Designs for Bioassays with Plate Location Effects

Brian Schlain, Hitendra S. Jethwa, Meena Subramanyam, Kevin Moulder, Bhavna Bhatt, and Michael Molloy

Standard 96-well microplate assays are prone to edge and bowl effects. Such plate location effects can badly bias potency estimates. Easy-to-use plate layouts enable scientists to adjust estimates for such effects when analyzing new production drug batches. *This article is indebted to the work reported by Russell Reeve (*BioPharm*, July 2000) and the subsequent letters from David Lansky and Reeve (*BioPharm*, September 2000). Find them online at www.biopharm-mag.com.*

Corresponding author ***Brian Schlain*** *is an independent consulting biostatistician to the health care industry, 46 Jamaica Road, #1, Brookline, MA 02445, 617.731.8812, fax 617.731.2456, bschlain@ultranet.com, www.ultranet.com/~bschlain.* ***Hitendra S. Jethwa***, ***Meena Subramanyam***, ***Kevin Moulder***, ***Bhavna Bhatt***, *and* ***Michael Molloy*** *are members of the cellular and molecular bioassay department at Biogen, Inc., Cambridge, MA.*

Standard 96-well microplates (of eight rows by 12 columns) are commonly used for estimating relative potencies in bioassays. Although many authors have described the use of parallel line and parallel logistic models in bioassays (1–5), few have discussed methods for addressing microplate location effects, which are fairly common with cell-based and antiviral assays and can bias relative potency estimates. Lansky describes a split-block design and analysis method to simultaneously address plate location, multichannel pipette, and dilution error effects (6–7). His method involves strict, statistical randomization of sample types to rows (or columns) and of dose levels to columns (or rows). Racine-Poon, Weihs, and Smith discuss relative potency estimation in the context of sequential dilution errors (8).

Here, we present systematic plate layouts that allow us to adjust potency estimates for some of the commonly occurring plate location biases such as edge and bowl effects. We assume that effects from dilution and multichannel pipettes are negligible. If that assumption is in question, dilution accuracy and precision experiments should be performed. Systematic designs that adjust for trends in place of randomization have been successfully used in the context of instrument validation and calibration (9–10) and in experimental design (11).

An online article (5) and other references (1,2) provide background for the bioassay concepts in this article.

Layouts for Location Effects

In our plate layouts, samples are allocated to columns, and dose levels are assigned to rows. The same method works for layouts in which samples are allocated to rows and doses to columns. Tables 1a and 1b present systematic plate layouts that adjust for linear and quadratic column effects in simultaneously estimating relative potencies for one and two test samples (12). The plate design in Table 1b is likely to be more useful, because one of its two test samples (denoted as s and t) can be delegated as a control. To prevent edge effects in these designs, rows A and H and columns 1 and 12 are not used. Historically, linear and quadratic trend-free designs have been applied to removing systematic trends from experimental treatment comparisons (11–12). Our section on the parallel logistic model (below) shows that trend-free designs can similarly be used for adjusting relative potency estimates for linear and quadratic plate location effects, because log-potency estimates can be considered log-ED50 differences between the test and reference samples (1).

Dilution Similarity

All the commonly used methods for estimating relative potency require that dilution similarity be assumed. With dilution similarity, a test sample acts as a dilution of its reference (1,5). For each fixed dose of reference (D_R), there is a dose of test sample (D_T) to which the bioassay responds equally:

$$D_T = \frac{D_R}{\rho}$$

where the factor ρ is the relative potency. If ρ is independent of the dose, then the samples exhibit dilutional similarity.

Parallel Logistic Model

Finney showed in 1976 how a logistic curve could be used to estimate relative potencies with radioimmuno- and immunoradiometric assays (13). He used the following parameterization of the logistic model:

$$Y = a + \frac{(d - a)}{1 + e^{-2(\alpha + \beta x)}}$$

in which Y is the assay response and x is the $\log_e$(dose).

Under the assumption of dilution similarity, Finney showed that

$$\log(\rho) = \frac{\alpha_T - \alpha_R}{\beta}$$

where α_T and α_R are the α-parameters for the test and the reference samples, respectively.

Table 1a and 1b. (a) Microplate layout for relative potency estimation of a single test sample; (b) microplate layout for simultaneous relative potency estimations of two test samples

(1a) Row	Column 3	4	5	6	7	8	9	10
B	s1	r1	r1	s1	r1	s1	s1	r1
C	s2	r2	r2	s2	r2	s2	s2	r2
D	s3	r3	r3	s3	r3	s3	s3	r3
E	s4	r4	r4	s4	r4	s4	s4	r4
F	s5	r5	r5	s5	r5	s5	s5	r5
G	s6	r6	r6	s6	r6	s6	s6	r6

(1b) Row	Column 3	4	5	6	7	8	9	10	11
B	r1	s1	t1	r1	t1	s1	r1	s1	t1
C	r2	s2	t2	r2	t2	s2	r2	s2	t2
D	r3	s3	t3	r3	t3	s3	r3	s3	t3
E	r4	s4	t4	r4	t4	s4	r4	s4	t4
F	r5	s5	t5	r5	t5	s5	r5	s5	t5
G	r6	s6	t6	r6	t6	s6	r6	s6	t6

Note: s = test sample in (1a) / first test sample in (1b); 1, 2, 3, 4, 5, 6 = doses; t = second test sample in 1b
r = reference sample

Assuming dilution similarity and no more than quadratic column effects across the plate, the following extension of Finney's parameterization of the logistic model can be fit to an entire microplate.

$$Y = a + \frac{(d-a)}{1+e^{-2\left(\alpha+\beta x+\sum_{i=1}^{k} w_i \delta_i+\tau_1 v+\tau_{11} v^2\right)}}$$

where Y is the assay response, x is the $\log_e$(dose), k is the number of test samples on the plate, w_i is 1 if the test sample is i or w_i is 0 if not the test sample i ($i = 1, \ldots, k$), and v is the coded column number (for example $-4, -3, -2, -1, 0, 1, 2, 3, 4$ corresponding to the nine columns used for the plate layout in Table 1b).

Note that dilution similarity under that logistic model is tantamount to the parameters a, d, and β being constant across all samples (5). The δ_i's are adjusted for linear and quadratic column effects both by the τ_1 and τ_{11} model terms and also by the near-orthogonality properties of the plate designs in Tables 1a–b. Here, the concept of near-orthogonality relates to low correlations of δ_i's with linear and quadratic column trend. With a completely orthogonal plate design, the δ_i's could be estimated as if the trend were absent. As will be seen later, less variation in the δ_i estimates reduces variation in the $\log(\rho_i)$ estimates. Some algebra and the assumption of dilution similarity show that $\log(\rho_i) = \frac{\delta_i}{\beta}$ for test sample i (2).

For the $\log(\rho_i)$'s to be estimated trend-free, the parameter β and the δ_i's must be estimated trend-free. We discuss more later about estimating β.

The usefulness of the parallel logistic model in addressing plate location effects can be evaluated by performing dilutions of identical samples in all columns to be used following the plate design given in either Table 1a or 1b. The procedure should then be repeated on multiple plates. To each column of each plate, the following four-parameter logistic model is fit to the assay response data:

$$Y = a + \frac{(d-a)}{1+e^{-2(\alpha+\beta x)}}$$

For each plate, each of the four estimated parameters (a, d, α, and β) can be statistically analyzed using multiple regression analysis for linear, quadratic, and higher-order column trends (14). The ideal case for applying plate designs in Tables 1a–b is one in which the parameters a, d, and β show no trends, and the parameter α shows no more than quadratic trends across the columns. That case can sometimes be more easily visualized by fitting the following logistic model to each plate.

$$Y = a + \frac{(d-a)}{1+e^{-2\left(\alpha+\beta x+\sum_{c=2}^{C} w_c \delta_c\right)}}$$

where δ_c corresponds to column c. For the plate layouts in Tables 1a–b to be useful, a plot of δ_c versus c should reveal a trend in columns of no higher order than quadratic.

These designs offer some robustness for relative potency estimation when there are only quadratic trends or in cases of negligible higher-order trends across the plate for the parameters a, d, and β. By removing average differences between the test and reference samples caused by trends in those parameters, the plate design preserves dilution similarity in an average sense. This works best when the plate design is symmetrically placed with respect to quadratic trends in the a, d, and β parameters.

Testing for Parallelism (Similarity)

The assumption of parallelism (or similarity) can be checked statistically by testing the parallel model against the nonparallel model, which is given in the equation:

$$= a + \frac{(d-a)}{1+e^{-2\left(\alpha+\beta x+\sum_{i=1}^{k} w_i \delta_i+\sum_{i=1}^{k} w_i x \gamma_i+\tau_1 v+\tau_{11} v^2\right)}}$$

The nonparallel differs from the parallel model only by the addition of the terms $\sum_{i=1}^{k} w_i x\gamma_i$ which are sample offset parameters to the slope (β) of the reference sample. We can test for nonparallelism by calculating:

$$F = \frac{\left[SSE(parallel) - SSE(nonparallel)\right] \div k}{SSE(nonparallel) \div (n-6-2k)}$$

(14). If F is larger than the critical value of an F-distribution with k numerator degrees of freedom (df) and $(n-6-2k)$ denominator df (in which n is the total number of observations used in the analysis, and k is the number of test samples), then the differences in the sum of squared errors (SSE) from fitting the equation are more than that attributable to random variation alone, and the assay fails parallelism.

If the above parallel and nonparallel models fail to account for all the location effects (for example, a column trend in the

upper asymptote parameter *d*), the estimated error variance will be inflated, resulting in less sensitive parallelism testing and wider fiducial intervals around the relative potency estimates. In our experience, the sensitivity of parallelism testing depends much less on unadjusted residual plate location effects than on the underlying assay variance, which can vary by many fold depending on the reagents and on the care taken in performing the assay. Provided that the column trends in the slope (β) and the asymptote (*a* and *d*) parameters are of no higher order than quadratic, the near-orthogonal properties of the plate design should provide some robustness for the parallelism *F*-test.

Independent observations. You may have noted that the parallelism *F*-test tacitly assumes that the observations are independent, when in fact they can contain various types of nesting effects and serial correlations related to multichannel pipetting or diluting. If those sorts of effects cannot be controlled (made negligible), then parallelism testing and potency estimation can be accommodated by a split-block design method, which involves strictly randomized samples to rows (or columns) and concentration levels to columns (or rows) (7).

Another method. Alternatively, parallelism can be tested through separate analyses of the γ_1 and γ_2 estimates from the nonparallel model using the averaging method, with each plate considered an experimental unit (5). The near-orthogonality properties of the plate designs (Tables 1a–b) can help remove location effects from the γ_1 and γ_2 estimates in the same way as discussed for estimating the β parameter. For example, with *J* replicate plates, the estimated γ_{1j}'s could be treated as a random sample of *J* observations for testing the parallelism of sample 1. Using a one sample Student's *t*-test or a two-sided confidence interval, the hypothesis $\gamma_1=0$ can be tested (15). This method, which is related to the averaging method for obtaining a confidence interval around a combined potency estimate (5), makes two important assumptions: The J plates are independent with respect to assay results; and the estimated γ_{1j}'s are distributed normally across plates. Rather than focusing on parallelism for a given plate, the averaging method addresses whether a given test sample is on average parallel with the reference sample. The appropriateness of this method also depends on there being consistency in the deviation from parallelism across plates for a given test sample.

Expanded Test: Dilution Similarity

The dilution similarity test can be expanded to include checking whether the upper asymptote is constant across sample types by statistically testing the parallel against the following expanded model.

$$Y = a + \frac{\left(d + \sum_{i=1}^{k} w_i\lambda_i - a\right)}{1 + e^{-2\left(\alpha + \beta x + \sum_{i=1}^{k} w_i\delta_i + \sum_{i=1}^{k} w_i x\gamma_i + \tau_1 v + \tau_{11} v^2\right)}}$$

The expanded model differs from the parallel model only by the addition of the terms $\sum_{i=1}^{k} w_i\lambda_i$ and $\sum_{=1}^{k} w_i\, x\gamma_i$, which are sample offset parameters to the upper asymptote *d* and to the slope β of the reference sample.

We can test for similarity by calculating

$$F = \frac{\left[SSE\left(parallel\right) - SSE\left(expanded\right)\right] \div 2k}{SSE\left(expanded\right) \div \left(n - 6 - 3k\right)}$$

(14), in which SSE(parallel) and SSE(expanded) are the error sum of squares from the fitted parallel and expanded models, and where *k* is the number of test samples. If *F* is larger than the critical value of an *F* distribution with 2*k* numerator df and $n-6-3k$ denominator df (where *n* is the total number of observations used in the analysis, and *k* is the number of test samples), then the differences in SSE are more than that attributable to random variation alone, and the assay fails similarity.

Fiducial intervals, which are similar in concept to confidence intervals, can be useful in both quantifying relative asymptote discrepancies and differentiating negligible from sizable upper asymptote effects. A well-known approximation to a 95% fiducial interval for $\frac{\lambda_i}{d}$ for $i = 1$ or 2 is the following (1):

$$\frac{\hat{\lambda}_i}{\hat{d}} - \sqrt{var\left(\frac{\hat{\lambda}_i}{\hat{d}}\right)}\, t_{.975;\, df} < \frac{\lambda_i}{d} < \frac{\hat{\lambda}_i}{\hat{d}} + \sqrt{var\left(\frac{\hat{\lambda}_i}{\hat{d}}\right)}\, t_{.975;\, df}$$

where $\hat{\lambda}_i$ and $\hat{d}$ are nonlinear least squares estimates of the parameters λ_i and *d*;

$$var\left(\frac{\hat{\lambda}_i}{\hat{d}}\right), \quad \frac{var\left(\hat{\lambda}_i\right) - 2\frac{\hat{\lambda}_i}{\hat{d}} cov\left(\hat{d}, \hat{\lambda}_i\right) + \left(\frac{\hat{\lambda}_i}{\hat{d}}\right)^2 var\left(\hat{d}\right)}{\hat{d}^2}$$

where $t_{.975;\, df}$ = upper .975 percentile of Student *t*-distribution with df degrees of freedom; and df = residual error degrees of freedom for the fitted expanded model. The variances var($\hat{\lambda}_i$) and var($\hat{d}$) and the covariance cov($\hat{d}$; $\hat{\lambda}_i$) can be obtained from the output of SAS NLIN, which is the nonlinear regression module for the SAS software system. Those equations approximate a 95% fiducial interval, which is considered adequate for most practical purposes, provided that (1)

$$\frac{\left(t_{.975}\right)^2 var\left(\hat{d}\right)}{\hat{d}^2} < 0.05$$

.

With nonnegligible violations of the assumption of independent observations, the expanded test of similarity can also be performed by using the split-block design method (7). Alternatively, each λ_i and γ_i from the expanded model can be tested using the previously explained averaging method with whole plates as the experimental units.

Estimating Relative Potency

$\text{Log}_e(\rho_i)$ for the test sample *i* is estimated by $\frac{\hat{\delta}_i}{\hat{\beta}}$, in which $\hat{\delta}_i$ and $\hat{\beta}$ are nonlinear least squares estimates of δ_i (with $i=1, \ldots, k$) and β from the fitted parallel logistic model. Under the assumption of no interassay variability, the following combined estimate of $\log(\rho_i)$, denoted as $\bar{M}_i$, can be pooled across multiple microplates.

$$\bar{M}_i = \frac{\sum_{j=1}^{J} \hat{\delta}_{ij}}{\sum_{j=1}^{J} \hat{\beta}_j}$$

in which δ_{ij} and β_j correspond to the *j*th microplate, and *J*=number of microplates. Using equations given in Finney's book (1), a fiducial interval can also be constructed for $\bar{M}_i$.

The implicit assumption for the combined log-potency estimate of

no interassay variability across J replicate plates can be tested using the following approximate chi-square test (4). For convenience, the subscript of $\bar{M}_i$ will be dropped.

$$\chi^2 = \frac{\sum_{j=1}^{J} \left(M_j - \bar{M}\right)\left(M_J - \bar{M}\right) \div \left(J - 1\right)}{avg\ var\left(M_j\right)}$$

where M_j= the log-potency estimate of the jth plate for a specific test sample;

$$\bar{M} = \frac{1}{J}\sum_{j=1}^{J} M_j;$$

$$avg\ var\left(M_j\right) = \frac{1}{J}\sum_{j=1}^{J} var\left(M_j\right)$$

and var(M_j) = estimated large sample variance of M_j obtained from fitting the parallel logistic model. If χ^2 is greater than the critical value from the chi-square distribution with $J-1$ degrees of freedom (15), we can judge interassay variability to be statistically detectable.

Extensive statistical analyses of our historical data have generally supported the assumption of no interassay variability. Moreover, we have found general agreement between the interplate standard deviations and the intraplate SAS NLIN estimated standard errors of log-potency estimates. If it is unsafe to assume no interassay variability in the log-potency estimates, then the averaging method with whole plates treated as experimental units can be used to obtain a pooled log-potency estimate and a corresponding confidence interval (5). The averaging method is accommodated by the near-orthogonality of the plate designs (Tables 1a–b) to linear and quadratic location effects.

Testing for Outliers

With the full model, whether that be the expanded or the nonparallel model, outliers can be detected using the maximum absolute value of the Studentized residuals (outputted by SAS NLIN) (16). In using that test, we like to maintain the critical nominal p-value at .01. The test, developed for single outliers, tends to lose sensitivity with multiple outliers, which can be identified visually as outlier candidates using residual against dose-level plots. Again, the validity of the test depends on the assumption of independent observations.

Removing Location Effects

Implementing the methodology presented in this article is straightforward using any good nonlinear least-squares software (such as SAS NLIN), but the method is potentially vulnerable to violations of the independent observations assumption, which can be caused by the use of multichannel pipettors and serial dilutions. On the other hand, the method should have some robustness to negligible multichannel and dilution effects. When nonnegligible violations of the independent observations assumption are unavoidable, nearly orthogonal plate designs can be combined with the split-block methodology (6–7) to determine relative potency estimations from assays. In the nearly orthogonal plate designs of Tables 1a–b, dose levels can be randomly assigned to rows, and sample types can be systematically rotated among column groups (denoted by r and s in Table 1a and by r, s, and t in Table 1b). In that way, some of the plate location effects can be removed from the split–block method error variances.

The plate designs given in Tables 1a–b attempt to systematically remove location effects from the data instead of propagating those effects into the error variances by randomizing samples to columns. Randomization can widen fiducial intervals or increase the numbers of plates. Certainly, randomizing samples into judiciously chosen column (or row) blocks can reduce the error variances. In our own work, for which we know the location effects to be quadratic, we have found nearly orthogonal plate designs to be more efficient than blocking. If location effects can not systematically be removed from the data, strict randomization of samples to columns becomes essential to prevent biased relative potency estimates. Randomization across location effects (instead of systematically removing the effects) can be a source of interassay variation, because location effects tend to be inconstant across plates.

When the underlying assay variability increases rapidly in dose, the methodology we've described can be extended to include weighting functions (1–2). However, unweighted analyses are known to be robust to mild violations of the constant variance assumption (1,5). A possible rule of thumb used in the analysis of variance is that an unweighted analysis be used, provided that the error variances remain within a 1:3 ratio (17).

References

(1) D.J. Finney, *Statistical Method in Biological Assay*, 2nd ed. (Charles Griffin and Company, Ltd, London, 1978).

(2) Z. Govindarajulu, *Statistical Techniques in Bioassay*, 2nd ed. (S. Karger, AG, Basel, Switzerland, 2001).

(3) "Design and Analysis of Biological Assays," *U.S. Pharmocopeia 24–NF 19* (The U.S. Pharmacopeial Convention, Inc., Rockville, MD, 1999), Ch. <111>, pp. 1837–1847.

(4) "Statistical Analysis of Results of Biological Assays and Tests," *European Pharmacopoeia: Supplement 2000* (Council of Europe and Maisonnevue SA, 57160 Sainte-Ruffine, France, 2000) Ch. 5.3, pp. 263–293.

(5) R. Reeve, "Two Statistical Methods for Estimating Relative Potency of Bioassays," *BioPharm* 13(7), 54–60 (2000). Available at www.pharmaportal.com/articles/bp/bp0700_54-60_reeve.pdf. See also "Letter to the Editor: Nice Summary, but . . . ," *BioPharm*13(9), 10–11 (2000). Available at www.pharmaportal.com/articles/bp/bp0900_10-12_LTE.pdf

(6) D. Lansky, "Validation of Bioassays for Quality Control," *Biological Characterization and Assay of Cytokines and Growth Factors; Developments in Biological Standards,* Vol. 97, F. Brown and A.R. Mire-Sluis, Eds. (S. Karger AG, Basel, Switzerland, 1999), pp. 157–168.

(7) D. Lansky, "Strip-Plot Designs, Mixed Models, and Comparisons Between Linear and Nonlinear Models for Microtiter Plate Bioassays," *Design and Analysis of Potency Assays for Biotechnology Products:Dev. Biol.*, Vol 107, F. Brown and A. Mire-Sluis, Eds. (S. Karger AG, Basel, Switzerland, 2001), in press.

(8) A. Racine-Poon, C. Weihs, and A.F.M Smith, "Estimation of Relative Potency with Sequential Dilution Errors in Radioimmunoassay," *Biometrics* 47, 1235–1246 (1991).

(9) C. Daniel, "Calibration Designs for Machines with Carry-Over and Drift," *J. Qual. Technol.* 7(3), 103–108 (1975).

(10) B. Schlain and J.S.Krouwer, "Multifactor Designs II: A Design for Identifying Instruments with Sample-to-Sample Carry-Over and Drift," *Clin. Chem.* 35(10), 2118–2120 (1989).

(11) C. Daniel and F. Wilcoxon, "Factorial 2^{p-q} Plans Robust Against Linear and Quadratic Trends," *Technometrics* 8(2), 259–278 (1966).

(12) D.R. Cox, "Some Systematic Experimental Designs," *Biometrika*, 38, 312–323 (1951).

(13) D.J. Finney, "Radioligand Assay," *Biometrics* 32, 721–740 (1976).

(14) N. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed. (John Wiley and Sons, New York, 1981), pp. 97–98.

(15) F.J. Massey and W.J. Dixon, *Introduction to Statistical Analysis*, 4th ed. (The McGraw-Hill Companies, New York, 1983).

(16) R.E. Lund, "Tables for an Approximate Test for Outliers in Linear Models," *Technometrics* 17(4), 473–476 (1975).

(17) H. Scheffe, *The Analysis of Variance* (John Wiley and Sons, New York, 1959). **BP**