Published in final edited form as:

Stat Biopharm Res. 2016 ; 8(1): 1–11. doi:10.1080/19466315.2015.1093019.

Transformation Model Choice in Nonlinear Regression Analysis of Fluorescence-based Serial Dilution Assays

Youyi Fong and Xuesong Yu

Vaccine and Infectious Diseases Division, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, yfong@fhcrc.org, xyu@fhcrc.org

Abstract

Many modern serial dilution assays are based on fluorescence intensity (FI) readouts. We study optimal transformation model choice for fitting five parameter logistic curves (5PL) to FI-based serial dilution assay data. We first develop a generalized least squares-pseudolikelihood type algorithm for fitting heteroscedastic logistic models. Next we show that the 5PL and log 5PL functions can approximate each other well. We then compare four 5PL models with different choices of log transformation and variance modeling through a Monte Carlo study and real data. Our findings are that the optimal choice depends on the intended use of the fitted curves.

Keywords

curve fitting; nonlinear calibration; heteroscedasticity; weighted least squares

1 Introduction

Nonlinear regression models are commonly used to fit dose-response curves to measure immune response biomarkers, and have important applications in immune correlates studies, which seek to identify immune response biomarkers associated with infectious disease transmission (Gilbert et al., 2008). For example, in order to quantify the amount of antibodies from a serum sample that shows specific binding to an antigen, an enzyme-linked immunosorbent assay (ELISA) is often performed on serial dilution of the serum sample. A dilution-response curve is then fitted to the optical density readouts at each dilution; the dilution that corresponds to the optical density just above the background level is called endpoint titer and is often used to quantify the amount of binding antibody in the sample.

Recent technological advancement ushered in new assays for measuring immune responses, e.g., binding antibody multiplex array (BAMA) (Tomaras et al., 2008). This assay shares the same principle as ELISA, but also possess some important differences (Elshal and McCoy, 2006). One difference is that BAMA is capable of simultaneously measuring antibodies that bind to dozens of different antigens using a small amount of serum sample. This feature allows BAMA to be used when the amount of clinical samples is limited, as is often the case in large human intervention trials. Another difference between BAMA and ELISA is in how the experimental readouts are detected. In ELISA, the detection reagent is an enzyme, such as alkaline phosphatase (AP) or horseradish peroxidase (HRP). HRP and AP catalyze reactions that produce a colored product, the amount of which is measured by optical density

using a spectrophotometer. In BAMA, detection occurs through fluorescently-labeled antibodies, and the experimental output is measured by fluorescence intensity (FI) using fluorometers. Fluorescence intensity readouts and optical density readouts are vastly different in scale and noise characteristics, and it is our goal to investigate the best curve-fitting statistical methods for fluorescence intensity-based immune assays such as BAMA.

There are two basic aspects to choosing a statistical model for nonlinear regression: mean function and variance function. In the choice of mean function, the four-parameter and five-parameter logistic functions have become the most widely used mean models since their introduction (Richards, 1959; Prentice, 1976; Rodbard et al., 1978; Finney, 1979). The development of variance function comes from the need to account for heteroscedasticity, i.e., non-constant variance in the responses at different mean response levels. Two common variance functions are the power model and the components-of-variance model (Rodbard and Cooper, 1970; Raab, 1981; Belanger et al., 1996).

Transformation is a modeling procedure that may affect both mean and variance functions. There are two types of transformation: transform outcome and transform both sides. Outcome transformation, such as Box-Cox transformation (Box and Cox, 1964), originates in linear regression; it can be done to stabilize variance or to provide a better fit for the mean function (e.g. Fine et al., 1998). In nonlinear regression, outcome transformation has been primarily done to convert a nonlinear regression problem into a linear regression problem (Section 1.7 Seber and Wild, 1989). Transforming both sides was proposed as a nonlinear regression method by Carroll and Ruppert (1984) to improve the modeling of the variance, and it does not affect the mean function.

In this paper, we seek to understand the effect of transformation on curve-fitting for fluorescence intensity-based immune assays. We consider three choices of transformation as listed below.

$$FI(x) = f_{5pl}(x, \theta) + \sigma g\{f_{5pl}(x, \theta), \gamma\} \varepsilon \quad \text{5PL} \quad (1)$$

$$\log FI(x) = \log f_{5pl}(x, \theta) + \sigma g\{f_{5pl}(x, \theta), \gamma\} \varepsilon \quad \log \text{ 5PL} \quad (2)$$

$$\log FI(x) = \log f_{5pl}(x, \theta) + \sigma \varepsilon \quad \log\text{-}\log \text{ 5PL} \quad (3)$$

In all three models, we choose to base the mean functions on five-parameter logistic (5PL) curves, denoted by $f_{5pl}(x, \theta)$ (Gottschalk and Dunn, 2005; Cumberland et al., 2014), where x denotes dose and θ denotes the vector of curve parameters. We use ε to denote a random variable with standard normal distribution, and σ as a positive scale parameter. The function g is a variance function with parameter γ that characterizes assay response variance in terms of the relationship with assay response mean; $g = 1$ corresponds to the constant variance assumption.

Model (1) is the most commonly used model in practice, and no transformation is applied; the mean model is a 5PL function, and g is commonly chosen to be the power variance function $g(f_{5pl}, \gamma) = (f_{5pl})^{\gamma/2}$. In model (2), a log transformation is applied to the left-hand side, and we refer to the mean model as a log 5PL function, i.e. we call y a log 5PL function of x if $\log(y) = f_{5pl}(x)$. We focus our attention on log transformation because it is a commonly used in the analysis of fluorescence intensity-based assay data, e.g. microarray data. It has been found that log transformation stabilizes the variance of fluorescence intensity readouts when they are above a certain intensity (Kerr et al., 2000; Cui et al., 2003).

We consider two choices of g : a constant function and a power function. In model (3), both the outcome and the logistic function are transformed, and the mean function is identical to that of the 5PL model (1). We call this model the log-log 5PL model. Estimation of models with non-constant variance functions is addressed in Section 2. Estimation of models with constant variance functions can be carried out using the R package `drc` and `nCal` (Ritz and Streibig, 2005; Fong et al., 2013), which contain specialized functions for choosing and trying multiple starting values for 5PL and log-log 5PL models.

The balance of the paper is organized as follows. In Section 2, we describe a generalized least squares-pseudolikelihood (GLS-PL) type algorithm for fitting 5PL and log 5PL models with non-constant variance functions. In Section 3, we study the distance between the 5PL curves and log 5PL curves. In Section 4, we perform Monte Carlo experiments to study the impact of model choice on both curve fitting and nonlinear calibration. In Section 5, we examine the optimal model choice using a set of real serial dilution assay data. We conclude with a discussion in Section 6.

2 A GLS-PL type algorithm for fitting logistic models to heteroscedastic dose-response data

The fitting of logistic models to heteroscedastic dose-response data has long been known to be a challenging task. For example, Seber and Wild (1989) wrote: "Bad ill-conditioning and convergence problems have often been experienced in fitting the Richards model (another name for logistic models) to data" (page 336); Ratkowsky (1990) referred to the model as "a particularly unfortunate model" (page 47); and Ratkowsky (1983) wrote: "The Richards model is the only one exhibiting a significant intrinsic curvature" (page 73). In our experience, the general purpose function `gnls()` in the R software system fails to fit most of the datasets, even after substantial efforts in tuning the control parameters, and reparameterization of the mean function (Fong et al., 2012) that reduces covariance of the parameter estimate.

Many specialized software tools have been developed to fit logistic curves when r is known, including the `drc` package (Ritz and Streibig, 2005) in R, and commercial platforms for analyzing fluorescence intensity assay data such as Bio-Plex Manager from Bio-Rad or MasterPlex QT from MiraiBio. The `drc` package works well in our experience, and its success is in part explained by self-start functions tailored to logistic models. A straightforward way to estimate variance component parameters using the `drc` package is to

perform a brute-force grid search in the variance component parameter space. This produces satisfactory fits, but is impractical for routine data analysis.

To fill the void of open-source, practical programming tools for analyzing heteroscedastic dose-response data with logistic models, we propose a new optimization procedure that embeds the R package drc in a GLS-PL type algorithm (Davidian and Giltinan, 1995; Belanger et al., 1996). The defining feature of a GLS-PL algorithm is to alternate between the GLS step, which solves a weighted least squares problem to estimate the mean model parameters, and the PL step, which does a maximum likelihood estimation of the variance component parameters. Our algorithm uses drc in the GLS step to find a good logistic model fit. In the PL step, working with the popular power variance function

$g\{f_{5pl}(x, \theta), \gamma\} = f_{5pl}^{\gamma/2}(x, \theta)$, the deviance of model (1) or (2) equals, sans constants,

$$dev = \sum_i \left[\log \left(f_{5pl}^{\gamma}(x_i, \theta) \sigma^2 \right) + \frac{\{y_i - f_{5pl}(x_i, \theta)\}^2}{f_{5pl}^{\gamma}(x_i, \theta) \sigma^2} \right], \quad (4)$$

where y can be either FI or $\log(FI)$. By profiling out σ^2 , we can find $\hat{\gamma}$ by solving a one-dimensional root-finding problem. The details of the algorithm are listed below.

1. Initialize $\hat{\gamma}$ to 2 and compute weights $w_i = y_i^{-\hat{\gamma}/2}$.
2. Solve the following weighted nonlinear least square problem to find $\hat{\theta}$ using the drc R package:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_i w_i^2 (y_i - f_{5pl}(x_i, \theta))^2 \quad (5)$$

3. Given $\hat{\theta}$, find $\hat{\gamma}$ by solving $\partial dev / \partial \gamma = 0$ and $\partial dev / \partial \sigma^2 = 0$, which leads to searching for the root of the following equation in the interval $(-8, 8)$:

$$\sum_i \log f_{5pl}(x_i, \hat{\theta}) \left(1 - \frac{\{y_i - f_{5pl}(x_i, \hat{\theta})\}^2}{f_{5pl}^{\gamma}(x_i, \hat{\theta}) n^{-1} \sum_i \left[\{y_i - f_{5pl}(x_i, \hat{\theta})\}^2 / f_{5pl}^{\gamma}(x_i, \hat{\theta}) \right]} \right) = 0.$$

Re-compute weights $w_i = f_{5pl}^{-\hat{\gamma}/2}(x_i, \hat{\theta})$, and go back to step 2, unless the maximum relative change in $\hat{\theta}$ from the previous step is less than a preset threshold (e.g. 10^{-3}), or the deviance starts increasing.

At the end of the iterations, σ^2 can be estimated by

$$\hat{\sigma}^2 = n^{-1} \sum_i \frac{\{y_i - f_{5pl}(x_i, \hat{\theta})\}^2}{f_{5pl}^{\hat{\gamma}}(x_i, \hat{\theta})}.$$

The initial value of $\hat{\gamma}$ in step 1 and the interval to be searched for $\hat{\gamma}$ in step 3 are based on prior knowledge that we gained through grid search. These values should be tested and changed if necessary in new situations to achieve optimal performance. Even with such knowledge, the above algorithm may fail for a small proportion of datasets. Because the deviance as a function of γ can be rather non-smooth, sampling a few more starting values for γ uniformly in the search interval is not likely to help in these cases. Instead we recommend performing grid search to find reasonable curve fits under such circumstances.

The above algorithm is a pseudolikelihood procedure (Carroll and Ruppert, 1988) because although step 3 minimizes the deviance, step 2 does not. PL algorithms may not converge well as noted by Carroll and Ruppert (1988), and can oscillate between local optima without stopping. To stop the algorithm without setting an arbitrary maximum number of iterations, we opt to stop the algorithm when the deviance starts increasing. In our numerical studies, this rule typically stops the algorithm in three iterations and most within five.

If we replace step 2 with a step that minimizes the deviance as a function of θ conditional on the variance component parameters, then we have a maximum likelihood algorithm. The trade-offs between GLS-PL algorithms and maximum likelihood algorithm have been well studied (Carroll and Ruppert, 1982; Davidian and Giltinan, 1995). The general conclusions are intuitive: if the data follows the working model, ML methods are more efficient, but if the working model is violated, GLS-PL methods may be preferable. The consensus in the field appears to favor the more robust GLS-type algorithms.

To ensure that the quality of curve fits obtained using our algorithm is on par with grid search, we apply both methods to a real dataset from the RV144 immune correlates study (Haynes et al., 2012). RV144 is a phase III, randomized controlled trial testing a HIV-1 vaccine, which showed a modest effect of protection (Rerks-Ngarm et al., 2009). To understand the potential protective mechanism, Haynes et al. conducted an immune correlates study to examine immune responses that are associated with infection risk in the vaccine recipients. In the first phase of the immune correlates study, a pilot study was performed to select and optimize experimental assays for measuring immune responses. As part of the pilot study, the HIV Vaccine Trials Network (HVTN) laboratory at the Fred Hutchinson Cancer Research Center performed experiments to quantify cytokine production by peripheral blood mononucleated cells using a multiplexed bead array based on the Luminex platform. Each serial dilution set contains 10 dilutions with 2 replicates at each dilution.

We use 100 sets of serial dilution assay data from this dataset to compare the GLS-PL algorithm with grid search in fitting 5PL models. We also apply the general purpose `gnls()` function from `nlme` package in R to these data, and it fails for all 100 sets. In the grid search, we choose 20 γ 's uniformly spaced between 0.2 and 4. For each γ , we compute weights $w_i = y_i^{-\gamma/2}$ and solve the minimization problem (5). The optimal γ is defined as the one with the smallest deviance as computed by (4). This procedure is on average five to six times slower than the GLS-PL algorithm, because the most time-limiting step in the algorithm is to solve (5). We then compare the minimal deviance found by grid search (e.g. Online Supplementary Figure A.1) with the minimized deviance arrived at by the GLS-PL

algorithm. Overall the performance of the two approaches is comparable, as can be seen in Figure 1(a). Figure 1(b) examines the distribution of the difference between each pair of deviances. GLS-PL has a smaller mean deviance by -0.13 and a smaller median deviance by -0.25 . Intuitively GLS-PL may perform better than grid search, because there is no limit to its resolution. Indeed, for 793 of the datasets, GLS-PL performs better than grid search. But in 213 of the datasets, grid search does better. We do not completely understand why GLS-PL may not perform as well as grid search. One possible explanation is GLS-PL may converge to a local optimum; another is that the GLS-PL has a well-known convergence issue.

3 5PL and log 5PL curves

The 5PL function has a long history, with much contribution from growth models literature (Seber and Wild, 1989, Chapter 7). As noted by Seber and Wild, there were two main approaches to the development of growth models: the empirical or statistical approach, and the theoretical or biological approach. And yet there was also a convergence of the two approaches, which occurred because theoretical models were often based on oversimplified assumptions and had to be relaxed in some empirical way to provide satisfactory model fit. This general trend fits the development of the 5PL model well. The four-parameter logistic model (4PL) was precursor to the 5PL model. The 4PL model can be derived from theoretical consideration of growth rate. Its usefulness in reality, however, is limited by the fact that many curves are not symmetric around its inflection point. Extension to 5PL (Richards, 1959; Prentice, 1976) creates a flexible family of curves, for which the inflection point can be located anywhere on the curve.

The log 5PL model can be viewed as another flexible, empirical model that provides a good fit to data exhibiting a sigmoidal trend. The model also has some physical interpretation. To see that, we assume the asymmetry parameter f is 1, and the model is reduced to a log 4PL model, which is expressed as follows in the g - h parameterization (Fong et al., 2012).

$$E \{ \log FI(x) \} = c + \frac{d - c}{1 + \exp \left\{ -\frac{4h}{d-c} (\log x - g) \right\}}$$

In this parameterization, c and d are the lower and upper asymptotes, respectively; g is the inflection point or mid point; and h is the slope of the curve, $d \log FI(x) / d \log x$, at the inflection point, also the maximum slope over the entire curve. Theoretically speaking, h should be close to 1. This is because when the background level is low, and the fluorescence intensity is far from both the background and the saturation levels, we expect to see that a doubling of analyte concentrations corresponds to a doubling of fluorescence intensity. Indeed we see that in many real datasets, the estimated h is close to 1.

Since both the 5PL and log 5PL models appear to fit sigmoidal data well, the space of all 5PL functions and the space of all log 5PL functions must be close, in the sense that it should be possible to approximate a 5PL function well with a log 5PL function and vice

versa. Establishing the degree to which this is true will help us choose between these two models.

Towards this goal, we first generate a dataset using the log transformation model (2) with $\sigma = 0$ to represent the log 5PL curve. We choose 200 dilutions uniformly spaced on the log scale. To find the 5PL function that best approximates the log 5PL function as measured by the L_2 distance in $\log(FI)$, we fit the dataset with model (3). For comparison, we also search for the 4PL functions that best approximate the target log 5PL functions. We repeat this procedure for three log 5PL functions specified with the τ -h parameterization (Cumberland et al., 2014), which is similar to the g-h parameterization, but replaces the inflection point with ED50 (denoted by $g=t$) as a parameter. The three curves are designed to have different degrees of asymmetry in a fixed range of concentrations. Their parameter values are listed in Online Supplementary Table A.2. The log 5PL curves and the best approximating 5PL and 4PL curves are plotted on the $\log(FI)$ scale in the first row and on the FI scale in the second row in Figure 2. The results show that when $f=1$, both the 5PL and 4PL curves can approximate the log 5PL curve well. One reason for this may be that on the FI scale the range of concentration does not cover much beyond the upper bending point, which is the point with minimal second derivative. When $f=3.2$, the 5PL curve provides a good approximation of the log 5PL curve, but not the 4PL curve. When $f=0.2$, it may be the most challenging case. The 5PL curve provides good approximation near the upper asymptote, but not near the lower asymptote. The 4PL curve fails to provide good approximation near either asymptote.

Secondly, we generate datasets to embody 5PL curves and look for log 5PL and log 4PL curves that best approximate the 5PL curves, also measured by the L_2 distance in $\log(FI)$. The parameters of the target 5PL curves are listed in the Online Supplementary Table A.2. The 5PL curves and the best approximating log 5PL and log 4PL curves are plotted on the FI scale in the first row and on the $\log(FI)$ scale in the second row in Online Supplementary Figure A.2. The results show that log 5PL can provide good approximation of the three 5PL curves. On the other hand, log 4PL functions do not provide as good a approximation of these 5PL or 4PL ($f=1$) curves.

We have shown that some log 5PL curves can be well approximated with 5PL curves and some 5PL curves can be well approximated with log 5PL curves. These results help explain our empirical observations that both the 5PL and log 5PL models provide good fits to many real datasets. On the other hand, the approximation performance may depend on the value of f , the range of dilutions, and how close the lower asymptote is to 0. As a result, the 5PL and log 5PL models may produce curve fits of different quality for some real datasets.

4 Optimal model choices in a Monte Carlo Study

In this section we study the optimal model fitting strategies through a Monte Carlo study. The optimal model choice depends on the ultimate goal of curve fitting. There are two common uses of fitted curves: (1) The curves are fitted using data generated from serial dilution of clinical samples. We call this type of dose-response curves dilution-response curves. Given a dilution-response curve, we are interested in extracting a summary feature to

quantify the variable of Interest in the corresponding clinical sample. For example, many humoral immune response assays belong to this group, including BAMA. (2) The curves are fitted using data generated from standard samples, whose analyte concentrations are known through other means. These curves are called standard curves. Given a standard curve and the experimental readout of clinical samples, often termed unknown samples, we can make an inverse prediction of analyte concentrations in the unknown samples. This process is known as nonlinear calibration. For example, in immune correlates studies, one way to gauge cellular immune responses is to measure cytokines concentrations in clinical samples through nonlinear calibration. These two scenarios for using fitted curves call for different optimality criteria in evaluating curve fitting methods, and we will look at both.

To specify the mean curve for the Monte Carlo study in a nonparametric way so as not to prejudice the comparison, we fit a sigmoidally constrained nonparametric maximum likelihood model (Schmoyer, 1984) to a real dataset that shows some degree of asymmetry (Online Supplementary Figure A.3a). The fitted curve is supported at the 10 dilutions in the dataset, and unspecified at the intervals between those dilutions. Each simulated dataset comprises two observations at each of the 10 dilutions. Empirically, the distribution of FI at a given dilution is skewed, thus we simulate each observation from a log normal distribution $\log N(\mu, \sigma^2)$. The values of σ^2 gradually decrease as μ increases (Table 1), and are estimated from real datasets. A sample dataset is shown in Online Supplementary Figure A.3(b). The simulation study is repeated 2000 times. Among the replicates, 6.93 fail to have proper 5PL fits using the GLS-PL algorithm (an example is shown in Online Supplementary Figure A.4), while all replicates have successful fits using the other models. In the following, we summarize the simulation results after excluding all curve fits from the 6.93.

We first look at the performance of different methods at estimating a summary feature of dilution-response curves. Figure 3(a) and Table A.3 present the mean squared error (MSE) of the fitted $\log(FI)$ versus the simulation truth at each dilution. We focus on $\log(FI)$ because this is often the scale at which immune responses are analyzed (Haynes et al., 2012; Yates et al., 2011; Tomaras et al., 2013). The 5PL model performs worse than the other three models for the more diluted samples, but better for the more concentrated (i.e. less diluted) samples (e.g. Figure 5a and 5b). Comparing the log 5PL model with and without power variance function, we see that their results are very similar. This could be due to (a) the power function does not adequately model the actual variances, and (2) the objective function in the log 5PL model with constant variance matches the performance criterion while the objective function in the log 5PL model with power variance does not. The log-log 5PL model performs similarly to the log 5PL models for the more diluted samples, and slightly better for the more concentrated samples. Aggregating over all 10 dilutions, the ranking of the models are: log-log 5PL, log 5PL with constant variance, log 5PL with power variance, and 5PL.

Another set of features of dilution-response curves commonly used in practice are dilutions, on the log scale, corresponding to some fixed experimental readout, e.g. endpoint titers (Frey et al., 1998). Given the nonparametric nature of our simulation setup, we can only evaluate titers corresponding to the true experimental readouts listed in Online Supplementary Table 1. We choose to look at the mean squared error of the log dilution corresponding to the true

experimental readout at dilution #7. The MSE for the 5PL model is 0.056, the MSE for the two log 5PL models are both 0.009, the MSE for the log-log 5PL model is 0.008.

We now compare different models by their nonlinear calibration performance. Figure 3(b) and Table A.4 present the results on nonlinear calibration performance. Here, we treat the fitted curves as standard curves and simulate 10 unknown samples at the same concentrations as the standard samples. For each unknown sample, we estimate the MSE of the estimated log concentrations versus simulation truth. For the more concentrated samples, the 5PL model performs slightly better than the other three methods. For the more diluted samples, however, the results are rather different: the 5PL model outperforms the other methods at dilution #9 and #8. The log 5PL model with power variance function performs similarly as the log 5PL model with constant variance. The log-log 5PL model performs similarly with the log 5PL models at the more diluted range, and performs in between the 5PL model and the log 5PL models at the more concentrated range.

To investigate whether these results depend on the modeling assumption that the experimental noise has a log normal distribution, we repeat this study with each observation generated from a normal distribution $N(\sigma^{\mu}, 0.0049 \times e^{2\mu})$, where the μ 's are as listed in Table 1. The results are shown in Online Supplementary Figure A.5. Similar to the results based on the log normally generated data, the 5PL model underperforms the log 5PL and log-log 5PL models in the low FI region, although to a much lesser degree, and outperforms in the high FI region, when we compare MSE in log(FI).

To better understand why the 5PL model has smaller MSE in estimated log concentrations at dilution #9 and #8, even though its curve fit quality is worse at those two dilutions as judged by MSE in the fitted log (FI), we plot the two components of MSE in estimated log concentrations separately in Figure 4(a) and 4(b). The figure shows that it is primarily the variance that is driving this difference. The uncertainty of the estimated concentrations comes from two sources: uncertainty in the fitted curves and uncertainty in the experimental readout for the unknown samples. Oftentimes, the latter would dominate the former. The effect of unknown sample measurement precision on the concentration estimates is inversely proportional to the local slope of the standard curve. The 5PL model often results in a steeper slope than other models near the lower asymptote (Figure 5a), leading to a smaller variability in the concentration estimates. This 'advantage' of 5PL model comes with an inherent cost: the percentage of times in the Monte Carlo study a method returns an estimated concentration of a background sample (simulated with $\mu = 2.80$ and $\sigma = 0.114$) greater than the most diluted standard sample is 513 for the 5PL model and 333 for the other three models.

In sum, if we are interested in the fitted FI, the model choice is relatively straightforward. Under the simulation settings in this section, the log-log 5PL model and the log 5PL models are the better choices in the more diluted range; in the more concentrated range, the 5PL model is the best choice. Over the full range, the log-log 5PL model has the best performance. On the other hand, if our goal is nonlinear calibration, the model choice is more nuanced. Under our simulation setting, in the low concentration range, if we believe there is signal, albeit weak, in the unknown samples, the 5PL model with power variance is

the best choice; but the other models are better at discriminating background noise from true weak signals. In the high concentration range, the calibration performances of the models are similar.

5 Real datasets

In this section, we compare the curve fits by different models using the pilot dataset from the RV144 immune correlates study. The full dataset contains 42×8 sets of serial dilution assay data, 8 plates for each of 42 analytes in a multiplexed experiment. We focus on comparing the quality of the curve fits as measured by the sum of squared distance between the observed $\log(FI)$ and the fitted values over a whole dose-response curve. In three datasets the 5PL model fails to fit, and in one dataset the log-log 5PL model fails. The failed fits are excluded from the comparison. We compute the sum of squared distance in $\log(FI)$ for each dataset and average over the plates for each analyte to arrive at a mean squared error. The comparison between different models is based on the 42 MSE values. Figure 6 shows that judging by the medians, the ranking of the methods are: log-log 5PL, log 5PL with constant variance, 5PL, and log 5PL with power variance. The paired twosample Wilcoxon and Student's t tests comparing log-log 5PL and any other method all return highly significant p-values (Table A.5). In addition, the paired two-sample Wilcoxon and Student's t tests comparing log 5PL with constant variance and log 5PL with power variance also return highly significant p-values (4.5×10^{-13} and 7.1×10^{-4}). This suggests that the residual heteroscedasticity after log transformation is better handled by assuming constant variance rather than power variance. This may be because the power variance model is not a very good model on the $\log(FI)$ scale in these datasets, or it may be because we use MSE in $\log(FI)$ to judge performance and the log 5PL model with power variance actually minimizes a different criterion function.

6 Discussion

In this paper we consider the choice of transformation in nonlinear regression for serial dilution assay. Comparison of the 5PL curves and log 5PL curves shows that they can approximate each other well. While both the 5PL and log 5PL functions are best viewed as empirical models for curve fitting, their 4PL versions have physical interpretation. While the development of 4PL model originated from a focus on the exponential growth near the lower asymptote, the slope parameter in the log 4PL model is quite helpful to look at for modern assays that have substantial linearity range in between the lower and upper asymptotes.

We develop a GLS-PL type algorithm for fitting logistic models to heteroscedastic doseresponse data and show that its performance is on par with grid search under most circumstances, but many times faster. The algorithm is implemented in a R package, and will be available from the Comprehensive R Archive Network. For a small percentage of datasets, the algorithm does have difficulty finding the optimal solution under the 5PL model. Unsatisfactory curve fits are easy to identify graphically, but our efforts to use lack-of-fit tests based on either cumulated residuals (Ritz and Martinussen, 2011) or modified ANOVA (Neill, 1988) to detect unsatisfactory fits only have limited success. More research in this area is warranted.

We compare different curve fitting models through Monte Carlo studies and real datasets. We find that the log-log 5PL model is the best overall choice for predicting features of doseresponse curves. In addition, the 5PL model produces better fits in the high FI region, and the models with transformation produce better fits in the low FI region. These conclusions are necessarily a function of our choice of simulation parameters (Table 1) and the datasets we encountered in our collaborative work. A limitation of our working models is that they assume either constant or power variance. When there is sufficient knowledge to inform better variance model, the curve fit can potentially be improved.

When the fitted curves are used in nonlinear calibration, bias variance trade-off may cause the model that fits worse to have smaller mean squared errors in the estimated log concentrations. However, we advise caution in adopting a model that has the worse curve fit, because it may not be as effective in discriminating samples with no signals from samples with weak signals.

Both the 5PL and log 5PL models can be viewed as special members of a class of transformation, such as the Box-Cox transformation (Box and Cox, 1964). Whether it is worth introducing further model complexity in the form of a transformation parameter (Carroll and Ruppert, 1988) in the analysis of serial dilution data under the common experimental designs is an open question. Given the small size of a typical serial dilution dataset, we suspect that the expanded model can be difficult to estimate and interpret as has been pointed out before in Bickel and Doksum (1981) and Box and Cox (1982).

Transformation and the choice between 5PL and 4PL are connected in some ways. First, one practical reason for using 4PL is the numerical instability of fitting 5PL models. An effect of transforming the fluorescence intensity is that this numerical difficulty largely dissipates. Second, in Section 3 we see that the class of 5PL functions and the class of log 5PL functions approximate each other rather well, but the same cannot be said of 4PL. A consequence of this is that more attention should be paid to transformation if one is using 4PL instead of 5PL. These two considerations further encourage the use of 5PL over 4PL beyond previous studies on this topic (Gottschalk and Dunn, 2005; Cumberland et al., 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank members of the Lab Data Operations at Statistical Center for HIV/AIDS Research & Prevention (SCHARP) for assay data quality control, and Stephen De Rosa, Nicole Frahm and members of the HVTN laboratory for performance of the assays. This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID) grant UM1-AI-068635 to HVTN, the cooperative agreement W81XWH-07-2-0067 between the Henry M. Jackson Foundation and the Department of Defense, and the NIAID grant AI104370 to Y.F.

References

Belanger B, Davidian M, Giltinan D. The effect of variance function estimation on nonlinear calibration inference in immunoassay data. *Biometrics*. 1996; 52(1):158–175. [PubMed: 8934590]

- Bickel PJ, Doksum KA. An analysis of transformations revisited. *Journal of the American Statistical Association*. 1981; 76(374):296–311.
- Box G, Cox D. An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*. 1982; 77(377):209–210.
- Box GE, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*. 1964; 26(2):211–252.
- Carroll, R.; Ruppert, D. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis; 1988. Transformation and Weighting in Regression.
- Carroll RJ, Ruppert D. A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *Journal of the American Statistical Association*. 1982; 77(380):878–882.
- Carroll RJ, Ruppert D. Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association*. 1984; 79(386):321–328.
- Cui X, Kerr MK, Churchill GA. Transformations for cDNA microarray data. *Statistical applications in genetics and molecular biology*. 2003; 2(1)
- Cumberland W, Fong Y, Yu X, Defawe O, Frahm N, De Rosa S. Nonlinear calibration model choice between the four and five parameter logistic models. *Journal of Biopharmaceutical Statistics* in press. 2014
- Davidian, M.; Giltinan, D. Nonlinear models for repeated measurement data. Vol. 62. Chapman & Hall/CRC; Boca Raton, FL, USA: 1995.
- Elshal MF, McCoy JP. Multiplex bead array assays: performance evaluation and comparison of sensitivity to elisa. *Methods*. 2006; 38(4):317–323. [PubMed: 16481199]
- Fine J, Ying Z, WEI L. On the linear transformation model for censored data. *Biometrika*. 1998; 85(4): 980–986.
- Finney D. Bioassay and the practice of statistical inference. *International Statistical Review/Revue Internationale de Statistique*. 1979; 47(1):1–12.
- Fong Y, Sebestyen K, Yu X, Gilbert P, Self S. nCal: a R package for nonlinear calibration. *Bioinformatics*. 2013; 29(20):2653–2654. [PubMed: 23926226]
- Fong Y, Wakefield J, De Rosa S, Frahm N. A robust bayesian random effects model for nonlinear calibration problems. *Biometrics*. 2012; 68(4):1103–1112. [PubMed: 22551415]
- Frey A, Di Canzio J, Zurakowski D. A statistically defined endpoint titer determination method for immunoassays. *Journal of Immunological methods*. 1998; 221(1):35–41. [PubMed: 9894896]
- Gilbert P, Qin L, Self S. Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Statistics in Medicine*. 2008; 27(23):4758–4778. [PubMed: 17979212]
- Gottschalk P, Dunn J. The five-parameter logistic: a characterization and comparison with the four-parameter logistic. *Analytical Biochemistry*. 2005; 343(1):54–65. [PubMed: 15953581]
- Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Sutthent R, Liao H-X, DeVico AL, Lewis GK, Williams C, Pinter A, Fong Y, Janes H, DeCamp A, Huang Y, Rao M, Billings E, Karasavvas N, Robb ML, Ngauy V, de Souza MS, Paris R, Ferrari G, Bailer RT, Soderberg KA, Andrews C, Berman PW, Frahm N, De Rosa SC, Alpert MD, Yates NL, Shen X, Koup RA, Pitisuttithum P, Kaewkungwal J, Nitayaphan S, Rerks-Ngarm S, Michael NL, Kim JH. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *New England Journal of Medicine*. 2012; 366(14):1275–1286. [PubMed: 22475592]
- Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. *Journal of computational biology*. 2000; 7(6):819–837. [PubMed: 11382364]
- Neill JW. Testing for lack of fit in nonlinear regression. *The Annals of Statistics*. 1988; 16(2):733–740.
- Prentice R. A generalization of the probit and logit methods for dose response curves. *Biometrics*. 1976; 32(4):761–768. [PubMed: 1009225]
- Raab G. Robust calibration and radioimmunoassay. *Biometrics*. 1981; 37(4):839–841.
- Ratkowsky, D. *Statistics* (Marcel Dekker). M. Dekker; 1983. Nonlinear Regression Modeling: A Unified Practical Approach.
- Ratkowsky, D. *Statistics, textbooks and monographs*. Marcel Dekker; New York: 1990. Handbook of nonlinear regression models.

- Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Prem Sri N, Namwat C, de Souza M, Adams E, Benenson M, Gurunathan S, Tartaglia J, McNeil JG, Francis DP, Stablein D, Birx DL, Chunsuttiwat S, Khamboonruang C, Thongcharoen P, Robb ML, Michael NL, Kunasol P, Kim JH. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *New England Journal of Medicine*. 2009; 361(23):2209–2220. [PubMed: 19843557]
- Richards F. A flexible growth function for empirical use. *Journal of Experimental Botany*. 1959; 10(2): 290–300.
- Ritz C, Martinussen T. Lack-of-fit tests for assessing mean structures for continuous dose-response data. *Environmental and Ecological Statistics*. 2011; 18(2):349–366.
- Ritz C, Streibig J. Bioassay analysis using R. *Journal of Statistical Software*. 2005; 12(5):1–22.
- Rodbard, D.; Cooper, J. Vienna International Atomic Energy Agency. Vienna Austria: 1970. A model for prediction of confidence limits in radioimmunoassays and competitive protein binding assays.
- Rodbard D, Munson P, DeLean A. Improved curve-fitting, parallelism testing, characterization of sensitivity and specificity, validation, and optimization for radioligand assays. *Radioimmunoassay and Related Procedures in Medicine*. 1978; 1:469–504.
- Schmoyer R. Sigmoidally constrained maximum likelihood estimation in quantal bioassay. *Journal of the American Statistical Association*. 1984; 79(386):448–453.
- Seber, G.; Wild, C. Wiley Series in Probability and Statistics. Wiley; 1989. Nonlinear Regression.
- Tomaras G, Ferrari G, Shen X, Alam S, Liao H, Pollara J, Bonsignori M, Moody M, Fong Y, Chen X, Nicholson C, Polin B, Zhang R, Lu X, Parks R, Kaewkungwal J, Nitayaphan S, Pitisuttithum P, Rerks-Ngarm S, Gilbert P, Kim J, Michael N, Montefiori D, Haynes B. Vaccine-induced plasma IgA specific for the C1 region of the HIV-1 envelope blocks binding and effector function of IgG. *Proceedings of the National Academy of Sciences*. 2013; 110(22):9019–9024.
- Tomaras GD, Yates NL, Liu P, Qin L, Fouda GG, Chavez LL, Decamp AC, Parks RJ, Ashley VC, Lucas JT, et al. Initial b-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin m (igm) and igg antibodies followed by plasma anti-gp41 antibodies with ineffective control of Initial viremia. *Journal of virology*. 2008; 82(24):12449–12463. [PubMed: 18842730]
- Yates NL, Lucas JT, Nolen TL, Vandergrift NA, Soderberg KA, Seaton KE, Denny TN, Haynes BF, Cohen MS, Tomaras GD. Multiple hiv-1-specific igg3 responses decline during acute hiv-1: implications for detection of Incident hiv infection. *AIDS (London, England)*. 2011; 25(17):2089.

Figure 1.
Comparison between a GLS-PL algorithm and grid search.

Figure 2. Three log 5PL curves with different f values and the best approximating 5PL and 4PL curves. Top row: $\log(FI)$ scale, bottom row: FI scale.

Figure 3.

(a) MSE of fitted $\log(FI)$ versus simulation truth at the ten dilutions. (b) MSE of estimated $\log(\text{concentration})$ versus simulation truth for ten unknown samples simulated at the same dilutions as the standard samples.

Figure 4.
 (a) Bias² and (b) variance of estimated log(concentration) versus simulation truth for ten unknown samples simulated at the same dilutions as the standard samples.

Figure 5.

An sample dataset from the Monte Carlo study. Dilution #5 to #10 are shown in panel (a) on the log(FI) scale, dilution #1 to #5 are shown in panel (b) on the FI scale.

Figure 6. MSE of fitted log(FI) versus observed values summed over a serial dilution curve. Each point corresponds to one different analyte.

Table 1

Simulation parameters in the Monte Carlo study.

Dilution #	10	9	8	7	6	5	4	3	2	1
x	0.5	1.5	4.6	13.7	41.2	123.5	370.4	1111.1	3333.3	10000
μ	2.84	2.91	3.01	3.57	4.94	6.45	7.66	8.73	9.4	9.48
σ	0.114	0.114	0.114	0.086	0.082	0.071	0.056	0.049	0.049	0.025