Fitting Brendan's Five-Parameter Logistic Curve

Paul G Gottschalk, PhD, and John R Dunn II, PhD, Brendan Scientific Corporation, 2236 Rutherford Road, Suite 107, Carlsbad, CA 92008 USA

Introduction

Biologists perform immunoassays to determine the concentration of an analyte in a sample. In Bio-Plex™ assays, this is typically done by measuring a response in the form of a signal that is proportional to the amount of analyte bound to the antibody on beads that have been incubated with the sample.

In order to quantitate the concentration of the analyte, the response must be compared to a calibration curve commonly called the standard curve. The unknown concentration of an analyte may then be determined by finding the concentration on the standard curve that produces the same response as that obtained from the unknown sample (Wild 1994, Dudley et al. 1985).

Ideally, the standard curve would be the true curve: the curve that expresses the concentration vs. response relationship without any distortion by errors. If an infinite number of concentrations were used, each with an infinite number of replicates, the resulting curve would be the true curve. Since, for practical reasons, only a limited number of samples can be run in an assay, the true curve must be estimated from a limited number of noisy responses.

From these noisy responses, the true response at each concentration must be estimated. Because there cannot be a standard at every concentration, a means of interpolating between standards is necessary. This is done by selecting a mathematical function that does a good job of approximating the true curve. This approximating function is called a curve model. Many functions have been employed as curve models of immunoassays, including lines, cubic splines, logistic functions, and lines in logit-log space.

Curve Fitting

Two steps must be taken to find a function that gives a good approximation of the true curve. First, the mathematical curve model must be selected. Then, the particular curve out of the entire family of possible curves that best explains the data must be determined by fitting the curve. This means that the parameters in the function must be adjusted until the function approximates the assay's true curve as well as possible.

In statistics, one of the most common ways to determine how well a candidate standard curve is fitting the true curve is to determine how likely (probable) it is for the candidate standard curve to have yielded the observed standard data under the assumption that the candidate standard curve is actually the true curve. The best-fitting curve is therefore the curve most likely to have given rise to the observed data. This curve is often called the maximum likelihood estimate of the true curve.

Statistical regression theory shows that finding the parameters of the maximum likelihood curve is equivalent to finding the curve whose parameters generate the smallest weighted sum of squared errors (*wSSE*) (Draper and Smith 1981, Bates and Watts 1988). The weighted sum of squared errors is the sum of all of the squares of the differences (Δ^2) between the observed standard responses (y_i) and the response predicted by the curve model ($\hat{y}_i$), weighted by the inverse variance (1/variance) of the standard responses at that concentration.

$$wSSE = \sum_{i=1}^{N} w_i [\, y_i - \hat{y}_i \,]^2 = \sum_{i=1}^{N} w_i [\, \Delta_i \,]^2$$

Regression, or fitting, is the process of minimizing the weighted sum of squared errors (*wSSE*). Therefore, the best-fitting curve is the one whose parameters produce the smallest *wSSE*. Note that the differences (Δ) are being called errors here, but the word "residuals" is often used, and *wSSE* could just as well be called the weighted sum of squared residuals.

Figure 1 diagrams the computation of *wSSE*, showing the errors and weights that factor into the computation.

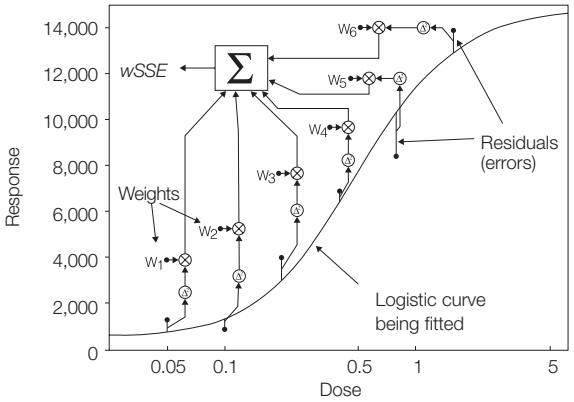

Fig. 1. Graphical rendition of the computation of *wSSE*. Shown are the residuals, how the weights are applied to them, and the summation of the weighted squared residuals to obtain *wSSE*. The direction and length of the line from the curve to the data point indicate the sign and magnitude of the residual. This residual is then squared, multiplied by a weighting factor, and summed to give *wSSE*.

Weighting

Choosing proper weights for the squared residuals is crucial for obtaining the best curve fit. According to regression theory, the weights should be set equal to the inverse variance of the responses at that concentration. This keeps the fitting procedure tighter around those standard responses with the smallest variance (error), that is, those with smallest responses, and looser around those standard responses with the largest variance, that is, those with largest responses. This yields the optimal use of information from the noisier and less noisy standards, and leads to the most accurate concentration estimates. Sample concentrations computed from unweighted curve-fitting procedures can differ from properly weighted curves by hundreds of percent.

The variance of a standard is a function of the magnitude of its response. It is common for the variance of a standard at the high-response end of a curve to be 3 or 4 orders of magnitude larger than those at the low-response end of the curve. There are two reasons for this response dependence. First, most signal detectors produce noise with a standard deviation that is proportional to the magnitude of the response. Second, the kinetics associated with antibody binding are nonlinear (Wild 1994), with the result that the kinetic variations in the reaction change with the magnitude of the response. The variance of the standards can usually be approximated by a power function of the response:

$$\text{variance} = A(\text{response})^B$$

where A is a function of the magnitudes of the responses and B falls in the range 1.0–2.0 for most immunoassays (Finney 1978). Data such as immunoassay responses in which the variances are not constant are called heteroscedastic data.

Since it is impractical to run enough replicates to reliably estimate the true variance function from a single assay, a pool of historical assay data can be used to compute this variance function (Finney and Phillips 1977, Raab 1981).

Residual Variance vs. r^2

Once the curve model has been fitted by adjusting its parameters to minimize *wSSE*, a means of assessing the quality of the curve fit is necessary because a poor curve fit will generate unreliable concentration estimates.

Perhaps the first quantity one might consider using as a fit metric is *wSSE*, since this is precisely what is being minimized by the curve-fitting algorithm to find the best fit. For any given set of standard data, a curve fit with a smaller *wSSE* is better than one with a larger *wSSE*.

The problem with using *wSSE* is that it cannot be compared directly with the *wSSE* of other assays if the number of data points is not the same. A quantity that is based on *wSSE* but that can be compared across assays that have differing amounts of data is called the residual variance. The residual variance is the *wSSE* divided by the number of degrees of freedom in the assay. The number of degrees of freedom in an assay is the number of data points in the standard curve beyond the number of parameters in the curve model. For example, a line has two parameters, and so requires two data points to uniquely determine it. If a line were being fitted to eight standard points, there would be 8 – 2 = 6 degrees of freedom. A five-parameter logistic (5PL) has five parameters and requires five data points to uniquely determine it. If a 5PL were being fitted to eight standard points, there would be 8 – 5 = 3 degrees of freedom. The residual variance normalizes the *wSSE* to properly account for differences in the number of data points.

The statistic r^2 is another measurement that is commonly used as a fit metric, especially for linear regression. Residual variance and r^2 are related, but they behave quite differently. If the squares of each of the mean-corrected responses are weighted and summed together, this total sum of squares (total SS) can be divided into two parts: the regression sum of squares (regression SS) is that portion of the total SS that is explained by the regression model, and the residual sum of squares (SSE) is that portion of the total SS that is left over. The value of r^2 is the fraction regression SS/total SS.

The problem with r^2 is that it is not a sensitive metric for assessing the quality of a curve fit. Even with bad curve fits, the vast majority of the total SS will be accounted for in the regression SS fraction. This leads to r^2 values above 0.95 even for a bad fit. Further, because it is not weighted, the majority of r^2 will have been contributed by the standards with the highest responses. The residual variance, on the other hand, is very sensitive to imperfections in the curve fit, when properly weighted.

Fit Probability

Residual variance allows the quality of fit to be measured. However, it provides no information about how good or bad a fit is. Under the assumption that the responses of the individual standard concentrations are approximately normally distributed, it can be shown that *wSSE* obeys a chi-square distribution with the number of degrees of freedom present in the assay (Draper and Smith 1981, Bates and Watts 1988). This distribution allows us to determine how likely it is that a

particular value of *wSSE* will occur. The chi-square probability is the fraction of assays, if performed under exactly the same conditions, that would be expected to have a worse curve fit, that is, a larger *wSSE*, than the curve fit of the assay under consideration. This probability obtained from the chi-square distribution is called the fit probability. Being a probability, the values of the fit probability range from 1 (perfect fit) to 0 (no fit).

The Advantages of a Good Fit

The goal of an assay designed to determine the concentration of an analyte in a sample is to be as accurate as possible. Improving the standard curve fit will improve the accuracy of all concentration estimates. In turn, improving the accuracy of the concentration estimates will extend the dynamic range of the assay. The dynamic, or reportable, range of an assay is the range over which the errors of the concentration estimates stay below the maximum acceptable error. Typically, a good fit will cause the dynamic range to be extended at the low concentration end of the assay, permitting lower concentrations to be accurately determined.

To improve the quality of the fit of the standard curve, the sources of fit error must be identified. In any regression, regardless of what curve model is used, there are two reasons that the curve will not fit the data perfectly. The first reason is the presence of random variation in the data. This kind of error is called pure error and can be reduced by increasing the number of replicates of each standard. The second reason is that the curve model may not approximate the true curve very well. This kind of error is called lack-of-fit error and cannot be reduced by increasing the number of standard replicates. For example, much immunoassay data has a sigmoidal, or "S", shape if data are taken over a wide enough range of concentrations. If a line were used as the curve model to fit such data, much of the *wSSE* would be due to lack-of-fit error. This is because a straight line simply cannot fit the sigmoidal shape of the data. In other words, the shape of the assay's true curve is not a straight line.

To fit the data as well as possible, a curve model must be chosen that does a good job of approximating the true curve. If this is not done, no amount of improving the assay process by reducing random noise will improve the fit beyond the limit allowed by the lack-of-fit component of the error.

The Four- and Five-Parameter Logistic Curve Models

The four-parameter logistic (4PL) model has historically been more successful at fitting immunoassay data than its predecessors, particularly the logit-log and mass action models. However, the 4PL model has a weakness: It is a symmetrical function, and most immunoassay data are not symmetrical. To rectify this shortcoming, the 4PL model was extended by adding a fifth parameter that controls the degree of asymmetry of the curve (Prentice 1976, Rodbard et al. 1978). With the extra flexibility afforded by its g parameter, the 5PL model is able to eliminate this lack-of-fit error from immunoassay curves. Figure 2 shows the result of fitting the same set of asymmetric data with a 5PL curve model and a 4PL curve model. It is apparent that the 5PL curve fits the data better and will provide more accurate concentration estimates than the 4PL fit. Table 1 shows the fit statistics for the two curve fits in Figure 2: *wSSE*, degrees of freedom, residual variance, and fit probability. The fit probabilities in Table 1 show that the quality of the 5PL fit is good, while the fit of the 4PL is very poor.

The 5PL and 4PL models are described by the following equation:

$$y = d + \frac{a - d}{\left[1 + \left(\frac{x}{c}\right)^{b}\right]^{g}}$$

Where:

a = estimated response at zero concentration
b = slope factor
c = mid-range concentration
d = estimated response at infinite concentration
g = asymmetry factor

In this equation, x is the concentration, y is the response, and a, b, c, d, and g are the five parameters of the 5PL model. The 4PL model is obtained by setting g = 1.

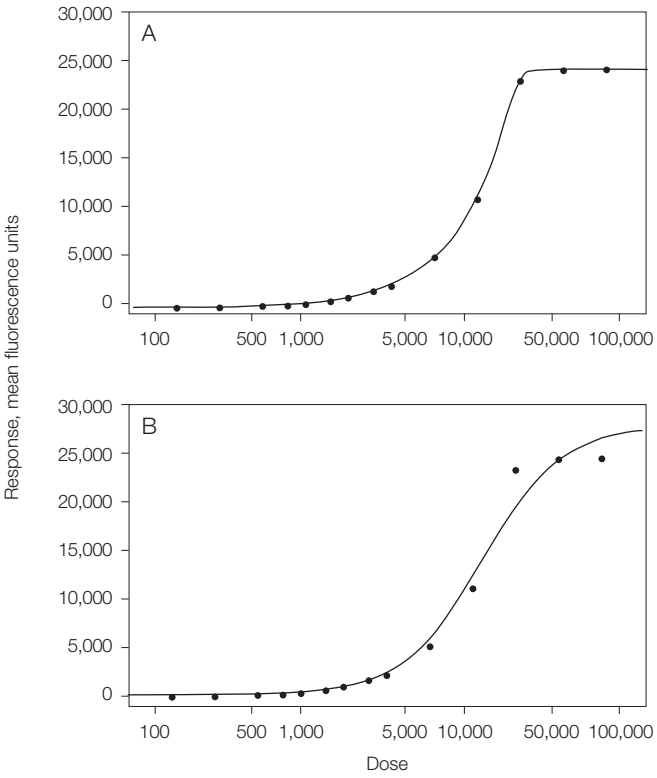

Fig. 2. Comparison of 5PL and 4PL curve fitting. A, the result of a 5PL fit on asymmetric immunoassay data; B, the result of a 4PL fit on the same data.

Table 1. Fit statistics for the 5PL and 4PL fits shown in Figure 2.

Statistic	5PL	4PL
wSSE	2.41	120.0
Degrees of freedom	9	10
Residual variance	0.269	12.0
Fit probability	0.983	<0.001

Table 2. The relationship between the order of a and d, the sign of b, and the slope of the monotonic 5PL function.

Case #	Order of a and d	Sign of b	Slope
1	a > d	b > 0	Down
2	a > d	b < 0	Up
3	a < d	b > 0	Up
4	a < d	b < 0	Down

Note that when g = 1 and the curve is actually a 4PL curve, pairs of cases can be combined into single cases, so that for 4PL, case #1 and case #2, or case #1 and case #3, generate the same functional forms. For the 5PL function, all cases produce distinct functional forms.

All curves generated by the 5PL equation are either monotonically increasing or decreasing, depending on the choice of parameters a, b, and d. Table 2 summarizes the effect of the parameters a, b, and d on the slope of the logistic function.

Figure 3 shows semi-log plots of a family of curves that can be generated from the 5PL equation. The figure makes several characteristics of the 5PL function apparent. The function approaches a horizontal asymptote as the dose approaches zero, and it approaches another horizontal asymptote as the dose approaches infinity. Between the asymptotic regions of the curve is a transition region. There is a single inflection point in the transition region. On either side of the inflection point, the curve will approach the left and right asymptotes at different rates unless g = 1.

Each parameter of the 5PL function has a different effect on the curve. Table 3 summarizes the effect of each parameter on the 5PL function.

Table 3. Geometric interpretation of the 5PL function's parameters.

Parameter	Effect on Curve
a	Controls the position of the asymptote. Order relative to d controls sign of slope. Along with b, magnitude relative to d controls magnitude of slope.
b	Magnitude controls the rapidity of the transition region; sign controls sign of the slope of the curve. Solely controls the rate of approach to the asymptote, and jointly with g controls the approach to the asymptote.
c	Controls the position of the transition region.
d	Controls the position of the asymptote. Order relative to a controls the sign of the slope of the curve. Along with b, magnitude relative to a controls magnitude of slope of the curve.
g	Jointly with b controls the rate of approach to the asymptote.

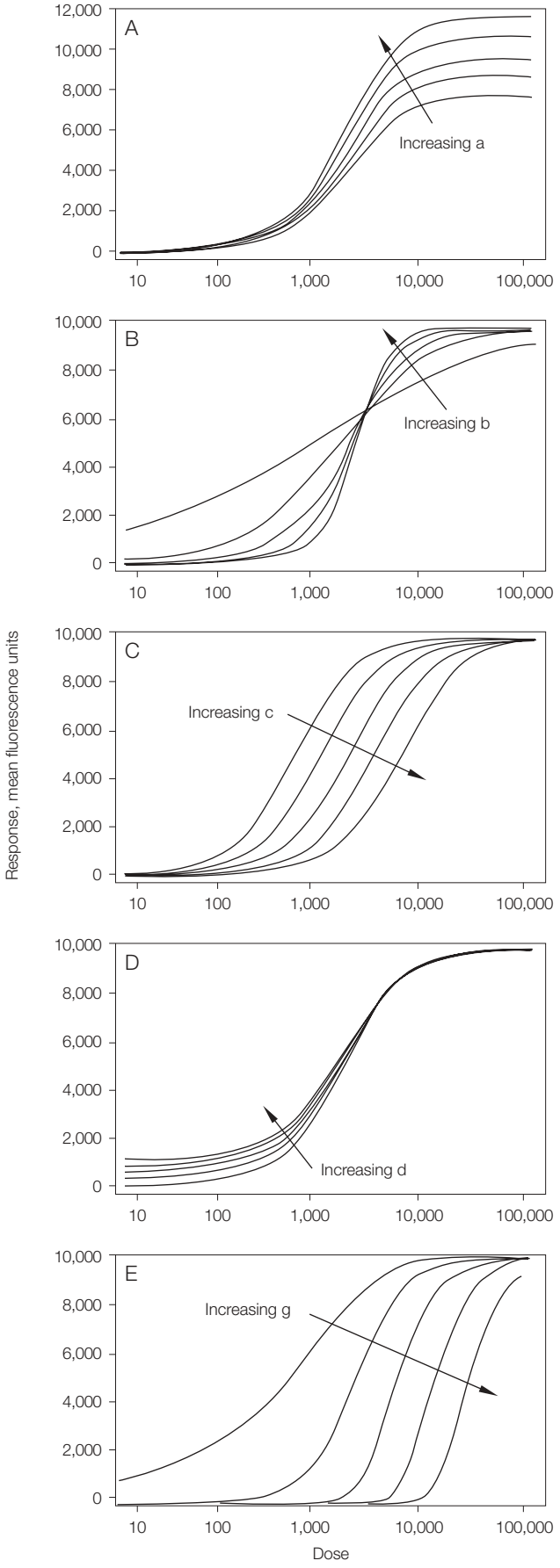

Fig. 3. Effects of varying the parameters of a 5PL function. Panels A–E, effects of varying the parameters a, b, c, d, and g, respectively. In this example, a > d and b < 0.

Fitting With the Five-Parameter Logistic Model

Asymmetric Curve Fitting

To handle data that were too asymmetric for the 4PL, many researchers were forced to use methods such as spline interpolations. The spline interpolation method does not attempt to find a minimum parameter model that approximates the true curve. Instead, the spline interpolation method puts a cubic spline through the data points using a parametric cubic function. Since a spline passes exactly through each data point, the number of parameters in a spline curve is always equal to the number of data points. The result is that there are always zero degrees of freedom in a spline fit. This means that a spline fit performs no averaging of the data to reduce random variation. Furthermore, a spline is not a good interpolating function for immunoassay data, because it often does not approximate the curve between the data points well. Splines are not always monotonic, and can oscillate up and down because of the random variation in every data point. Spline-based standard curves have substantial curve error because none of the random variation in the data points is averaged out, and therefore the concentration estimates contain a greater amount of error than a curve model with fewer parameters that can average out the random variation.

Another method sometimes used in lieu of a 5PL is the log 4PL method. This method takes advantage of the fact that taking the logarithm of the response of some asymmetric curves can make the data more symmetrical, and therefore better suited to a 4PL fit. This approach works when the low-response end of a sigmoidal curve has a shorter "tail" (approaches the asymptote faster) than the upper end of the curve (parameter g between 1.1 and 1.5). However, taking the log of the sigmoidal data when the low-response end of the curve has a longer tail than the high-response end makes the data more asymmetric. Since this type of behavior is encountered in immunoassay data more often than the former, this approach is not satisfactory for routine data reduction. Even when g fits into this optimal range, a 5PL model will almost always result in a better fit and more accurate concentration estimates.

When the 5PL model was first proposed as the means to obtain good fits to asymmetric assay data, many researchers recognized its potential and attempted to integrate the 5PL into their assay analyses. They found that fitting a 5PL model to assay data is much more difficult than fitting the 4PL model. The traditional numerical-fitting algorithms used for the 4PL failed at an unacceptably high rate; that is, they either returned an error or never finished. Worse still, in many of the assays for which the algorithm returned a fit, the fit was clearly not the best fit.

Why the 5PL Model Is Difficult to Fit

One reason that the 5PL model is so difficult to fit is that it is possible to wildly adjust its parameters in tandem so that the 5PL curve itself hardly changes in the places where the data are. When one can change the parameters of the curve so that the curve hardly changes relative to the data points, then the residuals also will hardly change. This results in the value of *wSSE* being largely unaffected by these tandem parameter changes. Figure 4 shows an example of a 5PL curve where changing the values of the c and g parameters of the 5PL substantially has little effect on the curve and therefore on the value of *wSSE*. This can be seen by the wide, flat valley in the plot where an algorithm could easily be fooled into stopping early. Also, the flat plain on the far left near the g = 0 axis could fool an algorithm into stopping with a much worse value of *wSSE* than in the valley. The values of c and g that minimize the value of *wSSE* in this example are c = 16,748 and g = 2,899.

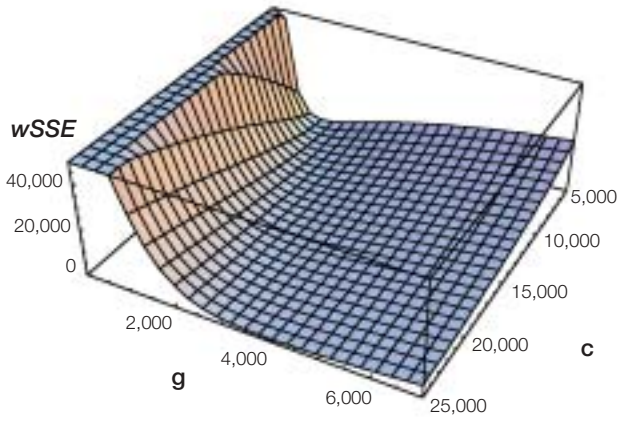

Fig. 4. A plot of the *wSSE* surface of a 5PL curve fit against the parameters c and g. The color of the plot corresponds to the slope of the surface. Blue hues are low slope, or "flat". The values c = 16,748 and g = 2,899 minimize *wSSE*. This minimum falls in the middle of a flat plain that can fool algorithms into stopping early or going haywire, or can cause other problems. Note also the plain on the far left near the g = 0 axis, where an algorithm could easily stop without realizing that the valley where the true answer lies even exists.

To facilitate the interpretation of this geometric view, imagine that *wSSE* is the result of fitting a two-parameter curve model instead of the 5PL curve model. A region in the parameter space where *wSSE* hardly changes even across wide changes in the parameter values would be like a wide flat plain. A plot of *wSSE* can also have "ravines" with steep sides but very level bottoms. Unlike a person standing on a landscape scanning the scene for a low spot, a fitting algorithm has only local information about *wSSE* and the history of where it has been to guide its next guess for the location of the minimum of *wSSE*. It has no way to "see" where to go.

The point in the parameter space that minimizes *wSSE* provides the parameters of the best-fitting curve. Such a point is at the bottom of a valley on the *wSSE* surface. If a wide plain surrounds such a valley, an algorithm that does not use advanced methods can be fooled into thinking that the plain is actually the floor of the best-fitting valley, causing it to stop well short of the true minimum. Also, on such a plain, second-derivative-based methods such as Gauss-Newton and Levenberg-Marquardt, the fitting methods most commonly used to fit 4PL curves, can be fooled into going in an entirely wrong direction because the matrix of second derivatives is very nearly singular. Such algorithms may hop around the

parameter space forever, or stop and report an error after their iteration limits are reached. Also, the near singularity of the second-derivative matrix can cause many numerical problems since it must typically be inverted to be used in algorithms that employ it. This can cause an algorithm to have unpredictable problems.

Another problem that happens to algorithms on such flat plains in parameter space is that they can be going somewhere, probably not even in the right direction, very, very slowly. The result of this is that the algorithm may also never return because it never thinks it is "finished".

Lastly, the *wSSE* surface is often not one big "bowl" such that all a fitting algorithm needs to do is "slide" down to the bottom. Almost any fitting algorithm can handle such a simple scenario. Real *wSSE* surfaces have many possible valleys, the bottoms of which may be the correct best minimum but more likely are not. A successful algorithm must be able to find the valley with the lowest bottom, and then successfully find the bottom. This requires that the algorithm be able to find a starting point for the fitting algorithm so that the starting point is within the region of convergence to the correct minimum of *wSSE* of the fitting algorithm being used. This is perhaps the most difficult task in developing a reliable algorithm for fitting the 5PL model. Many algorithms will end up in the wrong valley, and will report that they have found the best fit, even though they are nowhere near it.

The Brendan Logistic Module Fits the 5PL Better Than Other Software

A better 5PL fit means that concentration estimates will be more accurate. More accurate concentration estimates lead to wider dynamic ranges for assays. The Brendan logistic module, used in Bio-Plex Manager™ version 3.0 software, does a better job of fitting the 5PL than other data reduction software. The Brendan logistic module performs better because it:

- Uses optimal weighting with the 5PL curve model to correctly model the random error and thereby get the best possible fit
- Employs a number of sophisticated fitting algorithms that are able to traverse the "plains" and "level valleys" of the *wSSE* surface that can fool other algorithms
- Uses sophisticated logic to determine which algorithms are most effective to use during each phase of the fitting
- Uses advanced methods to ensure that the true global minimum of *wSSE* is obtained, not a false local minimum
- Is numerically stable and robust due to careful selection of appropriate numerical algorithms, careful customization and optimization of them, and carefully designed logic for handling exceptions

Summary

Weighted 5PL regression produces the most accurate concentration estimates of any method currently in use for asymmetric immunoassay data. Because of difficulties with fitting the 5PL curve model, the 5PL model is only now becoming widely used in conjunction with software like the Bio-Plex system's Brendan logistic module and StatLIA. Since no immunoassay data are perfectly symmetric, fitting with a 5PL model will almost always result in better concentration estimates than using a 4PL model. This in turn allows greater dynamic ranges to be used, and produces greater accuracy in the results. The Brendan logistic module is able to fit the 5PL model better than other software.

References

Bates DM and Watts DG, Nonlinear Regression Analysis and Its Applications, New York: Wiley (1988)

Draper NR and Smith H, Applied Regression Analysis, New York: Wiley (1981)

Dudley RA et al., Guidelines for immunoassay data processing, Clin Chem 31, 1264–1271 (1985)

Finney DJ, Statistical Methods in Biological Assays, 3rd ed, London: Charles Griffin (1978)

Finney DL and Phillips P, The form and estimation of a variance function, with particular reference to immunoassay, Appl Stat 26, 312–320 (1977)

Prentice RLA, A generalization of the probit and logit methods for dose response curves, Biometrics 32, 761–768 (1976)

Raab GM, Estimation of a variance function, with application to immunoassay, Appl Stat 30, 32–40 (1981)

Rodbard D et al., Radioimmunoassay and Related Procedures in Medicine 1, Vienna: International Atomic Energy Agency, 469–504 (1978)

Wild D (ed), The Immunoassay Handbook, New York: Stockton Press (1994)

About Brendan Scientific

Brendan Scientific develops and markets StatLIA software for advanced computational analysis and workflow management of immunoassay-related technologies. Brendan's core focus has been to combine proven scientific methods and robust mathematics with the processing speed and automation of software. This combination provides the highest level of accuracy and reliability with the workflow flexibility and automation required for today's laboratory.

BIO-RAD

Bio-Rad Laboratories, Inc.

Life Science Group

Web site *www.bio-rad.com* ***USA*** *(800) 4BIORAD* ***Australia*** *02 9914 2800* ***Austria*** *(01)-877 89 01* ***Belgium*** *09-385 55 11* ***Brazil*** *55 21 2527 3454* ***Canada*** *(905) 712-2771* ***China*** *(86-21) 63052255* ***Czech Republic*** *+ 420 2 41 43 05 32* ***Denmark*** *44 52 10 00* ***Finland*** *09 804 22 00* ***France*** *01 47 95 69 65* ***Germany*** *089 318 84-177* ***Hong Kong*** *852-2789-3300* ***Hungary*** *36 1 455 8800* ***India*** *(91-124)-6398112/113/114, 6450092/93* ***Israel*** *03 951 4127* ***Italy*** *39 02 216091* ***Japan*** *03-5811-6270* ***Korea*** *82-2-3473-4460* ***Latin America*** *305-894-5950* ***Mexico*** *55-52-00-05-20* ***The Netherlands*** *0318-540666* ***New Zealand*** *64 9 415 2280* ***Norway*** *23 38 41 30* ***Poland*** *+ 48 22 331 99 99* ***Portugal*** *351-21-472-7700* ***Russia*** *7 095 721 1404* ***Singapore*** *65-6415 3188* ***South Africa*** *00 27 11 4428508* ***Spain*** *34 91 590 5200* ***Sweden*** *08 555 12700* ***Switzerland*** *061 717-9555* ***Taiwan*** *(8862) 2578-7189/2578-7241* ***United Kingdom*** *020 8328 2000*