Appropriate Calibration Curve Fitting in Ligand Binding Assays

John W. A. Findlay[1,2] and Robert F. Dillard[3]

[1]Pharmacokinetics, Dynamics, and Metabolism, Pfizer Global Research and Development, Groton, CT
[2]Current address: Gilead Sciences Inc, 4 University Place, 4611 University Drive, Durham, NC 27707-3458
[3]BioStatistics and Data Management, Takeda Pharmaceuticals North America, Inc, Deerfield, IL

Abstract

Calibration curves for ligand binding assays are generally characterized by a nonlinear relationship between the mean response and the analyte concentration. Typically, the response exhibits a sigmoidal relationship with concentration. The currently accepted reference model for these calibration curves is the 4-parameter logistic (4-PL) model, which optimizes accuracy and precision over the maximum usable calibration range. Incorporation of weighting into the model requires additional effort but generally results in improved calibration curve performance. For calibration curves with some asymmetry, introduction of a fifth parameter (5-PL) may further improve the goodness of fit of the experimental data to the algorithm. Alternative models should be used with caution and with knowledge of the accuracy and precision performance of the model across the entire calibration range, but particularly at upper and lower analyte concentration areas, where the 4- and 5-PL algorithms generally outperform alternative models. Several assay design parameters, such as placement of calibrator concentrations across the selected range and assay layout on multiwell plates, should be considered, to enable optimal application of the 4- or 5-PL model. The fit of the experimental data to the model should be evaluated by assessment of agreement of nominal and model-predicted data for calibrators.

Introduction

Liquid chromatography/mass spectrometry has largely replaced immunoassay and other ligand binding assays (LBAs) as the preferred bioanalytical technique for determination of conventional low-molecular-weight drug candidates in biological matrices. However, rapidly increasing interest in the development of proteins and other products of biotechnology as potential therapeutics[1] has resulted in continued application of LBAs for the quantitation of these macromolecules in biological matrices. Validation and implementation of these methods have been reviewed[2-8] and have been the subject of several workshops sponsored by the American Association of Pharmaceutical Scientists (AAPS) and the US Food and Drug Administration (FDA),[4-6] among others, the latest of which was conducted in May 2006.[7] In addition, the FDA has issued a guidance for validation of bioanalytical methods.[8] Whether for quantitation of macromolecules or small molecules, sound calibration curves for these binding assays are central to their overall quality. A wide variety of algorithms and models is available for fitting of calibration curve data for LBAs. The intent of this article is to review the appropriateness of these approaches to calibration, examine the statistical basis of the models, and provide support for the most relevant of these for application to LBAs. This article was presented in summary form at the Quantitative Bioanalytical Methods Validation and Implementation: Best Practices for Chromatographic and Ligand Binding Assays workshop, held in Arlington, VA, in May 2006.[7]

LBA Characteristics

LBAs possess several characteristics that differentiate them from chromatographic assays.[2] Some of those related to the calibration curve are shown in Table 1. These include a response that is, generally, directly proportional to concentration for chromatographic assays but that may be directly or indirect proportional for LBAs, depending on whether the assay is competitive or noncompetitive. The response for a chromatographic assay is generally directly related to the amount of substance in the detector, while for LBAs the response is the result of interaction of the analyte with an antibody or other binding reagent. Precision for LBAs is generally poorer than for chromatographic assays, because of the role of biological reagents and reactions in LBAs, and the validated assay ranges are commonly considerably narrower. A key difference between LBAs and chromatographic assays is that for chromatographic assays the mean response is a linear function of the analyte concentration in most

Corresponding Author: John W. A. Findlay, Gilead Sciences Inc, 4 University Place, 4611 University Drive, Durham, NC 27707-3458. Tel: (919) 294-7556; Fax: (860) 493-5925; E-mail: john.w.findlay@gilead.com

Table 1. Key Differences Between Chromatographic Assays and Ligand Binding Assays Relating to the Calibration Curve

Chromatographic Assays	Ligand Binding Assays
Direct concentration-response relationship	Direct or inverse concentration-response relationship
High precision	Generally lower precision
Extended assay range	Limited assay range (frequent need for dilution)
Response generally a linear function of analyte concentration	Response generally a nonlinear function of analyte concentration

cases, while for LBAs this relationship is generally nonlinear. This property means that particular attention must be paid to the selection of appropriate algorithms for fitting of LBA calibration curve data.

Numerous data-fitting algorithms have been applied to experimental calibration curve data from LBAs. The properties of some of these algorithms have been reviewed by Rodbard and Frazier,[9] Haven et al,[10] and Dudley et al.[11] The basis of all of these data reduction models is an equation that describes the mean concentration-response relationship, in conjunction with another that describes the relationship between the mean response and the variance of replicate measurements.

SELECTION OF THE PREFERRED CALIBRATION MODEL

For LBAs the typical calibration curve is sigmoidal in shape, with a lower boundary (asymptote) near the background response (nonspecific binding) and an upper asymptote near the maximum response. The 4-parameter logistic (4-PL) model is generally acknowledged to be the reference model of choice for fitting calibration curves of this shape. This function provides an accurate depiction of the sigmoidal relationship between the measured response and the analyte concentration. The equation describing the 4-PL model is as follows:

$$Y = D + \frac{(A-D)}{\left(1+\left(\frac{x}{C}\right)^B\right)} \tag{1}$$

in which Y is the response, D is the response at infinite analyte concentration, A is the response at zero analyte concentration, x is the analyte concentration, C is the inflection point on the calibration curve (IC_{50}), and B is a slope factor. This model has several useful characteristics (Figure 1). The response is monotonic, increasing with concentration if $A < D$ and decreasing if $A > D$ (note that the same flexibility can be achieved by allowing B to be either positive or negative, but by convention B is usually assumed to be greater than 0). The calibration curve is symmetric around the IC_{50} concentration C, with a response at that concentration of $(A + D)/2$. The slope parameter, B, defines the steepness of

the curve. Since the curve is sigmoidal in shape, the slope (ie, first derivative) is changing throughout, but at the IC_{50} the slope is given by $B(D-A)/4C$.

Occasionally, the calibration model needs additional flexibility. In those situations a 5-PL model may work better.[12] This model allows for an asymmetric concentration-response curve by adding an additional parameter, G. The general equation is as follows:

$$Y = D + \frac{(A-D)}{\left(1+\left(\frac{x}{C}\right)^B\right)^G} \tag{2}$$

The asymmetry parameter allows the function to approach the asymptotes at different rates, effectively "stretching out" either the top or the bottom of the curve, depending upon the need (note, in this model $G > 0$, but to achieve maximum flexibility, B can be positive or negative). Generally we advise the use of this model only when the asymmetry is clear. In situations where the asymmetry is small, the addition of a fifth parameter can cause the fitting algorithm to become unstable.

Simpler models are often used. A popular choice is a linear model applied to transformed data, where the transformation used is log(response) vs log(concentration). This approach is an approximately linear transformation, where the quality of the fit depends in large part upon the assay range. Since the linearization is imperfect, an underlying
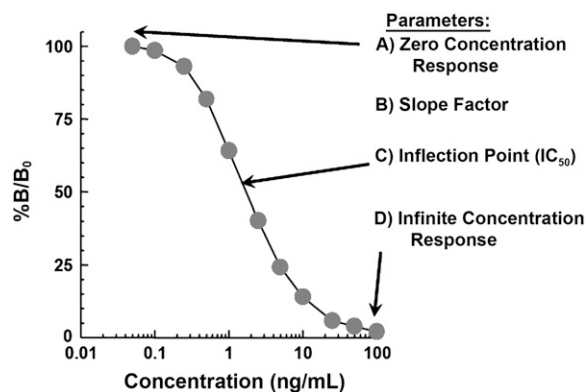

Figure 1. Typical 4-parameter logistic graph for a competitive-format immunoassay.

bias is introduced in which the magnitude of the bias is dependent upon concentration. Typically, the bias is greatest at the ends of the assay range, but it can also be substantial in the "heart" of the calibration curve, since the linear model has had to sacrifice some fit quality in the middle in order to accommodate the tails. Often, to achieve acceptable accuracy, the assay range is severely restricted, much more so than if one were to use a 4-PL model. This, of course, comes at the cost of having to do many more dilutions in order to read samples in the more restricted range. Figure 2 illustrates the point. This figure presents calibration curve data from a monoclonal antibody enzyme-linked immunosorbent assay fit to several different mathematical models. Visual inspection shows greatly varying

goodness of fit of data to the model. For this data set, 4-PL clearly provides the best fit across the entire calibration concentration range. Other models give reasonable fits of the data to portions of the calibration range but are generally characterized by poor fits in some areas, particularly at high and low concentration ranges. These areas of poor fit will have poorer accuracy.

FITTING THE MODEL

The calibration curve is usually fit to the data (concentration-response pairs) using a least squares approach. For nonlinear models such as the 4PL the algorithm is iterative but relatively easy to implement with modern software. The
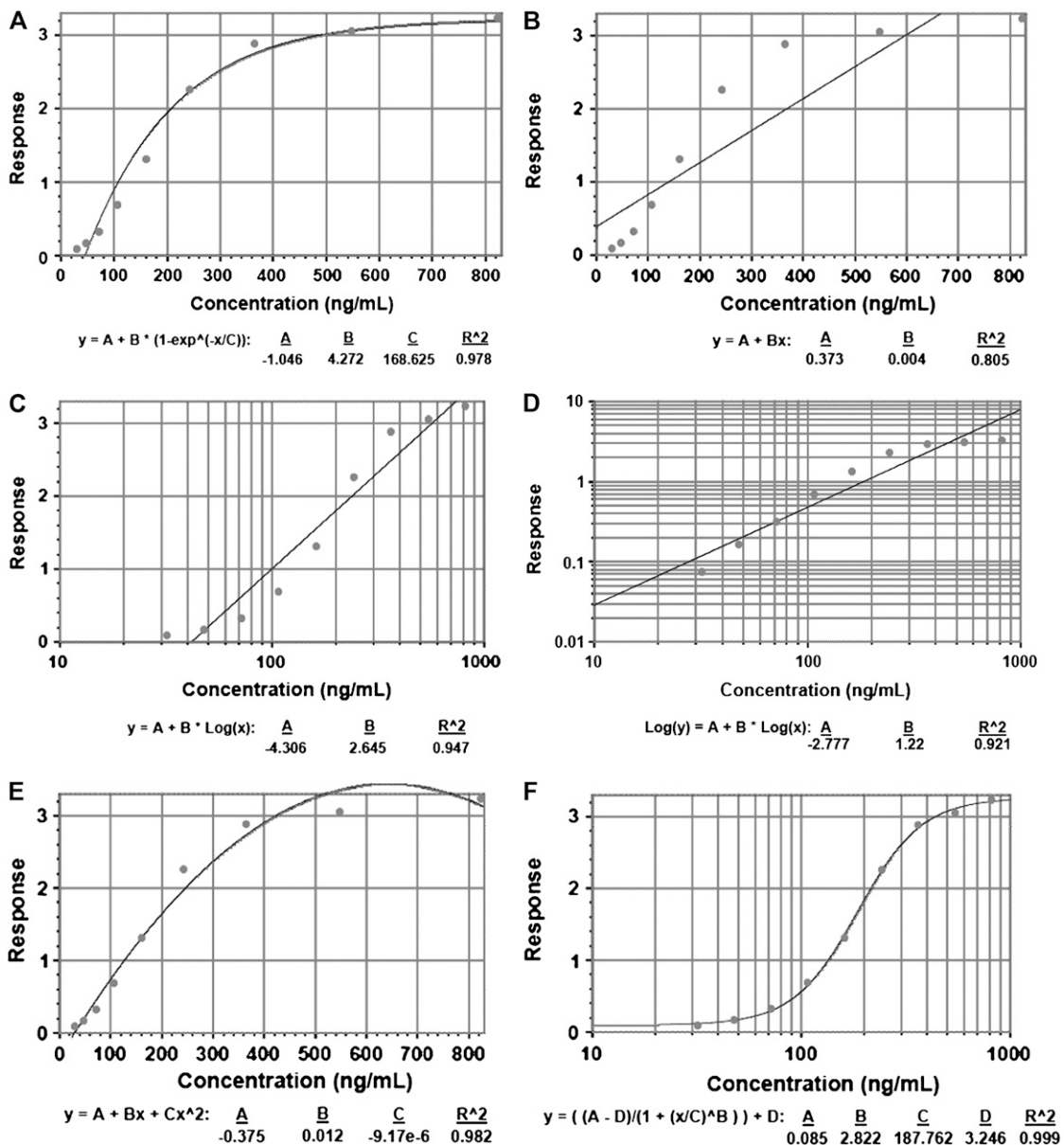

Figure 2. Fit of typical enzyme-linked immunosorbent assay data set for a monoclonal antibody to several mathematical models. Panel A = exponential model, panel B = linear-linear model, panel C = log-linear model, panel D = log-log model, panel E = quadratic model, panel F = 4-parameter logistic model.

algorithm does require starting estimates of the parameters and can be sensitive to poor choices for these starting values. The algorithm can fail to converge if the starting values are too far away from their true values. Usually, visual inspection of the calibration curve can supply reasonable choices. A and D should be chosen based on observed asymptotes. C should be chosen based on rough interpolation of the observed IC_{50} concentration, and B can be estimated by setting the slope (approximate first derivative) of the data points around the IC_{50} concentration equal to the theoretical first derivative (shown earlier) and solving for B. Note that the algorithm tends to be especially sensitive to the starting slope estimate. Care should be taken to get a good initial estimate of B. As a history develops, these estimates can be refined by averaging the observed parameter estimates across runs. For the most part, this work should be done during development and the starting values fixed thereafter.

There is another important aspect of the fitting process that can be somewhat challenging, but any time spent addressing it is time well spent. In LBAs the noise (variance) of the response is generally not constant but changes with the response. This characteristic is termed "heteroscedasticity." The quality of the calibration curve fit can be improved if heteroscedasticity is taken into account, with the essential idea being to place less "weight" on responses that exhibit higher variation. In other words, the calibration curve is more closely aligned to data of low variation and allowed to drift away somewhat from data of higher variation. Weighting in this way results in a broader assay range and demonstratively better accuracy and precision within the range.

To apply the weighting approach, the nature of the heteroscedasticity has to be determined. Fortunately, the variance-response relationship tends to be a smooth, predictable function. Usually the mean-variance relationship can be described by a simple power relationship, one that assumes that the variance, as estimated by the standard deviation, is proportional to some power, θ, of the mean response,[13-15] that is, $\sigma y \propto \mu \theta$. There are several ways to estimate θ. For example, one might determine this relationship based upon the residuals observed after fitting the calibration model, perhaps after using an iterative fitting process that updates the weights after each iteration (ie, generalized least squares). A more direct approach is to simply plot the log of the standard deviation observed from replicates of the calibration standards vs the log of the corresponding mean response (note that taking the log of both sides of the power function above gives a simple linear model with slope θ). The slope of the line is then an estimate of θ. This is not a process that should be repeated with each routine analytical run (ie, in-study). Rather, this relationship should be established during assay development and fixed thereafter.

Once the variance function has been determined, the inverse of the variance estimates should be used as weights in the least squares algorithm, that is, the weights are taken to be $(1/\mu^{2\theta})$. An alternative fitting algorithm that can be easier to implement is referred to as the transform both sides method.[16] In this approach a variance-stabilizing transformation is applied to both the response and the calibration model. The transformed data and model are then fit using the ordinary least squares method without weights. Table 2 shows the correspondence between the appropriate weights and transformations for various values of θ.

Typically, the variance increases in direct proportion to the mean response, that is, $\theta = 1$. This is equivalent to a constant CV across the concentration range. If the constant CV model is appropriate (at least approximately), then one can fit the calibration model by taking logs of both the responses and the model; that is, one fits Equation 3:

$$Log(Y) = Log\left(D + \frac{(A-D)}{\left(1 + \left(\frac{x}{C}\right)^B\right)} \right) \qquad (3)$$

The estimated parameters (A, B, C, and D) are then the best-fit estimates for the untransformed scale.

OPTIMAL ASSAY DESIGN FOR CALIBRATION

The quality of the calibration depends not only on the model and fitting algorithms used but also on the design (plate layout). The design includes the number and spacing of the calibrator concentrations, as well as the location of the calibrators on the plate. For the 4-PL model there are the following recommendations:

- A minimum of 5 calibration concentrations and not more than 8 should be used.
- The calibrators should be prepared and analyzed in duplicate or triplicate.
- The concentration progression should be logarithmic, typically of the power of 2 or 3.

Table 2. Correspondence Between Weights and Transformations

Relationship of σ_y and μ	Power of the Mean (θ)	Weight	Variance Stabilizing Transformation
$\sigma y \propto \mu^2$	2	$1/\mu^4$	Reciprocal
$\sigma y \propto \mu$	1	$1/\mu^2$	Log
$\sigma y \propto \mu^{0.5}$	0.5	$1/\mu$	Square root
σ_y constant	0	1	No transformation

- The midpoint concentration of the calibrators should be somewhat greater than the IC_{50}.
- Anchor concentrations outside the expected validated range should be considered for inclusion to optimize the fit.
- Suboptimal plate layouts should be avoided.

Several authors[17,18] have noted that the issue of optimal spacing of calibrators is essentially one of resource allocation. The goal is to minimize the space allocated to calibrators so as to maximize the space for unknowns, without sacrificing the quality of the calibration curve. Our recommendation of 5 to 8 calibrators ensures that there are enough calibrators to adequately estimate all 4 parameters in the model and still allow for an assessment of the fit quality (lack of fit). Inclusion of more than 8 calibration concentrations barely improves the fit and decreases the capacity of an individual assay for study sample analysis. In fact, if more than 8 calibration concentrations are required, a more flexible model is probably needed.

The use of duplicates at each concentration is recommended. This will help to reduce the noise in the parameter estimates and allow for the occasional missing replicate without adversely affecting the fit. One might even run triplicates, but there is generally little value in replication beyond that. This is especially true if plate layouts are not randomized, since additional replication will not resolve any induced biases caused by a fixed plate layout (see discussion below).

The choice of concentrations is driven, in part, by the practical considerations of ease of preparation and error-free replication from run to run. This suggests the use of serial dilutions and also suggests the use of a fixed dilution ratio (it is preferable for the calibrators to be diluted into the same matrix as the experimental samples to be analyzed). The result is calibrators spaced approximately evenly across the logarithm of the concentration range. Under these conditions Rocke and Jones[17] determined the optimal dilution ratio for calibrators. In their work they show that optimal dilution ratios typically work out to be 2:1 or 3:1. A 2:1 serial dilution with 6 calibrators would yield a series like 1x, 2x, 4x, 8x, 16x, 32x. For a typical decreasing response curve (D < A) Rocke and Jones[17] also showed that for optimal spacing, the midpoint of the calibrator concentrations should be somewhat greater than the IC_{50} concentration. This places more calibrators in the region of smaller variance. For the 2:1 dilution ratio noted above, x should be adjusted so that the desired IC_{50} concentration lies near the 8x dilution.

The inclusion of 1 or more concentrations outside the validated range, near the asymptotes, should be evaluated for potential improvement of the overall fit of the data to the model. These are referred to as anchor concentrations, since they often bring stability to the fit, particularly at the extremes of the acceptable concentration range. Thus, for example, in the 2:1 dilution sequence above, if either extreme (1x or 32x) of the calibration curve is exhibiting poor accuracy or precision, then adding a calibrator (or increasing the dilution level) beyond the concentration of the poorly fitting calibrator should be considered. Note that including a blank matrix is helpful as a quality check, but this sample should not be included in the calibration fit.

The assignment of calibrators and unknowns to wells on a plate also requires careful planning. Positional effects in immunoassays can be substantial, particularly in assays in 96-well plate format. In an ideal world, calibrators and unknown samples would be assigned randomly to wells. As stated in a recent US Pharmacopeia publication describing bioassays,[19] "The use of randomization results in systematic error becoming random error not associated with particular samples or a dilution pattern but distributed throughout the assay." This issue is as relevant for LBA layout as it is for bioassays. Randomization would eliminate any induced bias and better reflect the true underlying uncertainty in the estimated concentrations. One difficulty in reducing this consideration to routine practice is that data capture and reduction software sufficient to implement true randomization has not kept pace with the need. Fortunately, there are compromise designs that, although not completely random, can deftly deal with systematic effects. Figure 3 illustrates 2 disparate approaches to plate layouts. At left is a commonly used layout for an assay in which the calibrators are prepared in duplicate. In this plate configuration the calibrators are always located in the same wells on the upper right of the plate. This layout helps to ensure proper identification of calibrators, but it is a scheme that is susceptible to positional effects on the plate. The layout on the right is a much better choice. In this scheme the calibrators (as well as quality control [QC] samples and study samples) are distributed more widely on the plate, with one of the replicates positioned on the left side and the other on the right. The dilution direction
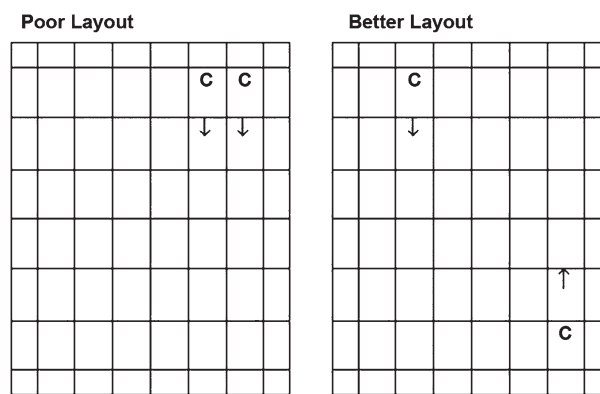

Figure 3. Potential plate layouts in a typical multiwell-plate assay. C indicates calibrator, with dilution increasing in the direction indicated by the arrow.

is also reversed, with increasing dilution going down the plate on the left side and up the plate on the right. Ideally, in this scheme, one would assign the columns to be used randomly each time a plate is run. In this way any location biases would be averaged out. The difficulty, of course, is keeping track of calibrators and other samples in the assay. This is an area where equipment manufacturers and software developers could add value.

EVALUATING THE FIT

A good model, design, and fitting algorithm do not by themselves guarantee the quality of the calibration curve—its suitability needs to be assessed. This assessment should happen early in method development, as a sound calibration curve is central to the development of sound assay characteristics (ie, accuracy, precision, in-process QC, etc).

The key metric is agreement of nominal calibrator concentrations with back-fitted concentrations read off the fitted calibration curve as if they were unknown samples.[20,21] These predicted calibrator concentrations can then be expressed as a percent recovery at each concentration level, 100(BC/NC), or alternatively as their associated percent relative error, %RE = 100(BC – NC)/NC, where BC and NC represent the back-calculated and nominal concentrations, respectively.

One can think of the back-fitted calibrators as surrogate validation samples. As such, the front-line check is whether the calibrators exhibit good accuracy and precision. Some bias and imprecision is inevitable, but the goal is to keep these metrics within an established acceptable range, typically ±15% (20% at the lower limit of quantitation).[3] It is important that this evaluation include data across several runs, as it is difficult to distinguish poor performance from the noise inherent in a single run. Note that the accuracy and precision associated with the calibrators will tend to underestimate the true bias and imprecision. The calibrator coefficient of variation values (CVs) in particular will be an underestimate of the true precision (in routine use the process will be noisier since now 2 measurements are involved, the sample and the calibrator). Nevertheless, the accuracy and precision of the back-calculated standards are a good first check of whether the method will support its requirements. If the method cannot achieve the goals for the back-calculated values, there is little hope of achieving the goals in validation or routine use.

The back-fitted concentrations should also be examined for lack of fit patterns, as poorly fitting models will exhibit a systematic pattern in the %RE with concentration. In fact, plots of the %RE against concentration can be a useful tool in evaluating competing models. Figure 4 illustrates an example of %RE patterns from 2 possible models. In the
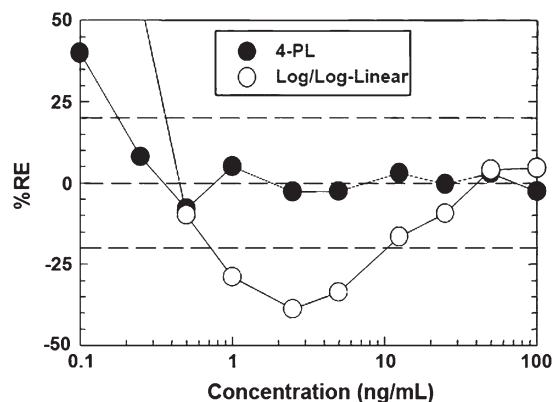

Figure 4. RE plots for 2 mathematical model fits of calibration data. RE indicates relative error; 4-PL, 4-parameter logistic.

figure, the plotted values are average %RE across several runs. The figure illustrates where each model is breaking down and clearly illustrates which model is a better choice.

Other assessment metrics can also be helpful. It is good practice to look at weighted or studentized residual plots.[11] As with %RE, these plots can be helpful in identifying regions where the model may not be fitting well. They can also indicate whether the weighting process is adequate. Ideally, the studentized residuals should not show any lack of fit patterns. There should not be any apparent curvature or any tendency to have some concentration regions exhibit more noise. In particular, hourglass-shaped patterns in the studentized residuals can be clues to poor choices in the weighting function. We do not recommend the use of R^2 to evaluate the fit. As many authors[21] have pointed out, this metric is not very useful, since it is possible to have a "good" R^2 and yet unacceptable bias. Figure 2 illustrates this point, as all of the models illustrated there have high values for R^2 but are clearly not of equal quality.

If poor calibration performance is seen, there are some approaches that can be tried to improve performance. If the performance problem is curvature in the %RE or residual plot, especially inflated %RE at the ends of the plot, an initial response may be to reduce the range of the calibrators. This may work, but at the cost of requiring more routine sample dilutions. A better approach is to expand the calibration range by adding anchor concentrations (see "Optimal Assay Design for Calibration" section above)—even if these points are outside the anticipated assay range. Often this will lead to better calibration performance within the anticipated assay range. Note that anchor concentrations are not required to meet accuracy and precision criteria pre-established for calibrator concentrations within the quantitation range of the assay. Although concentrations outside the range may have poor performance, that is acceptable if the performance within the quantitation range improves. The anchor concentrations are used

to better estimate the asymptotes and thereby improve the fit in the anticipated assay range. If the lack of fit persists, one might try more complex models, for example, the 5-PL model, but that approach is recommended somewhat reluctantly. One should strive first for the simplest model that provides good accuracy.

If imprecision is seen at 1 or more concentrations, then the weights used should be re-examined first. An inadequate weighting function will manifest itself as noisy back-calculated values. If imprecision in the calibrators persists, then a reduction in the assay range may be necessary. Imprecision may also indicate the need for additional replicates in both calibrator and unknown study sample analyses.

ROUTINE USE (IN-STUDY APPLICATION)

The industry standard for monitoring the routine use of an assay relies on results from QC samples placed with the run. If the QC sample results are sufficiently close to expectations, the run results are released. This process is, in part, a check on the calibration (as well as other aspects of the assay). But there are some parameters worth monitoring that speak directly to the quality of the calibration itself. One of the most useful parameters to track is the mean square error (MSE) of the fit. The MSE is a measure of the overall noise about the fitted line. As such, it can function as a first-line check for outliers in the calibrators. Other outlier checks include the range observed from the replicates at each concentration or, alternatively, %RE for each replicate, although these are generally better used as diagnostic measures after an MSE flag has been triggered.

Monitoring the MSE for long-term drift is also useful. Method changes will often show up as jump shifts in the trend lines, indicating that something has altered the characteristics of the assay. One could also monitor the calibration parameters (A, B, C, D) for long-term trends. Trends or shifts in A or D often indicate changes in assay limits of quantitation at the extremes of the range. Trends in B or C often point to fundamental changes in the binding characteristics. It is generally not a good idea to base the release of an assay on these results—they function better as long-term measures of calibration curve performance.

CONCLUSIONS

A wide range of mathematical models has been used in fitting of experimental calibration curve data for LBAs. Selection of a calibration curve model should take into account the characteristics of LBAs that differentiate them from chromatographic assays, in particular the fact that LBA calibration curves are inherently nonlinear. One of the major goals in calibration model selection is to optimize accuracy and precision across the maximum usable calibration range.

The accepted reference model for LBAs is the 4-PL model, with weighting, but this is sometimes extended to include a fifth parameter (5-PL model) to optimize fitting when calibration curve asymmetry is observed. Several assay design parameters, such as placement of calibrator concentrations across the selected range, should be considered to enable optimal application of the 4- or 5-PL models. The fit of the selected model to the experimental data should be evaluated primarily by assessing the %RE of the model-predicted data to nominal or theoretical values. It is important to assess this across the entire desired calibration range to determine the limitations of a particular model. The primary goal is to optimize accuracy and precision across the maximum usable calibration range.

REFERENCES

1. 2006 report, Medicines in development, biotechnology. PhRMA Website. 2006; Available at: http://www.phrma.org/files/Biotech%202006.pdf. Accessed January 14, 2007.

2. Findlay JWA, Smith WC, Lee JW, et al. Validation of immunoassays for bioanalysis: a pharmaceutical industry perspective. *J Pharm Biomed Anal*. 2000;21:1249-1273.

3. DeSilva B, Smith W, Weiner R, et al. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res*. 2003;20:1885-1900.

4. Shah VP, Midha KK, Dighe S, et al. Analytical methods validation: bioavailability, bioequivalence and pharmacokinetic studies. *Pharm Res*. 1992;9:588-592.

5. Miller KJ, Bowsher RR, Celniker A, et al. Workshop on bioanalytical methods validation for macromolecules: summary report. *Pharm Res*. 2001;18:1373-1383.

6. Shah VP, Midha KK, Findlay JWA, et al. Bioanalytical method validation—a revisit with a decade of progress. *Pharm Res*. 2000;17:1551-1557.

7. Vishwanathan CT, Bansal S, Booth B, et al. Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays. *AAPS J [serial online]*. 2007;9:E30-E42.

8. Food and Drug Administration. *Guidance for Industry: Bioanalytical Method Validation.* Rockville, MD: US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research; 2001.

9. Rodbard D, Frazier GR. Statistical analysis of radioligand assay data. *Methods Enzymol*. 1975;37:3-22.

10. Haven MC, Orsulak PJ, Arnold LL, et al. Data-reduction methods for immunoradiometric assays of thyrotropin compared. *Clin Chem*. 1987;33:1207-1210.

11. Dudley RA, Edwards P, Ekins RP, et al. Guidelines for immunoassay data processing. *Clin Chem*. 1985;31:1264-1271.

12. Gottschalk PG, Dunn JR. The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Anal Biochem*. 2005;343:54-65.

13. Box GEP, Hunter WG, Hunter JS, eds. *Statistics for Experimenters.* New York, NY: John Wiley & Sons; 1978.

14. Finney DJ, Phillips P. The form and estimation of a variance function, with particular reference to radioimmunoassay. *Appl Stat*. 1977;26:312-320.

15. Finney DJ, ed. *Statistical Methods in Biological Assay.* 3rd ed. London, UK: Charles Griffith; 1978.

16. Carroll RJ, Ruppert D, eds. *Transformation and Weighting in Regression.* London, UK: Chapman Hall; 1988.

17. Rocke DM, Jones G. Optimal design for ELISA and other forms of immunoassay. *Technometrics*. 1997;39:162-170.

18. Karpinski KF. Optimality assessment in the enzyme-linked immunosorbent assay (ELISA). *Biometrics*. 1990;46:381-390.

19. Singer R, Lansky DM, Hauck WW. Bioassay glossary. *Pharmacopeial Forum*. 2006;32:1359-1365.

20. Karnes HT, March C. Calibration and validation of linearity in chromatographic biopharmaceutical analysis. *J Pharm Biomed Anal*. 1991;9:911-918.

21. Smith WC, Sittampalam GS. Conceptual and statistical issues in the validation of analytic dilution assays for pharmaceutical applications. *J Biopharm Stat*. 1998;8:509-532.