	53.1	100	54.4	100	56.8	N/A
$-A$ ($\times 10^4$)	1.378	2.697	2.821	5.473	5.834	11.17	15.60	29.16	24.72
B	1	1	1	1	1	1	1	1	1
Max degree of quadratic, Q-factor ((ULOQ/LLOQ) × ($-A/B$))	0.138	0.270	0.141	0.274	0.146	0.279	0.156	0.292	0.247
Unit response at LLOQ	1.000	1.000	1.000	0.999	1.000	0.999	0.998	0.997	0.998
Unit response at ULOQ	0.862	0.730	0.859	0.726	0.766	0.721	0.844	0.708	0.753
Max % change in unit response from LLOQ to ULOQ	−13.8	−27.0	−14.1	−27.3	−23.3	−27.8	−15.5	−29.0	−24.5
Conc. corresponding to max positive bias	32	23	22	16	16	12	10	8	8
Conc. corresponding to max negative bias	N/A	1000	N/A	500	500	250	N/A	100	100

Notes: LLOQ, lower limit of quantitation; ULOQ, upper limit of quantitation; Conc., concentration in arbitrary unit; CS, calibration standard; N/A, not applicable; A and B are the constants before concentration square and concentration terms, respectively. Max values were calculated based on the acceptance criteria of ±15% for bias.

Fig. 3. Total bias squares *vs.* number of replicates for linear calibration using only the lower and higher limits of quantitation (based on experimental results of triamcinolone acetonide).

3.2. Comparison of different linear calibration schemes based on experimental results

The following three new bioanalytical methods were chosen for testing mainly because they were on-going projects in the lab when this research was conducted. The comparison was focused on the accuracy of quality controls. The precision was not compared because they were almost identical among different calibration schemes. Only minor difference was observed at LLQC and QC1 levels for the precision of reported concentrations. This is not surprising because the same analyte/IS ratios of the QC samples were used for all calibration schemes.

3.2.1. Triamcinolone acetonide method (good linearity)

As demonstrated in Table 4, the experimental results match those of simulations very well. Specifically, two-concentration calibration is better than eight-concentration and three-concentration calibrations. For example, the overall accuracy (as measured by the sum of bias squares for all QC types) of calibration scheme (7 + 7) is consistently better than that of calibration scheme (8 × 2) and even similar to or slightly better than that of the calibration scheme (8 × 7) for all the four scenarios, i.e. no deliberate error, −10% error in one replicate of CS1, −10% error in one replicate of CS8, and −10% error in one replicate of both CS1 and CS8. The overall accuracy of calibration scheme (7 + 7) is also consistently better than that of calibration scheme (5 + 5 + 5). Moreover, when no deliberate errors are introduced, the accuracy of calibration scheme (2 + 2), which uses only four CS replicates in total, is even slightly better or at least similar to that of calibration schemes (8 × 7) and (8 × 2), which use a total of 56 and 16 CS replicates, respectively. Similar conclusions can be made from the comparison between calibration schemes (8 × 1) and (4 + 4).

The comparison among duplicate-, singlet-, and partially duplicate-standard approaches also confirm the conclusion from the simulations. For instance, singlet calibration standard approach, i.e. (8 × 1), can lead to high % bias at low concentration end when there is an error with CS1, e.g. 10.6% for the LLQC. By duplicating CS1 and CS8, similar accuracy to duplicate CS approach, namely (8 × 2), can be obtained.

When two concentrations are used for calibration, the accuracy can be improved by increasing the number of replicates at each concentration (Fig. 3). However, when the number of replicates reaches four, further improvement by using more replicates is not significant. Hence, five or six replicates for each concentration strike a good balance between adequate accuracy and time- and cost-saving in implementation considering possible sample loss during

Table 4
Comparison of different calibration schemes based on experimental results of tramcinolone acetonide.

	Calibration scheme	8 × 7	8 × 2	8 × 1	8 × 1 + 2	5 + 5 + 5	7 + 7	4 + 4	2 + 2	4 + 4, T7
No error	LLQC bias (%)	−2.9	−2.3	−5.3	−1.9	−1.7	−1.4	−2.0	−1.8	−2.0
	QC1 bias (%)	−4.9	−5.2	−6.2	−5.1	−2.6	−2.9	−2.9	−3.5	−3.7
	QC2 bias (%)	3.0	2.1	2.0	2.1	6.0	5.5	5.8	4.7	4.4
	QC3 bias (%)	−6.2	−7.0	−7.1	−7.0	−3.5	−4.0	−3.6	−4.7	−4.9
	ULQC bias (%)	1.5	0.6	0.5	0.6	4.4	3.9	4.2	3.1	2.9*
	Sum of bias squares	82	86	121	83	77	71	77	69	70
	Slope ×10e4	2.541	2.563	2.565	2.563	2.469	2.482	2.474	2.500	2.507
	Intercept ×10e4	2.16	1.39	2.89	1.19	2.92	2.54	3.03	2.37	2.36
	R	0.9991	0.9995	0.9995	0.9995	0.9994	0.9994	0.9992	0.9993	0.9993
Error in CS1	LLQC bias (%)	−1.5	5.7	10.6	6.9	1.2	0.4	2.7	8.4	2.7
	QC1 bias (%)	−4.4	−2.8	−1.5	−2.4	−1.7	−2.3	−1.3	−0.1	−2.2
	QC2 bias (%)	2.9	1.7	1.2	1.7	6.0	5.5	5.8	4.7	4.4
	QC3 bias (%)	−6.2	−7.4	−7.8	−7.3	−3.5	−4.0	−3.6	−4.7	−4.9
	ULQC bias (%)	1.4	0.2	−0.3	0.2	4.4	3.9	4.2	3.1	2.9*
	Sum of bias squares	72	98	177	111	72	66	74	123	64
	Slope ×10e4	2.542	2.573	2.586	2.573	2.470	2.482	2.474	2.500	2.507
	Intercept ×10e4	1.40	−2.92	−5.72	−3.53	1.48	1.64	0.67	−2.69	0.01
	R	0.9990	0.9991	0.9985	0.9988	0.9989	0.9988	0.9986	0.9987	0.9985
Error in CS8**	LLQC bias (%)	−2.9	−2.4	−5.5	−2.0	−1.7	−1.4	−2.1	−1.8	−2.1
	QC1 bias (%)	−4.8	−4.9	−5.8	−4.5	−2.3	−2.6	−2.1	−1.0	−2.5
	QC2 bias (%)	3.1	2.8	3.0	3.2	6.5	5.8	7.1	8.8	6.4
	QC3 bias (%)	−6.1	−6.4	−6.2	−6.0	−3.0	−3.6	−2.5	−0.9	−3.1
	ULQC bias (%)	1.6	1.2	1.5	1.7	4.9	4.3	5.5	7.2	4.9*
	Sum of bias squares	81	80	113	73	84	74	96	134	85
	Slope ×10e4	2.539	2.547	2.542	2.536	2.457	2.473	2.444	2.406	2.459
	Intercept ×10e4	2.20	1.78	3.47	1.80	3.16	2.72	3.63	4.26	3.32
	R	0.9991	0.9993	0.9990	0.9991	0.9992	0.9992	0.9989	0.9985	0.9986
Error in both CS1 and CS8**	LLQC bias (%)	−1.5	5.6	10.5	6.8	1.3	0.4	2.7	8.7	2.7
	QC1 bias (%)	−4.4	−2.5	−0.9	−1.8	−1.4	−2.0	−0.5	2.5	−0.9
	QC2 bias (%)	3.0	2.4	2.2	2.8	6.5	5.8	7.1	8.8	6.4
	QC3 bias (%)	−6.2	−6.8	−7.0	−6.4	−3.0	−3.6	−2.5	−0.9	−3.1
	ULQC bias (%)	1.5	0.8	0.6	1.3	4.9	4.3	5.5	7.2	4.8*
	Sum of bias squares	71	90	164	100	79	70	95	211	83
	Slope ×10e4	2.541	2.558	2.563	2.546	2.458	2.473	2.444	2.406	2.459
	Intercept ×10e4	1.44	−2.53	−5.14	−2.92	1.72	1.83	1.27	−0.81	0.96
	R	0.9989	0.9989	0.9982	0.9985	0.9988	0.9987	0.9983	0.9988	0.9979

Notes:
The mean analyte/internal standard response ratio at the lower limit of quantitation is 5.18e−3.
The bias values are the mean of six replicates.
* Extrapolated values.
** For the calibration schemes (4 + 4, T7), the error was with CS7, instead of CS8.

sample processing or instrumental issues as well as the need to perform outlier removal. However, this number can be lowered if a bioanalytical method is very precise and robust. For example, when there is no deliberate error in CS response, two replicates are good enough.

3.2.2. Rosuvastatin method (good linearity)

The results of all four calibration schemes met the acceptance criteria (Table 5). Judged by the sum of bias squares, the overall accuracy of calibration scheme (10 × 1 + 2) is slightly better than that of calibration scheme (10 × 1) and the difference between the multi-concentration calibration scheme (10 × 1 + 2) and the two-concentration calibration scheme (2 + 2) is very small for both within- and between-run accuracies. This once again proves that two-concentration calibration is as good as ten-concentration calibration despite using only two replicates for each of the two concentrations in this case.

3.2.3. Valsartan method (quadratic response)

In this method, non-linearity exists due to high analyte concentrations and high detection sensitivity in mass spectrometer. As expected from simulations, high positive bias was observed in the lower middle part of the calibration range when only the LLOQ and ULOQ standards were used due to quadratic response (Table 6). For example, the bias at the concentration of 300 ng/ml is as high as 15.02% for calibration scheme (2 + 2). Since the biases with ten-concentration calibration, i.e. (10 × 2), (10 × 1) and (10 × 1 + 2), are all acceptable, one may conclude that ten-concentration calibration is better than two-concentration calibration in this regard. However, when the bias *vs.* concentration patterns are compared for calibration schemes (10 × 2) and (2 + 2) (Fig. 4), it is clear that ten-concentration calibration only shifted the bias-concentration curve down a little bit in comparison with two-concentration calibration, just as demonstrated earlier in simulation. This kind of improvement can be easily matched by adjusting the concentration of the second CS used in two-concentration calibration. For example, by substituting CS10 with CS7 in the two-concentration calibration, i.e. calibration scheme (2 + 2, T), the aforementioned unacceptable bias (i.e. 15.02%) not only becomes acceptable, but also the overall accuracy matches that of ten-concentration-calibration as demonstrated by the sum of bias squares (Table 6), which corroborates the simulation results presented earlier.

As shown above, extrapolation results are not always unacceptable, particularly for extrapolation at the high concentration end from linear calibration. It should be emphasized that there is a difference between extrapolation into a linearity-unknown concentration segment and that into a linearity-known one (calculated/intended extrapolation). The former one should be avoided because one does not know if the same function can be applied to the unknown concentration segment, such as beyond the ULOQ.

Table 5
Comparison of within-run (intra-run) and between-run (inter-run) accuracy among different calibration schemes for rosuvastatin (experimental results).

QC conc. (ng/ml)	Within-run mean bias[a]				Between-run mean bias[b]			
	10 × 1	10 × 1 + 2	2 + 2	2 + 1, T	10 × 1	10 × 1 + 2	2 + 2	2 + 1, T
0.10 (LLQC)	−9.9	−7.7	−6.7	−6.8	−0.7	−0.4	−0.3	−0.4
0.30 (QC1)	−4.1	−3.3	−3.6	2.6	−0.6	−0.5	−0.8	0.0
4.02 (QC4)	−1.9	−1.8	−2.7	−1.3	–[c]	–	–	–
50.30 (QC2)	−8.7	−8.7	−9.6	−8.2	−1.9	−1.9	−2.4	−1.2
80.47 (QC3)	−0.2	−0.2	−1.2	0.4[d]	0.1	−0.1	−0.4	0.8[d]
100.59 (ULQC)	0.9	0.9	−0.1	1.5[d]	–	–	–	(1.2)[d]
Sum of bias squares	195.0	150.0	158.8	124.3	4.5	4.0	6.7	2.2 (3.7)[e]

Notes:
[a] $n = 6$.
[b] $n = 36$, except for QC1 ($n = 35$ due to the removal of an outlier-contamination during sample processing) and ULQC ($n = 12$).
[c] Not done.
[d] Extrapolated values.
[e] The value in brackets includes the bias square of ULQC.

Table 6
Comparison of within-run (intra-run) accuracy among different calibration schemes for valsartan (experimental results).

QC conc. (ng/ml)	Mean % bias ($n = 6$) for different calibration schemes				
	10 × 2	10 × 1	10 × 1 + 2	2 + 2	2 + 2, T
15	10.2	11.9	9.5	10.0	9.2
30	−5.0	−3.9	−4.9	−2.0	−6.1
150	6.7	7.4	7.8	14.1	5.2
300	7.2	7.8	8.4	15.0 (15.02)	5.7
750	3.1	3.7	4.3	10.9	1.6
1500	0.0	0.6	1.2	7.7	−1.4
3000	3.5	4.1	4.7	11.4	2.0
9000	−7.4	−6.9	−6.3	−0.3	−8.8[a]
13,500	−6.9	−6.3	−5.7	0.3	−8.2[a]
15,000	−8.7	−8.2	−7.6	−1.6	−10.0[a]
Sum of bias squares	426	458	418	839	435

Note:
[a] Extrapolated values.

However, a calculated/intended extrapolation is different because it is already known that the response function from the LLOQ to the ULOQ can be approximated as linear, albeit not perfectly linear. Therefore, the utilization of a concentration less than the ULOQ for the second CS in two-concentration calibration for linear regression is acceptable for the original range, i.e. from LLOQ to ULOQ, particularly when additional ULQC samples are analyzed at the same time to monitor the accuracy.

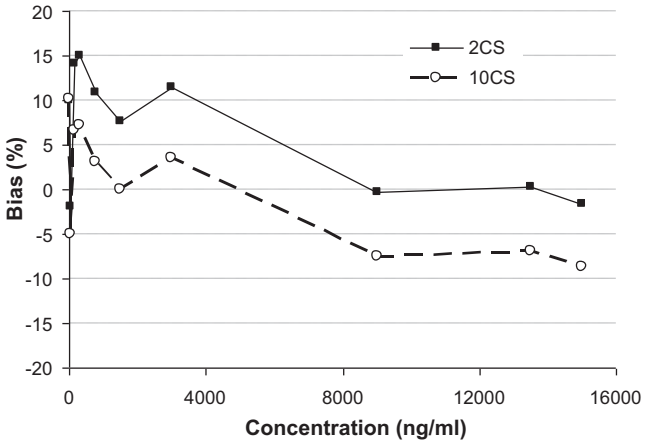

Fig. 4. Comparison of bias *vs.* concentration patterns when a given quadratic data set is regressed linearly using ten and two concentrations (based on experimental results of valsartan). The concentrations of CS1–CS10 are 15 ng/ml, 30 ng/ml, 150 ng/ml, 300 ng/ml, 750 ng/ml, 1500 ng/ml, 3000 ng/ml, 9000 ng/ml, 13,500 ng/ml, and 15,000 ng/ml, respectively. Duplicate standards were used for both two-concentration calibration (2CS) and ten-concentration calibration (10CS).

3.2.4. Retrospective data analysis of multiple validation projects

Six previous validation projects based on the calibration scheme (8 × 2) were selected for retrospective data analysis to compare different calibration schemes, such as (8 × 2), (8 × 1), (8 × 1 + 2), and (2 + 2). As to which CS replicates are to be included in the regression when not all replicates are needed, the same pre-defined selection procedures mentioned earlier in the experimental section for tramcinolone acetonide was followed. Three dynamic concentration ranges (1000-, 500-, and 100-fold) were involved and for each range two different methods were chosen, one method using deuterated internal standard while the other one utilizing structural analogue internal standard (a total of six validated methods), which represent typical different scenarios in LC–MS bioanalysis. Both within- and between-run accuracies were evaluated. Similar findings to those presented above were obtained (data not shown). It is particularly interesting to note that the results with two-concentration calibration (2 + 2) are all acceptable, which could have led to a lot of time and cost saving if it had not been specified in regulatory guidelines to use at least six concentrations for linear calibration. The between- and within-run biases of QC samples are respectively within ±3% and ±5% of absolute difference from the corresponding original values by calibration scheme (8 × 2) except one method, which is quadratic in nature due to a dynamic range of 500-fold and the utilization of structural analogue internal standard. However, by using CS6, instead of CS8 in the two-concentration calibration for this method, the absolute difference from calibration scheme (8 × 2) has also fallen within ±3% and ±5% for between- and within-run accuracy, respectively.

For those readers who are intrigued by the findings presented in this research and want further confirmation, they are encouraged to perform retrospective analysis of previous data. In case previous

calibration was based on singlet standard approach, i.e. singlet standard at each calibration concentration, the within-run accuracy validation run can be used, which has normally five or six replicates of LLQC and ULQC samples. In other words, one can consider combining some of these replicates with the existing singlet standard to satisfy the requirement of multiple measurements at each of the two selected calibration concentrations for two-concentration calibration.

3.3. More about two-concentration calibration

3.3.1. How does one know the response function is really linear if only two concentrations are used in calibration?

Up to this point, many would have been convinced that two-concentration calibration is indeed as good as or even better than multi-concentration calibration for linear calibration. Then, a resounding question/concern might be "how do you know it is linear if only two concentrations are used in linear calibration?" In fact, this one can be readily answered through the following.

Foremost, the linearity over a given concentration range can be checked during method development, such as using multiple concentrations for one or two batches. Secondly, the unit responses (response divided by concentration) at the two selected concentration levels of each batch give a good indication of linearity. For example, for triamcinolone acetonide, the unit response at the ULOQ (0.253 per ng/ml) is almost identical to that at the LLOQ (0.259 per ng/ml), a difference as small as 2.3%, which indicates good linearity. On the other hand, for valsartan, the unit response at the ULOQ (5.94 per μg/ml) is much lower than that at the LLOQ (7.30 per μg/ml), which represents a decrease of 18.6%. Since this difference is greater than the aforementioned maximum degree of quadratic response for two-concentration calibration using the LLOQ and ULOQ, i.e. 13.8% (Table 3), it explains why two-concentration calibration using the LLOQ and ULOQ concentrations was not successful for valsartan (Table 6). However, as it is still smaller than 27.0% (the allowable maximum degree of quadratic response for two-concentration calibration that does not use the ULOQ as the 2nd concentration), acceptable accuracy was still obtained when the concentration of the 2nd CS in two-concentration calibration was adjusted to 3000 ng/ml for valsartan. It should be pointed out that the response in zero standard samples should be deducted prior to the calculation of unit responses if there is minor interference from matrix compounds and/or the internal standard.

Thirdly, according to a recent research by Yuan et al. [31], there exists a thresh-hold value specific to each type of mass spectrometer. As long as the absolute response (peak height) at the ULOQ is below this value, linearity should be good independent of the compounds tested. For the API 4000 triple quadruple mass spectrometer, this value is about 1 $E+6$ cps for peak height. This rule of thumb holds true for the three methods presented in the above. Specifically, the peak heights at the ULOQ for triamcinolone acetonide and rosuvastatin were about 5e5 cps, i.e. less than 1e6 cps, and both have good linearity. On the other hand, the peak height for valsartan ULOQ was around 2e6 cps and quadratic response was observed.

Finally, since at least three different QC samples are analyzed in each batch (either during sample analysis or validation), e.g. low, medium, and high QC, and their concentrations are usually different from those of calibration standards, they can be plotted together with the calibration standards, such as in a LIMS or Excel table, to visually demonstrate the linearity over the range bracketed by the LLOQ and ULOQ or even perform other statistical analysis regarding linearity, if such is desired (note: just for visual demonstration or statistical analysis of linearity, not used in the regression calculation). However, in the authors' opinion, this visual demonstration or further statistical analysis of linearity should be optional because as long as the accuracy of QC samples is satisfactory, the linearity of the calibration curve is indirectly demonstrated or proven. After all, it is always the accuracy of quality controls (which represent unknown samples) that ultimately matters, not the demonstration of linearity.

3.3.2. It is more than just cost- and time-saving!

It is obvious that two-concentration linear calibration saves a lot of time and cost in the preparation and analysis of standard samples. Even when the same total number of calibration standard replicates are analyzed, such as the calibration scheme (4 + 4) vs. the calibration scheme (8 × 1), time and cost are still saved during the preparation of working standard solutions and spiking (preparation of calibration standards). However, using two-concentration calibration is not just a question of cost- and time-saving. It has several other important advantages. First of all, it is more robust. Unlike multiple-concentration linear calibration, there is no intrinsic variation due to the rejection of a concentration because the two concentrations will always be used. In comparison with multiple-concentration calibration, there are fewer calibration standards to be prepared and the prepared volume is usually larger, which can reduce the chances of potential errors in the spiking, such as mix-ups, contamination, and inaccuracy in adding working standard solutions (evidently, it is more accurate to add 200 μl than 20 μl). In addition, as several replicates are analyzed for the same concentration level, statistical procedure can be used to remove potential outliers, which improves not only the accuracy of calibration, but also the robustness. The aforementioned different aspects of robustness can be translated into reduced reassay rate and better reproducibility, e.g. ISR. Furthermore, it is easier to use two-concentration calibration. The responses of standard replicates at each concentration level can either be used individually or averaged; in the latter case no regression algorithm is necessary. There is no need to choose a weighting factor because the commonly used weighting factors (e.g. 1, $1/C$, $1/C^2$) do not make a difference for two-concentration linear calibration. The reason is as follows: even when each response is used individually, all individual responses at the same concentration level share the same nominal concentration, i.e. same weighting factor value, which can then be treated as a common factor during the various summations employed in the regression and can be cancelled out in the calculation of the slope. In other words, for these particular weighting factors (1, $1/C$, and $1/C^2$), regular regression calculation is still equivalent to taking the average of all responses at the same level for the two concentration levels and connecting the two averages by a straight line.

3.3.3. How should two-concentration calibration be best implemented?

As shown in Table 3, there are "tolerable" limits of non-linearity in linear calibration for different scenarios (concentration range and calibration scheme). For example, the maximal allowable decrease of unit response from LLOQ to ULOQ should be within 23.2% for linear calibration in the range of 1 to 1000 using eight calibration standards. Otherwise, satisfactory accuracy cannot be obtained, i.e. bias outside ±15%. Based on the data presented in Table 3 and by leaving some error margin in practice, the linearity of a given data set can be classified as good, fair, and bad if the difference of unit responses between the LLOQ and ULOQ falls in the ranges of <10%, 10–20%, and over 20%, respectively.

If the linearity is good, the LLOQ and ULOQ can be used directly as the two concentrations for calibration standards. Otherwise, measures should be taken to improve the linearity first, such as

obtaining cleaner extracts, use of a better internal standard, adjustment of internal standard concentration, reduction in injection volume, change of ionization mode, shortening of calibration range. If the linearity is still not good but it is within the fair category, the concentration of the 2nd CS can be lowered according to Table 3 to take the advantages of two-concentration calibration while obtaining similar accuracy as multiple-concentration linear calibration. In this case, a ULQC should be used to define the limit of extrapolation at the high concentration end. Of course, to make a unified approach, it is also possible to use this approach for the good linearity scenario (refer to the extrapolated calibration scheme, 4 + 4, T7 and 2+1, T in Tables 4 and 5, respectively). In fact, the results from the truncated two-concentration calibrations are even slightly better than those of the corresponding full range two-concentration calibrations for the good linearity scenarios (triamcinolone acetonide and rosuvastatin). The reason for this might be that they are not perfectly linear though their linearity is good. However, when the linearity is really bad, then quadratic regression should be used [35,36].

Finally, it should be pointed out that the data presented in Table 3 are indicative and they should be appropriately adjusted if the concentration range is not exactly the same as those presented in the table even though the dynamic range (ULOQ/LLOQ) is the same. For example, when the actual concentration range is 5 pg/ml to 5000 pg/ml (1000-fold), the concentration values listed in the table should be multiplied by a factor of 5. On the other hand, the values for the max degree of quadratic should be decreased by a factor of 5, e.g. 0.4638 instead of 2.319 for eight-concentration calibration. However, no adjustment is required for the max % decrease in unit response.

4. Conclusions

Linear calibration using two concentrations has been demonstrated as good as or even better than using multiple concentrations in terms of accuracy. Apart from significant time- and cost-saving, it has several other important advantages over the commonly used multiple-concentration calibration, such as better reliability, independence of the commonly used weighting factors (1, $1/C$, or $1/C^2$), and possibility to detect and remove outliers. Ideally, five or six replicates per concentration should be used in two-concentration calibration. However, acceptable and comparable results have been obtained using as less as two replicates per concentration when a method is very precise. The typical question of "how do you know it is linear if you use only two concentrations?" has also been properly addressed. In addition, it has been demonstrated that extrapolating a linear curve into a linearity-known region at the high concentration end is acceptable. Furthermore, the maximum degree of quadratic response that can be linearly regressed was also determined.

When a method is very precise, the difference between duplicate- and singlet-standard approaches is overall small. However, when the precision of a method is not very good, singlet-standard approach can lead to larger bias and variation at the low concentration end as well as higher batch failure rate. Hence, to improve the accuracy and robustness of singlet CS approach and to maintain its advantage in cost- and time-saving, it is preferable that both the LLOQ and ULOQ be duplicated, which is also good for practical reasons, e.g. reducing reassays due to batch failure or rejection of the LLOQ and/or ULOQ standards.

Despite the focus on LC–MS bioanalysis, in principle, the findings from this research should be applicable to GC–MS or even other linear calibrations in general in analytical and bioanalytical measurements.

Acknowledgements

We would like to thank the following bioanalytical experts for their insightful and challenging questions/comments when the preliminary results of this research were first presented at the 60th ASMS Conference on Mass Spectrometry and Allied Topics (May 20–24, 2012) held in Vancouver, Canada: Drs. Dongmei Zhou (Ardea Biosciences, CA, USA), John Yu (Boehringer Ingelheim, CT, USA), Xinping Fang (Xenobiotic Laboratories, Inc., NJ, USA), Zhongping John Lin (Frontage Laboratories, PA, USA). Moreover, we are grateful to Drs. Long Yuan and Guowen Liu (Bristol-Myers Squibb Co, NJ, USA), Lars B. Nilsson (AstraZeneca R&D, Mölndal, Sweden), Frank T. Peers (Jena University Hospital-Friedrich Schiller University Jena, Jena, Germany), Wenkui Li (Novartis Pharmaceuticals Corporation, NJ, USA), as well as Raymond Naxing Xu (Abbott Laboratories, IL, USA) for their sharing of the relevant research works with us. Finally, Xiaoyue Joyce Tan's help in grammatical reviewing of the manuscript is also greatly appreciated.

References

[1] C.J. Peer, S.D. Spencer, D.A.H. VanDenBerg, M.A. Pacanowski, R.B. Horenstein, W.D. Figg, J. Chromatogr. B 880 (2012) 132.
[2] W. Li, S. Luo, H.T. Smith, F.L.S. Tse, J. Chromatogr. B 878 (2010) 583.
[3] A. Tan, S. Hussain, F. Vallée, LC–GC North America 27 (2009) 414.
[4] A. Tan, P. Hang, J. Couture, S. Hussain, F. Vallée, J. Chromatogr. B 856 (2007) 9.
[5] M. Chen, C. Granvil, Q.C. Ji, Z.Y. Zhang, M.V. Padval, V.V. Kansra, J. Pharm. Biomed. Anal. 49 (2009) 1241.
[6] H. Wu, J. Zhang, K. Norem, T.A. El-Shourbagy, J. Pharm. Biomed. Anal. 48 (2008) 1243.
[7] R.N. Xu, L. Fan, G.E. Kim, T.A. El-Shourbagy, J. Pharm. Biomed. Anal. 40 (2006) 728.
[8] US FDA, Guidance for Industry: Bioanalytical Method Validation, 2001.
[9] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, M.J. Rose, J. Sailstad, V.P. Shah, J.P. Skelly, P.G. Swann, R. Weiner, AAPS J. 9 (2007) E30.
[10] European Medicines Agency. Guideline on Bioanalytical Method Validation, 2011.
[11] ANVISA (Brazil), Guide for Validation of Analytical and Bioanalytical Methods, 2003.
[12] L.B. Nilsson, G. Eklund, J. Pharm. Biomed. Anal. 43 (2007) 1094.
[13] L. Renman, D. Jagner, Anal. Chim. Acta 357 (1997) 157.
[14] B. Boulanger, P. Chiap, W. Dewe, J. Crommen, Ph. Hubert, J. Pharm. Biomed. Anal. 32 (2003) 753.
[15] E. Rozet, R.D. Marini, E. Ziemons, B. Boulanger, Ph. Hubert, J. Pharm. Biomed. Anal. 55 (2011) 848.
[16] W. Penninckx, C. Hartmann, D.L. Massart, J. Smeyers-Verbeke, J. Anal. At. Spectrom. 11 (1996) 237–246.
[17] F.T. Peters, H.H. Maurer, Anal. Chem. 79 (2007) 4967.
[18] A. Tan, I.A. Lévesque, I.M. Lévesque, F. Viel, N. Boudreau, A. Lévesque, J. Chromatogr. B 879 (2011) 1954.
[19] C.L. Bowen, J. Kehler, C.A. Evans, J. Chromatogr. B 878 (2010) 3125.
[20] A.Q. Wang, B.P. Wei, Y. Zhang, Y.J. Wang, L. Xu, K. Lan, J. Chromatogr. B 879 (2011) 2259.
[21] Global CRO Council (GCC), Bioanalysis 3 (2011) 2721.
[22] P. Sripalakit, B. Kongthong, A. Saraphanchotiwitthaya, J. Chromatogr. B 869 (2008) 38.
[23] P. van Amsterdam, M. Arnold, S. Bansal, D. Fast, F. Garofolo, S. Lowes, P. Timmerman, E. Woolf, Bioanalysis 2 (2010) 1801.
[24] L.D. Svensson, Bioanalysis 3 (2011) 2685.
[25] R. Bakhtiar, T.K. Majumdar, J. Pharmacol. Toxicol. Methods 55 (2007) 227.
[26] A. Bergeron, M. Furtado, F. Garofolo, Rapid Commun. Mass Spectrom. 23 (2009) 1287.
[27] A.M. Almeida, M.M. Castel-Branco, A.C. Falcão, J. Chromatogr. B 774 (2002) 215.
[28] P.J. Taylor, S.R. Brown, D.P. Cooper, P. Salm, M.R. Morris, P.I. Pillans, S.V. Lynch, Clin. Chem. 51 (2005) 1890.
[29] A. Tan, W. Jin, F. Deng, S. Hussain, A. Musuku, R. Massé, J. Chromatogr. B 877 (2009) 3673.
[30] S. Hussain, H. Patel, A. Tan, Bioanalysis 1 (2009) 529.
[31] L. Yuan, D. Zhang, M. Jemal, A.F. Aubry, Rapid Commun. Mass Spectrom. 26 (2012) 1465.
[32] T. Singtoroj, J. Tarning, A. Annerberg, M. Ashton, Y. Bergqvist, N.J. White, N. Lindegardh, N.P.J. Day, J. Pharm. Biomed. Anal. 41 (2006) 219.
[33] L.B. Nilsson, P. Skansen, Rapid Commun. Mass Spectrom. 26 (2012) 1399.
[34] G. Liu, Q.C. Ji, M.E. Arnold, Anal. Chem. 82 (2010) 9671.
[35] B. Chou, R.S. Adler, M. Meng, S. Percey, B. Dean, C.E.C.A. Hop, Y.G. Shin, J. Pharm. Biomed. Anal. 70 (2012) 354.
[36] J.M. Cunliffe, C.F. Noren, R.N. Hayes, R.P. Clement, J.X. Shen, J. Pharm. Biomed. Anal. 50 (2009) 46.