Assessing Goodness-of-Fit of Non-Linear Models During Calibration Inference in Biological Assay

Eloi P. Kpamegan, Ph.D., MSF
Director, Clinical & Nonclinical Biostatistics
Novavax Inc.
Rockville, MD
Abstract

Often to estimate unknown concentrations from an immunoassay data processing, non-linear models are applied to sculpt the relationship between concentration and response, i.e., calibration. In pharmaceutical drug development and clinical laboratories, the software used to process these data commonly relies on the coefficient of determination ($R^2$) to assess the goodness-of-fit of the non-linear models. Consequently, the results generated could be deceptive and therefore put the estimation at unforeseeable risk. This paper discusses a statistical method to assess a goodness-of-fit of non-linear models during calibration inference in immunoassays using a chi-square test. Empirical examples are given and are shown to yield better outcomes. In addition, the impact of the selection of variance function on the goodness-of-fit test is also addressed.

Key Words: Goodness of fit, Biological Assays, chi-square, coefficient of determination
Outline

• Introduction
• Reference Curve Model
• The 4-Parameter Logistic function
• Estimation of concentration-response function
• Assessing Goodness-of-fit
  – Statistical methods to assess goodness-of-fit
  – Weighting
  – Variance Function
• Examples
• Conclusion
Introduction

• Immunoassay
  – Standard problem in laboratory Science and drug development is the derivation of suitable assays for measuring levels of the analyte (e.g. antibody) in test samples.
  – Statistical challenge: establish the relationship between the measured response and the concentration (standard curve).
  – Commonly used assay techniques:
    • Radioimmunoassay (RIA)
    • Enzyme-Linked immunoabsorbent assay (ELISA)
    • etc.
Reference Curve Models

- Standard or reference curve would be the “TRUE” curve.
- To approximate the true curve, many standard models have been used
  - Linear
  - Cubic Splines
  - Lines in logit-log space
  - Logistic Model
- The four-parameter logistic function is a commonly used to model the concentration-response relationship
  - More successful than other previous models
  - It is a symmetrical function.
The 4-Parameter Logistic Function
The Definition of the Model

\[ E(Y) = d + \frac{(a - d)}{1 + \left(\frac{x}{c}\right)^b} \]

- \( E(Y) \): expected response
- \( E(Y) \rightarrow a \) when \( x \rightarrow 0 \)
- \( E(Y) \rightarrow d \) when \( x \rightarrow \infty \)
- \( b \): “slope” or shape parameter
- \( x_{\text{mid}} \): dose at which 50% of the maximal response is observed
The 4PL Parameters: The Lower Asymptote

\[ E(y) = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^b} \]
The 4PL Parameters: The Upper Asymptote

\[ E(y) = d + \frac{a - d}{1 + \left( \frac{x}{c} \right)^b} \]
The 4PL Parameters: The Slope

\[ E(y) = d + \frac{a - d}{1 + \left( \frac{x}{c} \right)^b} \]
The 4PL Parameters: \( EC_{50} \) or \( IC_{50} \)

\[
E(y) = d + \frac{a - d}{1 + \left( \frac{x}{c} \right)^b}
\]
Other 4PL Model Used

\[ y = d + \frac{a - d}{1 + \exp\{b[\log(x - c)]\}} \]

Where

- \( a \) = estimated response at zero concentration (lower asymptote)
- \( b \) = slope factor or shape parameter
- \( c \) = \( \log(\text{dose}) \) at which 50% of the maximal response is observed
- \( d \) = estimated response at infinity concentration (upper asymptote)
Estimation of Concentration-Response Function

• Steps to estimate the “Best” standard curve
  ➢ Select the mathematical curve model
  ➢ Fit the curve - find the particular curve out of the entire family of possible curves that best explains the data. (Often use the maximum likelihood approach)
  ➢ Assess the goodness-of-fit.
Statistics Methods to Assess Goodness-of-Fit

- Typical statistical methods used in software packages to assess goodness-of-fit of non-linear model (e.g. 4PL)
  - \( R^2 \geq 0.95 \)
  - \( CV\% \leq p\% \) (i.e. 20%)
- Assess goodness-of-fit using F-test
  - Test for lack of fit \( (F = \frac{MSE_{Pure\ Error}}{MSE_{Lack-Of-Fit}}) \)
  - Require replicates at each dilution
- Fit Probability
  - Use Chi-square distribution \( (DF=N-4) \)
Fit Probability

- Finding the parameters of the maximum likelihood curve is equivalent to finding the curve whose parameter generate the smallest weighted sum of square error (wSSE).

- The weighted sum of square is the sum of all of the squares of the differences ($\Delta^2$) between the observed standard response ($y_i$) and the response predicted by the curve model ($\hat{y}$), weighted by the inverse variance of the standard response at that concentration.

$$wSSE = \sum_{i=1}^{N} w_i [y_i - \hat{y}]^2 = \sum_{i=1}^{N} w_i [\Delta_i]^2$$
Measure goodness-of-fit Using Fit Probability

- Fit probability
  - Pearson goodness-of-fit theory:
    - $wSSE$ follow a Chi-square distribution: $wSSE \sim \chi(DF)$
  - Criteria for good-fit: Fit Probability $\geq 0.90$

If use 8 points to fit a 4PL and the $wSSE$ is estimated as 1.00, then Fit probability = 0.91

If use 8 points to fit a 4PL and the $wSSE$ is estimated as 6.25, then Fit probability = 0.10
Weighting $w_{SSE}$

- Assign weight to

$$w_{SSE} = \sum_{i=1}^{N} w_i [y_i - \hat{y}]^2$$

- Weight is the inverse variance (1/variance) of the response at the concentration
- Run enough replicates to reliably estimate the variance function;
  - Impractical, cost
  - a pool of historical data can be used to compute this variance function
Variance Functions

- Variance can usually be approximated by a power function of the response: (DJ Finney, 1978)
  \[ \beta_1 \times \text{Resp} \beta_2, \beta_2 \in (1.0, 2.0) \]
  - If \( \beta_2 = 1.0 \), response is proportional to the variance
  - If \( \beta_2 = 2.0 \), response is proportional to the \((\text{Variance})^{1/2}\) (Notice, CV = Std / Mean)
  - ELISA: \( \beta_2 \rightarrow 2.0 \); RIA \( \beta_2 \rightarrow 1.0 \) (O’Connell, Belanger)

- Variance Function parameters are estimated using SAS or R
Variance function for 4PL Model

> nls(Variance ~ A*(Response)^B, data = Rawdata.From.File, start = list(A=10, B=1.5));

Nonlinear regression model

model:  Variance ~ A * (Response)^B
data:  Rawdata.From.File

A         B
0.6992852 1.149336

Variance = 0.6993 \times response^{1.1493}

Weight = \frac{1}{Variance}
Comparison of Goodness-of-fit for Various Curves

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Degree Freedom</th>
<th>CV (%)</th>
<th>$R^2$</th>
<th>Fit Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1463</td>
<td>1.4884</td>
<td>0.0895</td>
<td>0.0889</td>
<td>20</td>
<td>14.38</td>
<td>0.9988</td>
<td>0.9907</td>
</tr>
<tr>
<td>4.3801</td>
<td>1.3632</td>
<td>0.1205</td>
<td>0.0884</td>
<td>20</td>
<td>12.76</td>
<td>0.9977</td>
<td>0.9185</td>
</tr>
<tr>
<td>4.3575</td>
<td>1.3732</td>
<td>0.1381</td>
<td>0.0581</td>
<td>20</td>
<td>11.71</td>
<td>0.9974</td>
<td>0.8335</td>
</tr>
<tr>
<td>4.2993</td>
<td>1.3080</td>
<td>0.0889</td>
<td>0.0935</td>
<td>20</td>
<td>14.47</td>
<td>0.9954</td>
<td>0.3303</td>
</tr>
<tr>
<td>4.2141</td>
<td>1.5104</td>
<td>0.0860</td>
<td>0.1251</td>
<td>20</td>
<td>15.05</td>
<td>0.9952</td>
<td>0.0766</td>
</tr>
</tbody>
</table>
Example of Good Fit (1/3)
Example of Good Fit (2/3)
Example of Good Fit (3/3)
Example of Bad Fit (1/2)
Goal of selecting “best” variance function (I)

- Underestimating variance function subject to Type I Error
Goal of selecting “best” variance function (II)

- Overestimating variance function subject to Type II Error
Estimation of Variance
## Estimation of Variance

<table>
<thead>
<tr>
<th>Assay</th>
<th>R²</th>
<th>CV(%)</th>
<th>Lack-of-Fit</th>
<th>Degree of Freedom</th>
<th>Fit Variance Function using 8 Dilution</th>
<th>Fit Variance Function using 6 Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100% (54/54)</td>
<td>100% (54/54)</td>
<td>87% (47/54)</td>
<td>20</td>
<td>81% (44/54)</td>
<td>96% (52/54)</td>
</tr>
<tr>
<td>B</td>
<td>100% (54/54)</td>
<td>100% (54/54)</td>
<td>50% (27/54)</td>
<td>20</td>
<td>98% (53/54)</td>
<td>98% (53/54)</td>
</tr>
<tr>
<td>C</td>
<td>100% (54/54)</td>
<td>100% (54/54)</td>
<td>61% (33/54)</td>
<td>20</td>
<td>52% (28/54)</td>
<td>93% (50/54)</td>
</tr>
<tr>
<td>D</td>
<td>100% (54/54)</td>
<td>100% (54/54)</td>
<td>39% (21/54)</td>
<td>20</td>
<td>91% (49/54)</td>
<td>74% (40/54)</td>
</tr>
</tbody>
</table>
Finney’s variance function fit to the ELISA assay
Conclusion

• Several software used for immunoassays data processing in the pharmaceutical laboratories are outputting $R^2$ as a statistics to assess goodness of fit for nonlinear model.

• Goodness of fit assessment using $R^2$ is not appropriate for nonlinear models. Its use could be deceptive and consequently put the assay development effort unforeseeable risk.

• This presentation addressed several statistical issues in implementation of fit probability algorithm. Notice the selection of variance function could potentially cause markedly impact the result of assessing goodness of fit.

• Pooling information across repeated assay runs to estimate variance function parameters will provide better estimates (Davidian and Giltinan, 1993).