Quantitative measures of the nonlinearity of an analytical method are defined as follows: the "(dimensional) nonlinearity" of a method is the square root of the mean of the square of the deviation of the response curve from a straight line, where the straight line is chosen to minimize the nonlinearity. The "relative nonlinearity" is defined as the dimensional nonlinearity divided by the difference between the maximum and minimum assayed values. These definitions may be used to develop practical criteria for linearity that are still objective. Calculation of the nonlinearity requires a method of curve-fitting. In this article, we use polynomial regression to demonstrate calculations, but the definition of nonlinearity also accommodates alternative nonlinear regression procedures.

**Mathematical Definitions and Concepts**

**Nonlinearity**

How should quantitative nonlinearity be defined? We propose that nonlinearity should be some measure of the average deviation of a function from an ideal straight line. This concept is illustrated in Figure 1, which shows a nonlinear function, \( f(x) \), within a domain bounded by \( x_L \) and \( x_U \), and an ideal straight line, \( g(x) \). For the moment, a discussion of precisely how \( g(x) \) is determined will be postponed. At any point along the curve, the deviation of \( f(x) \) from an ideal straight line is given by \( f(x) - g(x) \). Note that at some points along the curve, \( f(x) > g(x) \), while at other points along the curve, \( f(x) < g(x) \). The net effect is that the contributions of \( f(x) - g(x) \) along some portions of the curve tend to cancel out the contributions of \( f(x) - g(x) \) along other portions of the curve. In fact, if \( g(x) \) is chosen appropriately, the average value of \( f(x) - g(x) \) over the entire domain of \( x \) will be zero; that is:

\[
\begin{align*}
\int_{x_L}^{x_U} (f(x) - g(x)) \, dx &= 0 \\
\frac{x_U - x_L}{x_L - x_U} &= 0
\end{align*}
\]

Thus, the arithmetic mean of \( f(x) - g(x) \) is not an appropriate measure of linearity.

A more appropriate measure of nonlinearity would be the average of the absolute value of \( f(x) - g(x) \); that is:

\[
\begin{align*}
\int_{x_L}^{x_U} |f(x) - g(x)| \, dx \\
\frac{x_U - x_L}{x_L - x_U}
\end{align*}
\]

Conceptually, this definition makes sense. The contributions of \( f(x) - g(x) \) along the various portions of the domain will not cancel each other. Unfortunately, this definition is mathematically intractable—the appropriate function \( g(x) \) can be calculated only by a process of trial and error.

Another way of preventing the contributions of different portions of the domain from canceling each other is to average the square of \( f(x) - g(x) \). It turns out that this...
quantity is mathematically tractable but has the minor problem of being measured in squared units. Preferably, the measure of nonlinearity will have the same units as the function $f(x)$ itself. By taking the square root of the quantity after the average is taken, this problem is corrected. Thus nonlinearity ($L$) is defined as:

$$L = \sqrt{\frac{\int (f(x) - g(x))^2 \, dx}{x_U - x_L}} \quad (3)$$

The nonlinearity is said to be the “root mean square” (RMS) of the deviation of a function from an ideal straight line.

The RMS may, at first, appear to be a bizarre quantity, but it is used for many measurements with which the reader is probably familiar. To recapitulate, the two properties of the RMS that make it useful to measure nonlinearity are that contributions of opposite sign do not cancel each other, and that the quantity is mathematically tractable. The standard deviation (an RMS quantity) was adopted as the statistical parameter used to measure the width of the gaussian distribution because of these same two properties (3). An even more common example of the use of RMS quantities are the current and voltage from alternating current electrical power lines.

The final element needed to formulate a definition of nonlinearity is a criterion to determine $g(x)$. The most logical way to determine this function is by choosing $g(x)$ in such a way that the RMS deviation of $f(x) - g(x)$ is minimized. Therefore, if $f(x)$ is linear, $g(x)$ will be congruent with $f(x)$, and the nonlinearity will be zero.

A formal definition of nonlinearity may therefore be stated as follows:

The (dimensional) nonlinearity of a function, over a finite domain, is the square root of the mean of the square of the deviation of the value of that function from a straight line, where the straight line has been chosen to minimize the nonlinearity.

The optional qualifier, “dimensional,” may be used to emphasize the difference between the “nonlinearity” and the “relative nonlinearity” (defined below).

The implementation of the above definition is quite straightforward. Substituting $g(x) = a + bx$ into equation 3, the nonlinearity is calculated by the formula:

$$L = \sqrt{\frac{\int (f(x) - a - bx)^2 \, dx}{x_U - x_L}} \quad (4)$$

The parameters $a$ and $b$ must be chosen to minimize $L$. Minimization of $L$ is equivalent to minimization of $L^2$. Therefore, equation 4 is rewritten as:

$$L^2 = \int (f(x))^2 \, dx + a^2 \int dx + b^2 \int x^2 \, dx + 2ab \int xdx - 2a \int f(x)dx - 2b \int xf(x)dx \quad (5)$$

where it is understood that the limits on each of the integrals are $x_L$ and $x_U$. Minimization of $L^2$ is accomplished by setting to zero the partial derivatives with respect to $a$ and $b$. Therefore:

$$\frac{\partial (L^2)}{\partial b} = \frac{2a \int dx + 2b \int xdx - 2 \int f(x)dx}{x_U - x_L} = 0 \quad (6a)$$

$$\frac{\partial (L^2)}{\partial a} = \frac{2b \int x^2 dx + 2a \int xdx - 2 \int xf(x)dx}{x_U - x_L} = 0 \quad (6b)$$

Rearranging equation 6a yields:

$$a = \frac{\int f(x)dx - b \int xdx}{\int dx} \quad (7a)$$

Solving the simultaneous equations 6a and 6b for $b$ produces:

$$b = \left[ \int dx \right] \left[ \int xf(x)dx \right] - \left[ \int dx \right] \left[ \int f(x)dx \right] \quad (7b)$$

Once again stating the limits of integration explicitly, and solving the integrals where possible, we obtain:
Thus, for any arbitrary mathematical function, the constants a and b may be calculated by use of equations 8a and 8b, and, subsequently, the nonlinearity may be calculated from equation 4.

Relative Nonlinearity

Figure 2 shows two theoretical response curves, one for serum sodium, the other for serum lactate dehydrogenase. Visually, these curves appear very similar, and, in fact, except for the scales of the y-axes, the curves are identical. Notice, however, that the nonlinearities of the two curves are numerically different. This is because the range of the sodium response curve spans only 40 mmol/L, whereas the lactate dehydrogenase response curve spans 600 U/L.

To correct for this problem, we define a new quantity, the "relative nonlinearity," which expresses the degree of nonlinearity as a fraction of the full y-scale. This quantity is formally defined as follows:

The relative nonlinearity is the quotient of the linearity divided by the difference between the maximum and minimum values of the range of the function.

If the range of assayed values is bounded by \( y_U \) and \( y_L \), as shown in Figure 1, the relative nonlinearity may be written as:

\[
\lambda = \frac{L}{y_U - y_L}
\]

Notice that the relative nonlinearity of the two curves in Figure 2 are identical. The relative nonlinearity may be expressed as a fraction or as a percentage. It ranges from 0 to 0.5 (or 0 to 50%).

The relative nonlinearity is useful because it depends only on the shape of the curve; it is independent of the units used to measure the analyte of interest. We have found that curves that appear to us to be acceptably linear by visual interpretation have relative nonlinearities of <2.5%, whereas those that appear to be unacceptable have relative nonlinearities >2.5%. Thus, we have replaced a very subjective criterion (visual interpretation) recommended by NCCLS EP6-P with an objective criterion (relative nonlinearity <2.5%). Because relative nonlinearity is measured in dimensionless units, this criterion may be applied to all assays.

If this definition becomes widely adopted as a means of linearity evaluation, a consensus may develop within the clinical chemistry community with regard to the amount of nonlinearity that is acceptable. A single cutoff value for relative nonlinearity need not be applied to all assays. Conceivably, more-critical analytes (e.g., arterial blood pH) may be required to have a lower nonlinearity than other analytes (e.g., aspartate aminotransferase). Criteria for specific analytes could be stated in terms of the relative nonlinearity or the (dimensional) nonlinearity.

Practical (Statistical) Application

To calculate the nonlinearity, the analyst needs a function that describes the assay response curve. This function must be estimated from data obtained from the linearity experiment. Fortunately, numerous regression techniques are available to accomplish this. In the subsections that follow, we (a) describe a protocol for the linearity experiment, (b) define appropriate mathematical nomenclature for data obtained from the experiment, (c) describe how a stepwise polynomial regression model is used to obtain a function that represents the experimental data, and (d) compare the calculation of nonlinearity by use of the stepwise polynomial regres-
sion with the evaluation of linearity by the NCCLS EP6-P procedure.

We emphasize that the stepwise regression model is only one of many alternative methods of curve-fitting that can be used for the calculation of nonlinearity. We chose this particular model primarily for teleological reasons—the stepwise polynomial regression procedure uses the same "lack-of-fit test" (described below) that is used by NCCLS EP6-P. This facilitates the comparison between the two procedures. Two other advantages of this regression procedure are that it is the most commonly used method to obtain nonlinear functions, and that integration of polynomials is trivial.

The linearity experiment. The experimental protocol used to calculate nonlinearity is essentially the same protocol as outlined in the NCCLS EP6-P document. We include in the following description of this protocol a modification suggested by Draper and Smith (4), which the reader may consider optional:

A set of samples containing various concentrations of the analyte of interest must be prepared. The matrix of the samples should reflect the matrix of clinical samples to be assayed by the method (e.g., human serum). The true concentration of the analyte in each sample may be unknown, but the relative interval between the concentrations of each pair of samples must be known exactly, and samples must be assigned values that reflect these relative intervals. Samples having four different concentrations must be used in the study, but the use of at least five is preferred. The range of assayed values of the samples must reflect the clinically relevant range of the analyte of interest.

As an example, consider the preparation of samples for an experiment designed to assess the linearity of a serum sodium assay. The experimenter obtains samples from a hypernatremic patient, whose serum sodium concentration is ~160 mmol/L, and a hyponatremic patient, whose serum sodium is ~120 mmol/L. The hypernatremic serum is assigned a value of 5, and the hyponatremic serum is assigned a value of 1. A sample with an assigned value of 3 is prepared by mixing equal volumes of the hypernatremic serum and the hyponatremic serum. A sample with an assigned value of 2 is prepared by mixing equal volumes of the samples with assigned values of 1 and 3. Similarly, a sample with an assigned value of 4 is prepared by mixing equal volumes of the samples with assigned values of 3 and 5.

Finally, replicates (nominally quadruplicates) at each concentration are assayed. Although NCCLS EP6-P allows replicate assays to be performed on the same solution, Draper and Smith (4) point out that each of the replicates should be mixed separately and that the replicates must be assayed to assess run-to-run, not within-run error. By convention, we shall refer to replicates having the same assigned value as the same "sample," despite the fact that each replicate should be prepared separately.

Definitions and nomenclature. The variables and terms used in the next section are defined as follows:

\[ Q = \text{total number of assays performed in the experiment} \]
\[ r_i = \text{number of replicates of sample } i, \text{ where } i = 1, 2, 3, \ldots N \]
\[ X_i = \text{assigned value of sample } i, \text{ where } i = 1, 2, 3, \ldots N \]
\[ Y_i = \text{the mean assayed value of sample } i, \text{ where } i = 1, 2, 3 \ldots N \]
\[ Y_i^* = \text{the assayed value predicted by a regression equation for sample } i, \text{ where } i = 1, 2, 3 \ldots N \]
\[ Y_{ij} = \text{assigned value of replicate } j \text{ of sample } i, \text{ where } i = 1, 2, 3 \ldots N \text{ and } j = 1, 2, 3 \ldots r_i \]
\[ M = \text{the order of a regression equation, i.e., the highest value of } k, \text{ where the regression equation is written as the sum of terms in } x^k \]
\[ \beta_k = \text{the coefficient of the term } x^k \text{ in the regression equation, where } k = 0, 1, 2, \ldots M \]
\[ w_i = \text{the weight of each assayed value of sample } i, \text{ where } i = 1, 2, 3 \ldots N. \text{ Assuming that the variance of the assay at the concentration of sample } i \text{ is given by } \sigma_i^2, \text{ then the } w_i \text{ values are chosen such that:} \]
\[ w_1 \sigma_1^2 = w_2 \sigma_2^2 = w_3 \sigma_3^2 = \cdots = w_n \sigma_N^2 \]

When the assay is assumed to be homoscedastic (for example, when it passes the test for variance homogeneity described in NCCLS EP6-P), all \( w_i = 1 \).

From the above definitions, it is apparent that the following relationships are true:

\[ Y_i = \sum_{j=1}^{r_i} Y_{ij} / r_i \]
\[ Q = \sum_{i=1}^{N} r_i \]
\[ Y_i^* = \sum_{k=0}^{M} \beta_k X_i^k \]

Calculation of nonlinearity by using stepwise polynomial regression and the lack-of-fit test. Once the linearity experiment has been completed, the data may be fitted to a polynomial of any desired order from 1 to \( N - 1 \). Because of space limitations, it is impractical to reproduce the formulas for calculating the coefficients (\( \beta_i \)) here; the interested reader should consult the references cited (5-7). Some sources (5, 6) describe algorithms that are appropriate only for homoscedastic assays, but more general algorithms are also available (7).

The lack-of-fit test (5) requires the calculation of the error attributed to the imprecision of the replicates (mean square pure error; MSPE) and the error attributed to the lack of fit (mean square lack of fit; MSLF) of the data to the polynomial model. These quantities are calculated as follows:
The degrees of freedom associated with each of these quantities are \( Q - N \) for the MSPE and \( N - M - 1 \) for the MSLF. The basis for the lack-of-fit test is that the ratio MSLF/MSPE is an \( F \)-statistic. If MSLF/MSPE is greater than the 95th percentile of the \( F \)-distribution (for \( N - M - 1 \) and \( Q - N \) degrees of freedom), the regression model is said to “fail” the lack-of-fit test at the \( P = 0.05 \) level of significance. If MSLF/MSPE is less than the 95th percentile of the \( F \)-distribution, the model is said to “pass” the lack-of-fit test.

To obtain the appropriate polynomial, the analyst begins by fitting a first-order (\( M = 1 \)) polynomial to the data and applies the lack-of-fit test. If the first-order polynomial passes the test, this polynomial is used to calculate nonlinearity. Not surprisingly, the nonlinearity is 0 for any first-order polynomial. If the first-order polynomial fails the lack-of-fit test, the analyst proceeds to fit a second-order polynomial (\( M = 2 \)) to the data and again applies the lack-of-fit test. If the second-order polynomial passes the test, this polynomial is used to calculate nonlinearity. Otherwise, the analyst continues to fit successively higher order (\( M \rightarrow M + 1 \)) polynomials until obtaining one that passes the lack-of-fit test. When \( M = N - 1 \), \( Y^* = Y_i \) for all \( i \), and therefore MSLF = 0; this guarantees that, for any finite data set, a polynomial of order \( M \leq N - 1 \) will always pass the lack-of-fit test (5). The nonlinearity will be \( >0 \) whenever a polynomial of higher order than 1 is required to pass the lack-of-fit test.

Once the appropriate polynomial is obtained, this polynomial is substituted for \( f(x) \) in equations 4, 8a, and 8b, and the nonlinearity is calculated. If the nonlinearity is less than the maximum allowable for the assay of interest, the degree of nonlinearity is objectively determined to be “acceptable.” Otherwise, the nonlinearity of the assay is objectively determined to be “unacceptable.”

Comparison with the NCCLS EP6-P approach. This procedure is a special case of the lack-of-fit test, which is described in detail elsewhere (8). The analyst fits a linear (i.e., first-order polynomial) regression model to the data, tests this model with the lack-of-fit test, and stops there. If the linear model passes the lack-of-fit test, the assay “passes” the NCCLS EP6-P statistical linearity test. If the linear model fails the lack-of-fit test, the assay “fails” the EP6-P statistical linearity test. Note that an assay passes the NCCLS EP6-P statistical linearity test if and only if the calculated nonlinearity (calculated by using the stepwise polynomial regression model) is equal to 0.

**Examples**

In the following examples, we will assume that a laboratory considers the linearity of a method to be acceptable if and only if the relative nonlinearity is <2.5%, and that the 95th percentile of the \( F \)-distribution is used for the lack-of-fit test. The data used for all of the following examples are shown in Table 1. The data in all of the examples pass the NCCLS EP6-P test for homogeneity of variance, and, therefore, all are assumed to be homoscedastic.

**Example 1**

These data (Table 1) were obtained as part of an evaluation of an IgG assay in our laboratory (9). Calculation of the first-order (linear) regression yields the following regression equation and lack-of-fit \( F \)-test:

\[
y = 2.58 + 4.36x
\]

\[
F(3,15) = 3.05 < F_{crit} = 3.29
\]

Therefore, the linear model passes the lack-of-fit test. The nonlinearity is calculated to be 0. The assay also passes the NCCLS EP6-P statistical linearity test. In this case, the NCCLS test and the proposed test lead to the same conclusion.

**Example 2**

These data (Table 1) are taken from example 2 (Appendix B) in the NCCLS document (1). Calculation of the first-order regression model yields:

\[
y = 0.763 + 0.850x
\]

\[
F(3,15) = 5.44 > F_{crit} = 3.29
\]

The linear model fails the lack-of-fit test, and, therefore, also fails the NCCLS EP6-P statistical linearity test.

**Table 1. Data Used in Examples of Linearity Calculations**

<table>
<thead>
<tr>
<th>Assigned values</th>
<th>Replicate assayed values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example 1. Nephelometric IgG assay (g/L)</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.51</td>
</tr>
<tr>
<td>2</td>
<td>11.30</td>
</tr>
<tr>
<td>3</td>
<td>15.40</td>
</tr>
<tr>
<td>4</td>
<td>19.70</td>
</tr>
<tr>
<td>5</td>
<td>23.80</td>
</tr>
<tr>
<td>*<em>Example 2. Manual ammonia assay (μmol/L)</em></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>4</td>
<td>4.6</td>
</tr>
<tr>
<td>7</td>
<td>6.5</td>
</tr>
<tr>
<td>12</td>
<td>11.0</td>
</tr>
<tr>
<td>20</td>
<td>18.2</td>
</tr>
<tr>
<td><strong>Example 3. Nephelometric apolipoprotein A-1 assay (g/L)</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.04</td>
</tr>
<tr>
<td>2</td>
<td>1.49</td>
</tr>
<tr>
<td>3</td>
<td>1.89</td>
</tr>
<tr>
<td>4</td>
<td>2.31</td>
</tr>
<tr>
<td>5</td>
<td>2.94</td>
</tr>
</tbody>
</table>

* Data from Appendix B, example 2, of NCCLS EP6-P (1).
However, calculation of the second-order polynomial regression yields:

\[ y = -0.275 + 1.19x - 0.0157x^2 \]

\[ F(2,15) = 0.345 < F_{crit} = 3.68 \]

The second-order model passes the lack-of-fit test. Before the nonlinearity can be calculated, it is first necessary to calculate the parameters \( a \) and \( b \) from equations 8a and 8b, with \( x_L = 1, x_U = 20, \) and \( f(x) = -0.275 + 1.19x - 0.0157x^2. \) The results of these calculations (which are too long to show here) reveal that \( a = 0.98 \) and \( b = 0.856. \) These parameters are not very useful in and of themselves, but they are needed to calculate the nonlinearity from equation 4. Using the values of \( a, b, x_L, x_U, \) and \( f(x) \) shown above, the nonlinearity (L) is calculated to be 0.421 (again, these calculations are too long to show here). The relative nonlinearity is 2.6%. Because the nonlinearity is >2.5%, by the criterion stated above, the method would not be accepted. Again, the proposed test and the NCCLS test lead to the same conclusion.

[From the definition of the parameters \( a \) and \( b \) (equation 4), one might have expected that these parameters would be the same as the intercept and slope, respectively, of the first-order regression equation. Note that this is not the case.]

**Example 3**

These data (Table 1) were obtained as part of an evaluation of an apolipoprotein A-I assay in our laboratory (9). Calculation of the first-order regression model yields:

\[ y = 0.552 + 0.475x \]

\[ F(3,15) = 11.28 > F_{crit} = 3.29 \]

The first-order model fails the lack-of-fit test, and, therefore, the assay fails the NCCLS EP6-P statistical linearity test. Calculation of the second-order polynomial regression yields:

\[ y = 0.742 + 0.312x + 0.0271x^2 \]

\[ F(2,15) = 6.25 > F_{crit} = 3.68 \]

Thus, the second-order model also fails the lack-of-fit test. Calculation of the third-order model yields:

\[ y = 0.43 + 0.749x - 0.140x^2 + 0.0185x^3 \]

\[ F(1,15) = 2.27 < F_{crit} = 4.54 \]

The third-order model passes the lack-of-fit test. Calculation of the parameters \( a, b, \) and \( L \) yields \( a = 0.590, b = 0.456, L = 0.0394. \) The relative nonlinearity is 2.6%. Notice that despite the fact that a third-order polynomial is required to model the data, the relative nonlinearity is <2.5%; therefore, the assay is accepted by the proposed method. In this case, the conclusions drawn by the NCCLS EP6-P statistical linearity test and the proposed test are different.

The authors (9) who evaluated this assay concluded subjectively that it was sufficiently linear for clinical purposes, despite the fact that it failed the NCCLS EP6-P statistical linearity test. Our proposed method provides an objective means of accepting the assay.

**Discussion**

We have proposed a practical definition of a quantitative measure of nonlinearity and have presented an algorithm for calculating this quantity. Calculation of the nonlinearity requires a method of curve-fitting. We have shown how stepwise polynomial regression can be used within this framework, but other methods of curve-fitting may also be used. Alternative nonlinear regression models are less prone to collinearity problems and may require fewer fitted parameters than does polynomial regression (6).

The major advantage of a quantitative measure of nonlinearity is that it makes possible the establishment of an objective criterion for accepting methods that fail the statistical test recommended in NCCLS EP6-P. It is our hope that the concept of quantitative nonlinearity will be accepted by the clinical chemistry community, and that consensual standards for maximum dimensional and (or) relative nonlinearity will emerge for various assays.

An additional benefit is that the nonlinearity is a quantitative parameter that describes one aspect (the linearity) of the performance of an assay. This allows the analyst to compare the performance of two assays from the standpoint of linearity. Although clinical chemists measure precision (by the standard deviation) and accuracy (by systematic and proportional biases) quantitatively, assays have thus far been classified only as “linear” or “nonlinear.” Perhaps this is why so little emphasis has been placed on linearity until now. With a quantitative measure, nonlinearity may be incorporated into total medically acceptable error (10).

The major disadvantage of measuring nonlinearity quantitatively is that the calculations are quite tedious. However, this is not sufficient reason to exclude calculation of nonlinearity from a method evaluation. Using any of a number of spreadsheets (Microsoft Excel® or Lotus 1-2-3®, for example) or statistical packages available for microcomputers, one may perform these calculations with little effort. One of us (K.E.) has developed a customized software package that performs all calculations; no knowledge of the algorithm is required by the user.

Any test for linearity is prone to two types of statistical error: it may reject a method that is sufficiently linear for clinical purposes (in statistical parlance, type I error), or it may accept a method with unsatisfactory linearity (type II error). The NCCLS EP6-P protocol controls type II error at the expense of increased type I error. Furthermore, with increasing assay precision or
an increasing number of replicates per sample, type I error increases even more (accompanied by a further decrease in type II error). From this perspective, it is not surprising that type I error is such a problem when the NCCLS EP6-P linearity protocol is used with very precise assays.

The protocol we describe, calculating nonlinearity by using a stepwise polynomial regression for curve-fitting, controls type I error at the expense of type II error. However, both types of error decrease with increasing assay precision or with an increasing number of replicates per sample. The net effect is that neither type of error causes much of a problem with today’s high-precision analyzers. Furthermore, by changing the procedure for curve-fitting, the balance between type I and type II error may be altered. For example, the number of fitted parameters to be included in the regression model may be determined a priori rather than by the lack-of-fit test. In such a strategy, the balance between type I and type II error will lie between the NCCLS EP6-P linearity test and the stepwise polynomial approach. We have emphasized throughout this article that the definition of nonlinearity may be implemented by using any curve-fitting technique. We consider the method of curve-fitting used to calculate nonlinearity to be subject to future modification and debate.

Tetrau7 (11) has commented on the problem with type I error that is encountered when NCCLS EP6-P is used to evaluate high-precision analyzers. As an alternative to NCCLS EP6-P, Tetrau7 proposed using the correlation coefficient for the assigned values of each sample (X) vs the mean assayed value for each sample (Y) in much the same way that we propose using the relative nonlinearity. Although the correlation coefficient is simpler to calculate, it has several limitations in comparison with dimensional and relative nonlinearity. First, the correlation coefficient was not specifically designed to measure nonlinearity; thus, it lacks the theoretical foundation of the dimensional and relative nonlinearity. Second, the numerical value of the correlation coefficient is arbitrary; therefore, it is difficult to interpret and cannot be incorporated into an estimate of medically acceptable error. Third, the correlation coefficient is inflexible; there is no way to modify this measurement to accommodate specific goals of evaluation, such as a particular balance between type I and type II error. Finally, the correlation coefficient may be seriously underestimated if the sample concentrations are not evenly distributed; because this point is not obvious, we provide a simple example in the Appendix.

Appendix

Consider a purely hypothetical experiment involving samples with assigned values of –2, –1.0, 1.0, and 2, and mean assayed values of 0, 8, 10, 12, and 20, respectively. Furthermore, let us assume that each sample was assayed in quadruplicate, and that the MSPE = 0.67. For this experiment, stepwise polynomial regression yields the equation: y = 10 + x + x^2, and, therefore, L = 1.21 and λ = 6.1%. The relative nonlinearity greatly exceeds the acceptable cutoff of 2.5%, and thus the method would be rejected as nonlinear. By Tetraul7’s test, r^2 = 0.93. This is below Tetraul7’s proposed cutoff of 0.95, and, again, the method would be rejected as nonlinear.

Suppose the experiment were designed differently. Let us assume that samples having assigned values of –2, –1.9, –1.8, –1.0, 1.0, 1.8, 1.9, and 2 are assayed, and mean values of 0, 1, 2, 8, 10, 12, 18, 19, and 20, respectively, are obtained. The experimenter might do this thinking that the samples having the highest and lowest concentrations are very near the linear limits of the assay. If the linearity is not acceptable, the experimenter plans to discard one or both samples at the extremes of analyte concentration, recalculate the nonlinearity, and thus determine the linear range. Again, assuming the MSPE = 0.67 and that each sample is assayed in quadruplicate, the stepwise polynomial regression equation is: y = 10 + 1.11x + 0.996x^2, L = 1.20 and λ = 6.0%. Again, by our method, the nonlinearity is determined to be unacceptable. Unfortunately, r^2 = 0.97 so that, by Tetraul7’s test, the method would be accepted as linear.

Mathematically, the problem with using the correlation coefficient in this second experimental design is that the observations at the extreme concentrations are, in effect, inappropriately weighted. The samples with assigned values of –1.8 and –1.9 are not very different from the one with the assigned value of –2. Similarly, the samples with assigned values of 1.8 and 1.9 are not very different from the one with the assigned value of 2. The net effect is almost the same as if only five samples were assayed, as in the first experiment, but the samples with assigned values of –2 and 2 were given weights that were three times that of those with assigned values of –1, 0, and 1.

References