Troubleshooting

Selecting the Best Curve Fit

Melissa M. Kiser and John W. Dolan

Which curve-fitting function should be used?

In our laboratory, liquid chromatography (LC) is used primarily as a quantitative analytical tool for the determination of pharmaceuticals in biological samples, commonly called bioanalytical sample analysis. Like most laboratories in the bioanalytical business, we make extensive use of tandem mass spectrometry for detection (LC–MS-MS). The United States Food and Drug Administration (FDA) has set numerous regulations that must be followed to produce the analytical results that will pass the agency's scrutiny and thus be able to support the development of new drugs. Interpretation of these regulations is aided by published guidelines. For validation of bioanalytical methods, "Guidance for Industry: Bioanalytical Method Validation" (1), which we'll refer to as "the guidance," is a key document in laboratories such as ours.

One of the important factors in LC method development and validation is the selection of the proper weighting scheme to use with a standard curve, such that acceptable analytical results are obtained. The guidelines state, "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).

This one sentence contains three important requirements: "simplest model," "appropriate weighting," and "statistical tests." It slides off the tongue rather easily, but complying with these stipulations might not be so simple. The guidance also states that arbitrary curve weighting is not defensible. Almeida and colleagues (2) recently published an article pointing out some of the tests that can be used. They also stated that this treatment of data is nothing new for the statistician, but it might be unfamiliar territory for the analytical chemist. For example, the topic of curve weighting is treated quite clearly in statistics texts such as *Statistics for Analytical Chemistry* (3), which we've used to help support the following discussion. For this month's "LC Troubleshooting," we've taken some of these concepts and applied them to data from one of our recent validations to present a relatively simple method for complying with the requirements of the aforementioned guidelines.

A Traditional Approach

For the example in this presentation, data were obtained from Day 1 of an LC–MS-MS method validation to determine paclitaxel in pig serum. The run began and ended with separate standard curves (12 concentrations for each), with the validation samples run in between. The validation sample concentrations were based on recommendations in the guidelines: the lower limit of quantification, three times the lower limit of quantification, in the middle of the range, and at the upper limit of quantification. A single injection of six replicates of each validation sample was made for each concentration (four levels, $n = 6$ per level).

The simple, traditional way to test the data for curve fit is to start with a linear fit and see how it works. A plot of the data for the two 12-point standard curves is shown in Figure 1. The results look acceptable, fall on a linear trend line, and have a coefficient of determination (r^2) of almost 1. At this point, it might seem that we've found a simple fit with good statistics. But is this really true?

Residuals Plot

A plot of residuals is much more instructive than a calibration curve when the samples cover such a wide range of concentrations (3000-fold in this case). The residual plot is obtained by calculating the percent recovery of each sample and plotting it against sample concentration. Crowding at the lower end of the curve can be reduced for better viewing by plotting the concentrations on a logarithmic scale. If the correla-

John W. Dolan
LC Troubleshooting Editor

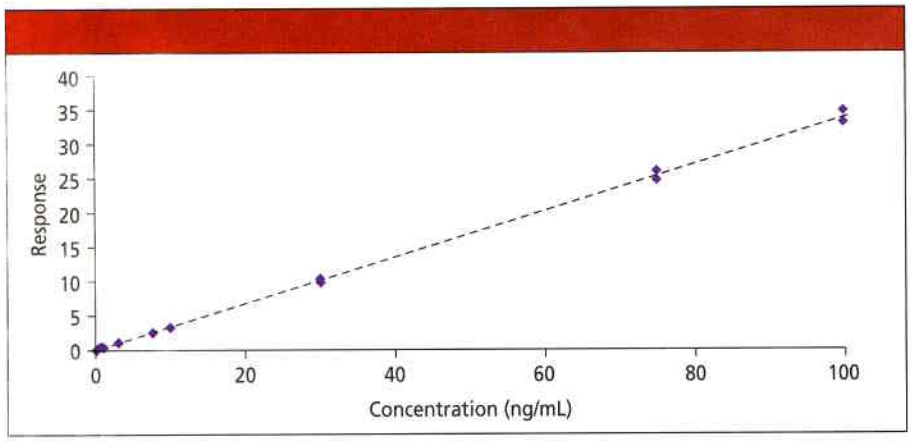

Figure 1: Plot of peak area ratio versus sample concentration with no curve weighting applied. Data points for two 12-point standard curves with trendline added; r^2 = 0.9990.

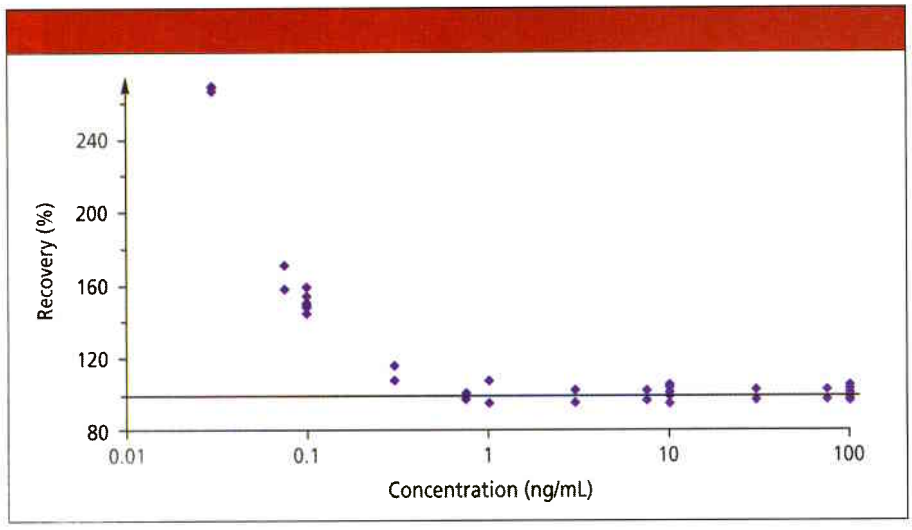

Figure 2: Residuals plot of Figure 1 data plus validation sample data (n = 6 samples at each of four concentrations).

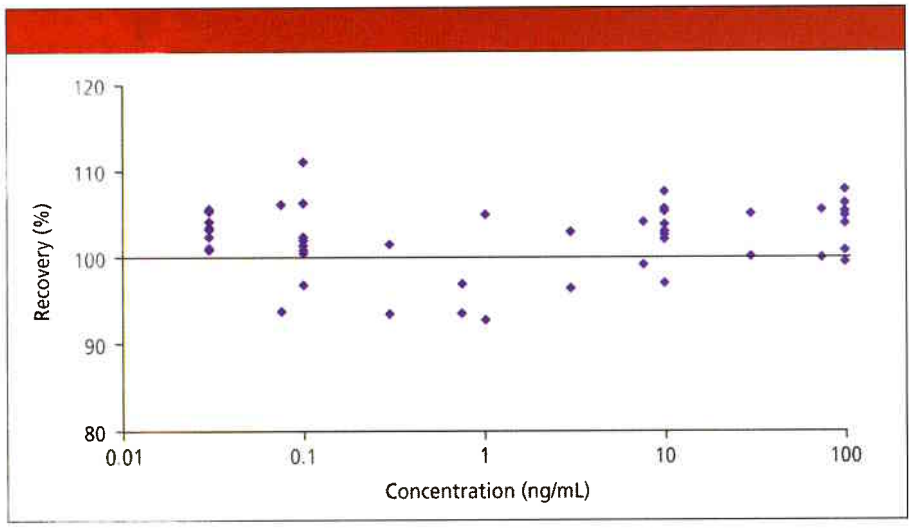

Figure 3: Residuals plot for Figure 2 data with $1/x^2$ weighting.

tion was perfect, all points would fall on the 100% line. Points below the line have less than 100% recovery, whereas points above the line have an apparent recovery of more than 100% of the theoretical amount. For a well-behaved data set, the data points should be evenly scattered on both sides of the 100% line.

Figure 2, which contains data from all 48 samples in the run, shows that the points at 0.03 ng/mL give 270% recovery. The guidelines suggest a limit of ±20% deviation (80–120%) at the lower limit of quantification and a limit of ±15% at all other concentrations. This plot of the data begins to show curvature below 1 ng/mL, and the 0.3-ng/mL samples fall just within the 20% limit. From the residuals plot of Figure 2, the data do not look nearly as good as they do in the linear regression plot of Figure 1. Even so, the data appear to support the use of the standard curve from 0.3 to 100 ng/mL.

However, before we take the data at face value, we need to ask if the data make sense. Are the plots of Figures 1 and 2 really good representations of the behavior of the method? Are the 0.03-ng/mL concentrations really 270% too high? For carefully prepared samples, most workers would agree that this is unlikely.

The Problem Source

If we look at the original data set again (Figure 1), we see a clue that the data are not behaving as well as they should. This plot is just for the standard curve samples, so each concentration has just two points. At the top of the curve, the points are visibly spaced, giving us a feel for the error involved. At the bottom end, however, the points are indistinguishable from each other. This tells us that the absolute error is larger at the top of the curve than at the bottom. This behavior tells us that the data are heteroscedastic, which means absolute error varies with sample concentration.

We can use the scedasticity of the data to determine whether weighting is needed. Homoscedastic data have standard deviations that are the same at all sample concentrations. That is, the error at the low end of the curve and at the high end of the curve are similar. In such cases, curve weighting is not appropriate, and calibration plots such as the unweighted linear regression of Figure 1 are appropriate.

With heteroscedastic data, the standard deviation increases with the sample concentration. The error is more or less proportional to concentration, so if we normalize

the data by calculating the relative standard deviation (RSD), we find that the RSD is fairly constant across the plot. In such cases, weighting usually will be beneficial and should be tested for improved curve performance.

The FDA guidelines call for statistical proof that the best weighting is used. Chemists can use the *F*-test to determine whether the data are homoscedastic or not. The *F*-test is a ratio of the variances (standard deviation squared) for two test sets. Data gathered at the upper end of the curve (upper limit of quantification) and the lower end of the curve (lower limit of quantification or three times the lower limit of quantification) are appropriate to use for this test. Because a standard deviation is required, more than the two data points are needed for each concentration on the standard curve. The validation samples, with $n = 6$ at each concentration, are appropriate.

Once the experimental *F*-value is calculated, we can look up the limiting *F*-value in a statistics book table for the appropriate degrees of freedom. (This test can also be done in Excel using one of the data analysis tools.) In the present case, with $n = 6$ for each set, $n - 1$ (or 5) is used for the degrees of freedom. We're testing at the 0.01 rejection level, or what is sometimes called the 99% confidence level. For the present data, the ratio of the variances is 72 million! This is certainly more than the rejection value of 10.97 from the table in reference 3. There is no question that the data are heteroscedastic, so curve weighting should be tested.

It is interesting that the RSD is nearly constant across the curve, with about a threefold difference between the lower and upper end of the curve — further evidence of heteroscedasticity.

Curve Weighting

The problem of a misleading regression coefficient in Figure 1 arises from the fact that the large standard deviations of the points at the top of the curve dominate the calculations. To give the points at the lower end of the curve equal consideration, we must apply weighting. A statistics text (3) will tell you that weighting factors should be determined based upon the calculated standard deviations of the data, but this approach adds complexity with little added benefit over standardized weighting approaches. It also requires more data points than are available from the standard curve data, so the technique is not very practical for routine use.

The most popular weighting schemes involve adjusting the data by a factor related to an inverse function of the concentration. Commonly, $1/x^0$ (no weighting), $1/x^{0.5}$, $1/x$, and $1/x^2$ are applied. The calculations are straightforward, but the equations are complex (2,3), so the use of computer software generally is the best approach. The data-system software for the mass spectrometers in our lab allows us to specify the curve-fitting function to use. For the present case, we set up an Excel spreadsheet that automatically calculated the effect of weighting and allowed us to plot the results graphically.

The effect of weighting can be dramatic, as shown in Figure 3, where all of the standard-curve and validation sample data are presented in a residuals plot using $1/x^2$ weighting. In this case, only one of 48 points exceeds 10% deviation from the nominal concentration, even at the 0.03-ng/mL level. Compare this with the results of the unweighted data in Figure 2 (note the different *y*-axis scale).

Selecting the Best Weighting

To determine which weighting factor is most appropriate to use, chemists should first calculate the results using different weighting factors. We used $1/x^0$, $1/x^{0.5}$, $1/x$, $1/x^2$, and $1/x^3$ for the data shown in Table I. Convert the calculated concentrations to percent recovery. Next, determine the difference between the calculated percent recovery and 100% to get the relative error. Take the sum of the absolute values of the relative error to get the sum of the relative error (Σ %RE in Table I). The weighting scheme that gives the smallest sum of the relative error is the best one to use. In the case of the present data, Table I shows $1/x^2$ as the best weighting.

Table I: Sum of relative errors for various curve-weighting values

Weighting	Σ %RE	r^2
$1/x^0$	18.02	0.9990
$1/x^{0.5}$	5.99	0.9992
$1/x$	2.07	0.9992
$1/x^2$	1.17	0.9978
$1/x^3$	1.45	0.9950
$1/s^2$	0.76	0.9985

For comparison, the ideal weighting using the standard deviation technique ($1/s^2$) is shown in Table I. You can see that the $1/x^2$ weighting gives a sum of errors less than twice as large as the ideal case. This result is pretty good, and it is doubtful that any practical difference exists between the two. Furthermore, the trial weighting values can be used with a single standard curve, whereas the standard deviation technique requires at least three sets of data.

What about the Coefficient of Determination?

Most of us have come to expect that the coefficient of determination will be part of the reported data. Many laboratories specify that the coefficient of determination for a bioanalytical calibration curve must be at least 0.95, and many workers expect values of r^2 to be greater than 0.98 for well-behaved methods. The last column in Table I lists the coefficients of determination for the different curve weightings. They're all greater than 0.99, and it is debatable whether one value is any better than the others. The data in Figure 2 showed us that even if the standard curve has a coefficient of determination of 0.9990, the fit will not necessarily be very good. This just tells us that the coefficient of determination is a poor measure of the curve-fit quality for data like these. Coefficients of determination are useful for homoscedastic data but can be misleading for heteroscedastic data.

Summary

We now have the tools to apply statistical tests to determine goodness of fit for our standard curves so that we can meet the guidelines and have confidence in our choice of weighting.

First, determine if the data are homo- or heteroscedastic. You can do this with the *F*-test, but an eyeball test on the data usually is sufficient. Furthermore, the nature of the data in bioanalytical calibration curves is such that it is very unlikely that the data are homoscedastic. It probably isn't worth taking the trouble to perform the *F*-test.

Next, calculate the results using various weighting schemes. Most likely, $1/x^0$, $1/x$, and $1/x^2$ are going to tell the story. This can be done with an Excel spreadsheet, and once the data are imported, it takes only a few seconds per weighting factor, so there is no reason not to try out several factors. Compare the sum of the relative errors to find the smallest value — this is the best fit. Now you have a statistical test that allows you to defend your choice of weighting factors.

There are a couple of other things to keep in mind. First, the residuals plot, in which the percent recovery is plotted against concentration, is very useful. These plots are most informative if you use a logarithmic scale for concentrations. With such plots, you can quickly see the performance improve at the bottom end of the curve as you add weighting. Again, this can be done quickly in Excel. Finally, don't be fooled into thinking that the coefficient of determination is giving you much useful information about the quality of the data.

References

(1) FDA Home Page. "Guidance for Industry: Bioanalytical Method Validation," http://www.fda.gov/cder/guidance/index.htm (May 2001).

(2) A.M. Almeida, M.M. Castel-Branco, and A.C. Falcao, *J. Chromatogr. B* **774,** 215–222 (2002).

(3) J.C. Miller and J.N. Miller, *Statistics for Analytical Chemistry* (John Wiley & Sons, Hoboken, New Jersey, 1984) pp. 107–112.

Melissa Kiser is a Project Manager for BASi Northwest Laboratory of McMinnville, Oregon. Her special interests include development and validation of bioanalyticial LC–MS-MS methods.

John W. Dolan

"LC Troubleshooting" Editor John W. Dolan is Vice-President of BASi Northwest Laboratory of McMinnville, Oregon; a Principal Instructor for LC Resources, Walnut Creek, California; and a member of LCGC's editorial advisory board. Direct correspondence about this column to "LC Troubleshooting," LCGC, Woodbridge Corporate Plaza, 485 Route 1 South, Building F, First Floor, Iselin, NJ 08830, e-mail John.Dolan@Bioanalytical.com.

For an ongoing discussion of LC troubleshooting with John Dolan and other chromatographers, visit the Chromatography Forum discussion group at http://www.chromforum.com.