

The five-parameter logistic: A characterization and comparison with the four-parameter logistic[☆]

Paul G. Gottschalk^{*}, John R. Dunn

Brendan Technologies, Inc., 2236 Rutherford Road, Suite 107, Carlsbad, CA 92008, USA

Received 4 February 2005

Available online 23 May 2005

Abstract

Improvements in assay technology have reduced the amount of random variation in measured responses to the point where even slight asymmetry of the assay data can be more significant than random variation. Use of the five-parameter logistic (5PL) function to fit dose–response data easily accommodates such asymmetry. The 5PL can dramatically improve the accuracy of asymmetric assays over the use of symmetric models such as the four-parameter logistic (4PL) function. Until recently, however, the process of fitting the 5PL function has been difficult, with the result that the 4PL function has continued to be used even for highly asymmetric data. Various ad hoc modifications of the 4PL method have been developed in an attempt to address asymmetric data. However, recent advances in numerical methods and assay analysis software have rendered easier the fitting of the 5PL routine. This paper demonstrates how use of the 5PL function can improve assay performance over the 4PL and its variants. Specifically, the improvement in the accuracy of concentration estimates that can be obtained using the 5PL over the 4PL as a function of the asymmetry present in the data is studied. The behavior of the 5PL curve and how it differs from the 4PL curve are discussed. Common experimental designs, which can lead to ill-conditioned regression problems, are also examined.

© 2005 Elsevier Inc. All rights reserved.

Keywords: 5PL; 4PL; Immunoassay; Bioassay; Data reduction; Statistical analysis; Dose–response curve; Curve model

Biologists perform immunoassays to determine the concentration of an analyte in a sample. Immunoassay techniques use antibody–antigen binding to quantify the concentration of an analyte. This is done indirectly by measuring a response that is proportional to the signal intensity of some type of label. Depending on whether the immunoassay is competitive or immunometric, the label is chemically attached to the analyte or the antibody, respectively [1]. Bioassay is a broader term which refers to any type of biological activity that is measured as a function of the dose level of some sub-

stance. Bioassays are often a central part of potency studies where the parallelism and relative potency of two dose–response curves is of primary interest [2].

To quantify 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 [3,4].

Ideally, the standard curve would be identical to the *true curve*: the curve that expresses the concentration versus response relationship without any degradation 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. For practical

[☆] This work was supported solely by Brendan Technologies, Inc.

^{*} Corresponding author. Fax: +1 760 929 7504.

E-mail addresses: pgottschalk@brendan.com (P.G. Gottschalk), jdunn@brendan.com (J.R. Dunn).

reasons, only a limited number of samples can be run in an assay, and the true curve must be estimated from a limited number of noisy responses.

Because there cannot be a standard at every concentration, a means of interpolating between standards is necessary to estimate the dose–response relationship between standards. 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*. The curve model, which typically describes a family of curves via one or more parameters, is then fitted to the assay data by adjusting the curve model's parameters to obtain the one curve from the family that best describes the assay data.

The goal of this process is to obtain concentration estimates that are as accurate as possible by improving the quality of fit of the standard curve. 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 have a sigmoidal, or “S”, shape, if data are taken over a wide enough range of concentrations. If a straight line is used as the curve model to fit such data, much of the error in fitting the data would be due to lack-of-fit error. This is because a straight line cannot fit the curving sigmoidal shape of the data.

The ideal curve model is able to accommodate the variations of the true curve with the smallest number of parameters. When a curve model is unable to follow the true curve, bias is introduced into the dose estimates and their accuracy is reduced. Furthermore, because overparameterized curve models tend not to represent the true curve very well between data points, accuracy is reduced in these models, even though such curve models may “fit” the data quite well. Overparameterized models have the further problem that they are more susceptible to noise in the responses, which reduces the precision of dose estimates.

A good curve model should possess three properties. First, such a curve model must do a good job of approximating the true curve. If the curve model does not do this, 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. Second, a good curve model must be able to average out as much of the random variation in the assay data as possible to produce concentration estimates that are

distorted by pure error as little as possible. Third, a good curve model must be able to predict concentrations well at points between the standard points and not just at the fitted data points.

Many functions have been tried as curve models for immunoassays, but few of them possess all of these properties. As mentioned above, the straight line curve model [3] cannot follow the curves of the sigmoidal shape of immunoassay true curves. Fitting lines in logit-log space [5] works better, but this technique effectively models only symmetrical data, and the logit transform worsens the character of the noise in the standard data. The four-parameter logistic (4PL)¹ function [6–8] is widely used in practice and is closely related to the linear logit-log model (a 4PL curve transforms to a straight line in logit-log space). However, like the logit-log model, the 4PL model cannot effectively model asymmetric data. The mass action model [9], the only physically based model, has many parameters, which makes it less able to average out noise than models with fewer parameters. It has also been shown that certain approximations to the mass action model are approximately equivalent to the logit-log model [10], which, as noted above, implies approximate equivalence to the 4PL curve. Like the 4PL, this approximation to the full mass action model is unable to fit asymmetric data.

Cubic splines [11,12], have been used when the data are too asymmetric to use a four-parameter fit. Cubic splines are strictly interpolative. That is, a cubic spline is designed to pass exactly through the data points instead of attempting to estimate the true curve. Since an interpolative 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 such a spline fit. This means that a spline fit performs no averaging of the data to reduce random variation. Some splines, called smoothing splines, do perform some averaging of the data, but the number of degrees of freedom in such spline models still tends to be large. Furthermore, splines, be they interpolative or smoothing, often do 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 a lot of 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 is able to average out the random variation.

A method sometimes used to handle asymmetry is the log 4PL method. This method takes advantage of the

¹ Abbreviations used: 5PL, five-parameter logistic; 4PL, four-parameter logistic; SSE, sum of squared errors.

fact that taking the logarithm of the response of some asymmetric assay data can make the data more symmetrical and therefore better suited to a 4PL fit. This approach can improve the fit when the low response end of a sigmoidal curve has a shorter “tail” than the upper end of the curve. However, taking the log of the responses of sigmoidal data where the responses on the low response end of the curve approach the lower asymptote more slowly than the responses on the high response end makes the data *more* asymmetric. Since this type of behavior is encountered in immunoassay and bioassay data more often than the former, this approach is not satisfactory for routine data reduction.

Many other curve models have been tried with varying degrees of success. Space does not permit further discussion here, but Rodbard [13] provides a useful overview of some of the dose–response models that have been discussed in the literature.

The 4PL model can be extended by adding a fifth parameter that controls the degree of asymmetry of the curve [14,15]. This model is called the five-parameter logistic (5PL) model [4,8] and is available commercially [16]. The authors have compared the 5PL to the 4PL in many thousands of dose–response curves from a wide variety of immunoassay and bioassay technologies. With the extra flexibility afforded by its asymmetry parameter, the 5PL model is able to virtually eliminate the lack-of-fit error that occurs when the 4PL is fitted to asymmetric dose–response data. This paper focuses on the use of the 5PL model for fitting standard curve data. The 5PL model provides an excellent compromise between overparameterized models that can fit data closely at the cost of a large variance in the predictions and underparameterized models that suffer from large lack-of-fit errors. This is because the fewest number of parameters that any general asymmetric sigmoidal function can possess is five: one for the upper asymptote, one for the lower asymptote one for the overall length of the function’s transition region, one for the location of the transition region, and one for the degree of asymmetry. It is unlikely that any function with less than five parameters will have the flexibility necessary to produce a high-quality fit to asymmetric sigmoidal dose–response data.

Curve fitting

Once a curve model has been chosen, the next step is to select the one curve that best explains the data out of the entire family of possible curves that the curve model can produce. This is accomplished by fitting the curve. Fitting a curve is the process of adjusting the free parameters of the function until the function’s parameters approximate the assay’s true curve better than any other parameter set.

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* (SSE) when the distribution of the responses at a specific concentration is approximately normal [17,18]. The weighted sum of squared errors is the sum of all of the squares of the differences between the observed standard responses (y_i) and the response predicted by the curve model (\hat{y}_i), weighted by the inverse variance ($w_i = 1/\text{Variance}(y_i)$) of the responses at that concentration.

$$\text{SSE} = \sum_{i=1}^N w_i [y_i - \hat{y}_i]^2. \quad (1)$$

Regression, or fitting, is the process of minimizing the SSE by adjusting the curve’s parameters. The best-fitting curve is the one whose parameters produce the smallest SSE. Note that the differences $y_i - \hat{y}_i$ are being called errors here but the word “residuals” is often used, and SSE could just as well be called the “weighted sum of squared residuals.”

Weighting

Choosing proper weights for the squared residuals in Eq. (1) 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. When this is done, the best-fitting curve, i.e., the one that minimizes SSE, will also be the optimal maximum likelihood estimate of the true curve. This mathematical result stands to reason: weighting the squared errors in this way causes the fitting procedure to adjust the curve to be tighter around those standard responses with the smallest variance (error), typically those with the smallest responses, and looser around those standard responses with the largest variance, typically those with the largest responses. By weighting this way, the information from the noisier and less noisy standards is optimally combined, leading to the most accurate estimate of the true curve and therefore resulting in the most accurate concentration estimates. That is why 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 three or four 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 [1], with the result that the kinetic variations in the reaction change as the ratio of analyte to tracer and binder changes.

The variance of the standards can usually be approximated by a power function of the response,

$$\text{variance} = V(r) = A(\text{response})^B, \quad (2)$$

where A is a function of the magnitudes of the responses and the average noise level, and B falls in the range 1.2–2.0 for most immunoassays [4,10] and the range 0.8–2.2 for bioassays (unpublished observation). Data from dose–response curves where the response variances are not constant (heteroscedastic data) will typically have $B > 0$ in Eq. (2). 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 [19–21].

It should be noted that the range of applicability for the variance model described by Eq. (2), and for the curve model itself often will not include extreme concentrations much beyond the minimum and maximum detectable concentrations [22]. These regions often become nonmonotonic beyond the random fluctuations predicted by the variance function. Furthermore, the fact that the form of Eq. (2) is such that the predicted variance approaches zero as the response approaches zero is clearly not physically correct. In some cases, adding a constant minimum variance parameter to Eq. (2) will improve the variance fit of very low responses [23].

Fit probability

Once the curve model has been fitted by adjusting its parameters to minimize SSE, a means of assessing the quality of the curve fit is necessary. The residual variance is the metric typically used to assess curve fits. The residual variance is the SSE divided by the number of degrees of freedom (number of data points minus number of parameters). Although the residual variance allows the quality of fit to be measured, it provides no information about how good or bad a fit is relative to other assays in the (hypothetical) infinite population of assays performed under exactly the same conditions. Under the assumptions that the responses of the individual standard concentrations are approximately normally distributed and that the regression's expectation surface is approximately

linear in the neighborhood of the best fit, it can be shown that SSE obeys a χ^2 distribution with the number of degrees of freedom equal to the number of curve points minus the number of parameters [18]. This distribution allows us to determine the probability that curve fits having a particular value of SSE, or worse, will occur.

The p value of the SSE can be viewed as the fraction of an infinite number of assays that, if performed under exactly the same conditions, would be expected to have a worse curve fit, i.e., a larger SSE, than the curve fit of the assay under consideration. This probability, which can be obtained from the cumulative χ^2 distribution, can be called the *fit probability*.

Five-parameter logistic

The need for a curve model that accommodates asymmetry has been necessitated by improvements in instrument and laboratory technology. The development of sandwich assays led to dose–response curves that tend to be more asymmetric than earlier types of assays. Additionally, because of improvements in signal-to-noise ratios, asymmetry is an issue even for assays whose dose–response relationships are not as highly asymmetric. The reason for this is that even modest levels of lack-of-fit error caused by fitting mildly asymmetric data to a symmetric model can dominate the pure error due to random variation in low-noise modern assays.

For symmetric immunoassay and bioassay data, it can be argued that no curve model has been as successful as the four-parameter logistic function. Despite its utility, the 4PL function is generally not an adequate curve model for much of the asymmetric response data commonly observed in immunoassay and bioassay applications. The five-parameter logistic function, which includes a fifth parameter, permits asymmetry to be effectively modeled.

The 5PL function is given by

$$y = f(x; \mathbf{p}) = f(x; a, b, c, d, g) = d + \frac{(a - d)}{\left(1 + \left(\frac{x}{c}\right)^b\right)^g}, \quad (3)$$

where $\mathbf{p} \equiv [a, b, c, d, g]$ is the parameter vector of the 5PL logistic and where the domain of the parameters is restricted so that $c > 0$ and $g > 0$. The 4PL function is obtained by setting $g = 1$.

After the 5PL model has been fitted to standard data, estimates for single-dilution unknown concentrations x can be obtained from unknown responses y using the inverse of Eq. (3):

$$x = f^{-1}(y; \mathbf{p}) = f^{-1}(y; a, b, c, d, g) = c \left(\left(\frac{a - d}{y - d} \right)^{\frac{1}{g}} - 1 \right)^{\frac{1}{b}}. \quad (4)$$

Advantages of the 5PL model compared to the 4PL model

The 5PL model is able to eliminate most of the lack-of-fit error present in fitted 4PL models. Fig. 1 shows the result of fitting the same set of asymmetric data with a 5PL curve model and a 4PL curve model. From the graph, it is apparent that the 5PL curve fits the data better. Table 1 shows the fit statistics for the two curve fits in Fig. 1: SSE, 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 quality of the 4PL fit is very bad.

Properties of the 5PL model

All curves generated using Eq. (3) 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.

Because the 4PL function ($g = 1$) is point symmetric on semi-log axes about its midpoint ($c, a + d/2$), it can be shown that, of the cases in Table 2, cases 1 and 4 of are equivalent and cases 2 and 3 are equivalent by making the substitutions $a \rightarrow d$, $d \rightarrow a$, and $b \rightarrow -b$. Because of this, the literature has adopted two conventions for eliminating this redundancy in the parameterization of the 4PL function: either fix $a > d$ and allow the sign of

Table 1

Fit statistics for the 5PL and 4PL fits shown in Fig. 1

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

Table 2

Relationship between the order of a and d , the sign of b , and the slope of the monotonic 5PL function

Case No.	Order of a and d	Sign of b	Slope
1	$a > d$	$b > 0$	–
2	$a > d$	$b < 0$	+
3	$a < d$	$b > 0$	+
4	$a < d$	$b < 0$	–

b to determine the slope of the logistic or fix $b > 0$ and allow the ordering of a and d to determine the slope of the logistic. In the 4PL case, neither of these conventions restricts the range of functions that can be produced.

In contrast, the 5PL ($g \neq 1$) function has no symmetry. Therefore, all the cases of Table 2 yield distinct functional forms. Cases 1 and 4 of Table 2 are both suitable for modeling decreasing dose–response data, while cases 2 and 3 of Table 2 are both suitable for modeling increasing dose–response data. Since they are distinct, one of the cases will usually be a better model of the dose–response relationship than its counterpart, and, to find the “best” model, both should be considered.

Fig. 2 shows semi-log plots of the families of curves that can be generated from Eq. (3) by varying one parameter at a time. The figure makes several characteristics of the 5PL function apparent. The function approaches a horizontal asymptote as the dose approaches zero, and it approaches a horizontal asymptote as the dose approaches infinity. Between the asymptotic regions of the curve is a transition region which contains a single inflection point. On either side of the inflection point the curve will approach the left and right asymptotes at different rates unless $g = 1$.

Consideration of Fig. 2 gives some insight into how the five parameters of the 5PL function affect the resulting curves. Parameters a and d control the position of the curve’s horizontal asymptotes. However, examining the behavior of the curve in its asymptotic regions provides additional insights, especially into the roles of b and g . Table 3 provides approximations of the 5PL function near its asymptotes. These approximations are derived in Appendix A. From Table 3 it is evident that the 5PL curve behaves like a power curve as it approaches its asymptotes. When approaching the “ a ” asymptote, only parameter b controls the rate of approach to the asymptote. However, when approaching

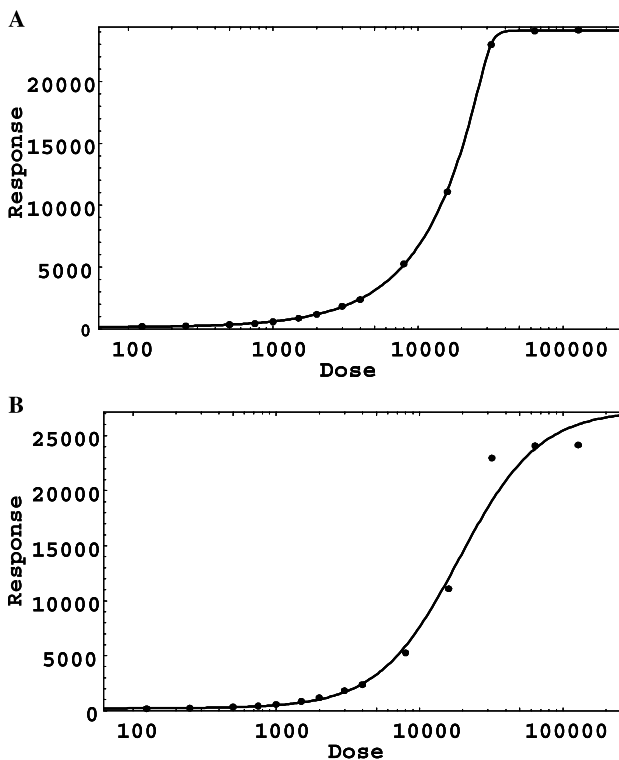


Fig. 1. (A) Result of a 5PL fit on asymmetric immunoassay data; (B) result of a 4PL fit on the same data.

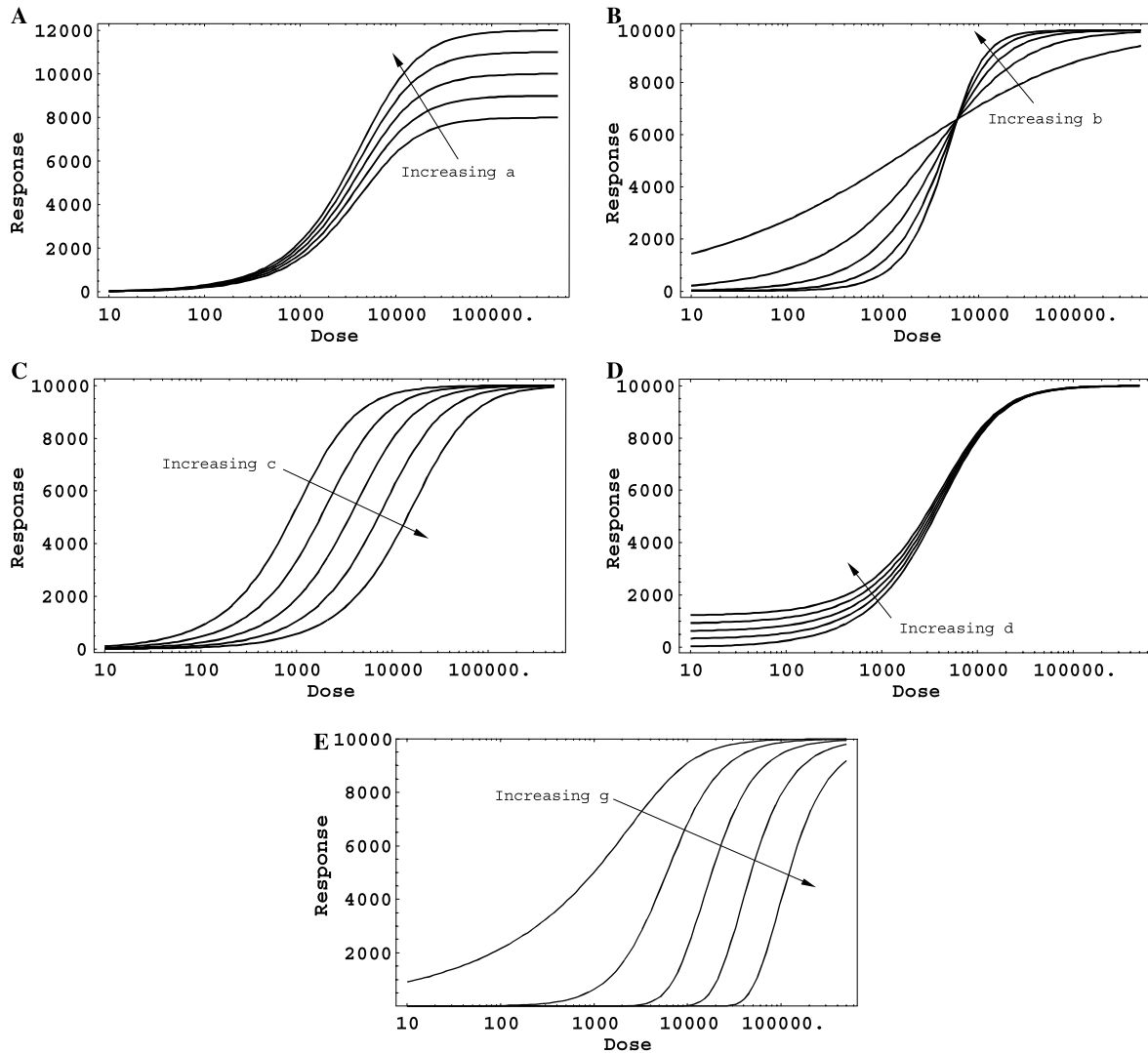


Fig. 2. (A)–(E) Effects of varying the a , b , c , d , and g parameters, respectively, of a 5PL function. In this example, $a > d$ and $b < 0$. In (B), the common point of the curves occurs when the dose level equals parameter c .

Table 3
Asymptotic behavior of the 5PL function

Region	Approximation
“ a ” asymptote	$y \approx a - g(a - d)(x/c)^b$
“ d ” asymptote	$y \approx d + (a - d)(x/c)^{-bg}$

the “ d ” asymptote, the rate of approach is controlled by the product bg . This leads to the result that if h/b is substituted for g in the 5PL formula, Eq. (3), the resulting parameterization with regard to of the five parameters a , b , c , d , and h permits the rate of approach to the two asymptotes to be controlled independently using parameters b and h . This can improve the performance of some fitting algorithms. A further observation is that when all the data fall in the asymptotic region of the curve, a three-parameter approximation to the logistic curve

$$y \approx r + sx^t \tag{5}$$

can be used to model the data instead of a 5PL or 4PL curve.

Table 4 summarizes the effect of each parameter on the 5PL function. The geometric interpretation of parameters a and d as the positions of the 4PL curve’s asymptotes can also be applied to the 5PL curve. However, the 4PL geometric interpretations of b and c do not carry over to the 5PL.

In the 4PL curve, c is the location of the curve’s ED₅₀ point, the location of its inflection point (point of greatest slope), and the location of its point of symmetry (the ED _{α} dose of a logistic is the dose where the curve crosses the response that is α percent of the distance between the “ d ” asymptote to the “ a ” asymptote and is given by $x = c((100/\alpha)^{1/g} - 1)^{1/b}$). The ED₅₀ dose for a 5PL curve is then $x = c(2^{1/g} - 1)^{1/b}$. Thus, the ED₅₀ dose for a 5PL can be larger or smaller than c , depending on the values

Table 4
Effects of the 5PL parameters

Parameter	Effect on 5PL curve
a	Controls the position of the “ a ” asymptote. With b , size relative to d controls slope of curve. The curve approaches the “ a ” asymptote for small concentrations when $b > 0$ and for large concentrations when $b < 0$
b	Magnitude controls the rapidity of the curve’s transition between asymptotes. Its sign, along with the order of a and d , controls slope of the curve. Solely controls the rate of approach to the “ a ” asymptote and jointly with g controls the approach to the “ d ” asymptote
c	Controls position of the transition region in dose
d	Controls the position of the “ d ” asymptote. With b , size relative to a controls slope of curve. The curve approaches the “ d ” asymptote for small concentrations when $b < 0$ and for large concentrations when $b > 0$
g	Jointly with b controls the rate of approach to the “ d ” asymptote

of b and g . Similarly, the dose at which the inflection point of the 5PL curve is viewed in logarithmic dose coordinates is $x = c(1/g)^{1/b}$, which can also be larger or smaller than c depending on the values of b and g . These facts, along with the fact that the 5PL curve has no point of symmetry, implies that none of the common geometric interpretations that apply to parameter c of a 4PL curve apply to parameter c of a 5PL curve.

Parameter b is often called the “slope parameter” of the 4PL curve because when a 4PL curve is plotted in normalized response coordinates $y' = \frac{y-d}{a-d}$ against log dose, then the slope of the 4PL curve at dose c is $-b/4$. By comparison, the slope of the normalized 5PL curve at dose c is $-bg/2^{1+g}$, which depends more strongly on g than on b . Thus, while the slope at c is proportional to b , it would be misleading to call b the “slope parameter” of the 5PL curve since g has at least as much influence on the slope as b .

Fitting the 5PL model

The 5PL curve can be much more difficult to fit to data than the 4PL curve. This inability of standard algorithms to reliably fit the 5PL curve has undoubtedly limited the use of the 5PL curve for modeling immunoassay data.

Fitting a curve to data is equivalent to minimizing the SSE of Eq. (1). If the 5PL model, Eq. (3), is substituted into Eq. (1),

$$\text{SSE}(\mathbf{p}) = \text{SSE}(a, b, c, d, g) = \sum_{i=1}^N w_i \left[y_i - \left(d + \frac{(a-d)}{\left(1 + \left(\frac{x_i}{c}\right)^b\right)^g} \right) \right]^2 \quad (6)$$

is obtained. Fitting the 5PL model amounts to finding the values of the parameters a , b , c , d , and g that minimize $\text{SSE}(a, b, c, d, g)$ in Eq. (6).

The reason that fitting a 5PL curve is so much more difficult than fitting a 4PL curve is that the 5PL regression problem is more likely to be ill-conditioned than its

4PL counterpart. Ill-conditioning in a regression context occurs when $\text{SSE}(a, b, c, d, g)$ is close to having infinite minima instead of a single minimum, so that numerical algorithms, which have limited numerical precision, may not be able to distinguish the multitude of near minima from a true minimum. Since most algorithms cannot by themselves distinguish a local minimum from a global minimum, the same phenomenon can occur wherever $\text{SSE}(a, b, c, d, g)$ locally looks like a minimum. With regard to the SSE of Eq. (1), this means that for some region around the true minimum, \mathbf{p}_{\min} , there is a subspace of the parameter space where the function $\text{SSE}(\mathbf{p})$ hardly changes. For example, if $\text{SSE}(\mathbf{p})$ were a two-dimensional function instead of a five-dimensional function, one could say that the valley in which \mathbf{p}_{\min} is located has a very flat bottom.

The most common reason that a 5PL model can become ill-conditioned is that it is sometimes possible to wildly adjust the 5PL parameters so that the 5PL curve itself hardly changes in the places where the data are, even though the curve may be drastically different in places where there are no data. When the parameters of the curve can be changed so that the curve hardly changes relative to the data points, the residuals also hardly change. The value of SSE is then largely unaffected by such parameter changes.

Such a situation occurs when an assay’s data fall in the region of the logistic curve where the asymptotic approximations of Table 3 apply. A common example occurs when $a > d$, the curve is upsloping, i.e., $b < 0$, and the data fall within the lower asymptotic region of the curve. From Table 3, the equation of the curve in its lower asymptotic region is approximately $d + (a-d)(x/c)^{-bg}$, which can be rewritten as $d + \left(\frac{a-d}{c^{-bg}}\right)x^{-bg}$. This has the form $r + sx^t$, where we make the identifications $r = d$, $s = \frac{a-d}{c^{-bg}}$, and $t = -bg$. As long as r , s , and t do not change, the curve will stay the same. Note that d is unambiguously determined since any changes to it will change the curve. However, *any* b and g can be chosen so long as their product remains constant since doing so will prevent s and t from chang-

ing. Thus, b could be very small and g very large, or vice versa, and anything in between. Similarly, to keep s from changing, any a and c can be chosen such that the quantity $\frac{a-d}{c-bg}$ remains constant. In this case, if no constraints are placed on their values, it is not uncommon for fitting algorithms to adjust a and c to become extremely large and sometimes overflow, as the fitting algorithms attempt to obtain minute improvements in SSE. It is worth noting that this situation can also occur when using a 4PL model.

Fig. 3 shows three curves that are nearly identical in the vicinity of the assay data, with the middle curve being the best-fitting curve. The data fall entirely in the asymptotic region of these curves. In the asymptotic region, the three curves in Fig. 3 have nearly identical values of r , s , and t . Because the curves are nearly identical near the data points, $SSE(a, b, c, d, g)$ will also be nearly identical. However, the parameters a , b , c , and g are very different for each of these curves. The differences in the parameters is apparent if a larger range of doses is plotted, as in Fig. 3B. Indeed, there is a two-dimensional nonlinear subspace contained in the five-dimensional parameter space of all possible values of a , b , c , d , and g where $SSE(a, b, c, d, g)$ is nearly constant, leading to an ill-conditioned regression.

Fig. 4 shows another example of a common situation where a 5PL regression is ill-conditioned. In this case, it is because b and g can be varied such that the product bg

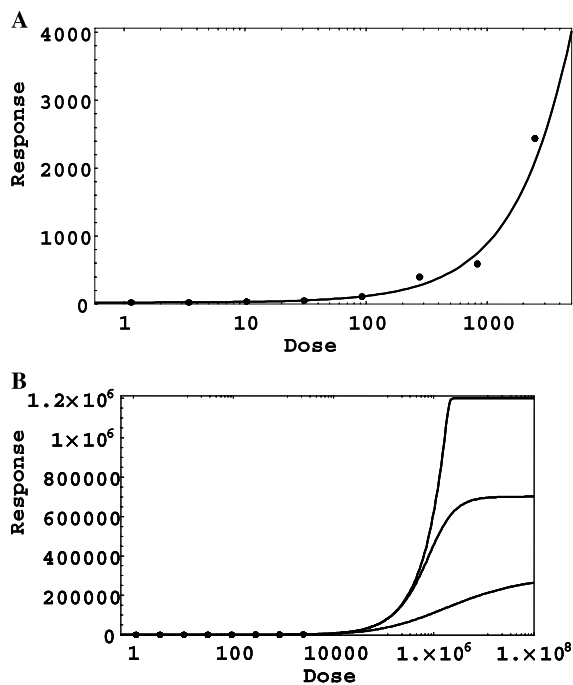


Fig. 3. (A) and (B) Each plot three curves with the same assay data sets. The best-fitting curve is the middle curve plotted in (B). The values of the parameters (a, b, c, d, g) of the curves are bottom (300000, -0.5 , 465451, 20.2283, 1.9046054), middle (700832.2, -1.7960017 , 1134595.4, 20.22823, 0.53023486), and top (1200000, -20 , 1995777.868, 20.22823, 0.047615135).

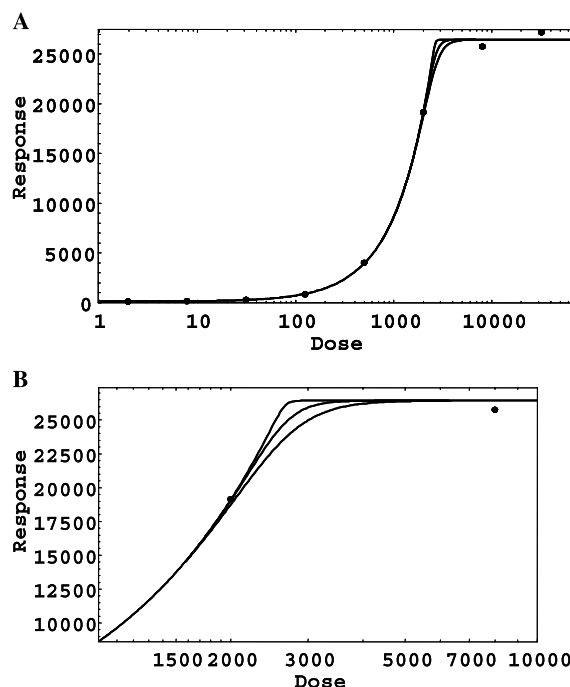


Fig. 4. (A) and (B) Each plot three curves from the same assay data set. Parameters a , c and d are the same but the values of b and g are varied such that bg is constant. The parameters (a, b, c, d, g) of the curves are bottom (26435.374, -7 , 2648, 114.42905, 0.16519860), middle (26435.374, -12 , 2654.1668, 114.42905, 0.09636585), and top (26435.374, -50 , 2654.1668, 114.42905, 0.02312780).

is constant, thereby keeping the lower portion of the curve invariant. In addition, a “knee” in the curve exists between two data points, permitting the sharpness of the knee, which is controlled in this case by the value of b , to be varied without changing the curve appreciably in the vicinity of the data points on either side of the knee.

For the assay data shown in Fig. 4, the transition between the upper and the lower asymptote is quite rapid. Hence, the knee in the curve is a reasonably accurate representation of the behavior of the underlying true response curve. However, if the placement of standards is not well designed, the best-fitting 5PL curve may have a knee, but the true dose–response curve may not. In such a situation the knee is artificial and does not represent the true behavior of the response curve. For example, given positively sloping data where all but the highest point of the standards are reasonably well fitted by the asymptote of the 5PL curve and the last standard, due to random variation, happens to be lower than the extrapolation of the asymptotic approximation, the best-fitting 5PL curve will have an artificial knee because a curve with a knee can fit the last data point perfectly without affecting the shape of the asymptotic part of the curve that fits the rest of the data well. Reducing the spacing of the standards at the high-noise end of the curve is the best way to reduce the likelihood of this behavior.

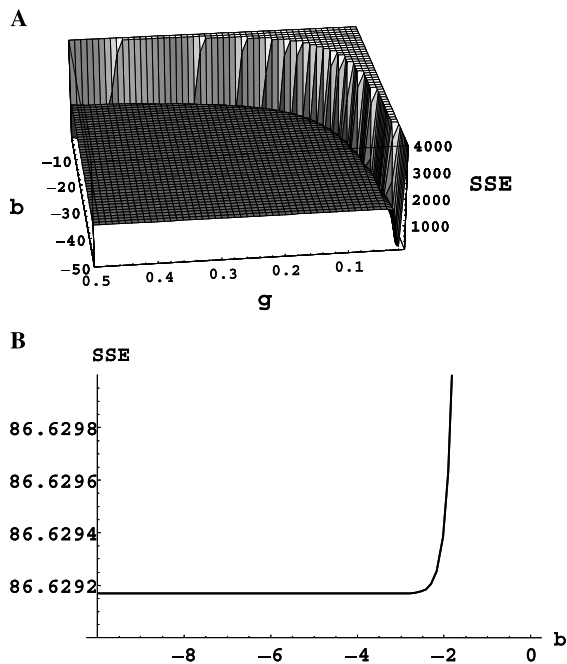


Fig. 5. (A) Plot of SSE versus the 5PL parameters b and g for the assay of Fig. 4. The other 5PL parameters, a , c , and d , have been held constant at the values that minimize SSE. (B) Plots the minimum value attained by SSE for each value of b , i.e., the depth of the “trench” in (A) as a function of b .

Fig. 5A shows a plot of SSE for the data in Fig. 4. The parameters a , c , and d are held constant at the values that minimize SSE, and the parameters b and g are varied. There are three main features apparent in the plot: two “plains” where SSE is nearly constant and a hyperbola-shaped “trench” separating the two planes. The shape of the trench corresponds to the curve $bg = K$, where K is some constant. As we have seen, the lower asymptote of the curve in Fig. 4 will stay largely invariant if the product bg is invariant, so the hyperbolic shape of the trench stands to reason.

The topology of the SSE surface shows the importance of finding reasonable starting points for fitting algorithms. If a starting point that is on one of the plains is chosen, most fitting algorithms will have difficulty locating the trench and, therefore the minimum that lies within it, since the plains have no local information that a fitting algorithm can use to locate the trench.

The problem of ill-conditioning is illustrated in the plot shown in Fig. 5B. There, the value of SSE at the lowest point of the trench is plotted as a function of b . It is apparent from the plot that the value of SSE appears to be nearly constant for $b < -3$. In fact, within the limits of machine precision, the curve continues to descend out to about $b = -4$, but the scaling of the plot in Fig. 5B does not allow this to be seen. Thus, there is no unique minimum of SSE in this case.

The behavior of fitting algorithms in response to an assay for which regression is ill-conditioned is unpredict-

able since it depends specifically on how the algorithm is coded to handle the various numerical instabilities that can occur. Common behavior falls into one of five categories: algorithm failure, lack of convergence, never finishing (“hanging”), returning with an incorrect result, or returning with the correct result. Of these, returning with an incorrect result is the most serious because there is no way to distinguish this case from a successful correct result.

Simulation experiments

Simulation experiments were run with the aim of quantifying the actual improvement in concentration accuracy that accrues from using a 5PL curve model versus a 4PL curve model as a function of the level of asymmetry in the data. In these experiments, data were simulated by drawing from a zero mean, unit variance normal random number generator and computing the simulated response from the formula

$$R = f(x; \mathbf{p}) + N(0, 1)\sqrt{V(f(x; \mathbf{p}))}, \quad (7)$$

where R is the random variable representing the response at dose x , $f(x; \mathbf{p})$ is the response of the simulated underlying 5PL curve of Eq. (3) at concentration x , $V(r)$ is the variance model as a function of mean response from Eq. (2), and $N(0, 1)$ represents a zero-mean, unit variance random variable that is produced using the computer’s random number generator. This method is more completely described in [2].

There are many design parameters that control the precise nature of the data obtained from the assay simulation: five for the shape of the underlying 5PL curve, two for the power law variance model, and several more for describing the number and locations of the simulated standards. Thus, the number of parameters is well over a dozen. For the simulation to be useful, it must represent a significant fraction of the space of realizable assays. However, to explore such a high-dimensionality space is not practical. Fortunately, many of the dimensions of this space are illusory and can be normalized out. Those that remain are very constrained by observational data and can be set to typical values. The result is that the simulation below can be taken to represent a wide range of realistic assays. Furthermore they can be extrapolated to more extreme situations since the underlying principles leading to the results do not change.

The simulated assays consisted of nine standards. The highest standard had a concentration of 500 and each succeeding standard had a concentration 1/2 of the concentration of its predecessor. The parameters of the underlying curve were chosen so that parameter a of the 5PL was 1. This can be done without loss of generality since the precise units of the vertical scale are irrel-

evant so long as the variance is properly scaled. The size of parameter d relative to parameter a was chosen to be typical. The size of parameter A of the variance model was chosen so that the resulting CVs of the standards were also within typical ranges. Parameter B of the variance model was set to 1.5 [19].

Parameters b and c , which control the scaling and translation of the 5PL curve, were chosen so that the curve's ED_{10} point and the curve's ED_{99} point occurred at the concentrations of the lowest standard and the highest standard, respectively. This procedure kept the transitional section of the curve positioned over the standards so that the standard concentrations did not have to be adjusted as parameter g was changed for each simulation.

For each of 101 logarithmically equally spaced values of parameter g between 0.1 and 10 (the range most assay curves fall within), 200 data sets were simulated using Eq. (7) in the procedure above. For each data set, the best-fitting 5PL curve and the best-fitting 4PL curve were determined. The mean concentrations of each standard were estimated by computing the sample mean of the backfit concentrations from the 200 simulated

curves. The error of the mean backfit concentration, relative to the actual standard concentration, was then computed and plotted as a function of parameter g for each standard for both the 5PL and the 4PL curves. The results of this procedure are shown in Fig. 6, where the error for each of the standards is plotted against parameter g of the underlying curve.

Except for the regions of the plots of Std 7, Std 8, and Std 9 where there is too much noise to draw any conclusions, Fig. 6 shows clearly that the accuracy of concentration estimates obtained using the 5PL curve model is superior to the accuracy of the 4PL concentration estimates when asymmetry is present. When there is no asymmetry present ($g = 1$), the plots show that the error in the estimates becomes nearly identical, as one would expect. In practice, symmetrical curves are not usually observed.

Many of the plots show that the 4PL error reaches a maximum in the interval [1, 1]. For example, the 4PL error of Std 1 reaches a maximum of about 17.5% in the neighborhood of $g = 0.3$ before decreasing as g decreases. The reason for this is that, as g gets very small, the part of the 5PL curve containing the standards be-

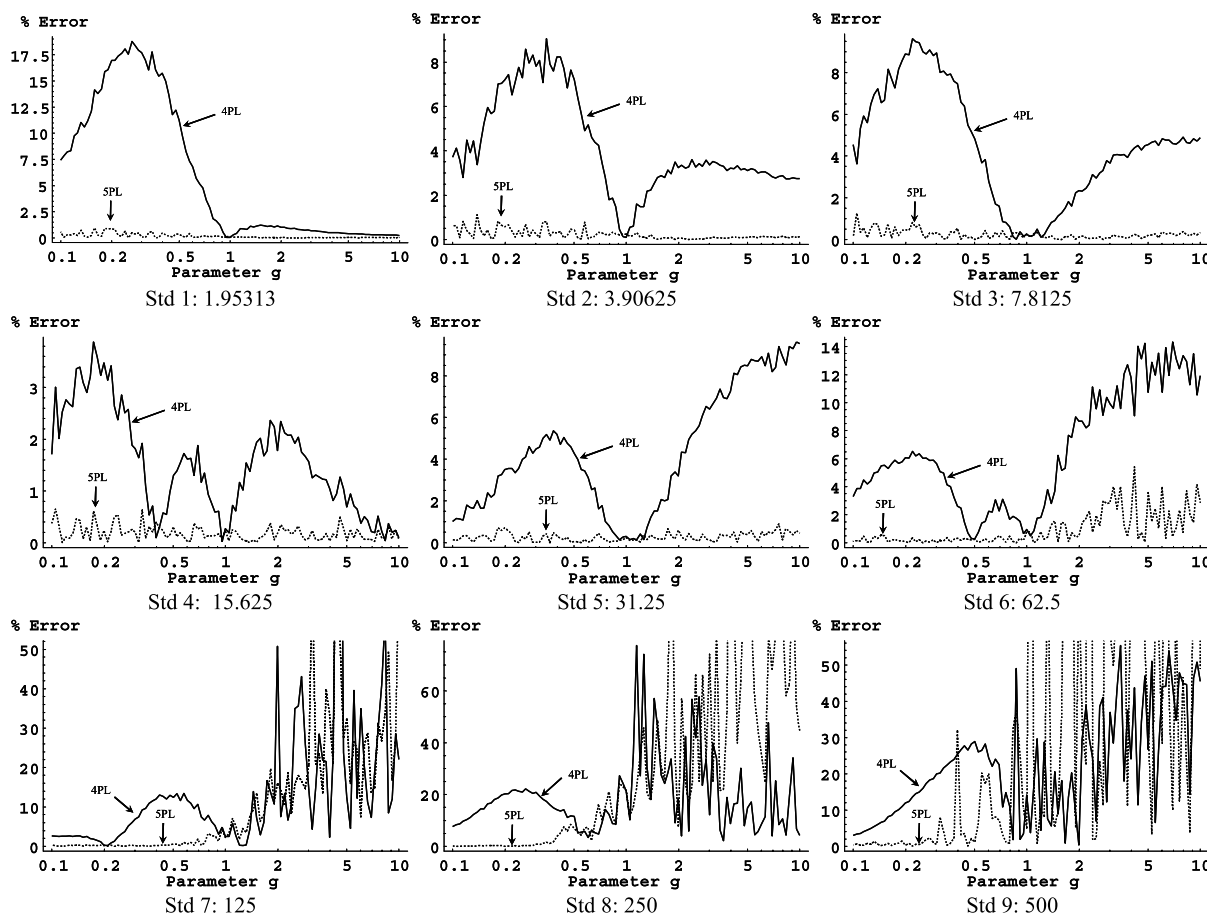


Fig. 6. Plots of the mean % error of the calculated backfit concentration value as a function of parameter g of the underlying curve are shown for each of the nine standards. The actual concentration values are displayed next to the standard number.

comes progressively confined to the lower asymptotic region of the curve. In this region, when the data are essentially one-sided at the low end, the asymmetry in the data becomes less apparent and the 4PL model will begin to approach the performance of the 5PL model. Most immunoassays and bioassays have asymmetry such that $g < 1$.

The reason that the plots of Std 7–9 have a lot of noise in the region where $g > 1$ is that the curve precision of the concentration estimates become very poor (P.G. Gottschalk and J.R. Dunn, unpublished). The precision of the concentration estimates is low for these standards as g increases because the slope of the fitted curves tends to be small, which amplifies response variation. In addition, the higher standards in this case have higher responses, which implies that they will be the most variable.

The extra valleys in the error plots of Std 6 and Std 4 occur because, at the values of parameter g at which they occur, the underlying asymmetrical curve and the curve obtained by fitting the 4PL model to the data intersect near these two standards. The fact that these two curves cross implies that many of the simulated curves will also intersect in the vicinity of these standards. The result is that the mean error for the 4PL will, for these specific values of g , approach that of the 5PL curve model.

Conclusions

When immunoassay data are asymmetric, substantial improvements in the accuracy of concentration estimates can be obtained by using a 5PL curve model instead of a 4PL curve model. The degree of improvement increases toward the extremes of the reportable range, the low and high concentrations. Also, levels of asymmetry where the penalty for using a 4PL curve is largest occurs in the interval [0.2, 0.6] of parameter g , where a large percentage of immunoassays curves are observed.

However, when the data fall entirely in one asymptotic region of the dose–response curve, first principle arguments, supported by the simulation data of Fig. 6, show that a 4PL model or even the 3PL model of Eq. (5), will be nearly as accurate but more precise than the 5PL. This is due to the greater averaging of noise resulting from the additional degree(s) of freedom.

Appendix A. 5PL Asymptotic approximation

There are two cases that exist: $(x/c)^b \rightarrow 0$ and $(x/c)^b \rightarrow \infty$. When $(x/c)^b \rightarrow 0$, $y \rightarrow a$, and when $(x/c)^b \rightarrow \infty$, $y \rightarrow d$.

When $(x/c)^b \ll 1$, Eq. (3) can be written as

$$y = d + (a - d)(1 + \delta)^{-g}, \quad (8)$$

where $\delta = (x/c)^b \ll 1$. When $\delta \ll 1$, Taylor's theorem gives $(1 + \delta)^{-g} \approx 1 - g\delta$. Substituting into Eq. (8) gives $y = d + (a - d)(1 - g\delta) = a - g(a - d)(x/c)^b$.

When $(x/c)^b \gg 1$, it dominates the denominator in Eq. (3), so $1 + (x/c)^b \approx (x/c)^b$. This gives $y \approx d + \frac{(a-d)}{(x/c)^{bg}} = d + (a - d)(x/c)^{-bg}$, which is the second result in Table 3.

References

- [1] Wild, David (Eds.), *The Immunoassay Handbook*, Stockton, New York, 1994.
- [2] P.G. Gottschalk, J.R. Dunn, Measuring parallelism, linearity and relative potency in immunoassay and bioassay data, *J. Bioform. Stat.* 15 (2005) 437–463.
- [3] D.J. Finney, *Statistical Methods in Biological Assays*, third ed., Charles Griffin, London, 1978.
- [4] R.A. Dudley, P. Edwards, R.P. Ekins, D.J. Finney, I.G.M. McKenzie, G.M. Raab, D. Rodbard, R.P.C. Rodgers, Guidelines for immunoassay data processing, *Clin. Chem.* 31 (1985) 1264–1271.
- [5] D. Rodbard, P.L. Rayford, J.A. Cooper, G.T. Ross, Statistical quality control of radioimmunoassays, *J. Clin. Endocrinol.* 28 (1968) 1412.
- [6] M.J.R. Healy, Statistical analysis of radioimmunoassay data, *Biochem. J.* 130 (1972) 107–210.
- [7] D. Rodbard, D.M. Hutt, *Statistical Analysis of Radioimmunoassays and Immunoradiometric labeled antibody Assays: A Generalized Weighted, Iterative, Least-Squares Method for Logistic Curve Fitting*. Radioimmunoassay and Related Procedures in Medicine 1, Vienna: Int. Atomic Energy Agency (1974) 165–192.
- [8] D. Rodbard, P.J. Munson, A. DeLean, Improved Curve Fitting, Parallelism Testing, Characterization of Sensitivity and Specificity, Validation, and Optimization for Radioimmunoassays 1977. Radioimmunoassay and Related Procedures in Medicine 1, Vienna: Int Atomic Energy Agency (1978) 469–504.
- [9] G.M. Raab, Comparison of a logistic and a mass action curve for radioimmunoassay data, *Clin. Chem.* 29 (1983) 1757–1761.
- [10] A.A. Fernandez, G.W. Stevenson, G.E. Abraham, N.Y. Chiarnori, Interrelations of the various mathematical approaches to radioimmunoassay, *Clin. Chem.* 29 (1983) 284–289.
- [11] V. Guardbasso, D. Rodbard, P.J. Munson, A model-free approach to estimation of relative potency in dose–response curve analysis, *Am. J. Physiol.* (1987) E357–E364.
- [12] J. Vekselsoe, Spline-fits: an alternative route to computer treatment of data from competitive protein-binding assays, *Acta Endocrinol. Suppl.* 147 (1973) 100.
- [13] D. Rodbard, Data processing for radioimmunoassays: an overview, *Clin. Immunochem.* (1978) 477–494.
- [14] R.L.A. Prentice, A generalization of the probit and logit methods for dose–response curves, *Biometrics* 32 (1976) 761–768.
- [15] M. Straume, J.D. Veldhuis, M.L. Johnson, Model-independent quantification of measurement error: empirical estimation of discrete variance function profiles based on standard curves, *Methods Enzymol.* 240 (1994) 121–150.
- [16] Brendan Technologies, Inc. *StatLIA 3.2 Reference Manual*. Carlsbad CA: Brendan Technologies (www.brendan.com), 2004.
- [17] D.M. Bates, D.G. Watts, *Nonlinear Regression Analysis and Its Applications*, Wiley, New York, 1988.
- [18] N.R. Draper, H. Smith, *Applied Regression Analysis*, third ed., Wiley, New York, 1998.

- [19] D.J. Finney, P. Phillips, The form and estimation of a variance function, with particular reference to immunoassay, *Appl. Stat.* 26 (1977) 312–320.
- [20] G.M. Raab, Estimation of a variance function, with application to immunoassay, *Appl. Stat.* 3 (1981) 32–40.
- [21] M. Davidian, R.J. Carroll, W. Smith, Variance functions and the minimum detectable concentration in assays, *Biometrika* 75 (1988) 549–556.
- [22] P.G. Gottschalk, J.R. Dunn, Determining the error of dose estimates and minimum and maximum acceptable concentrations from assays with nonlinear dose–response curves. Submitted for publication.
- [23] P.B. Daniels, The fitting, acceptance, and processing of standard curve data in automated immunoassay systems, as exemplified by the Serono SR1 analyzer, *Clin. Chem.* 40 (1994) 513–517.