A Statistical Method to Account for Plate-to-Plate Variability in Multiple-Plate Bioassays

Bioassays are increasingly important in the development and manufacture of biopharmaceutical products for both performing bioequivalency testing and quality monitoring. Standard 96-well microplates are used to gather data to fit the dose–response curve characterizing a particular test specimen and, subsequently, to estimate the specimen’s relative potency against a standard or reference. But statistical analysis of the data can be misleading, if plate-to-plate variability isn’t included.

In a multiple-plate sampling scheme, the assay is repeated for the reference and test specimens on more than one plate, usually two or three. Different plates, however, frequently yield quite different results, even if they come from the same supplier. Experience has shown that plate heterogeneity, if ignored, can distort the statistical analysis of a bioassay from a multiple-plate scheme. Consequently, subsequent inferences about parallelism or relative potency may be misleading. A statistical method that accounts for plate-to-plate heterogeneity is needed — and presented here.

Statistics don’t lie, but if you don’t include appropriate data, the resulting statistical analyses can be misleading. In biopharmaceutical development, if the variability from several different microplates is not addressed, assays testing the relative potency against a standard can be inaccurate. A statistical method that accounts for plate-to-plate heterogeneity is needed — and presented here.

The Four-Parameter Logistic Model
One of the most common statistical models used to analyze bioassay data is the four-parameter logistic function, given by

\[ y = A + \frac{D - A}{1 + \left( \frac{z}{C} \right)^B} \]  

in which \( y \) is the measured response (such as optical density), \( z \) is the dose, \( A \) is the upper asymptote parameter, \( D \) is the lower asymptote parameter, \( C \) is the \((ED_{50})\) parameter (the dose required to elicit 50% response), and \( B \) is the “rate” parameter. Other forms of the four-parameters logistic model are given in Reeve (1) and Schlain (2).

Testing for parallelism. In bioequivalency studies, testing for parallelism between the test specimen and the reference specimen is required. When the asymptote and rate parameters \((A, B, \text{and} D)\) of both the test and reference curves are the same, the curves are said to be parallel. Parallelism tests using the sum of squares for error (SSE) are usually accomplished using the full and reduced model F-test in the format

\[ F = \frac{\text{SSE}_{\text{parallel}} - \text{SSE}_{\text{nonparallel}}}{3} \frac{3}{\text{SSE}_{\text{nonparallel}}(n-8)} \]

which has an \( F \)-distribution with a 3 numerator and \( n-8 \) denominator degrees of freedom \((df)\), where \( n \) is the combined total number of observations. The full model (denoted by “nonparallel” in the formula) estimates eight parameters in all, four for the reference and four for the test. The reduced model (denoted by “parallel”) forces the asymptotes and rate parameters to be the same for both the reference and test curves, while allowing different \( EC_{50} \) parameters. If the test and reference curves are parallel, then the reduced model is appropriate, and the relative potency can subsequently be estimated. Using the reduced model, the estimated relative potency \( \hat{\rho} \) is calculated as

\[ \hat{\rho} = \frac{\hat{C}_R}{\hat{C}_T} \]

where \( \hat{C}_R \) is the estimated reference and \( \hat{C}_T \) is estimated test \( EC_{50} \) values, respectively.

James D. Williams, Jeffrey B. Birch, and Steven Walfish

Corresponding author James D. Williams is a doctoral student and statistical consultant, and Jeffrey B. Birch is a professor of statistics in the department of statistics at Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0439, 540.552.0328, fax 540.231.3863, jdwilliams@vt.edu; and Steven Walfish is the manager of process statistics at Human Genome Sciences in Rockville, MD.
Relative potency is a scientific benchmark value used to compare the potency of a test specimen against a standard or reference specimen. If the test and reference curves are found to be nonparallel, then the relative potency is different for any given dose, whereas for parallel curves, the relative potency is the same for every dose. For example, for nonparallel curves, the relative potency of the test specimen compared with the reference specimen may increase with increasing dose, thus giving a different relative potency for any given dose. We recommend that researchers not calculate the relative potency if the parallelism test fails. For nonparallel curves, inferences about the relative potency can only be made conditional on a given response value, which may not be of much interest.

**Although** it is often assumed that plate effects influence the test and reference specimens in the same way, that may not always be the case. Random plate variability can affect the test and reference curves differently.

Multiple-plate assays. A multiple-plate sampling scheme is one in which the assay is repeated on different microplates. In a three-plate assay, for example, the same experiment is repeated on three plates, then a test curve and a reference curve are fit to the data from each of the three plates, resulting in a total of six curves. Analysis of the data can proceed in one of two ways: data from each plate are analyzed separately, and the results are compared for uniformity; or the data from all plates are combined into one data set, and the combined data are analyzed as a whole.

The first method (analyzing the data separately) ignores the plate-to-plate variability, and it is not clear how to proceed with the analysis if one or more plates fails the parallelism test, but the other plates do not. The second method (gathering all data in one data set), however, accounts for plate heterogeneity by considering the plates to be factors. The factors can then be included in the statistical model as the “plate effect.” Because the plates chosen for a particular experiment are usually chosen at random from a population of plates, the plate effect can be considered a “random” effect.

**Fixed versus random effects.** In statistical models, a fixed effect is one in which the effect is an unknown constant that needs to be estimated. In the four-parameter logistic model, the four parameters A, B, C, and D are considered “fixed” effects because they capture the mean response as a function of dose across the population of plates. The mean, therefore, is a function of the four fixed-effect parameters, and unknown constant values will need to be estimated.

On the other hand, a random effect is a random variable for which the levels of the factor were chosen at random from the population of factor levels. For multiple-plate bioassays, the plates are the levels of the random effect when the plates for a given experiment are randomly chosen. If the experiment were to be repeated, a different random selection of plates would be chosen, which is different from a fixed effect where the levels of the fixed effect (the plates) were chosen a priori and remain fixed throughout repetitions of the experiment.

**The distribution variance.** Because the plate random effect is a random variable, it has a statistical distribution. The variance of this distribution is called the variance component of the random factor and must be estimated from the data. If the plates used in any particular experiment are not exactly a random sample of the plates, it is still conceivable to think of the plates as a random effect. The reason for this is because we desire that the subsequent inferences about the test and reference curves not be limited to the particular plates chosen for the experiment but rather for the average behavior across plates. By considering plates as a random effect, we allow inferences to be made for the relative potency estimate across the population of plates. Otherwise, if plates are considered a fixed effect, we limit our conclusions to only those few plates used for the experiment, not across the population of plates. If the researcher wants to limit the inferences to the plates at hand, then the plate effect should be considered a fixed effect.

Statistical models can include either fixed effects, random effects, or both. Those models that include both fixed and random effects are called **mixed models**; Laird and Ware discuss these linear mixed models (3). The four-parameter logistic model has four fixed-effect parameters. If an additional random effect for the plates is added, the model becomes a mixed model. Because the four-parameter logistic curve is a nonlinear function, our model is called a **nonlinear mixed model (NLMM)**.

**Nonlinear Mixed Models**

NLMMs are contemporary statistical models that allow nonlinear random effects as well as fixed effects to be incorporated. Schabenberger and Pierce have published a good introduction to NLMMs (4). Davidian and Giltinan (5) and Vonesh and Chinchilli (6) offer thorough discussions of NLMMs. More technical details can be found in Davidian and Giltinan (7) and Lindstrom and Bates (8).

NLMMs are common in pharmacokinetic data analyses. For example, Littell et al. (9) describe an NLMM used to analyze data from a clinical trial described by Grasela and Donn (10). Because NLMMs are flexible and sophisticated, their popularity and use are increasing in biological applications.

In applying NLMMs, the random effect for the plates can affect several different parameters in the four-parameter logistic model. For example, the plate might affect the upper

<table>
<thead>
<tr>
<th>Model</th>
<th>Random Effect</th>
<th>(-2(\text{LogLikelihood}))</th>
<th>(X^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>none</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.** The random effect chosen and the \(-2(\text{LogLikelihood})\) and \(X^2\) values for Equations 1–5
asymptote $A$ or the $EC_{50}$ parameter $C$. To incorporate the plate random effect into the four-parameter logistic model, an appropriate random effect for the data is added to the model in Equations 2–5

\[
y = A + a + \frac{D - (A + a)}{1 + \left(\frac{z}{C}\right)^b} \tag{2}
\]

\[
y = A + \frac{D - A}{1 + \left(\frac{z}{C+b}\right)^b} \tag{3}
\]

\[
y = A + \frac{D - A}{1 + \left(\frac{z}{C+c}\right)^b} \tag{4}
\]

\[
y = A + \frac{(D + d) - A}{1 + \left(\frac{z}{C}\right)^b} \tag{5}
\]

in which $a$, $b$, $c$, and $d$ are plate random effects. In Equation 2, the $a$ random effect is added to the model; in Equation 3, the $b$ random effect is added, and so forth. So if the plate variability is manifest in the upper asymptote parameter $A$, then $A$ is simply replaced with $(A+a)$ in the model, where $a$ is the random effect. We use the SAS (formerly called statistical analysis software) package to analyze data using NLMMs. An example of the SAS code for Equation 2 is given in the “Using SAS for NLMMs” box. The estimates are found using the maximum likelihood (ML) method.

**Different random effects for test and reference.** Plate effects are generally assumed to influence the test and reference specimens in the same way. This, however, may not always be the case: The random plate effect might not affect the test curve in the same way that it affects the reference curve. If that happens, an additional random effect can be added to the model: one for the test curve and one for the reference curve for each plate. For an example of how to fit two random effects into Equation 2, see the “Two Random Effects Per Plate” box.

**Random effects on two or more parameters.** It is possible for the random effect of a plate to influence more than one parameter, affecting both $A$ and $B$, for example. Theoretically, the NLMM approach can accommodate this situation. From a practical viewpoint, however, computational problems can result from estimating these additional variance components. In our experience, we were fortunate to find that these types of models were statistically insignificant.

**The Appropriate Random Effect**
Choosing the appropriate model for the data is important. As discussed, if the random effect of the plate is made manifest in the upper asymptote parameter $A$, then the $a$ random effect should be added to the model (Equation 2); if the random effect of the plates are made manifest in the rate parameter $B$, then the $b$ random effect should be added to the model (Equation 3), and so on.

**The likelihood ratio.** For a particular bioassay, the random effect to choose (if any) for the final model is not always clear. To find the random effect model that is most appropriate for the data, several different selection methods can be used. One such method is the likelihood ratio test (LRT). To compare two competing models that are “nested” in the parameters, we can use the LRT. Nested models have one model as a subset of the other. In other words, the two models are identical except that one model has additional parameters. The model with the greater number of parameters is called the full model, and the model with fewer parameters is the reduced model. For example, Equation 1 is nested in both Equation 2 and Equation 3. The LRT equation has the format shown in (a) in the “Using LRTs” sidebar, which also shows: (b) how to test the significance of a random effect and (a) how to use the LRT to test for parallelism.

**Estimating Relative Potency**
If, using LRT, the test and reference curves are found to be parallel, the relative potency estimate can be calculated using the parameter estimates from the reduced model. The estimated relative potency $\hat{\rho}$ is derived from the equation

\[
\hat{\rho} = \frac{\hat{C}_R}{\hat{C}_T}
\]

which can be computed in SAS code using the “Estimate” statement. The confidence interval associated with the estimate of
Using LRTs, Equation (a) is the equation for determining the likelihood ratio test. It can also be used to test for parallelism in nonlinear mixed models (NLMM). Equation (b) is used to test for significance; in this example it tests whether the random effect of Equation 2 is significant.

### Choosing the Appropriate Random Effect

The $-2\log \text{Likelihood}^{\text{reduced model}}$ is the $-2\log \text{Likelihood}$ value obtained from fitting the reduced model. The $-2\log \text{Likelihood}^{\text{full model}}$ is the $-2\log \text{Likelihood}$ value obtained from fitting the full model. The $-2\log \text{Likelihood}$ values come as standard output for most statistical packages equipped to fit NLMMs.

### Determining Significance

The LRT statistic has an approximate $X^2$ distribution, with degrees of freedom equal to the difference between the number of parameters in the full model and the number of parameters in the reduced model. The second LRT format in the “Using LRTs” box provides an example for testing whether Equation 2 (the a random effect) is significant.

Because Equation 2 has one more parameter than Equation 1, the number of degrees of freedom is one. For an $\alpha=0.01$ test, the 99% upper tail cutoff of an $X^2$ distribution with one degree of freedom, is 6.635. So if $X^2$ is greater than 6.63, Equation 2 is significant. This process may be repeated for Equations 3–5 as well (Table 1). If more than one model is significant, choose the model with the largest $X^2$ statistic.

### The Parallelism Test

Once an appropriate random effect has been chosen for the model, the parallelism between the test and reference curves must be tested. As discussed, a four-parameter logistic model is tested for parallelism by using the full and reduced model $F$-test, which is necessary when all effects are fixed. In NLMMs, parallelism is tested using the LRT (as shown by the first model in the “Using LRTs” box). The full model contains the appropriate random effects and has different asymptote and rate parameters ($A$, $B$, and $D$) for both the test and reference curves; the reduced model contains the same random effects and has the same asymptote and rate parameters for both the test and reference curves.

As in the tests for choosing the appropriate random event, the $-2\log \text{Likelihood}^{\text{reduced model}}$ is the $-2\log \text{Likelihood}$ value obtained from fitting the reduced, or parallel model. Similarly, $-2\log \text{Likelihood}^{\text{full model}}$ is the $-2\log \text{Likelihood}$ value obtained from fitting the full, or nonparallel model. This test statistic has an approximate $X^2$ distribution with three degrees of freedom because the difference between the number of parameters in the full model and the number of parameters in the reduced model is three. For an $\alpha=0.01$ parallelism test, the 99% upper tail cutoff of an $X^2$ distribution with three degrees of freedom is 11.34. So if the $X^2$ statistic is greater than this cutoff, the null hypothesis of parallelism is rejected. The “Testing Parallelism” box shows an example of how to use SAS code and PROC NLMIXED to test parallelism in Equation 2.

### Initial Parameter Values

In all nonlinear models, starting values (sometimes called initial values) for all the parameters to be estimated must be supplied to find the maximum likelihood estimates. The reason is because no closed-form solution for ML parameter estimates exists. An iterative procedure must be applied, starting with the initial values and iteratively updating them until convergence.

Finding starting values for the NLMM is similar to finding initial values for the fixed-effects logistic model. The only difference is that — in addition to the initial values for

---

**Using SAS code.** SAS code can be used in Equation 2. Input the data in a way that allows PROC NLMIXED to consider plate as a factor. That amounts to “stacking” the data by plate. The following SAS code will stack the data and create the data set “bioassay.” Once the data set is created, the data must be sorted by plate for this code to work properly.

```sas
data bioassay;
  input sample$ dose plate1 plate2 plate3;
  array pl{3} plate1 plate2 plate3;
  do i = 1 to 3;
    input sample$ dose plate1 plate2 plate3;
    datalines;
    ... the rest of your data goes here ...
  end;
  proc sort data=bioassay;
    by plate;
  run;

  The following PROC NLMIXED code will fit the full or nonparallel model with the a random effect included.
  proc nlmixed data=bioassay;
    parms A1=10 B1=-1 C1=2 D1=0
       A2=10 B2=-1 C2=2 D2=0
       s2a=1000 s2e=10000;
    if sample = 'ref' then
      themodel = A1+a + (D1-(A1+a))/(1+(dose/C1)**B1);
    if sample = 'test' then
      themodel = A2+a + (D2-(A2+a))/(1+(dose/C2)**B2);
    model resp ~ normal(themodel,s2e);
    random a ~ normal(0,s2a)
      subject=plate;
  run;
```
Helpful SAS Code for Determining Plate-to-Plate Variability

## Estimating Relative Potency

The first PROC NL MIXED statement fits the full or nonparallel model, and the second fits the reduced or parallel model. Included in the output is the relative potency estimate and its associated confidence interval.

### SAS Code

```sas
proc nlmixed data=bioassay;  
parms A=10 B1=-1 C1=2 D=0 s2a=1000 s2e=10000; 
if sample = "ref" then 
  themodel = A + (D-(A+a_rand))/(1+(dose/C1)**B); 
if sample = "test" then 
  themodel = A + (D-(A+a_rand))/(1+(dose/C2)**B); 
model resp ~ normal(themodel,s2e); 
random a_rand ~ normal(0,s2a) subject=plate; 
estimate 'Rel Pot' c1/c2; 
run;
```

### Output

| Label       | Estimate | Error  | DF | t Value | Pr>|t| | Alpha | Lower | Upper |
|-------------|----------|--------|----|---------|------|------|-------|-------|
| Standard    | 0.9836   | 0.0583 | 53 | 6.87    | <0.0001 | 0.05 | 0.8667 | 1.1005 |

The relative potency estimate is 0.9836 and the 95% confidence interval is (0.8667, 1.1005). Because the interval covers one, the dose-response curves of the test and reference specimen are not statistically different.

### Two Random Effects Per Plate

This SAS code fits the full (nonparallel) model with two random effects per plate: one random effect for the test, and one for the reference.

```sas
proc nlmixed data=bioassay;  
parms A1=10 B1=-1 C1=2 D1=0 a11=1000 a21=10000; 
if sample = "ref" then 
  themodel = A1+a11 + (D1-(A1+a11))/(1+(dose/C1)**B1); 
if sample = "test" then 
  themodel = A2+a22 + (D2-(A2+a22))/(1+(dose/C2)**B2); 
model resp ~ normal(themodel,s2e); 
random a ~ normal(0,s2a) subject=plate; 
run;
```

### Output

<table>
<thead>
<tr>
<th>X2</th>
<th>5.6.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2</td>
<td>5.6.</td>
</tr>
<tr>
<td>Pr&gt;</td>
<td>t</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01</td>
</tr>
<tr>
<td>Lower</td>
<td>0.05</td>
</tr>
<tr>
<td>Upper</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Because 5.6 is less than 11.34, the α=0.01 cutoff from an X2 distribution with three degrees of freedom, declare that the reference and test curves are parallel.

## Testing Parallelism

The first PROC NL MIXED statement fits the full (nonparallel) model, and the second fits the reduced (parallel) model.

### SAS Code for Nonparallel Models

```sas
proc nlmixed data=bioassay;  
parms A1=10 B1=-1 C1=2 D1=0 A2=10 B2=-1 C2=2 D2=0 s2a=1000 s2e=10000; 
if sample = "ref" then  
  themodel = A1+a11 + (D1-(A1+a11))/(1+(dose/C1)**B1); 
if sample = "test" then 
  themodel = A2+a22 + (D2-(A2+a22))/(1+(dose/C2)**B2); 
model resp ~ normal(themodel,s2e); 
random a ~ normal(0,s2a) subject=plate; 
run;
```

### SAS Code for Parallel Models

```sas
proc nlmixed data=bioassay;  
parms A=10 B=-1 C=2 C2=2 D=0 s2a=1000 s2e=10000; 
if sample = "ref" then  
  themodel = A+a + (D-(A+a))/(1+(dose/C)**B); 
if sample = "test" then 
  themodel = A+a + (D-(A+a))/(1+(dose/C)**B); 
model resp ~ normal(themodel,s2e); 
random a ~ normal(0,s2a) subject=plate; 
run;
```

### Two Random Effects Per Plate

This SAS code fits the full (nonparallel) model with two random effects per plate: one random effect for the test, and one for the reference.

```sas
proc nlmixed data=bioassay;  
parms A1=10 B1=-1 C1=2 D1=0 A2=10 B2=-1 C2=2 D2=0 s2a=1000 s2e=10000; 
if sample = "ref" then 
  themodel = A1+a11 + (D1-(A1+a11))/(1+(dose/C1)**B1); 
if sample = "test" then 
  themodel = A2+a22 + (D2-(A2+a22))/(1+(dose/C2)**B2); 
model resp ~ normal(themodel,s2e); 
random a11 a22 ~ normal(0,s2a) subject=plate; 
run;
```

### Output

<table>
<thead>
<tr>
<th>X2</th>
<th>5.6.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2</td>
<td>5.6.</td>
</tr>
<tr>
<td>Pr&gt;</td>
<td>t</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01</td>
</tr>
<tr>
<td>Lower</td>
<td>0.05</td>
</tr>
<tr>
<td>Upper</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Because 5.6 is less than 11.34, the α=0.01 cutoff from an X2 distribution with three degrees of freedom, declare that the reference and test curves are parallel.

## Working at Original Dose Scale

The form of the four-parameter logistic function that we have described allows researchers to work in the original scale of the doses, rather than in the log(dose) scale, which is often used for the fixed-effects logistic model. Fitting the curve to the data and estimating the relative potency do not require that log(dose) be used instead of the dose at original scale. We recommend that the user analyze the data in the scale in which subsequent inferences are to be made.

Continued on page 54
For example, if the EC\textsubscript{50} parameter is to be reported in the original dose scale, then the analyst should use the dose in the model to estimate the EC\textsubscript{50}, not log(dose).

We have proposed a contemporary statistical method to analyze data from multiple-plate bioassays. The four-parameter logistic model for biological dose–response data has been used successfully for many years. When multiple plates are used in an experiment, the plate effect must be determined, which can be done by adding a random effect for the plate to the appropriate fixed-effect parameter in the logistic function. In this way, a nonlinear mixed model is formulated. Statistical theory for the estimation and practical use of NLMMs have been well developed for several years (7,8), and its use is spreading in the biopharmaceutical industry. **BPI**

**References**


