

## FitProb offers an *effective* alternative to the ineffective $R^2$ metric for nonlinear curve fitting

### Fit Probability:

- Designed for nonlinear curve fitting
- Measures how closely the curve fit matches the true curve
- Provides reliable metric for curve fit quality

*R<sup>2</sup> can be a lousy measure of goodness-of-fit, especially when it is misused...*  
**Charles Annis, P.E., Statistical Engineering**

- Bioassays
- Immunoassays
- Quantitative Tests
- Qualitative Tests
- Multiplexed assays
- Parallelism
- Immunogenicity

A principle requirement for an effective metric to analyze nonlinear curve fits is the ability to distinguish a good curve fit from a poor one. With results usually  $> 0.9$  for even obviously poor fitting curves (see Figure 3), the metric  $R^2$  is unable to achieve this basic objective.  $R^2$  is a measure of whether the response and concentration are related, which they are, and is not appropriate for assessing the quality of nonlinear fits.

In the literature<sup>1,2,3,4</sup>, residual variance, which is the residual sum of squared errors (RSSE) per degree of freedom, is the metric used for this purpose. Residual variance, unlike  $R^2$ , measures how well the curve fits the individual data points. If the data is accurately weighted, the RSSE is  $\chi^2$  distributed and its probability makes an effective statistical metric, the FitProb. The StatLIA® TrueFit™ Data Reduction

System pools the lab's previously run historical assays to determine the correct weighting.

An effective curve fit metric must be able to make the following determinations:

- Assess the quality of the entire curve in a manner that is able to discern even small, but significant, differences in curve fit quality (*RSSE, Residual Variance*).
- Measure the difference between the proposed curve fit dose/response relationship and the observed data (*RSSE*).
- Include the number of data points into the calculation to normalize the varying number of points in different curves. (*Degrees of Freedom*).
- Report how closely the observed dose/response relationship matches the true curve (*Fit Probability [FitProb]*).

Figure 1

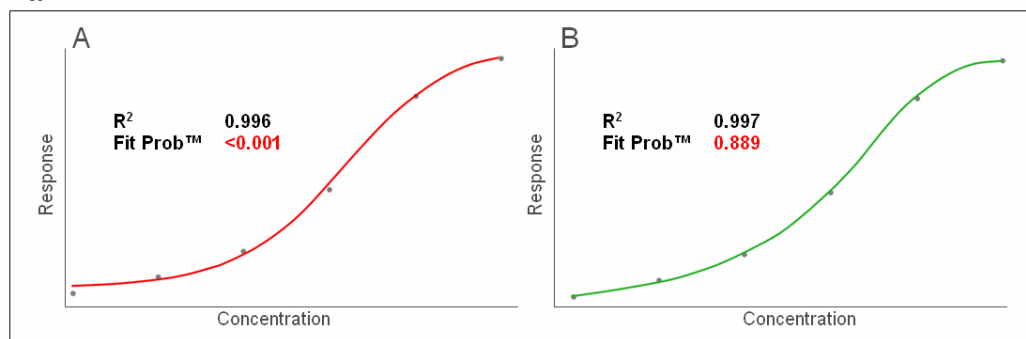
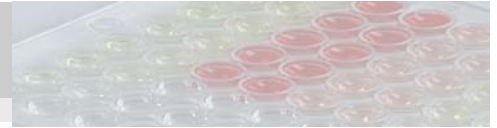


Table 1

Statistic	StatLIA® TrueFit™ 5PL	4PL
rSSE	10.2	28,568
Degrees of freedom	1	2
Residual variance	10.2	14,284
Fit Prob™	0.889	<0.001
$R^2$	0.997	0.996

Figure 1 shows two curve fits with visibly different quality, yet  $R^2$  failed to detect it. The reason is that  $R^2$  was designed to measure the quality of fit for linear regression methods and it is not sufficiently sensitive to measure residual error in nonlinear regressions.

Table 1 summarizes the curve fit statistics for the two curves illustrated in Figure 1.



### What is the biological question?

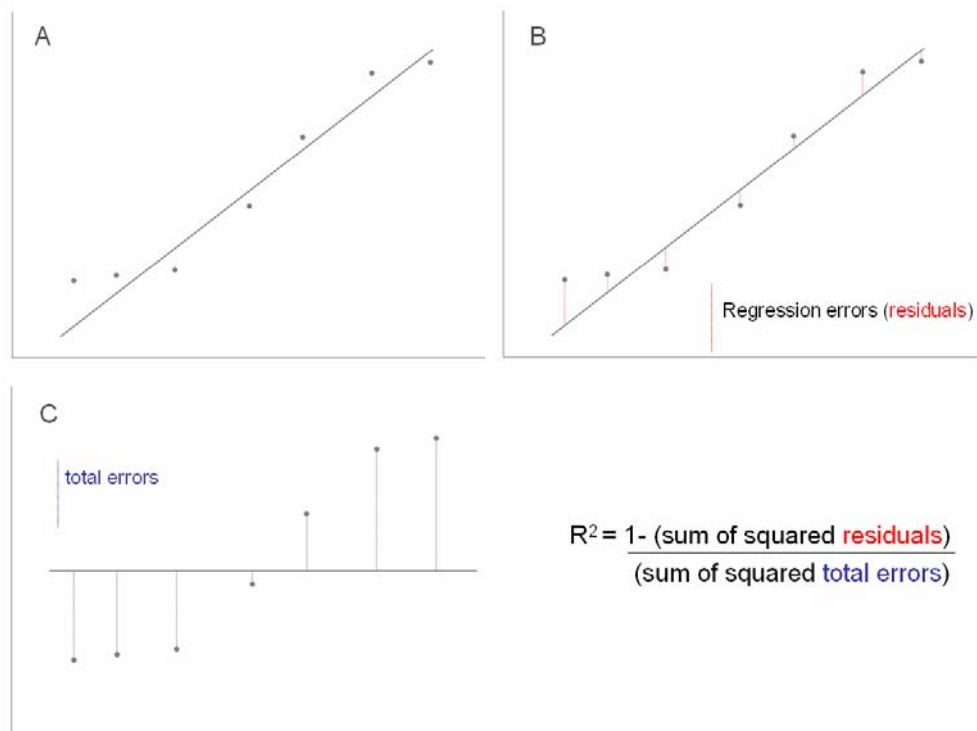
The biological question being assessed is how well the candidate standard curve matches the true dose-response curve of the biological reaction represented by the data points. By looking at how  $R^2$  and FitProb are calculated, it clarifies why FitProb measures the accuracy of nonlinear curve fits and  $R^2$  does not.

Figure 2

### Residual Variance = Weighted Sum of Squared Residual Errors (RSSE) / Degrees of Freedom

Regression theory states that the candidate standard curve that best represents the observed data must be the closest to the true curve. The StatLIA® TrueFit™ Data Reduction system uses a weighted  $RSSE$ , where the residual error is multiplied by the weight (the reciprocal of the

**Figure 2** explains the lack of sensitivity of  $R^2$  by how the metric is calculated. The calculation depends on a fraction of the sum of squared residual errors ( $RSSE$ ) over the sum of squared total errors. As shown in Figure 2C, the problem lies with the calculation of the sum of squared total errors, which measures the difference in magnitude between the data points and a horizontal line drawn through the mean of the data points. The sum of squared residual errors from the regression is small versus the sum of the squared total errors, thus resulting in poor sensitivity to residual error.



$$R^2 = 1 - \frac{\text{(sum of squared residuals)}}{\text{(sum of squared total errors)}}$$

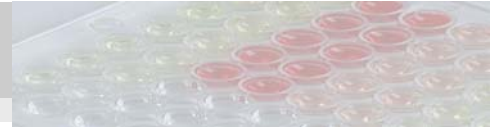
The squared total errors value dwarfs the  $RSSE$  value, which pushes the  $R^2$  value to 1.

### $R^2 = 1 - \text{Sum of Squared Residual Errors (RSSE) / Squared Total Errors}$

The problem with  $R^2$  lies with the calculation of the sum of squared total errors, which measures the difference in magnitude between the data points and a horizontal line drawn through the mean of the data points (Figure 2C). The sum of squared residual errors from the regression is small versus the sum of the squared total errors, thus resulting in poor sensitivity to residual error. This explains why the  $R^2$  values of even bad fits are usually  $>0.9$ . The squared total errors value dwarfs the  $RSSE$  value, which pushes the  $R^2$  value to 1.

variance). Weighting is performed to normalize the magnitude of the variance to the magnitude of the responses along the curve. (*Residual variance and FitProb™ require accurate weighting to be effective.*) The  $RSSE$  provides a quantifiable metric that can be monitored during curve fitting. The curve fit that yields the lowest  $RSSE$  (and therefore the highest FitProb) must be the one that most resembles the true curve.

While the  $RSSE$  is a good measure of fit, it does not by itself allow one to compare between curve fits that use different numbers of data points. Thus, one way to normalize the  $RSSE$  value is to divide it by the degrees of freedom in

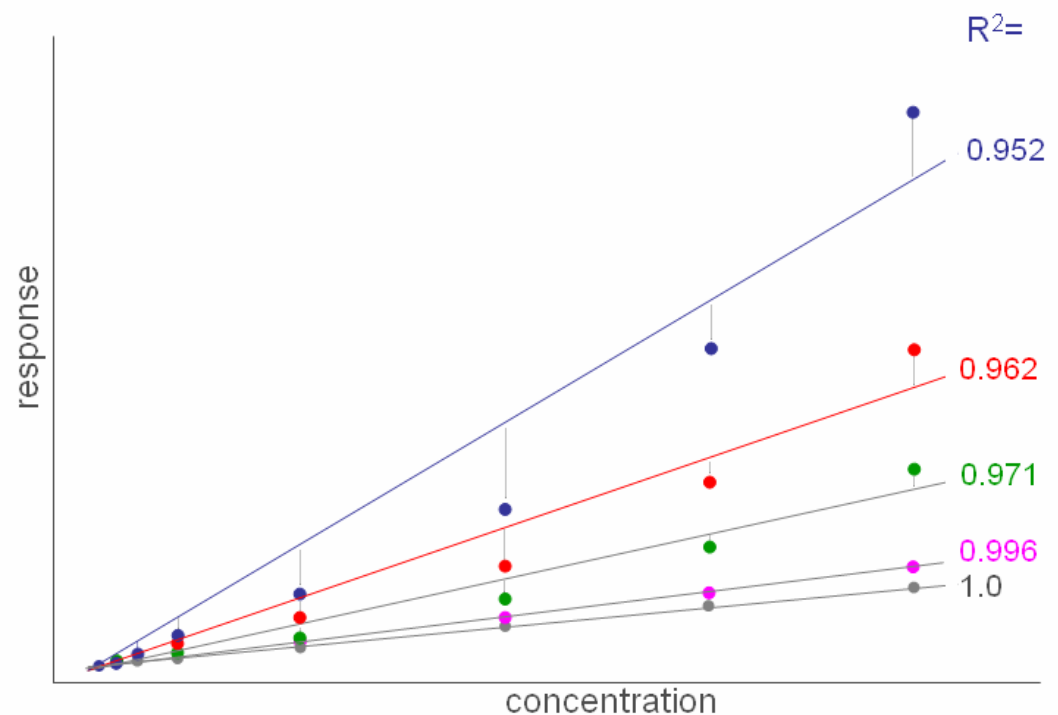


the data set. The degrees of freedom are the number of data points in the data set minus the number of parameters used in the curve fit model. For instance, if a data set with nine points is fitted with a 5PL, the degrees of freedom would be nine minus five, or four degrees of freedom. Residual variance equals the *RSSE* divided by the degrees of freedom.

lower quality curve fit <sup>5</sup>. As a percentage, this number is referred to as the *Fit Probability*, and is expressed as 0 (no fit) to 1 (perfect fit). See Table 1 and Figure 1. **Fit Probability =  $\chi^2$  Dist (RSSE, df)**

Figure 3

Figure 3 shows a range of plotted data sets that increasingly drift from linear to nonlinear (logarithmic). As the data sets become more nonlinear, the regression errors increase markedly, yet the  $R^2$  values remain greater than 0.95 for all the data sets.



#### How good is good enough? (FitProb)

The residual variance is a sensitive metric for curve fit comparisons, but does not tell us the likelihood that it is a good fit. The biological question that needs to be answered now is whether the residual variance between the candidate standard curve and the true curve is statistically valid. The hypothesis is that the candidate standard curve is the true curve (null hypothesis). To test this hypothesis, a statistical test is needed that can measure the differences between populations. Since the residual variance is a  $\chi^2$  distributed value, the  $\chi^2$  test is the appropriate test. The  $\chi^2$  test of the *RSSE* tells the percentage of assays that, if performed under the same conditions, would have a lower *RSSE*, or

#### References

- (1) Bates, D.M.; Watts, D.G. Comparing Models. Nonlinear Regression Analysis and Its Applications; Wiley: New York 1988; 103-108.
- (2) Dudley, R.A.; Edwards, P.; Ekins, R.P.; Finney, D.J.; McKenzie, I.G.M.; Raab, G.M.; Rodbard, D.; Rodgers, R.P.C. Guidelines for Immunoassay Data Processing. *Clinical Chemistry* 1985, 31 (8), 1264-1271.
- (3) Finney, D.J.; Phillips, P. The Form and Estimation of a Variance Function, with Particular Reference to Immunoassay. *Applied Statistics* 1977, 26, 312-320.
- (4) Rodbard, D.; Munson, P.J.; DeLean, A. Improved Curve Fitting, Parallelism Testing, Characterization of Sensitivity and Specificity, Validation, and Optimization for Radioimmunoassays. In *Radioimmunoassay and Related Procedures in Medicine 1*, Vienna, Italy 1977; Int. Atomic Energy Agency 1978; 469-504.
- (5) Gottschalk, P.G.; Dunn, J.R. The Five Parameter Logistic: A Characterization and Comparison with the Four-Parameter Logistic. *Analytical Biochemistry*, 2005; 343:54-65.



Visit [brendan.com](http://brendan.com) for these technical reports from Brendan Technologies:

- Improve the accuracy of your results with StatLIA® TrueFit™ Data Reduction System... *Where the biology defines the curve fit*
- “FitProb offers an *effective* alternative to the *ineffective*  $R^2$  metric for analyzing curve fits from nonlinear data
- “Consistent, Accurate, Reliable Parallelism Analysis...*even for your most difficult assays*”
- StatLIA® SmartQC™ Statistical Assay Analysis: Take the guesswork out of assay performance and result analysis.

Visit [brendan.com](http://brendan.com) for published manuscripts:

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

**ELSEVIER** **ANALYTICAL BIOCHEMISTRY**

The five-parameter logistic: A characterization and comparison with the four-parameter logistic<sup>17</sup>

Paul G. Gottschalk\*, John R. Dunn

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

Received 4 February 2005  
Available online 15 May 2005

**Abstract**

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

© 2005 Elsevier Inc. All rights reserved.

**Keywords:** FPL; FPL; Immunoassay; Bioassay; Data reduction; Statistical analysis; Dose response curve; Curve model

Bioassays 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 immuno-sandwich, 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 substance. 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 must 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

Computer Methods and Programs in Biomedicine (2005) 80, 324–335

**ELSEVIER** **Computer Methods and Programs in Biomedicine**

Determining the error of dose estimates and minimum and maximum acceptable concentrations from assays with nonlinear concentration versus response curves

Paul G. Gottschalk\*, John R. Dunn<sup>1</sup>

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

Received 22 February 2005; received in revised form 9 July 2005; accepted 12 August 2005

**KEYWORDS**  
Dose response curve;  
Monte Carlo method;  
True estimate;  
Five parameter logistic;  
FPL;  
Fractalon profile;  
Weighted regression

**Summary** A method is described here that uses a modified Monte Carlo method to provide an improved estimate of the confidence bounds of concentration estimates. This method accommodates even strongly nonlinear curve models, such as the five parameter logistic model, in contrast to the common but often poor approach of linearizing the regression problem and using linear theory to obtain the confidence bounds. The method uses an interpolation technique to reduce artifacts in the precision profile due to small simulation sample sizes and proximity to horizontal asymptotes in the curve model. The paper also describes how to define and calculate the minimum and maximum acceptable concentrations of dose-response curves by finding the concentration where the size of the error, defined in terms of the size of the concentration confidence interval, exceeds the threshold of acceptability determined by the application.

© 2005 Elsevier Inc. All rights reserved.

**1. Introduction**

Immunoassay techniques use antibody-antigen binding to quantify the dose of an analyte. This is done indirectly by measuring the quantity of some type of label. Depending upon whether the immunoassay is competitive or immuno-sandwich, the label is chemically attached to either the analyte or the antibody, respectively [1]. In order to quantify the concentration of the analyte, a response produced by the sample is compared to a calibration curve, commonly called the standard curve. The unknown concentration of an analyte is then estimated by finding the concentration on the standard curve that produces the same response as that obtained from the unknown sample [2,3]. In some applications, unknown dilution curves are compared to standard curves to determine the degree of parallelism or linearity and relative potency [4]. The standard curve is an estimate of the true curve; the curve that expresses the dose-response

Journal of Biopharmaceutical Statistics, 15: 447–463, 2004  
Copyright © Taylor & Francis, Inc.  
ISSN: 1533-0346 print/ISSN 1744-5019 online  
DOI: 10.1081/BIP-20040112

**Taylor & Francis**

MEASURING PARALLELISM, LINEARITY, AND RELATIVE POTENCY IN IMMUNASSAY AND BIOLUMINESCENCE DATA

Paul G. Gottschalk and John R. Dunn  
Brendan Technologies, Inc., Carlsbad, California, USA

There is often a need to determine parallelism or linearity between pairs of dose-response data sets for various biological applications. This article describes a technique based on a modification of the well-known extra-sum-of-squares principle of statistical regression. The standard extra-sum-of-squares method uses an F-distribution ratio as a statistic, and as a consequence this statistic on the parallelism test. It is shown here that this metric does not directly measure the parallelism between the two curves, and can often vary in opposition to actual parallelism. To overcome this problem, a metric based on a likelihood ratio test applied directly on the likelihood-based extra-sum-of-squares statistic is developed, which is shown to correspond directly to parallelism. This parallelism metric does not suffer from the shortcomings of the conventional F-ratio-based metric, and is a more reliable and appropriate measure of parallelism. The article also shows that the choice of curve model has a large effect on the sensitivity of relative potency, and that using an asymmetric model, such as the asymmetric five-parameter logistic function, a generalization of the commonly used symmetric four-parameter logistic function, is necessary when working with asymmetric dose-response data. The effect of noise, as well as the importance of correct weighting on the parallelism metric will be relative potency, is also studied.

**Key Words:** Bioassay; Five parameter logistic; Four parameter logistic; Immunoassay; Linearity; Nonlinear weighted regression; Parallelism; Relative potency

**1. INTRODUCTION**

The determination of parallelism or linearity between sets of dose-response data plays an important role in a number of biological applications. Such applications include drug comparison, analyte confirmation, cross-reactivity, interfering substances, matrix compensation, concentration estimation, and stability studies. These applications can be reduced to two purposes: to determine if the biological responses to two substances are similar (indicating the similarity of the substances), or to determine whether two different biological environments will give similar dose-response curves to the same substance. The difference in the biological environments are often different matrices, but can be differences between any component of the reaction mixture or the incubation condition. These latter applications are generally referred to as linearity or recovery studies (Kilfi et al., 2000; Krieger and Schmitt, 1993; Thoden, 1992). They are the

Received August 15, 2004; Accepted November 15, 2004  
Address correspondence to Paul G. Gottschalk, Brendan Technologies, Inc., 2236 Rutherford Road, Suite 107, Carlsbad, CA 92008, USA. Email: [pgottschalk@brendan.com](mailto:pgottschalk@brendan.com)

Visit [brendan.com](http://brendan.com) to download a demo version of StatLIA®.

Copyright © 2010 Brendan Technologies, Inc. All rights reserved.

StatLIA is a registered trademark and TrueFit and SmartQC are trademarks of Brendan Technologies, Inc., 1947 Camino Vida Roble, Suite 215, Carlsbad, California, 92008, USA, 760-929-7500 (v), 760-929-7504 (f), [sales@brendan.com](mailto:sales@brendan.com).