Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Pharmaceutical and Biomedical Analysis 41 (2006) 219–227
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

A new approach to evaluate regression models during validation of bioanalytical assays

T. Singtoroj [a,1], J. Tarning [b,1], A. Annerberg [a], M. Ashton [b], Y. Bergqvist [c], N.J. White [a,d], N. Lindegardh [a,d,*], N.P.J. Day [a,d]

[a] *Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand*
[b] *Department of Pharmacology, Sahlgrenska Academy at Göteborg University, Sweden*
[c] *Dalarna University College, Borlänge, Sweden*
[d] *Nuffield Department of Clinical Medicine, Centre for Tropical Medicine, University of Oxford, Oxford, UK*

Received 6 August 2005; received in revised form 4 November 2005; accepted 6 November 2005
Available online 5 December 2005

Abstract

The quality of bioanalytical data is highly dependent on using an appropriate regression model for calibration curves. Non-weighted linear regression has traditionally been used but is not necessarily the optimal model. Bioanalytical assays generally benefit from using either data transformation and/or weighting since variance normally increases with concentration. A data set with calibrators ranging from 9 to 10 000 ng/mL was used to compare a new approach with the traditional approach for selecting an optimal regression model. The new approach used a combination of relative residuals at each calibration level together with precision and accuracy of independent quality control samples over 4 days to select and justify the best regression model. The results showed that log–log transformation without weighting was the simplest model to fit the calibration data and ensure good predictability for this data set.

Keywords: Bioanalytical assays; Calibration; Transformation; Weighting; Piperaquine; Linear regression; Liquid chromatography

1. Introduction

In recent years great efforts has been made to standardize international validation procedures for bioanalytical assays. Different European and American authorities such as the FDA, ICH and ISO continuously develop validation guidelines and directives about experimental design and data evaluation in the field of bioanalytical method validation [1–4]. A first attempt at harmonization and standardisation was the conference held in Washington in 1990 to discuss what a validation of bioanalytical methods should consist of, i.e. which analytical parameters (bias, precision, etc.) need to be documented to validate a method. The resulting Washington Conference Report and publications related to the conference are generally viewed as the basis for bioanalytical method validation [5,6]. However, the usefulness of some of the recommendations is questionable, particularly given the lack of advice for the practical execution of a validation study. In the light of this critique, a new SFSTP (Société Francaise des Sciences et Techniques Pharmaceutiques) committee was founded in 1995 to develop guidance for validation of bioanalytical methods. The SFSTP validation guide of chromatographic methods for drug bioanalysis was published in 1999 by Hubert et al. and illustrated the same year by Chiap et al. [7,8]. The guide has recently been updated by the introduction of the concept of an accuracy profile [9]. The accuracy profile utilises a "*β-expectation tolerance interval*" to visually discriminate between acceptable and non-acceptable regression models during pre-validation. The "*β-expectation tolerance interval*" is constructed using estimates of the bias and the standard deviation of the intermediate precision obtained from validation standards or back calculated concentrations of calibration standards analysed in replicate series [10–15]. The concept and content of the two validation phases (i.e. pre-validation and validation) is substantially covered in the literature [10–12,15–18]. Boulanger et al. state that: "During the 'pre-validation', the model to be used

* Corresponding author at: Wellcome Unit, Faculty of Tropical Medicine, Mahidol University, 420/6 Rajvithi Road, Bangkok 10400, Thailand.
E-mail address: niklas@tropmedres.ac (N. Lindegardh).
[1] Authors have contributed equally to this work.

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.11.006

as calibration curve will be identified and the quality of fit will be assessed only at this stage. The experiments proposed are designed to consistently evaluate the adequacy of the model. In the second phase, called 'validation', the objective is to mimic the routine practice that is envisaged. The model will be used as is – the parameters will of course be estimated based on the new data – and no more investigation specific to the quality of fit will be conducted, the same way it should be carried out during routine. In this second step, the experiments are designed to focus on the estimation of the bias and precision of the method, not on the calibration curve." [18]. The present paper suggests a new approach to choose an optimal regression model. Instead of fixate the regression model during pre-validation the final choice should be based on all available data from the validation phase.

A good regression model is the foundation for accurate and reproducible quantification over the whole calibration range. A linear model is commonly preferred since the complexity increases with the use of non-linear regression. FDA guidelines state that: "Standard curve fitting is determined by applying the simplest model that adequately describes the concentration–response relationship using appropriate weighting and statistical tests for goodness of fit" [1]. These requirements sound very clear and straightforward. However, complying with the stipulations might in reality not be so simple. The simplest and most commonly used parameter to define the degree of association between two variables as a straight line is denoted by the coefficient of determination (r^2). Many analysts depend entirely on the value of r^2 being greater than 0.99 as an acceptance criterion when evaluating regression model and linearity. However, r^2 alone is not adequate to demonstrate linearity since r^2 values above 0.999 can be achieved even when the data show signs of curvature [19].

The most common approach to fit a calibration curve to data points (x, y) is by ordinary linear regression (OLR) using least squares calculation. This approach presupposes that each data point in the range has a constant absolute variance (i.e. homoscedastic data). Most bioanalytical assays usually have to cover a broad concentration range and the variance is more likely to increase with concentration (i.e. heteroscedastic data) [19–24]. A consequence of using OLR is that deviations at high concentrations will influence the regression line more than deviations at low concentrations. Thus the use of OLR with heteroscedastic data will lead to impaired accuracy despite an acceptable r^2 value, particularly at the lower end of the concentration range [23].

All bioanalytical assays could benefit from a regression model more complex than OLR. Alternative models include weighted linear regression (WLR) and/or data transformations [20–25]. These models will normally generate a better curve fit (i.e. smaller sum of residuals and random scatter in residual plots) than OLR. They will also minimise time-dependent variation (i.e. minimise variation in slope and intercept for standard curves obtained over several days) and increase accuracy over the whole concentration range.

Traditionally the regression model is chosen in the pre-validation phase by evaluating 3–5 series of calibration curves and comparing the total sum of residuals for each tested regression model [17,24,26,27]. Some reports have also incorporated predictability by looking at the accuracy of independent quality control (QC) samples before choosing the final model [25].

We propose a strategy that will enable the analyst to choose the regression model that gives the optimal overall performance over time. This approach is based on parameter ranking of data generated during several days (4 days in the present paper) to mimic the actual conditions during routine bioanalysis instead of only one day of pre-validation data. The curve fit was evaluated by minimising the residuals at each calibration level rather than just the total sum of residuals. Accuracy and precision were also incorporated for three independent QC levels during several days of analysis before the final regression model was chosen. Nineteen different regression models were evaluated using data obtained during the validation of a liquid chromatographic assay for piperaquine (PQ) quantification in urine using a 1000-fold concentration range (9–10 000 ng/mL) [28].

2. Experimental

2.1. Background

2.1.1. Homoscedasticity

The first step during an evaluation of regression models should include a test for homoscedasticity. The two most common ways to evaluate homoscedasticity are to conduct an F-test (i.e. test for significant difference in variance) or to visually examine a residual versus concentration plot [24]. If the variance is constant (i.e. homoscedastic data) over the calibration range the residual versus concentration plot should show residuals randomly distributed around the x-axis [21]. In the F-test the experimental F-value (F_{exp}) is expressed as the ratio between the variance at the lowest and at the highest concentration in the calibration range as proposed by the International Organization for Standardization [29]. If the F_{exp} value is greater than the tabulated F-value (F_{tab}) at a chosen confidence level the variances are significantly different (i.e. the data are heteroscedastic) [24,30–33].

2.1.2. Ordinary and weighted linear regression (OLR/WLR)

OLR assumes homoscedasticity and associates the dependent variable y with the independent variable x. The regression line is constructed so as to minimise the squared sum of the vertical distance (sum of squared residuals, SSR) between the observations and the constructed regression line [24]. One method of dealing with heteroscedastic data is to apply a weighted regression model. The principle of weighting is to give more importance to data points with a low variance and less importance to data points with high variance. Weighted models are particularly suitable for assays where the relative standard deviation (R.S.D.) is constant (i.e. S.D. increases proportional to concentration) throughout the concentration range. An optimal weighted model will balance the regression line to generate an evenly distributed error throughout the calibration range. The most commonly used weights are the empirical weights $1/x$, $1/x^2$ and $1/x^{1/2}$.

2.1.3. Polynomial regression

Most bioanalytical data sets are heteroscedastic as the S.D. increases as concentration (x) increases. Applying a weighted regression line can often circumvent this problem. Thus, if the weighting is successful the relative residual versus concentration plot should generate a scatter of points randomly distributed around the x-axis (i.e. around zero). If the average relative residual (ΔRR) tends to increase or decrease with increasing x the data set is non-linear or curved. The simplest method applicable to non-linear data is the quadratic form of polynomial regression. The equation for quadratic regression is $y = a + bx + cx^2$, where y is dependent both on the x variable and its square (x^2). Polynomial regression, in particular with a higher order than quadratic, should be used with caution as most bioanalytical techniques should produce linear data. A curved data set can be detector-related (e.g. GC-ECD or LC-ESD), but can also be an indication of problems with the bioanalytical assay such as adsorption, solubility or ion suppression.

2.1.4. Transformation methods

An alternative approach to overcome heteroscedatic data is to transform x and/or y before constructing the regression line. Two common approaches are logarithmic or square root transformation of both x and y before OLR [7,8,22,23]. A different approach is the power model originally proposed by Box and Cox in 1964 and illustrated by Kimanani et al. in 1998 [34–36]. The power model transforms the data using an optimal power value (λ_{opt}) which minimises the sum of squared residuals (SSR). A given set of concentrations (x) and responses (y) are transformed by an assigned power value (λ) as expressed in Eqs. (6) and (7), and evaluated by SSR as expressed in Eq. (8). The optimal power value is determined empirically through iteration towards a minimum value of SSR. OLR is thereafter applied on the transformed data to construct the regression line. All equations are as follows:

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda-1}}, \quad \text{if } \lambda \neq 0 \tag{6}$$

$$y^{(\lambda)} = \dot{y} \ln y, \quad \text{if } \lambda = 0 \tag{7}$$

$$\mathrm{SSR}(\lambda) = \sum (y_i{}^{(\lambda)} - a - b x_i^{(\lambda)})^2 \tag{8}$$

$$y^{(\lambda)} = a + b x^{(\lambda)} + \varepsilon \tag{9}$$

where λ is the power value, $\dot{y}$ the geometric mean of the responses, $y^{(\lambda)}$ the transformed response, $x^{(\lambda)}$ the transformed concentration and ε is the uncorrelated random error.

2.2. Methods

2.2.1. Sample preparation and chromatographic conditions

Data from the validation of a sensitive and specific bioanalytical method for determination of PQ in urine by automated solid-phase extraction (SPE) and liquid chromatography (LC) was used in this study [28]. Briefly, buffered urine samples (containing internal standard) were loaded onto mixed phase (cation-exchange and octylsilica) SPE columns using an ASPEC XL SPE robot. Chromatographic separation was achieved on a Chromolith Performance RP-18e (100 mm × 4.6 mm ID) LC column with phosphate buffer (pH 2.5; 0.1 mol/L)–acetonitrile (92:8, v/v). PQ was analysed at a flow rate of 3 mL/min with UV detection at 347 nm.

2.2.2. Data analysis

The peak height ratio of PQ to internal standard (IS) was used as response (y) for a concentration range (x) from 9 to 10 000 ng/mL using six calibration levels. The standard curve was assayed in five replicates on day 1 and then in single determination for four consecutive days. Standard curves were constructed using 19 different regression models. The unit of the calibrators (μg/mL and ng/mL) was also evaluated to investigate if it would alter the results for each model. A strategy diagram of the different regression models is presented in Fig. 1. Back calculated concentrations (x_{new}) for the different regression models were calculated using Eqs. (10)–(13).

Back calculated concentration for linear regression (Eq. (10)):

$$x_{new} = \frac{y - a}{b} \tag{10}$$

where y is the PQ/IS height response, a the y-intercept and b is the slope.

Back calculated concentration for quadratic regression (Eq. (11)):

$$x_{new} = \frac{-b + \sqrt{b^2 - 4a(c - y)}}{2a} \tag{11}$$

Back calculated concentration for Box–Cox power transformation (Eqs. (12) and (13)):

$$x_{new} = \sqrt[\lambda_{opt}]{(\lambda_{opt} x^{(\lambda_{opt})} \dot{x}^{\lambda_{opt}-1}) + 1}, \quad \text{if } \lambda_{opt} \neq 0 \tag{12}$$

$$x_{new} = e^{x^{\lambda_{opt}}/\dot{x}}, \quad \text{if } \lambda_{opt} = 0 \tag{13}$$

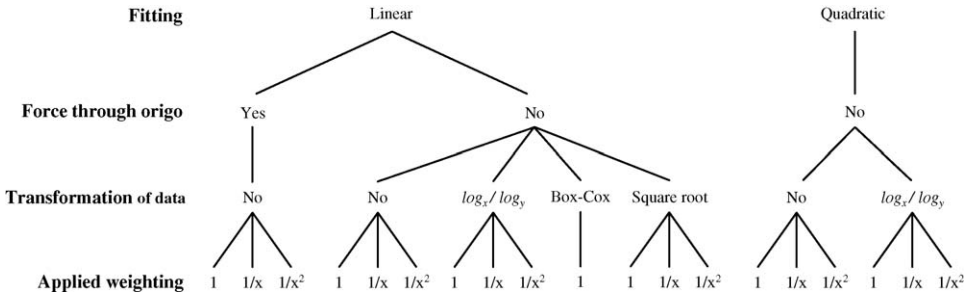

Fig. 1. Strategy diagram of the 19 evaluated regression models.

where λ_{opt} is the optimal power value and $\dot{x}$ is the geometric mean of concentration values.

The relative residuals (RR) were calculated based on back calculated concentration (x_{new}) and nominal concentration ($x_{nominal}$) according to the following equation:

$$RR = 100 \times \frac{x_{new} - x_{nominal}}{x_{nominal}} \quad (14)$$

Three quality control (QC) levels (50, 500, 5000 ng/mL) were prepared to validate accuracy and precision (i.e. predictability). The accuracy and precision of the method were estimated by analysis of five replicates of each QC level for 4 days. Concentrations were predicted for each regression model using a calibration curve prepared on the same day as the QC samples. Average accuracy and total-assay precision for all replicates at each level were used in the evaluation. All regression models were fitted to data using model options in LaChrom Elite software (VWR International, Darmstadt, Germany) or by manual calculation in Windows Excel® (Microsoft Corporation).

2.2.3. Data evaluation

Two approaches to evaluate regression models were compared in this paper. The first approach is based on the evaluation of five series of calibration curves analysed during one analytical run (i.e. during pre-validation). This is the traditional approach where the total sum of relative residuals (SRR) for all the calibration standards guides the choice of regression model [24]. The second is a new approach that uses both calibration curve fit and calibration curve predictability to evaluate the regression model. This approach uses data generated during 4 days instead of a single day to ensure that the evaluation process incorporates time-dependent reproducibility. Average relative residuals (ΔRR) at each calibration level served as markers for calibration curve fit. Accuracy and precision for the three independent QC levels served as markers for predictability. All the regression models investigated were initially ranked for calibration curve fit and predictability separately. The rank sum of these two markers generated a final ranking for the regression models (Table 4). The regression model with lowest rank sum represents the optimal model amongst those evaluated. The aim was to find the simplest regression model that had the best overall characteristics over the whole calibration range.

3. Results

The initial *F*-test confirmed that the data set was heteroscedastic as expected. The variance at the highest (10 000 ng/mL) calibrator level was significantly higher than the variance at the lowest (9 ng/mL) calibrator level.

3.1. Traditional approach using SRR (i.e. pre-validation)

The five replicates of calibrators were fitted to the regression models individually and as a mean of the replicates. The total SRR for the evaluated regression models are summarised in Table 1. It is obvious from the results that the simplest model OLR generates much higher residuals than all the other models. It is not surprising considering that the data was already shown to be heteroscedastic. As expected, back calculated concentrations from the individual curves presents lower SRR than when the mean curve is used to predict the replicates. However, the relative difference between the 19 models remained constant suggesting that either individual or mean fitting can be used for model evaluation. An additional requirement is that the models have to meet the acceptance criteria set up by the FDA [1]. At least four out of six data points should have accuracy and precision <15%. All replicates at the LLOQ and the highest concentration should have accuracy and precision <20 and <15%,

Table 1
Total sum of relative residuals for evaluated regression models fitted to the regression models individually and as mean of replicates

Fitting	Regression model			Sum of relative residuals (SRR)		Criteria meet
	Transformation	Weight	Forced through origo	Individual curve (%)	Mean curve (%)	
Linear	No	No	No	2788	2801	N
Linear	No	$1/x$	No	419	423	N
Linear	No	$1/x^2$	No	268	277	N
Linear	No	No	Yes	540	545	N
Linear	No	$1/x$	Yes	518	522	N
Linear	No	$1/x^2$	Yes	481	481	N
Linear	log–log	No	No	107	114	Y
Linear	log–log	$1/x$	No	108	117	Y
Linear	log–log	$1/x^2$	No	108	119	Y
Linear	Box–Cox	No	No	86	101	Y
Linear	Square root	No	No	313	320	N
Linear	Square root	$1/x$	No	200	206	N
Linear	Square root	$1/x^2$	No	181	189	Y
Quadratic	No	No	No	887	909	N
Quadratic	No	$1/x$	No	271	279	N
Quadratic	No	$1/x^2$	No	188	199	Y
Quadratic	log–log	No	No	68	94	Y
Quadratic	log–log	$1/x$	No	71	98	Y
Quadratic	log–log	$1/x^2$	No	75	104	Y

respectively. Nine out of the 19 tested models met this requirement and could theoretically be chosen. The calibration curve fit and the impaired accuracy at the lower concentration range for some of the models are illustrated in Fig. 2. FDA guidelines state that the simplest model that adequately describes data should be chosen and that weighting and transformation needs justification [1]. The linear and quadratic models with log–log transformation and the Box–Cox transformation model all showed comparable

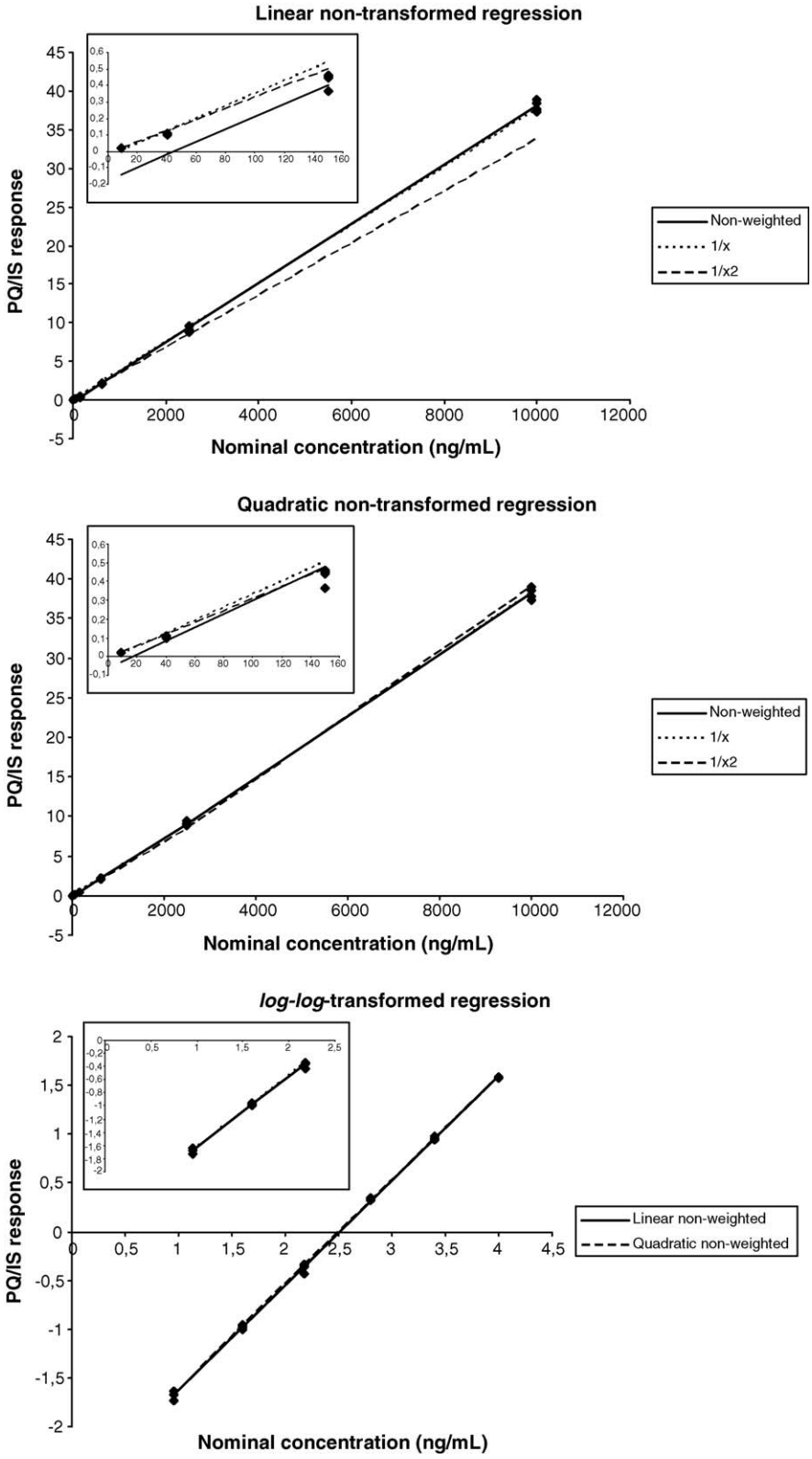

Fig. 2. Selected regression models fitted to mean of five replicates at each calibration level.

SRR. Based only upon the SRR results in Table 1, a quadratic log–log transformed calibration model would be the model of choice. However, whether the presented data alone are powerful enough to justify that choice when using this approach is a valid point for discussion. The *F*-test has already shown the data to be heteroscedastic, thus justifying a more complicated model than OLR. The main risk when choosing a complicated model is over-fitting which could lead to an impaired predictive capacity. The challenge of choosing between the best models based on a good compromise between performance (i.e. SRR) and simplicity still remains.

3.2. New approach using parameter rank

The relative residuals for each calibration level over 4 days were used to rank the models according to calibration curve fit (Table 2). The accuracy and precision for three independent QC levels were used to rank the models according to calibration curve predictability (Table 3). Calibration standards were best described by a log–log transformed linear regression, whereas precision and accuracy for the QC samples were best described by the square root transformed $1/x$ weighted linear regression. The final ranking of the models was based on a combination of calibration curve fit and calibration curve predictability (Table 4). The best model with respect to overall characteristics was the linear log–log transformed model. This choice is based on 12 different parameters covering the whole calibration range. Furthermore the data were generated over four different days thus reflecting the actual performance during routine use of the assay. The requirement in the FDA guidelines to justify any other model than OLR originates from the concern of over-fitting the calibration curve when using a more complex model. The proposed approach will minimise the likelihood of over-fitting the calibration curve as it incorporates curve independent QC samples. The linear log–log models with weighting are ranked second best with respect to the calibration curve fit. However, they are only ranked in 9th and 11th place with respect to predictability of the QC samples. The square root transformation with $1/x$ weighting and Box–Cox transformation models are ranked as the two best models with respect to predictability of the QC samples but only come in at 7th and 13th place when ranked for calibration curve fit (Table 4). The unit of the calibrators (μg/mL or ng/mL) did affect the results for the log–log transformed models when weighting was used but not for any of the other models.

4. Discussion

Data transformations (e.g. log–log) are a good way to improve calibration curve fit and ensure a robust calibration model. The reason for the improvement compared to OLR is simply that the calibration range is significantly reduced after transformation. The square root transformation might be a better choice if a more narrow calibration range is used considering that the reduction is less drastic than for the log–log transformation. The use of $\log x$ and $\log y$ decreases the distance between the lowest and highest point in the calibration range, thus leading to a more compressed regression line. This stabilises the variance over the concentration range and equalises the influence of each point on the regression line [33]. The Box–Cox transformation shows good predictability for the QC samples but a poor fit for the calibration standards. This is likely to be

Table 2
Total relative error for the calibration curve (range 9–10 000 ng/mL)

Fitting	Regression model			Mean relative residuals over all four validation days												Rank sum
	Transformation	Weight	Forced through origo	9 ng/mL		40 ng/mL		150 ng/mL		625 ng/mL		2500 ng/mL		10 000 ng/mL		
				ΔRR (%)	Rank	ΔRR (%)	Rank	ΔRR (%)	Rank	ΔRR (%)	Rank	ΔRR (%)	Rank	ΔRR (%)	Rank	
Linear	No	No	No	281.9	19	62.6	19	15.6	19	7.4	16	4.4	7	0.3	3	83
Linear	No	$1/x$	No	33.4	16	14.9	12	11.2	16	9.0	17	4.3	5	1.2	6	72
Linear	No	$1/x^2$	No	4.1	4	16.8	14	5.7	13	3.5	3	10.2	17	9.6	18	69
Linear	No	No	Yes	21.6	14	27.4	17	14.8	18	10.1	19	5.2	11	0.3	4	83
Linear	No	$1/x$	Yes	21.1	13	27.1	16	14.3	17	9.6	18	4.3	6	1.3	7	77
Linear	No	$1/x^2$	Yes	10.7	10	17.7	15	3.4	6	4.5	7	14.7	19	14.2	19	76
Linear	log–log	No	No	5.8	8	9.3	5	2.9	3	3.7	4	4.8	8	2.4	10	38
Linear	log–log	$1/x$	No	4.2	5	10.0	6	2.7	1	3.8	5	5.7	14	2.6	11	42
Linear	log–log	$1/x^2$	No	4.2	5	10.0	6	2.7	1	3.8	5	5.7	14	2.6	11	42
Linear	Box–Cox	No	No	11.2	11	12.3	9	4.8	10	4.6	9	5.7	13	2.8	13	65
Linear	Square root	No	No	37.6	17	3.9	1	4.4	9	5.2	11	3.9	4	0.9	5	47
Linear	Square root	$1/x$	No	19.3	12	10.3	8	5.5	12	5.7	12	3.4	3	1.9	8	55
Linear	Square root	$1/x^2$	No	7.4	9	13.0	10	4.4	8	3.2	1	7.5	16	6.1	16	60
Quadratic	No	No	No	235.1	18	29.2	18	3.1	4	7.3	15	0.6	1	0.01	1	57
Quadratic	No	$1/x$	No	28.2	15	14.9	13	9.4	15	6.8	14	3.0	2	0.2	2	61
Quadratic	No	$1/x^2$	No	3.1	3	13.8	11	3.2	5	4.8	10	10.7	18	2.3	9	56
Quadratic	log–log	No	No	4.6	7	9.2	4	3.8	7	3.5	2	5.0	9	3.3	14	43
Quadratic	log–log	$1/x$	No	2.5	2	8.9	3	5.3	11	4.5	8	5.4	12	5.1	15	51
Quadratic	log–log	$1/x^2$	No	1.3	1	8.1	2	7.2	14	5.9	13	5.1	10	7.6	17	57

ΔRR: mean relative residuals.

Table 3
Predictability of QC samples (range 50–5000 ng/mL)

Fitting	Regression model			Accuracy and precision of QC samples over all four validation days												Rank sum
	Transformation	Weight	Forced through origo	50 ng/mL				500 ng/mL				5000 ng/mL				
				Accuracy (%)	Rank	Precision (%)	Rank	Accuracy (%)	Rank	Precision (%)	Rank	Accuracy (%)	Rank	Precision (%)	Rank	
Linear	No	No	No	16.0	16	49.8	19	−1.0	4	5.8	15	9.7	6	4.6	17	77
Linear	No	$1/x$	No	−10.1	15	7.4	1	−3.1	6	5.6	10	10.1	8	4.0	14	54
Linear	No	$1/x^2$	No	−9.5	14	8.2	5	4.9	11	5.3	6	19.9	18	2.9	4	58
Linear	No	No	Yes	−20.2	18	9.5	16	−4.6	10	6.5	19	9.3	4	5.1	19	86
Linear	No	$1/x$	Yes	−19.8	17	8.9	9	−3.9	7	5.8	14	10.2	11	4.0	13	71
Linear	No	$1/x^2$	Yes	−9.1	12	8.6	8	8.9	16	5.5	9	24.9	19	3.2	7	71
Linear	log–log	No	No	−0.3	2	9.2	11	7.0	12	5.2	3	10.1	9	3.6	11	48
Linear	log–log	$1/x$	No	−1.0	4	9.3	12	7.3	13	5.3	5	11.5	14	3.3	9	57
Linear	log–log	$1/x^2$	No	−1.4	6	9.3	13	8.5	15	5.5	7	14.4	16	2.8	1	58
Linear	Box–Cox	No	No	−1.0	3	8.2	4	4.6	9	5.8	13	10.2	10	2.9	3	42
Linear	Square root	No	No	1.7	7	9.4	14	1.4	5	5.1	1	10.0	7	4.0	15	49
Linear	Square root	$1/x$	No	−4.1	9	8.1	3	0.7	1	5.2	2	11.0	13	3.5	10	38
Linear	Square root	$1/x^2$	No	−5.9	11	8.4	6	4.0	8	5.3	4	16.4	17	3.0	5	51
Quadratic	No	No	No	25.2	19	10.2	18	−0.9	3	5.7	12	8.9	3	2.9	2	57
Quadratic	No	$1/x$	No	−9.3	13	7.7	2	−0.8	2	6.0	18	10.7	12	3.2	8	55
Quadratic	No	$1/x^2$	No	−5.6	10	8.5	7	9.6	17	5.9	16	14.2	15	3.2	6	71
Quadratic	log–log	No	No	0.2	1	9.2	10	8.1	14	5.5	8	9.4	5	3.8	12	50
Quadratic	log–log	$1/x$	No	1.1	5	9.5	15	9.9	18	5.7	11	8.5	2	4.1	16	67
Quadratic	log–log	$1/x^2$	No	2.4	8	9.9	17	11.6	19	5.9	17	7.0	1	4.9	18	80

Accuracy = (mean predicted concentration/nominal value) − 1; precision = mean relative standard deviation.

Table 4
Final ranking for the regression models (range 9–10 000 ng/mL)

Regression model				Ranking of calibration standards and QC samples			
Fitting	Transformation	Weight	Forced through origo	Ranking of calibration standards	Ranking of QC samples	Rank sum	Final ranking
Linear	No	No	No	18	17	35	18
Linear	No	$1/x$	No	15	7	22	12
Linear	No	$1/x^2$	No	14	11	25	14
Linear	No	No	Yes	18	19	37	19
Linear	No	$1/x$	Yes	17	14	31	17
Linear	No	$1/x^2$	Yes	16	14	30	16
Linear	log–log	No	No	1	3	4	1
Linear	log–log	$1/x$	No	2	9	11	5
Linear	log–log	$1/x^2$	No	2	11	13	6
Linear	Box–Cox	No	No	13	2	15	7
Linear	Square root	No	No	5	4	9	3
Linear	Square root	$1/x$	No	7	1	8	2
Linear	Square root	$1/x^2$	No	11	6	17	8
Quadratic	No	No	No	9	9	18	9
Quadratic	No	$1/x$	No	12	8	20	11
Quadratic	No	$1/x^2$	No	8	14	22	12
Quadratic	log–log	No	No	4	5	9	3
Quadratic	log–log	$1/x$	No	6	13	19	10
Quadratic	log–log	$1/x^2$	No	9	18	27	15

caused by day-to-day variation in the calibration curve. The Box–Cox transformation determines an optimal power value through the process of iteration. A requirement is therefore that several calibration curves are analysed during several days in order to have enough data to generate an optimal power value which is reproducible. The challenge is to find and fixate a power value during validation that can be used during routine analysis (i.e. several standard curves over several days are required).

Our result also showed that in the case of log–log transformation, weighting had less influence on the regression line compared to non-scaled models. Weighting factors have an important effect on the SRR for non-transformed models and results in a considerable improvement in predictability, especially in the low concentration region. The slope was marginally affected by weighting but the intercept dropped by more than one order of magnitude. Karnes et al. argued that forcing a regression line through zero would improve the overall fit [21]. Forcing the regression line through zero significantly improved the SRR in this investigation compared to OLR with intercept. However, with respect to the calibration curve fit ranking and the final ranking, forcing through zero turned out to be the least favourable model in the present investigation.

Although the calibrator units (μg/mL or ng/mL) made no difference for most of the models, it is important to note that the unit actually has an impact on the model performance when using log–log transformation and weighting. log–log transformation of decimal numbers less than one form negative values after transformation and therefore also different relative weights in the regression compared to transformed values with decimal numbers larger than one.

The quantification limits are generally overestimated when heteroscedastic data are treated by the OLR approach [23]. This can have a significant impact on pharmacokinetic studies when accurate quantification at low concentrations is required. This would in particular affect pharmacokinetic studies of drugs, such as PQ, with a long terminal elimination phase and low plasma concentration [37,38].

5. Conclusion

Traditionally OLR has been used in quantitative assays of drugs in biological matrices. We suggest that OLR should not be the primary method of choice; on the contrary it should generally be avoided especially when a broad concentration range is used. Other regression models that are capable of correcting for an increased variance throughout the calibration range should be evaluated. Transformation models have traditionally been used less often than weighted calibration models. The present investigation shows that transformation models can be a beneficial alternative compared to weighted models. The design of the proposed strategy will provide enough support for the analyst to justify a choice of a more complex regression model than OLR in accordance with the requirements from FDA guidelines. We recommend including evaluation of independently spiked QC samples together with calibration standards assayed over several days when choosing calibration model during validation.

Acknowledgements

This investigation was part of the Wellcome Trust – Mahidol University – Oxford tropical Medicine Research Programme, supported by The Wellcome Trust of Great Britain. This research was also partly supported by the Swedish International Development Cooperation Agency (SIDA).

References

[1] Guidance for industry, Bioanalytical Method Validation, U.S. Department of Health and Human Services, Food and Drug Administration, Rockville, 2001.

[2] Text on validation of analytical procedure: definitions and terminology (Q2A), Tripartite International Conference on Harmonisation Text (ICH), ICH Tech. Coordination, London, 1994.

[3] Text on validation of analytical procedure: methodology (Q2B), Tripartite International Conference on Harmonisation Text (ICH), ICH Tech. Coordination, London, 1995.

[4] ISO 5725, Accuracy (trueness and precision) of measurement methods and results, Parts 1–4, 6, ISO, Geneva, Switzerland, 1994.

[5] V.P. Shah, K.K. Midha, S. Dighe, I.J. McGilveray, J.P. Skelly, A. Yacobi, T. Layloff, C.T. Viswanathan, C.E. Cook, R.D. McDowall, K.A. Pittman, S. Spector, Pharm. Res. 9 (1992) 588–592.

[6] C. Hartmann, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 12 (1994) 1337–1343.

[7] P. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, Anal. Chim. Acta 391 (1999) 135–148.

[8] P. Chiap, P. Hubert, B. Boulanger, J. Crommen, Anal. Chim. Acta 391 (1999) 227–238.

[9] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (2004) 579–586.

[10] B. Streel, A. Ceccato, R. Klinkenberg, P. Hubert, J. Chromatogr. B 814 (2005) 263–273.

[11] O. Rbeida, B. Christiaens, P. Hubert, D. Lubda, K.S. Boos, J. Crommen, P. Chiap, J. Pharm. Biomed. Anal. 36 (2005) 947–954.

[12] O. Rbeida, P. Chiap, D. Lubda, K.S. Boos, J. Crommen, P. Hubert, J. Pharm. Biomed. Anal. 36 (2005) 961–968.

[13] M. Feinberg, B. Boulanger, W. Dewe, P. Hubert, Anal. Bioanal. Chem. 380 (2004) 502–514.

[14] A. Ceccato, R. Klinkenberg, P. Hubert, B. Streel, J. Pharm. Biomed. Anal. 32 (2003) 619–631.

[15] B. Christiaens, M. Fillet, P. Chiap, O. Rbeida, A. Ceccato, B. Streel, J. De Graeve, J. Crommen, P. Hubert, J. Chromatogr. A 1056 (2004) 105–110.

[16] C. Hartmann, J. Smeyers-Verbeke, D.L. Massart, R.D. McDowall, J. Pharm. Biomed. Anal. 17 (1998) 193–218.

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

[18] B. Boulanger, P. Chiap, W. Dewe, J. Crommen, P. Hubert, J. Pharm. Biomed. Anal. 32 (2003) 753–765.

[19] M. Mulholland, D.B. Hibbert, J. Chromatogr. A 762 (1997) 73–82.

[20] E.L. Johnson, D.L. Reynolds, D.S. Wright, L.A. Pachla, J. Chromatogr. Sci. 26 (1988) 372–379.

[21] H.T. Karnes, G. Shiu, V.P. Shah, Pharm. Res. 8 (1991) 421–426.

[22] G.K. Szabo, H.K. Browne, A. Ajami, E.G. Josephs, J. Clin. Pharmacol. 34 (1994) 242–249.

[23] M.A. Castillo, R.C. Castells, J. Chromatogr. A 921 (2001) 121–133.

[24] A.M. Almeida, M.M. Castel-Branco, A.C. Falcao, J. Chromatogr. B 774 (2002) 215–222.

[25] N.V. Nagaraja, J.K. Paliwal, R.C. Gupta, J. Pharm. Biomed. Anal. 20 (1999) 433–438.

[26] J.R. Lang, S. Bolton, J. Pharm. Biomed. Anal. 9 (1991) 357–361.

[27] J. Wieling, G. Hendriks, W.J. Tamminga, J. Hempenius, C.K. Mensink, B. Oosterhuis, J.H. Jonkman, J. Chromatogr. A 730 (1996) 381–394.

[28] J. Tarning, T. Singtoroj, A. Annerberg, M. Ashton, Y. Bergqvist, N.J. White, N. Lindegardh, N.P.J. Day, J. Pharm. Biomed. Anal., in press.

[29] K. Baumann, Process Control Qual. 10 (1997) 75–112.

[30] M.J. Cardone, S.A. Willavize, M.E. Lacy, Pharm. Res. 7 (1990) 154–160.

[31] C. Hartmann, W. Pennickx, Y.V. Heyden, P. Vankeerberghen, D.L. Massart, R.D. McDowall, Bioavailability, Bioequivalence and Pharmacokinetic Studies, Medpharm Scientific Publishers, Stuttgart, 1995.

[32] Water Quality—Calibration and Evaluation of Analytical Methods and Estimation of Performance Characteristics. Part 1. Statistical Evaluation of the Linear Calibration Function, International Standard-ISO 8466-1:1990 (E), ISO, Geneva, 1990.

[33] P. Armitage, G. Berry, Statistical Methods in Medical Research, Blackwell Science Ltd., London, 1994.

[34] G.E.P. Box, D.R. Cox, J. Roy. Stat. Soc. B 26 (1964) 211–252.

[35] E.K. Kimanani, J. Pharm. Biomed. Anal. 16 (1998) 1117–1124.

[36] E.K. Kimanani, J. Lavigne, J. Pharm. Biomed. Anal. 16 (1998) 1107–1115.

[37] T.Y. Hung, T.M. Davis, K.F. Ilett, H. Karunajeewa, S. Hewitt, M.B. Denis, C. Lim, D. Socheat, Br. J. Clin. Pharmacol. 57 (2004) 253–262.

[38] J. Tarning, N. Lindegardh, A. Annerberg, T. Singtoroj, N.P.J. Day, M. Ashton, N.J. White, Antimicrob. Agents Chemother., in press.