StatLIA® from Brendan

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^2 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[1,2,3,4], 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 thet curve fits the individual data points. If the data is accurately weighted, the *RSSE* is χ^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

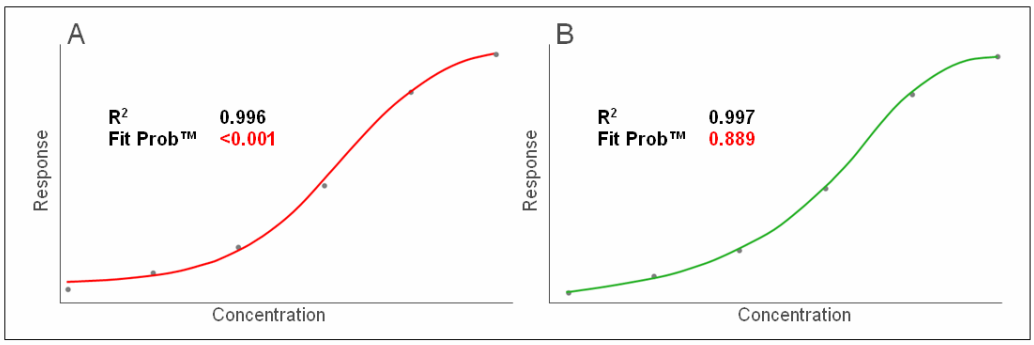


Table 1

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

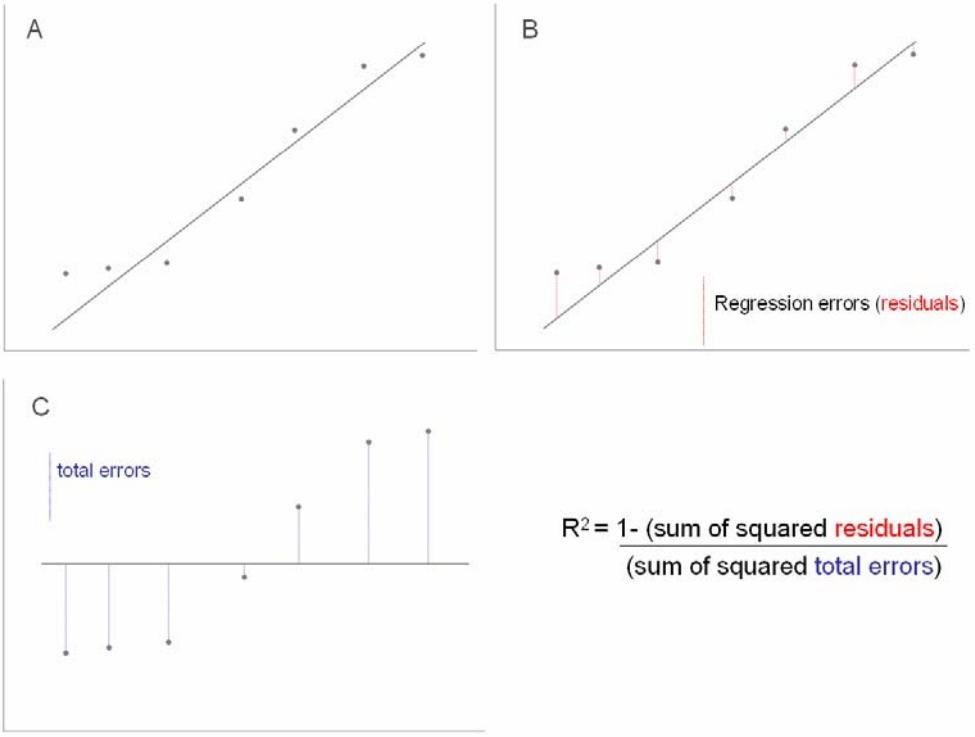


Figure 2 explains the lack of sensitivity of R^2 by how the metric is calculated. The calculation depends on a fraction of the sum of squared residual errors (*RSSE*) over the sum of squared total errors. As shown in Figure 2C, the problem lies with the calculation of the sum of squared total errors, 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.

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

R2 = 1 – 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.

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 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.

Figure 3

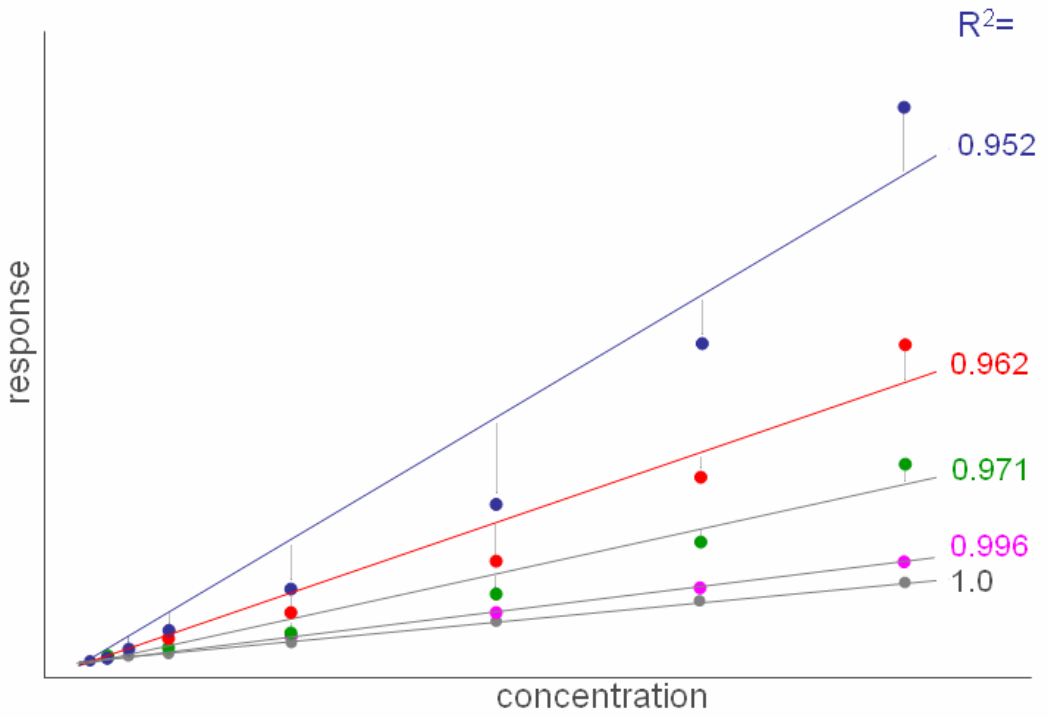


Figure 3 shows a range of plotted data sets that increasingly drift from linear to nonlinear (logarithmic). As the data sets become more nonlinear, the regression errors increase markedly, yet the R^2 values remain greater than 0.95 for all the data sets.

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 χ^2 distributed value, the χ^2 test is the appropriate test. The χ^2 test of the *RSSE* tells the percentage of assays that, if performed under the same conditions, would have a lower *RSSE*, or lower quality curve fit [5]. 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 = χ^2 Dist (RSSE, df)**

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.

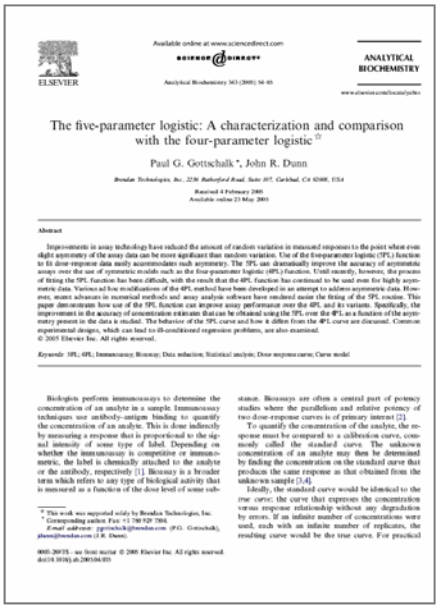
(4) Rodbard, D.; Munson, P.J.; DeLean, A. Improved Curve Fitting, Parallelism Testing, Characterization of Sensitivity and Specificity, Validation, and Optimization for Radioimmunoassays. *In Radioimmunoassay and Related Procedures in Medicine 1, Vienna, Italy* **1977**; *Int. Atomic Energy Agency* **1978**; 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 for these technical reports from Brendan Technologies:

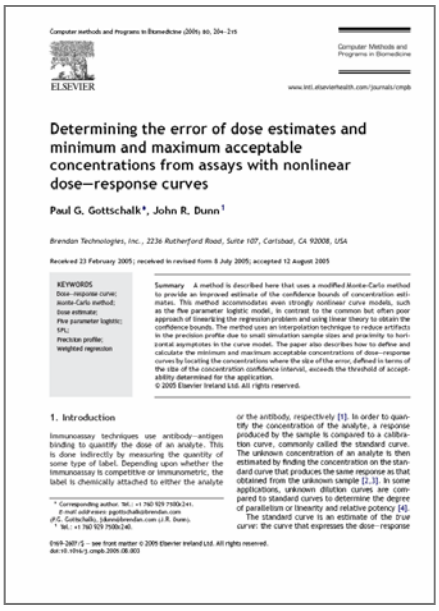
- Improve the accuracy of your results with StatLIA® TrueFit™ Data Reduction System… *Where the biology defines the curve fit*
- "FitProb offers an *effective* alternative to the *ineffective* R^2 metric for analyzing curve fits from nonlinear data
- "Consistent, Accurate, Reliable Parallelism Analysis…*even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

Visit brendan.com for published manuscripts:



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

Paul G. Gottschalk *, John R. Dunn



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

Paul G. Gottschalk*, John R. Dunn[1]

MEASURING PARALLELISM, LINEARITY, AND RELATIVE POTENCY IN BIOASSAY AND IMMUNOASSAY DATA

Paul G. Gottschalk and John R. Dunn

Visit brendan.com to download a demo version of StatLIA®.