Journal of Pharmaceutical & Biomedical Analysis
Vol. 9, No. 6, pp. 435–442, 1991
Printed in Great Britain

0731–7085/91 $3.00 + 0.00

Analytical Survey

A comprehensive method validation strategy for bioanalytical applications in the pharmaceutical industry — 2. Statistical analyses

J.R. LANG† and S. BOLTON*‡

† *Department of Clinical Pharmacokinetics, Development Division, Glaxo Inc., Five Moore Drive, Research Triangle Park, NC 27709, USA*
‡ *College of Pharmacy and Allied Health Professions, St John's University, Jamaica, NY 11439, USA*

Abstract: The first paper in this two-part series described [Lang and Bolton, *J. Pharm. Biomed. Anal.* **9**, 357–361 (1991)] an overall validation strategy for bioanalytical methods. This second paper focuses on the statistical analyses performed on the validation data that will allow the analyst to evaluate the reliability and reproducibility of a bioanalytical method.

Based on the validation results, acceptance criteria for the quality control concentrations are established and used during the study proper to determine if the analytical run is valid. After analysing the clinical study samples and accepting the analytical runs, the quality control results are incorporated into databases to update their acceptance limits. This continuous validation process enables the analyst to monitor the method's performance over time and be confident that accurate sample concentrations are being reported.

It is important to emphasize that the statistical analyses of the data provide information that should be considered from a practical point of view by the analyst. The analyst should use sound judgement in evaluating the reliability of the method.

Keywords: *Statistical analyses; standard curve; limit of quantitation; regression analysis; quality control samples; control charts; confidence limits; analysis of variance (ANOVA); linearity; outliers; acceptance criteria.*

Introduction

Many potential problems can be encountered in the validation process that relate to the statistical evaluation of bioanalytical data. The statistical analyses are meant to be flexible and allow for modifications and additions that may be required as each situation demands. The following statistical analyses and data interpretation are discussed: standard curve raw data analysis, limit of quantitation, regression analysis for the standard curve, quality control sample data analysis, control charts for QC samples and outliers. A flowchart describing these processes is shown in Fig. 1. In these analyses, a significance level of 5% indicates that an analytical problem may exist. Significance at the 1% level requires the analyst to carefully examine the data relating to the significant effect and to make an appropriate decision regarding further action, such as deleting outliers or repeating analyses.

Standard Curve Raw Data Analysis

The daily standard curve consists of standards prepared and analysed in duplicate. The raw data may be presented in a format similar to that shown in Table 1.

Analysis of variance (ANOVA)

The purpose of this analysis is to assess the within- and between-day consistency of the calibration data. The ANOVA is performed on a logarithmic (log) transformation of all calibration data and includes "days", "replicates" and "concentration" as factors. The log transformation equalizes the variance of the observations that have relatively constant standard deviations. The terms of interest in the ANOVA are "replicate" and "replicate × concentration" interaction (Table 2). A significant "replicate × concentration" effect in the ANOVA can discredit the validation. The data must be examined for errors or outliers before

* Author to whom correspondence should be addressed.

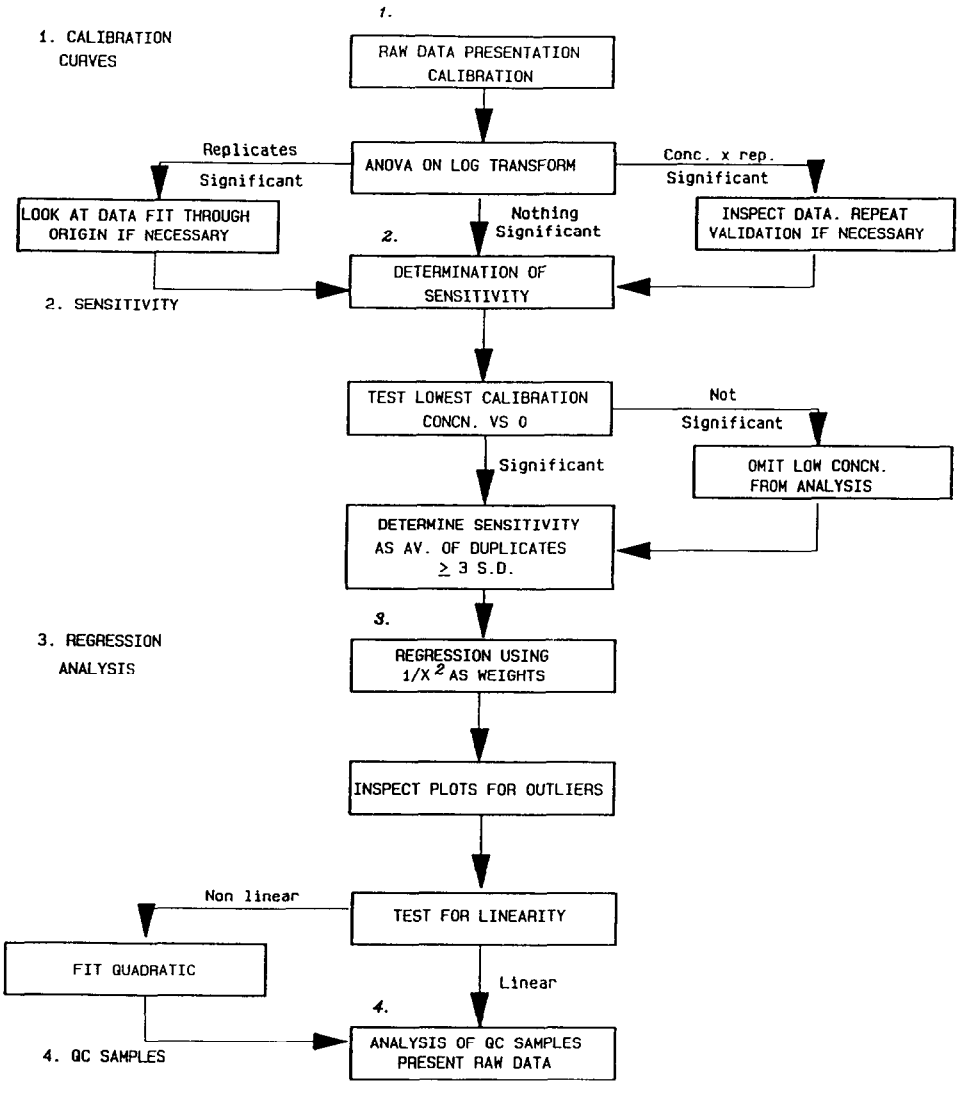

Figure 1
Flowchart for statistical analysis.

Table 1
Representative calibration data

	Peak area			
Theoretical concentration	Rep. 1	Rep. 2	Average	RSD (%)
Day 1				
1	770316	800684	785500	2.73
2	1613058	1599858	1606458	0.58
5	5024084	3962054	4493069	16.71
10	10003846	8664332	9334089	10.15
20	18182916	16979228	17581072	4.84
40	36696072	35977024	36336548	1.40
60	54620832	54699344	54660088	0.10

initiating a re-validation. Outliers should be eliminated from the data set and the statistical analysis repeated. One way of monitoring the variability of points around the standard regression line is to construct a control chart for the standard deviation computed from the residual sum of squares from the least-squares plot. If a standard curve shows a standard

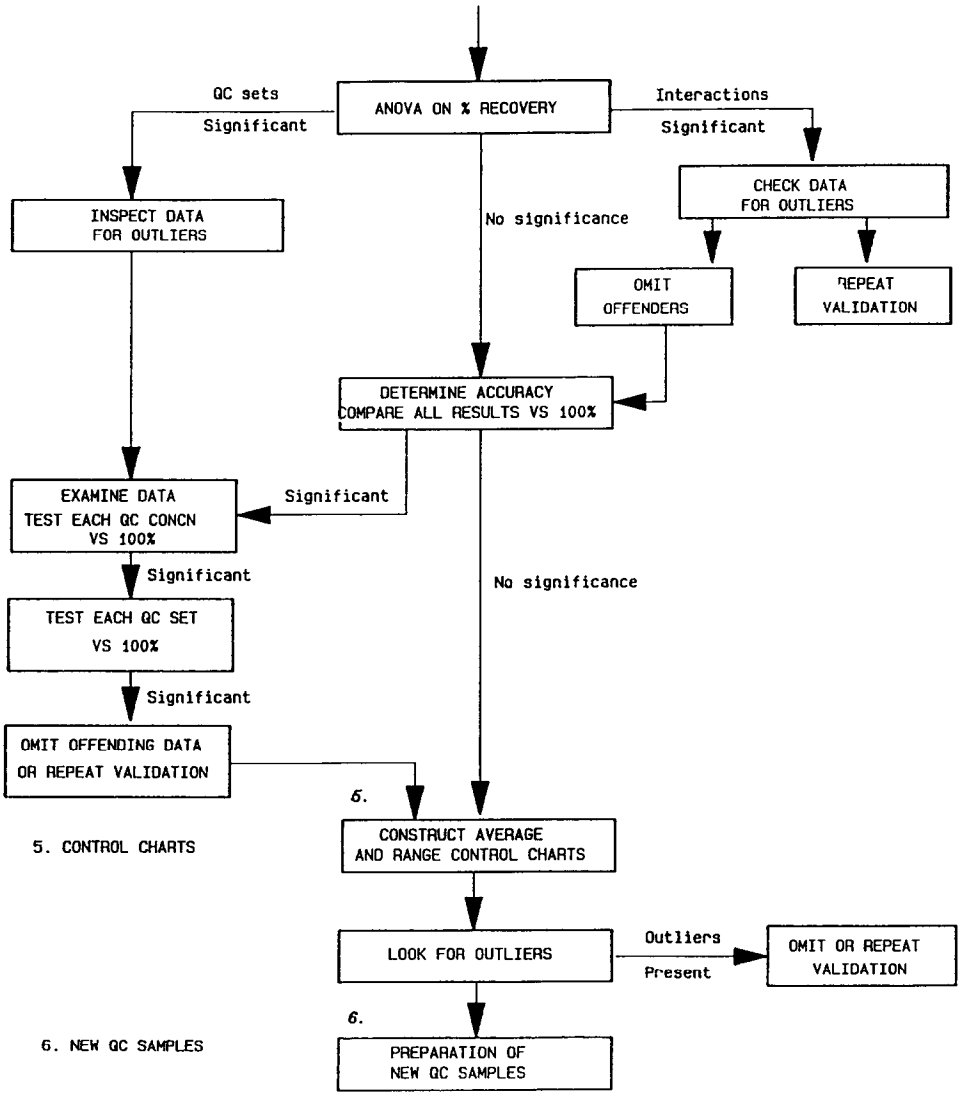

Figure 1
Continued.

Table 2
ANOVA table for cefuroxime response variables: log (area ratio)

Source	d.f.	SS	MS	*F*-ratio	$P > F$
A (Day)	3	3.5531590	1.1843860		
B (Replicate)	1	0.0004380	0.0004380	1.38	0.3255
C (Conc.)	8	298.5447000	37.3180800		
AB	3	0.0009553	0.0003184		
AC	24	0.0922100	0.0038420		
BC	8	0.0119132	0.0014890	0.24	0.9782
ABC	24	0.1476226	0.0061500		
Total	71	302.3510000			

deviation out of limits, the plot should be inspected and the outlier or outliers eliminated. The remaining data can then be re-analysed.

Test of linearity

A statistical test of linearity is performed for each curve separately using a weighted ANOVA. An *F*-test comparing the deviation

mean square of the means from the line compared with the within mean square from the duplicates ($P < 0.05$) are tested. Because separate tests are performed for each analytical run, linearity is considered to exist if (a) all four calibration curves show linearity or (b) three show linearity and one curve shows non-linearity at the 0.05 level, but is not significant at the 0.01 level.

If the fit shows non-linearity, not attributable to the lowest or highest concentration, a quadratic equation may be considered. If either the lowest or highest concentration is causing the non-linearity, the data are re-analysed, omitting these outliers.

The following calculations are used to compute the deviation and within mean squares (MS) for the ANOVA. The weighted sum of squares (SS) will be computed from the duplicates at each concentration.

$$\text{Weighted SS} = \frac{(Y_1 - Y_2)^2}{2X^2}, \qquad (1)$$

where Y_1 is the concentration of one duplicate, and Y_2 is the concentration of the other duplicate.

These values are summed for each concentration and divided by the number of different calibrator concentrations. This gives the within mean square. The deviation SS is obtained by calculating the error SS from the regression analysis and subtracting the within SS calculated. The deviation MS is the deviation SS divided by number of calibrator concentrations minus 2. An F-test of (deviation MS)/(within MS) can then be conducted.

The within and deviation sums of squares are calculated as follows:

$$\text{weighted within SS} = \sum wy^2 - \frac{(\sum wy)^2}{\sum y}. \qquad (2)$$

In the present case

$$w = 1/X^2, \qquad (3)$$

$$\text{weighted within SS at each concentration} = \frac{(Y_1 - Y_2)^2}{2X^2}, \qquad (4)$$

where, X = concentration, Y_i = observed response.

Using the data from Table 1,

$$\text{at } X = 1,\ Y_1 = 770316 \text{ and } Y_2 = 800684, \qquad (5)$$

$$\text{weighted SS} = \frac{(770316-800684)^2}{2 \times 1^2} = 4.6111\text{E+}08. \qquad (6)$$

Repeat at each concentration, X, and sum:

Concentration	Weighted SS
1	4.6111E+08
2	2.1780E+07
5	2.2558E+10
10	8.9715E+09
20	1.8111E+09
40	1.6157E+08
60	8.5613E+05
Total	3.3986E+10 (within SS)

$$\text{deviation SS} + \text{within SS} = \text{error SS}, \qquad (7)$$

$$\text{deviation SS} + (3.3986\text{E+}10) = (4.1640\text{E+}10), \qquad (8)$$

$$\text{deviation SS} = 7.6544\text{E+}09, \qquad (9)$$

$$\text{deviation MS} = \frac{7.6544\text{E+}09}{5} = 1.5309\text{E+}09, \qquad (10)$$

$$\text{within MS } \frac{3.3986\text{E+}10}{7} = 4.8551\text{E+}09, \qquad (11)$$

$$F_{5,7} = \frac{\text{deviation MS}}{\text{within MS}} = 0.315 \quad (F_{5,7} = 3.97 \text{ at } 0.05 \text{ level}). \qquad (12)$$

Limit of Quantitation

The limit of quantitation is defined as the smallest concentration included in the standard curve and is used to interpolate unknown sample concentrations. The criteria for determining this concentration is based on two factors involving background interferences (signal-to-noise) and the reproducibility of the response. The response difference between the lowest concentration and the background sample is determined for each analytical run. The mean of these response differences is

tested statistically against the mean background response. If this difference is not significant, this calibrator will not be included in the standard curve. Additionally, the variability of the response is evaluated by comparing the mean response of the lowest concentration to the standard deviation calculated from the responses at that concentration. If the mean response is not equal to or greater than 3 standard deviations, this concentration is not accepted as the limit of quantitation.

From Table 3, the mean response at the low concentration is 0.00425, with a standard deviation of 0.00155. A *t*-test with 3 degrees of freedom shows this is significantly different from 0 ($t = 2.353$ at $P = 0.05$). Three standard deviations are: $3 \times 0.00155 = 0.00465$. Therefore, this concentration was not accepted as the limit of quantitation, and cannot be included in the standard curve.

If the value equal to 3 standard deviations is considerably smaller than the mean average response at the lowest calibrator concentration, the analyst can test the limit of quantitation at a lower concentration and repeat the analysis (Table 4).

Table 3
Calculation of limit of quantification for cefuroxime (0.05 μg ml^{-1})

Day	Average response
1	0.0035
2	0.0040
3	0.0065
4	0.0030
Average	0.00425
SD	0.00155

$T = (0.00425)/(0.00155/\sqrt{4}) = 5.47$

Table 4
Calculation of limit of quantification for GR43175C (1 ng ml^{-1})

Day	Average response
1	785500
2	1183261
3	1119620
4	1230865
Average	1079812
SD	201431

$T = (1079812)/(201431/\sqrt{4}) = 10.72$
$3 \times SD = 604293$

Regression Analysis for the Standard Curve

Regression analysis is performed using weighted least-squares, with weights equal to $1/X^2$, where X is the theoretical concentration. The slope, intercept, correlation coefficient, variance and interpolated concentration are recorded for each standard curve. A representative standard curve is included in the validation report. A visual inspection of these daily plots is used to identify trends or outliers that could cause problems in the data analysis.

Quality Control Data Analysis

During the validation, two sets of quality control samples are analysed in duplicate from different sample preparations at three different concentrations. The data may be presented as shown in Table 5.

Table 5
Interpolated quality control results

	Theoretical concentration (ng ml^{-1})	Mean concentration Rep. 1	Rep. 2	(% accuracy)*	RSD (%)
Day 1					
QC set I	3	2.8	2.9	2.85	2.5
		(93.3)	(96.6)	(95.0)	
	15	14.8	14.9	14.85	0.5
		(98.6)	(99.3)	(99.0)	
	50	49.8	49.4	49.6	0.6
		(99.6)	(98.8)	(99.2)	
QC set II	3	3.2	3.3	3.25	2.2
		(106.6)	(110.0)	(108.0)	
	15	15.3	14.5	14.9	3.8
		(102.0)	(96.6)	(99.3)	
	50	50.9	49.9	50.4	1.4
		(101.8)	(99.8)	(100.8)	

*% Accuracy is determined by taking a ratio of the calculated to theoretical concentration and expressing it as a percentage.

Determination of accuracy

An ANOVA is performed on the quality control (QC) data using per cent accuracy with "days", "QC set", and "concentration" as factors (Table 6).

If no significant results other than "day" is apparent in the ANOVA, a *t*-test is constructed to compare the overall mean accuracy to 100% as shown below:

$$t = \frac{|\text{overall average} - 100|}{\sqrt{\text{days MS}/N}} . \qquad (13)$$

A significant effect means a consistent bias exists in the quality control sample results. If the *t*-test shows significance, the data should be examined to see if transcription errors or outlying data are responsible. For clarification, an ANOVA is performed for each quality control concentration separately with factors "days" and "QC sets". The predicted versus theoretical (100%) should be tested for each concentration [equation (13)].

A significant effect for "QC set" indicates that one of the QC sample sets is giving consistently biased results. If "QC set" in the overall ANOVA (Table 6) is significant, the raw data should be examined for errors or outliers, the suspect data removed and the remaining data reanalysed. If no outliers or errors are apparent and "QC set" is significantly different, a *t*-test should be performed for each QC set at each concentration separately, comparing the mean result from the four runs to the theoretical concentration.

If any of the results of these tests are significant, the data should be evaluated and outlying values rejected or the validation repeated according to the analyst's judgement. If no significant difference is observed, the sample results are accepted.

Control Charts for Quality Control Samples

Control charts should be constructed using the quality control sample results from the validation. Two control charts will be constructed for each quality control concentration, an average chart which will monitor within- and between-day accuracy and a range chart to monitor assay reproducibility.

Control chart limits

Average chart limits are determined using the principles based on control charts for individuals [2]. The chart is constructed using the daily quality control averages and an average range based on the moving range of size 2. The average range (R) is calculated individually for each quality control sample. The average range for determining limits for the average chart, R, is the average of R_1 and R_2, where the subscripts refer to QC sample sets 1 and 2, respectively. The limits for the average chart are generated using $\bar{X} \pm 3R/1.28$. Sample data and calculations are shown in Table 7.

In this example, the average moving range is 14.1/6 = 2.35. The overall average result is 100.8. The limits are 100.8 ± 3(2.35)/1.128 = 100.8 ± 6.3. The initial average control chart will have a mean of 100.8, with lower and upper limits of 94.5 and 107.1, respectively (Fig. 2). The average range from the eight sets of duplicates above is 3.1. The lower limit is 0 and the upper limit is 10.1 (3.10 × 3.27) (Fig. 3) [3].

With these charts, there is a basis to reject a quality control sample after completing four runs. If duplicate samples fall within both the average and range limits, the data are considered acceptable. If the control charts show outliers, these values will be eliminated from

Table 6
ANOVA table for response variables: QC accuracy

Source	d.f.	SS	MS	*F*-ratio	$P > F$
A (Day)	3	0.0112500	0.0037510	4.03	0.0187
B (QC set)	1	0.0093000	0.0093000	5.24	0.1060
C (Conc.)	2	0.0003963	0.0001981	0.15	0.8602
AB	3	0.0053220	0.0017740	1.90	0.1558
AC	6	0.0077000	0.0012830	1.38	0.2636
BC	2	0.0033010	0.0016500	0.72	0.5240
ABC	6	0.0137300	0.0022890	2.46	0.0538
Error	24	0.0223500	0.0009314		
Total	47	0.0733600			

Table 7
Control chart calculations: medium QC (15 ng ml^{-1}), accuracy (%)

Validation day	Set I					Set II				
	Rep. 1	Rep. 2	Average	Range	Moving range	Rep. 1	Rep. 2	Average	Range	Moving range
1	98.6	99.3	99.0	0.7		102.0	96.6	99.3	5.4	
2	100.0	102.0	101.0	2.0	2	100.0	110.0	105.0	10.0	5.7
3	100.0	102.0	101.0	2.0	0	100.6	102.0	101.3	1.4	3.7
4	100.6	99.3	100.0	1.3	1	100.6	98.6	99.6	2.0	1.7

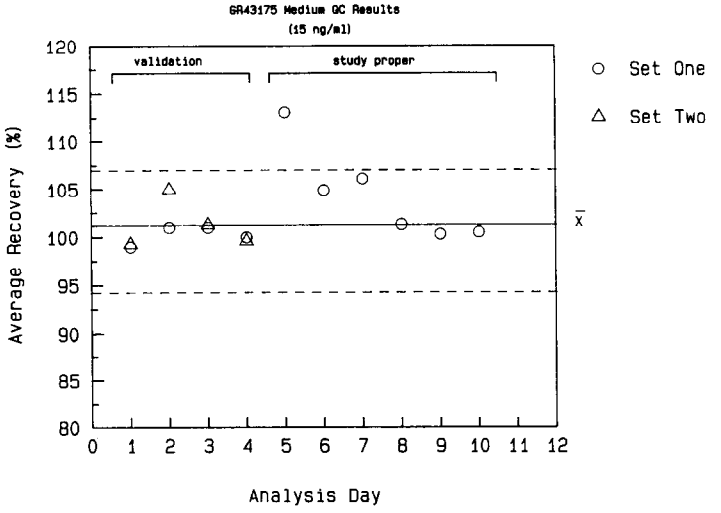

Figure 2
Average chart.

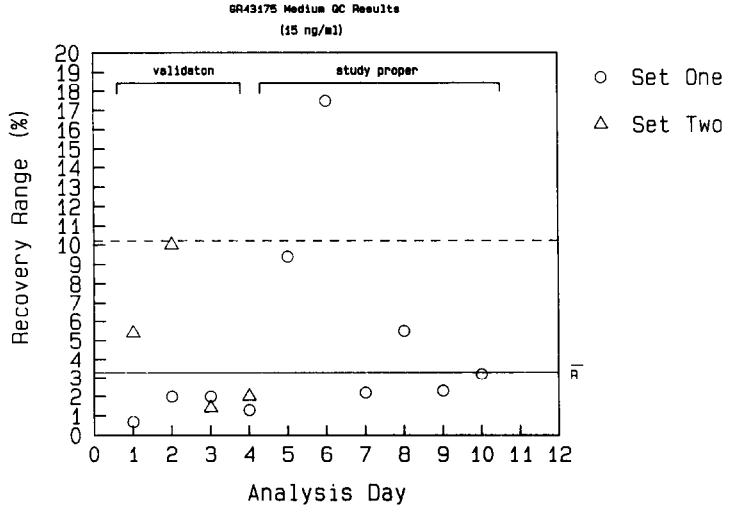

Figure 3
Range chart.

the analysis and new charts prepared. For the validation run to be accepted, there must be at least one quality control value reported at each concentration with no more than one outlier per QC data set. After removing an outlier, the data set is unbalanced. This requires more caution in analysing the remaining data. It should be noted that these estimates of the standard deviation and range are approximate but may be used to initiate the control charts.

Modifying control charts

The initial control chart limits generated during the method validation should be used to determine if the quality control results from the first study proper are acceptable. The average and range control chart limits are to be modified after accepting the analytical runs and quality control results from subsequent clinical studies.

Preparation of new quality control samples

If the two QC sample sets prepared for the method validation show no significant differences, samples from either of the two sets can be used during the study proper. If the two QC sets show significant differences and both are used during the study proper, the duplicates during the study proper will consist of one from each of the two QC sample sets. If new QC samples are prepared, they must be analysed prior to use to determine if they are statistically equivalent to the QC set being used in the study proper. If a single QC set has been used during the study proper, the new quality control samples may be prepared by a single analyst. If two QC sample sets have been used during the study proper, the new quality control samples are to be prepared by two different analysts. A two-tailed *t*-test will be performed to show the equivalence of the old and new QC sample sets. If the test shows a significant difference between QC sample sets, a new set of quality control samples should be prepared and the analysis repeated.

Outliers

After completing the method validation, outliers can be identified either by significant effects in the ANOVA or by the control charts. If aberrant data are observed, the analyst should examine the calibration curve and quality control results to determine if the anomaly can be explained. The analyst will use judgement in deciding the seriousness of the problem and if the method validation or specific runs must be repeated. During the study proper, the control charts generated during the method validation will be the basis for rejecting quality control results.

Conclusions

The method validation presented provides a scheme in which statistical analyses of the bioanalytical data are used to determine the reproducibility and reliability of the method. Consideration has been given to emulate study proper analysis conditions to understand the method's limitations and performance expectations.

Acceptance criteria for the quality control samples and the assay are established based on the validation results and used during the study proper. After accepting the analytical runs from a clinical study, the quality control results are incorporated into databases to modify their acceptance limits. This continuous validation process enables the analyst to monitor the performance of the method over time and be confident that valid sample concentrations are being generated.

It is important to emphasize this statistical approach in evaluating validation data is a more tangible concept than using arbitrary acceptance criteria. Our statistical design allows the bioanalytical data to reveal how reproducible the method should be on a routine basis. The analyst should interpret the statistical data using sound judgement in determining the reliability of the method and deciding if it should be used to support clinical studies and other applications.

Acknowledgements — Elsbeth van Tongeren and Lizbeth Liefer are gratefully acknowledged for their administrative assistance.

References

[1] J.R. Lang and S. Bolton, *J. Pharm. Biomed. Anal.* **9**, 357–361 (1991).
[2] A. Duncan, *Quality Control and Industrial Statistics*, 5th edn, p. 501. Irwin, Homewood, IL (1986).
[3] S. Bolton, *Pharmaceutical Statistics*, Table V.10, p. 486. Marcel Dekker, New York (1984).

[Received for review 19 October 1990; revised version received 21 January 1991]