ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
20 (1999) 433–438

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Choosing the calibration model in assay validation [1]

N.V. Nagaraja, J.K. Paliwal, R.C. Gupta *

Pharmacokinetics and Metabolism Division, Central Drug Research Institute, PO Box 173, Lucknow 226 001, India

Received 27 November 1997; received in revised form 5 May 1998; accepted 21 May 1998

Abstract

Data transformations and weighting schemes are normally used to obtain the best-fit of standard curves in bioanalysis and the calibration model is usually selected during prevalidation. In the present study, a comparison has been made between unweighted and weighted ($1/x$, $1/x^2$, and $1/\sqrt{x}$) regression models with or without an intercept in achieving the best-fit for the standard curve of CDRI compound 81/470, a new anthelmintic agent, in cow milk. Validation samples in milk at the LLOQ, medium, and high concentrations were also analysed by each of the calibration models. An unweighted regression equation with an intercept overestimated the concentrations at the LLOQ. An unweighted equation without intercept and weighted equations with or without an intercept significantly minimized the bias at the LLOQ without distorting the results at higher concentrations. Hence, an unweighted equation for a straight line passing through the origin was found to be the best model for a standard curve of 81/470 in milk. Similar results were obtained for 81/470 and UMF-078 in serum and plasma, respectively. Bioanalysts should routinely test these models to obtain the best fit model for their calibration curves as part of their assay validation not during prevalidation. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Assay validation; Calibration model; Residue structure; CDRI-81/470

1. Introduction

A fully validated, accurate and reproducible bioanalytical method is an important prerequisite in a pharmacokinetic/biopharmaceutical study. The quality of bioanalytical data are highly dependent on the quality of the standard curve and the calibration model used to generate it. It is imperative to establish a relationship between the analyte concentration and the detector response. Unlike pharmaceutical analysis, the concentration range in the bioanalysis test samples is dynamic and broad, normally of the order of three or more. Although using two or more standard curves with different calibration ranges is not uncommon, a single standard curve that encompasses the entire dynamic concentration range in a pharmacokinetic study is of great use during routine analysis. Most of our pharmacokinetic studies currently employ a high-performance liquid chromatographic (HPLC) method of analysis. With the use of UV or fluorescence detection in HPLC, a linear increment in detector response can be expected over a broad analyte concentration range.

* Corresponding author.

[1] CDRI communication number 5705.

0731-7085/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0731-7085(98)00256-8

Variance function of the response is not uniform over the calibration range in bioanalysis [1] and hence, the data are said to be heteroscedastic [2]. In such cases, an ordinary least square linear regression equation ($y = mx + c$, where y is the response at concentration x, m is slope and c is the y-intercept), by virtue of minimizing the residuals, gives less importance to the concentrations at or near the lowest limit of quantitation (LLOQ) (since variance is less at these levels) and gives more importance in minimizing residuals at higher concentrations. This might result in incorrect measurements of unknown samples near the LLOQ and thus wrongly question the validity of the assay method. Data transformations or the application of suitable weights are generally employed to overcome heteroscedasticity [1–4]. However, the applicability of a linear equation without an intercept ($y = mx$) for the best-fit of standard curve has not been advocated in recent reports [1–5]. Also, it is the current practice to select the calibration model during the pre-validation stage [6]. In this paper, we share our experience of selecting an appropriate calibration model and the stage of its selection during assay validation.

2. Materials and methods

The assay validation data of CDRI compound 81/470, a new anthelmintic compound in cow milk [7], was used in this study. Milk samples were treated with acetonitrile for protein precipitation followed by extraction with diethyl ether. The compound 81/470 was chromatographed on a C_{18} column and quantitated using a fluorescence detector [7]. In total, seven standard curves in milk were constructed (range 10–1000 ng ml^{-1}) from independently spiked standards, and 120 validation samples, independent from calibration standards, at LLOQ (10 ng ml^{-1}), medium (100 ng ml^{-1}) and high (1000 ng ml^{-1}) concentrations were analysed in four batches. The peak heights (μV) of all calibration standards were pooled and used for an initial calibration model selection. Initial observations showed good linearity between the responses and the concentrations ($r > 0.995$). Linear regression equations $y = mx + c$ and $y = mx$ (line passing through origin) were fitted with $1/x$, $1/x^2$ and $1/\sqrt{x}$ weighting schemes on Microsoft Excel software (Version 5). Individual standard curves were also fitted using these equations/weights. In addition to this, 120 validation samples in the milk matrix were read using each of the models applied to the standard curve of the corresponding batch. Table 1 summarizes the notations used to refer the models in the following text.

3. Results and discussion

Preliminary analysis of standard curves showed linearity between the response (y) and the concentration (x) ($r > 0.995$). Model 1A—the most widely used equation for a standard curve—gave an adequate coefficient of correlation, however, the value of the y intercept was often higher than the response at the LLOQ. Residues at medium and high levels were within acceptable limits. Pooled standard curve data ($n = 7$ at each of the six points) were also fitted with different models (Table 1, Models 1A–4B). The difference in the residual structure among the calibration models was significant at low concentrations, whereas, the difference was moderate at medium and only nominal at higher concentrations. The residue scatter at lower concentrations (10–100 ng ml^{-1}) with each of the models is presented in Fig. 1.

With Model 1A, the intercept was not zero (negative side) and the value was at least 50% of

Table 1
Notations for calibration models with different equations and weighting schemes

Model notation	Equation	Weight
1A	$y = mx + c$	1
1B	$y = mx$	1
2A	$y = mx + c$	$1/x$
2B	$y = mx$	$1/x$
3A	$y = mx + c$	$1/x^2$
3B	$y = mx$	$1/x^2$
4A	$y = mx + c$	$1/\sqrt{x}$
4B	$y = mx$	$1/\sqrt{x}$

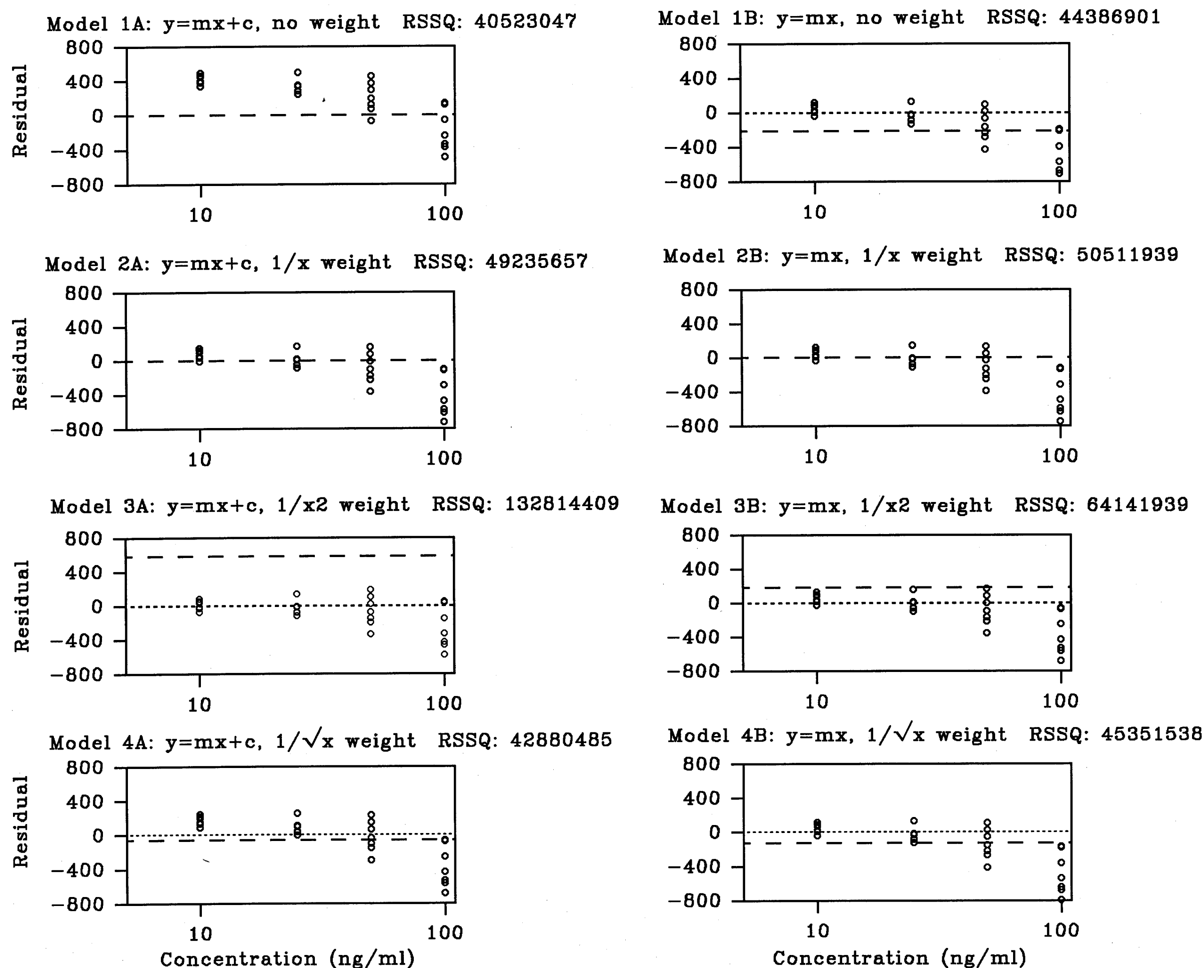

Fig. 1. Comparison of residue scatter at low concentrations with different calibration models. The line (– – –) represents the observed mean.

the response at the LLOQ and the residuals were always on the positive side of the mean of residuals (zero) at 10 and 25 ng ml^{-1}. With Model 1B, the mean of the residuals was not zero, and though the residuals at 10 and 25 ng ml^{-1} were on the positive side of the observed mean, they were scattered around the ideal mean, zero. Although this resulted in underestimation at 50 and 100 ng ml^{-1}, the deviations were within 5%. Other models also resulted in bringing the residuals around zero at 10 and 25 ng ml^{-1} and Model 3A was found to be the best one at lower concentrations (Fig. 1). A comparison of the different models showed that Model 1A yielded least RSSQ, however, this model resulted in overestimation at lower concentrations. Also, RSSQ for the models without an intercept were higher than the corresponding models with an intercept, except with the $1/x^2$ weight. Thus, by analysing the pooled data, it was not possible to choose an appropriate calibration model.

At this stage, it was decided to fit the individual sets of the calibration standards with Models 1A–4B and read the concentrations of validation samples from the corresponding standard curves. An example of regression parameters for one set of calibration standards is given in Table 2. Model 1A was the best model in terms of RSSQ.

The curves without an intercept resulted in higher RSSQ than the corresponding curves with an intercept. The sum of residuals was zero in Models 1A, 2A and 2B, and non-zero in other models. Validation samples were analysed in four batches during inter- and intra-batch variation studies, freeze–thaw cycling studies and in-process residue stability studies. The calibration curves of each batch were constructed using each of the models (Model 1A–4B) and the concentrations of 81/470 in each of the 120 validation samples (at the LLOQ, medium and high concentrations) were estimated using the corresponding standard curves. A subset of the results at the three concentration levels is presented in Table 3. The detector responses (peak heights) of these samples corresponded with those of calibration standards (milk matrix) and the analytical standards (recovery of 81/470 was greater than 90%). The overestimation of samples corresponding to the LLOQ (10 ng ml^{-1}) using unweighted $y = mx + c$ equation (Model 1A) was readily evident (Table 3). Percent bias was considerably greater than acceptable limits of $\pm 20\%$ [5], exceeding 100% in some instances. While, at higher levels, %bias decreased and it was only nominal at 1000 ng ml^{-1}. Model 1B ($y = mx$, unweighted) significantly improved the results at all concentration levels. Percent bias in Model 1B was within 10% at the LLOQ, less than Model 1A at 100 ng ml^{-1} and comparable at 1000 ng ml^{-1}. With $1/x$ weight, Model 2B (without intercept) gave better results, in terms of %bias, than the model with an intercept (Model 2A). The positive residue scatter at the LLOQ in Model 4A ($y = mx + c$ with $1/\sqrt{x}$ weight, Fig. 1) was reflected in overestimation of validation samples at the LLOQ in Table 3. Also, in this weighting scheme, the equation without an intercept (Model 4B) yielded better results than the one with an intercept (Model 4A). At medium and high concentration levels, the differences in the estimated concentrations were within 10%. Dunnet's multiple comparisons of estimated concentrations also showed that the differences in the calculated concentrations at high concentration (1000 ng ml^{-1}) were insignificant ($P > 0.05$). Since it is the convention to choose the simplest algorithm that can satisfactorily explain the relationship between the x and y variables, we selected Model 1B (unweighted $y = mx$) for constructing standard curves of 81/470 in cow milk [7]. This equation was used for the analysis of milk samples of dairy cows after a single oral dose (20 mg kg^{-1}) of 81/470 (7).

While data transformations and weighting schemes have been suggested for linear regression in recent literature, the applicability of standard curve passing through an origin or 'forced' through an origin has not been mentioned [1–5]. Is this because the $y = mx$ equation employs an algorithm whereby the sum of residuals is not equal to zero? If so, weighted equations also do not always achieve this criterion (Table 2). Or is it taken for granted that both $y = mx + c$ and $y = mx$ equations are tested for suitability, without mentioning it?

Our observation was not limited to milk samples. Overestimation at the LLOQ was more pronounced for the estimation of 81/470 in serum for the clinical pharmacokinetic studies (unpublished

Table 2
Regression parameters of a representative standard curve

Model	Slope	Intercept	Sum of residuals	Residual sum of squares
1A	73.0	−298	0	2119923
1B	72.6	0	−1117	2452471
2A	72.2	−61.3	0	2679469
2B	71.9	0	0	3004854
3A	72.1	−60.0	55.4	2731143
3B	70.4	0	2609	8630138
4A	72.6	−119	−320	2294024
4B	72.3	0	−676	2538821

Table 3
Concentration of 81/470 in validation samples calculated through different calibration models

Peak height (μv)	Unweighted				Weight: $1/x$				Weight $1/x^2$				Weight $(1/\sqrt{x})$			
	$y=mx+c$ (1A)		$y=mx$ (1B)		$y=mx+c$ (2A)		$y=mx$ (2B)		$y=mx+c$ (3A)		$y=mx$ (3B)		$y=mx+c$ (4A)		$y=mx$ (4B)	
	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias	Conc. (ng ml^{-1})	% bias
10 ng ml^{-1} (LLOQ)																
694	19.1	91.5	9.6	−3.7	10.3	2.7	9.8	−1.6	10.0	0.3	10.0	−0.5	12.0	20.4	9.7	−2.7
747	20.4	104.2	10.2	1.5	11.9	19.4	10.4	4.0	10.9	8.8	10.8	8.4	14.2	42.0	10.2	2.5
692	25.6	155.5	9.7	−2.7	11.5	14.9	10.1	1.0	9.4	−6.2	10.5	4.9	15.5	54.8	9.9	−1.2
752	26.4	163.8	10.6	5.8	12.4	23.6	11.0	9.8	10.3	3.3	11.4	14.0	16.3	63.3	10.7	7.3
768	14.6	46.0	10.6	5.8	11.5	14.9	10.7	6.8	11.5	14.8	10.9	9.1	12.2	22.3	10.6	6.2
722	14.0	39.7	9.9	−0.6	10.9	8.5	10.0	0.4	10.8	8.4	10.3	2.6	11.6	15.9	10.0	−0.2
100 ng ml^{-1} (medium)																
7014	105.7	5.7	97.3	−2.7	99.7	−0.3	99.4	−0.6	100.4	0.4	100.6	0.6	100.2	0.2	98.3	−1.7
6918	111.3	11.3	97.3	−2.7	101.9	1.9	101.0	1.0	108.1	8.1	104.8	4.8	103.4	3.4	98.7	−1.3
6920	111.3	11.3	97.3	−2.7	101.0	1.0	101.0	1.0	108.1	8.1	104.9	4.9	103.5	3.5	98.8	−1.2
7221	115.4	15.4	101.6	1.6	106.3	6.3	105.4	5.4	112.9	12.9	109.4	9.4	107.7	7.7	103.1	3.1
7066	100.9	0.9	97.3	−2.7	98.8	−1.2	98.2	−1.8	98.8	−1.2	100.4	0.4	99.0	−1.0	97.7	−2.3
7186	102.5	2.5	99.0	−1.0	100.4	0.4	99.9	−0.1	100.5	0.5	102.0	2.0	100.7	0.7	99.3	−0.7
1000 ng ml^{-1} (high)																
74276	1027	2.7	1030	3.0	1052	5.2	1053	5.3	1062	6.2	1065	6.5	1039	3.9	1041	4.1
71241	965.1	−3.5	968.2	−3.2	987.4	−1.3	991.5	−0.9	1032	3.2	1034	3.4	974.2	−2.6	977.2	−2.3
72676	1016	1.6	1022	2.2	1057	5.7	1061	6.1	1150	15.0	1101	10.1	1032	3.2	1037	3.7
73792	1015	1.5	1016	1.6	1023	2.3	1026	2.6	1024	2.4	1048	4.8	1019	1.9	1020	2.0
71059	977.5	−2.3	978.7	−2.1	985.6	−1.4	987.8	−1.2	986.1	−1.4	1009	0.9	981.0	−1.9	982.3	−1.8
72988	1004	0.4	1005	0.5	1012	1.2	1014	1.5	1013	1.3	1037	3.7	1008	0.8	1009	0.9

data) wherein, the analytical concentration range was broad (10–10 μg ml^{-1}). Overestimation of 81/470 at low concentrations with the least-squares calibration line (Model 1A) would have altered the pharmacokinetic model of 81/470 in humans because of significant change in the slope of the terminal phase of the concentration–time curve. Also, the standard curves of a new antifilarial compound methyl (±)-[5-(α-*p*-fluorophenyl)-1H-benzimidazol-2-yl]carbamate (UMF-078) and four of its metabolites (10–500 ng ml^{-1}) could be successfully fitted (unpublished assay, method developed and validated in this laboratory) to unweighted straight line equation passing through origin (Model 1B). For these analytes also, Model 1B improved the results to the same extent as the weighted regression equations.

A second aspect that has drawn our attention is the time taken to choose the calibration model. It is the current practice to analyse replicates of calibration standards in a single batch prior to assay validation, and treat the pooled data through different transformations and weighting schemes [6]. However, it has been found that the calibration model can be rightly chosen only after the process of assay validation is completed. This procedure involves the analysis of independently spiked validation samples and individual calibration curves (six to eight data points) in several batches which actually simulate the real conditions of a routine analysis. If we were to select a calibration model prior to validation, we would have chosen either Model 1A based on RSSQ or Model 3A based on residue scatter (Fig. 1). This would ultimately lead to an inappropriate estimation of the concentrations in the validation samples, thus questioning the authenticity of the validation process. Hence, we recommend that the calibration model should be chosen only after analysing all the validation samples, and not during prevalidation, as is the current practice.

Another question that arises in this context is whether choosing the calibration model is limited to new assay methods, or is it a requirement for validation of an existing method also. In our opinion, since the scatter/bias of points on a calibration curve might change with analyst, system, laboratory, and period, the appropriate calibration model should be chosen whenever a revalidation is attempted.

Acknowledgements

The authors thank Drs S.K. Mandal and Dr M. Srivastava of Biometrics Division, CDRI for their valuable suggestions. Thanks also to WHO for supplying UMF-078 and its metabolites under Macrofil Project. N.V. Nagaraja thanks Council for Industrial and Scientific Research, New Delhi for providing him with a senior research fellowship during the study.

References

[1] C. Hartmann, W. Penninckx, Y.V. Heyden, P. Vankeerberghen, D.L. Massart, R.D. McDowall, in: H.H. Blume, K.K. Midha (Eds.), Bioavailability, Bioequivalence and Pharmacokinetic Studies, Medpharm Scientific Publishers, Stuttgart, Germany, 1995, pp. 331–346.

[2] Analytical Methods Committee, Analyst, 119 (1994) 2363–2366.

[3] A.R Buick, M.V. Doig, S.C. Jeal, G.S. Land, R.D. McDowall, J. Pharm. Biomed. Anal. 8 (1990) 629–637.

[4] J.N. Miller, Analyst 116 (1991) 3–14.

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

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

[7] N.V. Nagaraja, S.K. Singh, R.C. Gupta, J. Pharm. Biomed. Anal 15 (1997) 1175–1180.