Title
Optimal Design for ELISA and other forms of Immunoassay

Journal
Technometrics, 39

Authors
Rocke, David
Jones, Geoffrey

Publication Date
1997

Peer reviewed

Optimal Design for ELISA and Other Forms of Immunoassay

David M. ROCKE and Geoffrey JONES

Graduate School of Management
University of California
Davis, CA 95616
(dmrocke@ucdavis.edu)

In enzyme-linked immunosorbent assay (ELISA), as well as in many other kinds of immunoassay, a log-logistic or similar-shaped calibration curve is fit using standards at a series of known levels and then used to transform the measured values for the unknowns into estimated concentrations. The choice of the number of standards, the concentration of the standards, and the number of replicates of the standards and of the unknowns all affect the precision of the measurement. This article develops an optimal design paradigm for this type of problem and shows how optimal choices can be calculated so that the system achieves the maximum precision of which it is capable. Although exact calculation of optimal designs requires use of a computer program, close approximations to the optimum can be derived from simple rules for hand calculation.

KEY WORDS: Calibration; Log-logistic curve; Precision.

Enzyme-linked immunosorbent assay (ELISA) is a form of chemical analysis that can be used for detection or quantitation. Originally developed for clinical use, it has been broadened to other areas, including environmental measurement. The principle on which it works is that antibodies developed to the target compound or a related compound cause inhibition, which leads to a measurable response. In ELISA, this is usually in the form of a color change that is measured by an optical reader.

ELISA's have been developed in many configurations and for many applications. One widespread application is for the determination of low molecular weight analytes in unknown samples by using calibration curves. These immunochemical analytical systems are often configured for 96-well microplates because of the ease of analysis and capacity for handling large numbers of samples. Such systems have been described in many areas, including clinical analysis for hormones (Rajkowski, Hanquez, Bouzoumou, and Cittanova 1989) and drugs (Laurie, Manson, Rowell, and Seviour 1989); analysis of foods for toxins, natural (Chu et al. 1987) or synthetic (Dixon-Holland and Katz 1988); monitoring of human exposure to toxicants (Niewola, Hayward, Symington, and Robson 1985; Bjercke et al. 1986); and analysis of environmental samples for agricultural chemicals (Hall, Deschamps, and Krieg 1989; Jones et al. 1994; Jones, Wortberg, Kreissig, Hammock, and Rocke 1995; Wortberg, Jones, Kreissig, Rocke, and Hammock 1995; Wortberg, Kreissig, Jones, Rocke, and Hammock 1995) and hazardous wastes (Vanderlaan, Stanker, Watkins, Petrovic, and Gorbach 1988).

For the routine analysis of unknown samples, such methods typically specify a protocol that includes a fixed 96-well template. These templates designate a specified number of wells for preparation of a calibration curve, either with each batch of plates or, more commonly, on each plate—Figure 1 shows an example template. The number of calibration wells and the known concentrations used for the calibration curve affect the precision of the determinations of the unknowns, as does the choice of the number of replicates of each unknown. This amounts to a problem of resource allocation. Within the framework of a 96-well plate, how does one maximize the number of samples analyzed while maintaining the best possible accuracy and precision? We present here an analysis of how the calibration and replication parameters may be selected to give the maximum precision of which the system is capable and to allow comparison among different protocols. The statistical optimization procedure described here is generalizable to virtually any calibration-curve type and to n-well plates.

Suppose that a plate with n_w wells is divided between n_s samples, each replicated k times, and n_{cal} wells devoted to calibration. An example is the ELISA protocol presented by Harrison, Braun, Gee, O'Brien, and Hammock (1989), which was used as an assay for the direct analysis of the herbicide molinate in rice-field water. A 96-well plate was used, with 15 unknown samples, quadruply replicated. The remaining 36 wells were devoted to calibration. There was one zero-concentration sample and one method blank, which should theoretically correspond to infinite concentration. (For the type of immunoassay used as an example, water placed in a well of the microplate provides a reading similar to that of a very large concentration, which would bind all of the coloring agent.) There were seven other calibration samples at a series of dilutions from 500 ppb by a factor of two (i.e., 500, 250, 125, etc.). Each of these nine calibration samples was quadruply replicated.

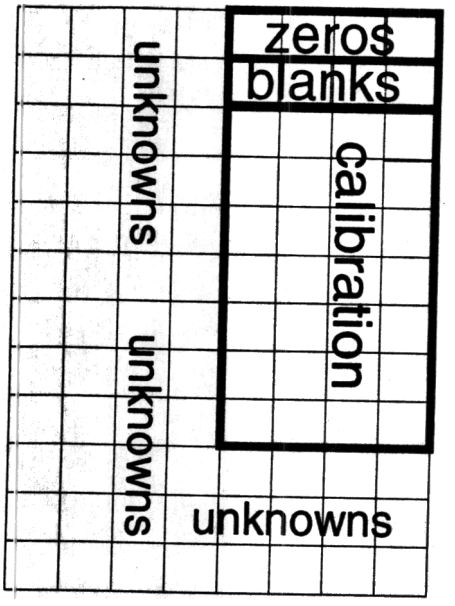

Figure 1. Typical Immunoassay Template.

1. STATISTICAL METHODS FOR CALIBRATION IN IMMUNOASSAY

From the calibration samples, a four-parameter log-logistic calibration can be fit by a variety of methods. The most widely used is apparently least squares or weighted least squares, although with most commercial implementations it is not easy to determine the method because it is rarely documented. We use maximum likelihood for its good statistical properties; least squares with a variance function chosen to match maximum likelihood will yield asymptotically identical results. Optical density readings for unknowns can then be transformed, using this curve, into estimated concentrations. The equation for this curve relating concentration x to optical density y is

$$y = f(x) = \frac{A - D}{1 + (x/C)^B} + D.$$

where A is the response (absorbance) at zero dose, B is the slope or curvature parameter, C is the concentration (ppb) giving 50% inhibition, and D is the response at infinite dose.

Suppose that true concentrations $x_1, x_2, \ldots, x_n$ are the concentrations chosen for calibration. Replications are separately listed here. Before we develop the maximum likelihood estimation procedure for the parameters of the calibration curve, it is necessary to address the question of the dependence of the variance of the response on the mean response.

There are two major approaches to this problem. In the first, the transform-both-sides model of Carroll and Ruppert (1988), there is assumed to exist a transformation $t(\ \ )$ such that the correct model is

$$t(y) = t\left(\frac{A - D}{1 + (x/C)^B} + D\right) + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$. This is the approach used in this article. The second approach is to assume the existence of a variance function $v(\ \ )$ such that

$$y \sim N(\mu, v(\mu)\sigma^2)$$

$$\mu = \frac{A - D}{1 + (x/C)^B} + D$$

(Raab 1981; Carroll and Ruppert 1982; Davidian and Carroll 1987; Davidian, Carroll, and Smith 1988).

In general, it may be necessary to determine either the transformation $t(\ )$ or the variance function $v(\ )$ from the data. For the purpose of this article, we will treat this as fixed because this issue is extensively dealt with in the previously cited literature. In future work, this could be incorporated also in the estimation and design problems that are addressed in this article. We will develop the remainder of the model using the transform-both-sides method, with the formulas being for a general monotonic transformation $t(\ )$. Note that the development easily follows also for the variance-function formulation.

For the ELISA protocol we are using for illustration in this article, we determined that the variance of the y values was stable on the log scale so that we would take $t(y) = \ln(y)$, but we develop the model first for a general monotonic transformation, which will usually consist of one of the Box–Cox (1964) class. This class includes both the logarithmic transformation (widely used in ELISA) and the square root transformation (widely used in radioimmunoassay). Thus the full specification of the model in general is

$$z = t(y) = t\left(\frac{A - D}{1 + (x/C)^B} + D\right) + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$.

The log-likelihood for this model is

$$L = -.5n\ln(2\pi) - n\ln(\sigma) - .5\sigma^{-2}\sum_{i=1}^{n} r_i^2$$

$$r_i = t(y_i) - t(f_i)$$

and

$$f_i = \frac{A - D}{1 + (x_i/C)^B} + D.$$

The gradient g of the log-likelihood of the ith data point with respect to the parameters (A, B, C, D) is given by

$$g_1 = \frac{\partial L}{\partial A} = r_i h_i t'(f_i)\sigma^{-2}$$

$$g_2 = \frac{\partial L}{\partial B} = -(A - D)r_i h_i^2 t'(f_i)\left(\frac{x_i}{C}\right)^B \ln\left(\frac{x_i}{C}\right)\sigma$$

$$g_3 = \frac{\partial L}{\partial C} = (A - D)r_i h_i^2 t'(f_i)\left(\frac{x_i}{C}\right)^B \left(\frac{B}{C}\right)\sigma^{-2}$$

$$g_4 = \frac{\partial L}{\partial D} = r_i(1 - h_i)t'(f_i)\sigma^{-}$$

where

$$h_i = \frac{}{1 + (x_i/C)^B}$$

If we write the gradient of log-likelihood of the ith point as

$$u_i = \tilde{u}_i r_i / \sigma^2,$$

where the components of $\tilde{u}_i$ are given by

$$\begin{aligned}
\tilde{u}_{i1} &= h_i t'(f_i) \\
\tilde{u}_{i2} &= -(A - D) h_i^2 t'(f_i) \left(\frac{x_i}{C}\right)^B \ln\left(\frac{x_i}{C}\right) \\
\tilde{u}_{i3} &= A - D) h_i^2 t'(f_i) \left(\frac{x_i}{C}\right)^B \left(\frac{B}{C}\right) \\
\tilde{u}_{i4} &= (1 - h_i) t'(f_i),
\end{aligned}$$

then the observed information is

$$E\left(\sum_i r_i^2 \sigma^{-4} \tilde{u}_i \tilde{u}_i'\right) = \sigma^{-2} \sum_i \tilde{u}_i \tilde{u}_i' = W.$$

Let $V = -W^{-1}$. Then this represents the asymptotic variance of the parameter vector (Cox and Hinkley 1974). We will use as an estimate of σ^2 the approximately unbiased version

$$s^2 = (n - 4)^{-1} \sum_{i=1}^{n} r_i^2$$

rather than the maximum likelihood estimator version with divisor n.

2. CALIBRATION ERRORS

Now suppose that $y_1, y_2, \ldots, y_k$ are replicate measurements of an unknown. Because of the specification of the model, the correct summary of these replicates is

$$\tilde{y} = t^{-1}(\bar{z}) = t^{-1}\left(k^{-1} \sum z_i\right),$$

where $z_i = t(y_i)$. The difference between the estimated value of $w = \ln(x)$ given by

$$\hat{w} = \ln(\hat{C}) + \hat{B}^{-1} \ln\left(\frac{\hat{A} - \tilde{y}}{\tilde{y} - \hat{D}}\right)$$

and the true value w has two parts. The first is due to the difference between $\tilde{y}$ and the value it would have taken on with an infinite number of replicates and the second is due to use of estimated coefficients.

One approach to separation of these effects is the delta method (Stuart and Ord 1987). We have a vector $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ with estimated variance V and, independently, an observation $\bar{z}$ which has variance σ^2/k. The calibration can be considered as a transformation

$$\begin{aligned}
\hat{w} \quad & G(\bar{z}, \hat{A}, \hat{B}, \hat{C}, \hat{D}) \\
& \ln(\hat{C}) + \hat{B}^{-1} \ln\left(\frac{\hat{A} - \tilde{y}}{\tilde{y} - \hat{D}}\right)
\end{aligned}$$

If $\tilde{g}$ is the gradient of this function with respect to the parameters $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$, then the delta-method variance of $\hat{w}$ is

$$\tilde{g}' V \tilde{g} + \left(\frac{\partial G}{\partial \bar{z}}\right)^2 \frac{\sigma^2}{k} \tag{2.1}$$

The components of this gradient are

$$\begin{aligned}
\tilde{g}_1 &= \frac{1}{B(A - \tilde{y})} \\
\tilde{g}_2 &= -\ln\left[\frac{A - \tilde{y}}{\tilde{y} - D}\right] / B^2 \\
\tilde{g}_3 &= (\\
\tilde{g}_4 &= \frac{1}{B(\tilde{y} - D)}
\end{aligned}$$

Note that this method separates the effect of the number of replicates from the effect of the calibration design. For a given number of calibration points to be used, one may optimize a criterion that depends only on $\tilde{g}' V \tilde{g}$, which is the only part of (2.1) that is affected by the choice of calibration values. Note also that, because this is a first-order approximation, the results from any transformation $t(y)$ apply equally to the variance function method with $v(\mu) = [t'(\mu)]^{-2}$.

Many previous analyses have tended to use only the second term of (2.1), which implicitly assumes that the calibration curve is estimated very precisely (e.g., Rodbard 1981) and which also assumes that the choice of calibration values is irrelevant (but see Bunch, Rocke, and Harrison 1990 and Rocke 1995). The term due to estimation of the curve should not be neglected, however, because it may form a significant proportion of the uncertainty. Figure 2 shows the asymptotic variance of the log concentration as a function of the concentration for a curve with true parameters $A = .501, B = .872, C = 105.8$, and $D = .151$ as given by Harrison et al. (1989). [Note that, to the order considered, the asymptotic variance of the log concentration is equal to the square coefficient of variation of the concentration, which is (the square of) a popular measure of accuracy among practitioners.] The calculations for this graph were made assuming the use of four dilutions, each replicated four times, and one observation each of the zero-dose and method blank; unknowns were replicated. The transformation $t(y) = \ln(y)$ was used. The dashed line is using Rodbard's assumption that the curve is known exactly, and the solid curve incorporates the curve-uncertainty factor explicitly. As can be seen, there is often a substantial difference between the commonly used expression for the coefficient of variation and the real accuracy.

Simulation-derived variances due to replication, curve-uncertainty, and total are also shown on this figure. The close agreement between the simulated and asymptotic values, especially in the central region where the precision of the method falls within a reasonable range, supports the use of the asymptotic expressions in deriving suitable calibration designs.

3. OPTIMAL DESIGN CRITERIA

There are many criteria for optimal design that have been proposed (Atkinson and Donev 1992). The majority of these focus on the variability of parameter estimates, or combinations thereof. Criteria that focus on predictions include V optimality, which minimizes the average variance of a prediction, and G optimality, which minimizes the maximum variance of a prediction.

For calibration problems, coefficient estimation is secondary; the primary purpose of the exercise is to produce precise estimated concentrations for one or more unknowns. In this case, it would make sense to focus on the precision of an estimated concentration (Buonaccorsi 1986). In fact, we will focus on the precision of the estimated log concentration for a number of reasons. First, many measurement methods have standard deviations that rise with the concentration of the analyte (Rocke and Lorenzato 1995). This makes the log scale a natural one for quantitative analytical chemistry. Second, it is the logarithm of the concentration that is presumed to have a logistic relationship with the response.

Expression (2.1) measures the uncertainty at any fixed value of $\ln(x)$, the log concentration of the unknown. Because this is, by definition, not known in advance, one cannot choose the calibration concentrations to minimize the variance. Conceptually, one could imagine minimizing the average variance over some prior distribution of x, but this is generally unavailable. An unweighted average is not feasible because the variance is unbounded. This leaves at least two possibilities. One is to use a weighted average variance with a user-chosen weight function (this includes restricting the range as a special case). The weights would probably be large in the region where precise estimation would be possible if the curve were known and small outside that area. This approach was pursued by Bunch et al. (1990). Because this suffers from the disadvantage of introducing a somewhat arbitrary weighting scheme, a different tack is taken here.

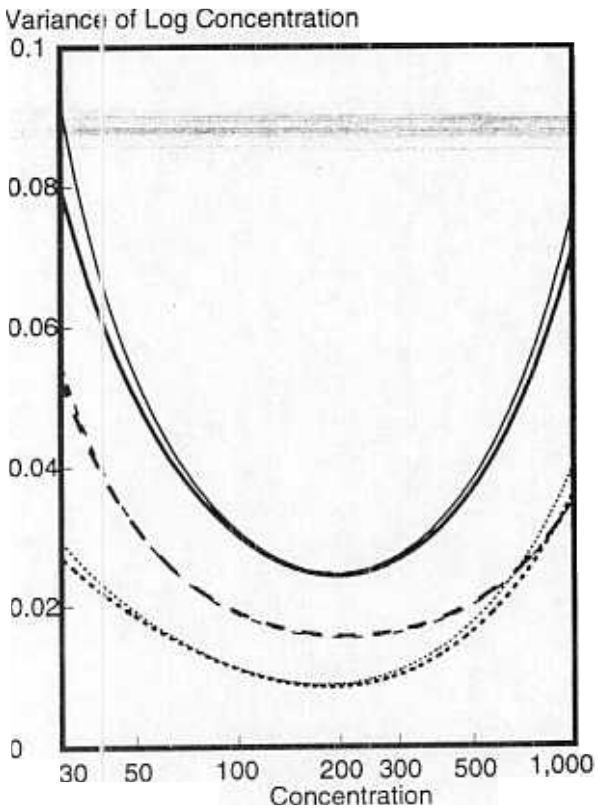

Figure 2. Variance of the Estimated Log Concentration as a Function of the Concentration. The dashed lines represent the variance due to measurement errors in the response for the unknown. The dotted lines represent the variance due to errors in the calibration values leading to an inaccurate calibration curve. The solid lines show the total variance. Heavy lines are from asymptotics; light lines are from a simulation.

Instead, we will maximize the *total precision* of $\hat{w} = \widehat{\ln(x)}$, where the precision is defined as the reciprocal of the variance and where total signifies an unweighted integral over the reals with respect to w. We will use the theoretical precision defined for each w using the value of y corresponding to the expected response $y = (A - D)(1 + C^{-B}e^{Bw})^{-1} + D$. Theorem 1 shows that this criterion is well defined.

This criterion should capture at least a rough measure of the goodness of the calibration. For the total precision to be high, there must be a wide region on which the assay is accurate. This region cannot be any larger than the region on which the assay would be accurate if the calibration curve were known exactly, but it can be much smaller if bad choices are made for the calibration values. This criterion, then, will encourage the assay to be as accurate as it can be, given the existing budget for calibration.

Theorem 1. The precision of an ELISA calibration at a log concentration $w = \ln(x)$, given by the reciprocal of (2.1), is integrable with respect to w over the real line. (Assumptions on parameter values are $0 < D < A$, $B > 0$, and $C > 0$.)

Proof. The first term in (2.1) involves the asymptotic variance V of the estimated parameters, which does not depend on x, and the gradient of the calibration function. The issue in the integrability of this function is the behavior as the concentration x goes to 0 or ∞. First, as x gets large, y tends to D. The difference $y - D$ tends to $x^{-B}(A - D)C^B$. For the gradient, then,

$$\tilde{g}_1 \quad B^{-1}(A - D)^{-}$$

$$\tilde{g}_2 \rightarrow \quad B^{-1}\ln(x/C) = -B^{-1}(w - \ln(C))$$

$$\tilde{g}_3 \rightarrow$$

and

$$\tilde{g}_4 \quad B \quad A - D)^{-1}C^{-B}x^B$$

$$= B \quad A - D)^{-1}C^{-B}e^{Bw}$$

so that the dominant term in $\tilde{g}'V\tilde{g}$ is of order e^{2Bw}.

The second term in the expression for the variance of

$$\frac{\partial G}{\partial \tilde{z}}\Big)^{z} \frac{\sigma^2}{k} = \left[-\frac{y(A - D)}{B(A - y)(y - D)}\right]^2 \frac{\sigma^2}{k}. \qquad (3.5)$$

When $x \quad \infty$, this tends to

$$\frac{x^{2B}D^2\sigma^2}{kB^2(A-D)^2C^{2B}} \tag{3.6}$$

so that this too is of order $x^{2B} = e^{2Bw}$ as $w \to \infty$.

Because the whole expression for the variance at w is of order e^{2Bw}, the precision, which is the reciprocal of the variance, is of order e^{-2Bw}, as $w \to \infty$, and consequently can be integrated as the upper limit of integration goes to ∞.

Similarly, when $x \to 0$, the variance of w tends to a term of order e^{-Bw} so that the dominant term in $\tilde{g}'V\tilde{g}$ is also of order e^{-2Bw}. Then the precision is of order e^{2Bw}, and consequently can be integrated as the lower limit of integration goes to $-\infty$.

4. AN EXAMPLE

In this section an application is presented to an assay for molinate (Harrison et al. 1989). As in any optimal design problem, the solution depends on the true parameters. Unlike the usual case in experimental design, however, the scientist will have had extensive experience with an assay before standardizing a protocol so that typical values for the parameters will be known, at least within some degree of certainty. From a set of 56 molinate assay calibrations, typical values of $A = .501, B = .872, C = 105.8$ (parts per billion), and $D = .151$ were chosen. These values, along with an average replication standard deviation of the log absorbance of $\sigma = .045$, are used for this illustration. Later, we show that variations from these figures within the range of occurrence in the 56 datasets did not importantly reduce the quality of the designs.

Although, in theory, the optimal design problem is to choose independently the location of all the calibration values, certain practical considerations restrict the possible solutions. The set of concentrations needs to be easily prepared by a technician; therefore, we consider only the case in which the concentrations are serial dilutions with the same dilution ratio. These may each be replicated several times. Added to this, we will have (replicated) zeros (corresponding to zero concentration), which are obtained by running pure water through the analysis. In addition, we will assume (replicated) method blanks, in which pure water is placed directly on the plate reader. For some assays, such as the molinate assay under consideration, this corresponds to infinite concentration. For assays where this is not the case, a very high concentration would be used. Note

Table 1. Design Factors for ELISA

n_w	Number of wells per plate
n_s	Number of samples (unknowns) per plate
	Number of replicates per sample
n_d	Number of dilutions used for calibration
k_m	number of replicates of each calibration dilution
k_0	Number of zero-concentration replicates
k_∞	Number of method blanks
n_{cal}	Total number of wells used for calibration = $k_0 + k_\infty + k_m n_d$
M	Maximum concentration used for calibration
m	Midpoint of the dilution series
α	Calibration dilution factor

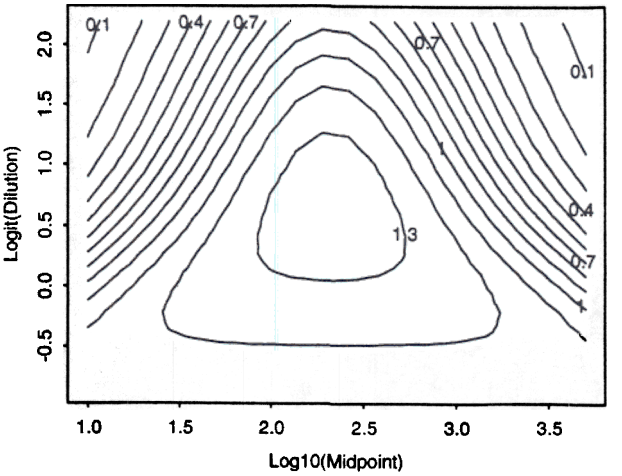

Figure 3. Contour Plot of the Total Precision for Differing Values of the Dilution Midpoint and Ratio. The horizontal axis is the base-10 logarithm of the dilution midpoint. The vertical axis is the logit of the dilution ratio.

that we are ignoring for the purposes of this article the dilution error. This could be incorporated into the analysis as in the work of Racine-Poon, Weihs, and Smith (1991).

For this formulation, this leaves several decision variables as given in Table 1. We will consider the optimization problem in three stages—varying the continuous variables with the integer variables held fixed, varying the integer variables within a fixed allocation of calibration wells while maintaining always the continuous variables at their optimum, and then varying the allocation of calibration wells.

First, for fixed values of the other values, consider varying m, the midpoint of the calibration series, and α, the dilution ratio. Figure 3 shows a typical contour plot for a case in which $n_w = 96, n_s = 15, k = 4, n_d = 10$, and $k_m = k_0 = k_\infty = 3$. The horizontal axis is the base-10 logarithm of the dilution midpoint, and the vertical axis is the logit of the dilution factor α. The response is the total precision. The peak of the "hill" is at a dilution midpoint of 210.4 ppb and a dilution factor of .666, which corresponds on the plot to 2.32 on the horizontal axis and .69 on the vertical axis. The response at the peak is 1.36.

From this plot we can see that the choice of dilution midpoint and ratio can be critical. Especially using too large a value of α (corresponding to placing the concentrations too close together) can give bad results. Using too small a value for α is less risky. The optimal midpoint of 210.4 may be somewhat surprising because it is larger than the 50% inhibition point $C = 105.8$, and large departures from this value cause considerable deterioration of the precision. This somewhat counterintuitive result may be explained by the assumption that the variance of y increases with y. This means that more points are needed above the midpoint C than below it. The bad results when the points are too closely spaced suggest that the idea of calibration only in the so-called linear region of the curve is not a very good idea, even though this is sometimes thought by practitioners to simplify the problem.

Suppose now that we fix the number of samples (unknowns) and the number of replicates of each, and consider

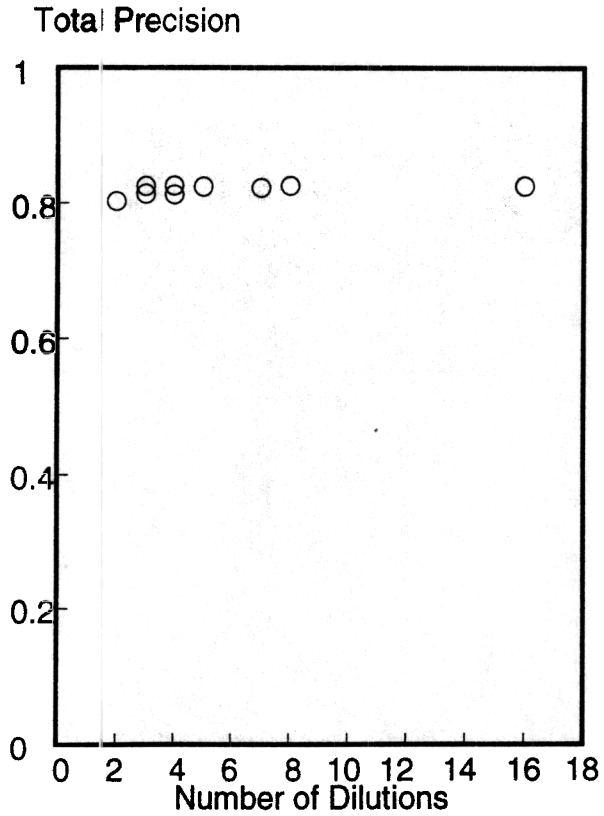

Figure 4. Total Precision of Several Choices of Number of Dilutions and Other Combinatorial Parameters. In each case, the total number of points devoted to calibration is 18; there are 36 unknown samples, each replicated twice. For each choice of combinatorial parameters within these constraints, optimal values for the dilution ration and midpoint were used.

varying the arrangement of points within the total of calibration points, while always maintaining the optimal choice of midpoint and dilution ratio. For example, with 36 calibration points we could have 2 each of zeros and blanks and 2 replicates of 16 dilutions, or 3 each of zeros and blanks and 6 replicates of 5 dilutions. Figure 4 shows a plot of the total precision for a case with 39 samples and $k = 2$ replicates. After optimizing the dilution midpoint and factor, the arrangement otherwise is without much importance. This allows these to be chosen for convenience without much loss of efficiency.

The final important factor is the number of samples (unknowns) to be analyzed and the number of replicates of each. Clearly, analyzing more samples with the same number of replicates decreases the precision of each because the number of calibration runs is reduced. The effect of increasing the number of replicates with the number of samples held fixed is less clear. A larger number of replicates reduces the error due to response uncertainty but, by reducing calibration data, increases the error due to curve uncertainty. Figure 5 illustrates this phenomenon and shows that the optimal number of replicates depends on the number of samples that one wishes to analyze. Note that the precision values given in this figure are for the optimal choice of the dilution midpoint and dilution factor. The choice of the number of samples to be analyzed depends on the user's preference for precision versus cost.

It may be of some interest to see in what way the results depend on knowing exact values for the parameters A, B, C, and D. To investigate this issue, we considered an additional 16 parameter sets, which consisted of the corners of the hypercube at the center $\pm$ two standard deviations for each parameter as given by Harrison et al. (1989). We then computed the true optimal design criterion at that corner, as well as the precision of the design chosen using the center point but evaluated at the corner. The logarithm of the ratio of these quantities provides an estimate of the loss due to lack of precise advance knowledge of the coefficient values. Naturally, there is always a loss, but it is usually quite small. The average value of the log ratio over the 16 corners was $-.01$, corresponding to about a 1% loss in precision. The worst of the 16 had a log ratio of $-.04$, and the second worst had a log ratio of $-.02$. Note that this assessment is a kind of worst case because most of the actual distribution of the parameters lies on an ellipsoid well inside the hypercube (because the estimates are correlated).

5. APPROXIMATE RULES FOR OPTIMAL DESIGNS

5.1 Equivariance Results

It turns out that the optimal transformation can be chosen by performing the calculations with a simpler model and then transforming that optimum in a simple way to obtain optimal values for the given model. To do this, we need an alternative formulation of the model that is obtained by making a simple linear data transformation. Instead of $w =$

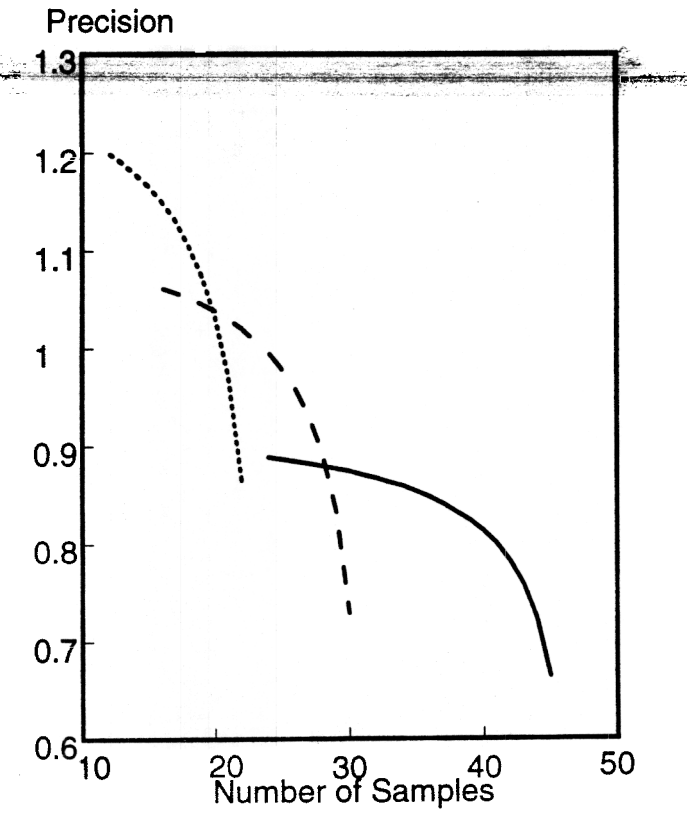

Figure 5. Effect of Varying the Number of Unknown Samples and the Number of Replicates of Each: ———, $k = 2$; — —, $k = 3$; - - - -, $k = 4$.

$\ln(x)$ and y, formulate the model in terms of $w^* = B(w - \ln(C)), x^* = \exp(w^*)$, and $y^* = (y - D)/(A - D)$. In this case, we can express the model as

$$t^*(y^*) = t^* \quad \Big) \qquad \text{(illegible)}$$

where $t^*(u) = t((A - D)u + D)$ and $\varepsilon \sim N(0, \sigma^2)$. This concentrates all of the parameter dependence in the transformation function, which enters into the optimal design calculation then only in terms of the influence on V, the covariance matrix of the estimated parameters.

If we restrict attention to the class of power transformations $t(u) = u^\lambda$ [or equivalently the Box–Cox (1964) class of transformations], we obtain as the model

$$D + (A \quad D)y^*)^\lambda = \left(D + (A \quad D)\,\frac{1}{1 \cdot \quad -^*}\right) \quad + \varepsilon$$

and

$$(1 + (A/D - 1)y^*)^\lambda$$

$$+ (A/D - 1)\,\frac{x}{1 + x^*}\Big) \qquad (5.1)$$

where $\varepsilon^* \sim N(0, \sigma^2/D^{2\lambda})$. This means that the optimal designs depend in a certain sense only on $\delta = A/D$. Thus, one can always calculate optimal designs for the case with altered parameters $A^* = \delta, B^* = C^* = D^* = 1$ in terms of $\{x_i^*\}$ and then transform to the original scale using the data transformation $x_i = C(x_i^*)^{1/B}$ (using the original parameters B and C); the pattern of the observations on the curve is equivariant to linear changes in w.

One special case is of interest. If the variance of the optical densities is constant, so that $t(u) = u$ (corresponding to $\lambda = 1$), then (5.1) becomes

$$\frac{}{+ x^*} + \varepsilon \qquad (5.2)$$

which corresponds to the case $A = B = C = 1, D = 0$. Thus, there is only one set of optimality computations to be done. We now use these observations to derive some approximations for hand calculation.

5.2 Empirical Analysis of the Optima

As observed previously, all of the analysis can be done in terms of

$$B| \qquad (5.3)$$

Table 2. Optimal and Predicted Midpoints

λ	A/D	$\bar{m}/\hat{m}$	$s_m/\hat{m}$
.0	3	.999	.004
.0	10	.999	.003
.0	20	.999	.003
.5	3	1.000	.003
.5	10	1.000	.004
.5	20	.998	.005
1.0	—	1.000	0

$$x^* = \exp(u$$

and

$$y^* = (y - D)/(A - D). \qquad (5.5)$$

This leaves a single parameter, $\delta = A/D$, to determine the optimum. Given δ, one obtains the optimal dilution ratio and midpoint, then transforms back to the original scale. Because $w = w^*/B + \ln(C)$ and $x = \exp(w)$, the midpoint m on the original scale is given in terms of the midpoint m^* on the standardized scale by

$$m = C(m^*)^{1/B} \qquad (5.6)$$

and the dilution ratio α on the original scale is given in terms of the dilution ratio α^* on the transformed scale by

$$\alpha = (\alpha^*)^{1/B}. \qquad (5.7)$$

This much is exact. As it turns out, quite accurate approximations can be given for the optimal values of m and α, regardless of the choice of the other parameters. We computed the optimal choice of m and α for all reasonable choices of combinatorial design factors in Table 1 with $D = 1, B = 1$, and $C = 1$ (the standardized values), and for $\delta = 3, 10$, and 20. (Reasonable means that no more than half the wells are devoted to calibration, that there are at least two calibration dilutions, and that the number of replicates of each calibration dilution, zero, and blank is at least 2.) We did this both for the value $\lambda = 0$, which is appropriate for ELISA, and for $\lambda = .5$, which is appropriate for RIA. In addition, calculations were done for the constant variance case $\lambda = 1$.

When $\lambda = 0$, it turns out that the optimal midpoint is remarkably close to $\hat{m} = \sqrt{A/D}$ in every case. For $\lambda = .5$, a more complex relationship emerges. Here the predicted optimal midpoint is given by $\hat{m} = .946 + .330\ln(A/D)$. When $\lambda = 1$, the optimal midpoint is always exactly at the center of symmetry C of the logistic curve. Table 2 shows the mean and standard deviation of the optimal midpoint divided by this prediction over all the cases for each dilution ratio.

Second, there is a strong linear relationship between the number of dilutions chosen and $\alpha/(1 - \alpha)$, where α is the dilution ratio. Table 3 shows the slope, intercept, and residual root mean squared error (RMSE) of the regression of $\alpha/(1 - \alpha)$ on n_d. In interpretation of the residual RMSE, note that the ratio $\alpha/(1 - \alpha)$ varies over a range from about .15 to about 5. If we approximate the equation by

$$- \alpha) = -.37 + .25n$$

and rearrange, we obtain

$$\alpha = \frac{n_d \quad 1.48}{n_d + 2.52}, \qquad (5.9)$$

which is conveniently approximated by

$$\alpha = \frac{n_d - 1.5}{n_d + 2.5}. \qquad (5.10)$$

In terms of the choice of combinatorial parameters, the results do not depend greatly on the distribution of design

Table 3. Optimal and Predicted Dilution Ratios

λ	A/D	Intercept	Slope	Residual RMSE
.0				.048
.0				.041
.0				.038
.5				.048
.5				.044
.5				.040
1.0				.038

points within a fixed n_{cal}, but in general it is better to have k_0 and k_∞ close in size and relatively small (2 or 3). It is also theoretically slightly better to use a few dilutions with many replicates, but this has significant practical problems in cases in which the curve fits imperfectly.

5.3 Some Rules of Thumb

1. From preliminary runs, identify approximate values for the parameters A, B, C, and D.
2. Choose a moderate number of dilutions n_d, perhaps 3–5, for convenience. Choose a number of replicates k_m for the calibration runs (2 to 3) and several replicates k for the unknowns (2 to 3).
3. Let the calibration midpoint be $\hat{m} = C(m^*)^{1/B}$, where $m^* = \sqrt{A/D}$ for ELISA and $m^* = .946 + .33\ln(A/D)$ for RIA.
4. Let the dilution ratio be

$$\alpha = \left(\frac{n_d - 1.5}{n_d + 2.5}\right)^{1/B} \tag{5.11}$$

6. CONCLUSIONS

In this article, a new method was developed of evaluating the effectiveness of an immunoassay protocol by its integrated precision (reciprocal variance). Three conclusions emerged. First, the choice of dilution midpoint and dilution factor can be extremely important; a poor choice can lead to greatly reduced precision. Second, given that the dilution midpoint and factor are optimally chosen, the particular arrangements of calibration points among zeros and blanks and the number of different dilutions are not important. For example, one may wish to use only a few dilutions to reduce technician effort. Third, there is an inevitable trade-off between the number of samples analyzed and the precision of each determination. The method given in this article allows this trade-off to be quantified and helps determine the correct number of replicates of each unknown.

This method can be used to determine optimality and judge convenient but nonoptimal arrangements for any form of immunoassay in which a log-logistic curve is appropriate. The error structure (in the sense of the transform-both-sides model or the variance-function model) is quite general, and the computations are not difficult given a good integration routine and a good numerical optimization routine. Convenient approximations are given that require only hand calculation and approximate the results of the exact analysis quite well.

ACKNOWLEDGMENTS

The research reported in this article was supported by National Science Foundation grants 93-01344, 94-06193, and 95-10511 and National Institute of Environmental and Health Sciences, National Institutes of Health grant P42 ES04699. The authors are grateful to Bruce Hammock and collaborators for the data used in this article.

[Received September 1992. Revised September 1996.]

REFERENCES

Atkinson, A. C., and Donev, A. N. (1992), *Optimum Experimental Designs*, Oxford, U.K.: Oxford University Press.

Bjercke, R. J., Cook, G., Rychlik, N., Gjika, H. B., Van Vanukis, H., and Langone, J. J. (1986), "Sterospecific Monoclonal Antibodies to Nicotine and Cotinine and Their Use in Enzyme-Linked Immunosorbent Assays," *Journal of Immunological Methods*, 90, 203–213.

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Ser. B, 26, 211–252.

Bunch, D. S., Rocke, D. M., and Harrison, R. O. (1990), "Statistical Design of ELISA Protocols," *Journal of Immunological Methods*, 132, 247–254.

Buonaccorsi, J. P. (1986), "Design Considerations for Calibration," *Technometrics*, 28, 149–155.

Carroll, R. J., and Ruppert, D. (1982), "Robust Estimation in Heteroscedastic Linear Models," *The Annals of Statistics*, 10, 429–441.

——— (1988), *Transformation and Weighting in Regression*, London: Chapman and Hall.

Chu, F. S., Fan, T. S. L., Zhang, G. S., Xu, Y. C., Faust, S., and McMahon, P. L. (1987), "Improved Enzyme-Linked Immunosorbent Assay for Aflatoxin B1 in Agricultural Commodities," *Journal of the Association of Official Analytical Chemists*, 70, 854–857.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Davidian, M., and Carroll, R. J. (1987), "Variance Function Estimation," *Journal of the American Statistical Association*, 82, 1079–1091.

Davidian, M., Carroll, R. J., and Smith, W. (1988), "Variance Functions and the Minimum Detectable Concentration in Assays," *Biometrika*, 75, 549–556.

Dixon-Holland, D. E., and Katz, S. E. (1988), "Competitive Direct Enzyme-Linked Immunosorbent Assay for Detection of Sulfamethazine Residues in Swine Urine and Muscle Tissue," *Journal of the Association of Official Analytical Chemists*, 71, 1137.

Hall, J. C., Deschamps, R. J. A., and Krieg, K. K. (1989), "Immunoassays for the Detection of 2,4-D and Picloram in River Water and Urine," *Journal of Agricultural Food Chemistry*, 37, 981–984.

Harrison, R. O., Braun, A. L., Gee, S. J., O'Brien, D. J., and Hammock, B. D. (1989), "Evaluation of an Enzyme-Linked Immunosorbent Assay (ELISA) for the Direct Analysis of Molinate (Ordram) in Rice Field Water," *Food and Agricultural Immunology*, 1, 37–51.

Jones, G., Wortberg, M., Kreissig, S. B., Bunch, D. S., Gee, S., Hammock, B., and Rocke, D. M. (1994), "Extension of the Four-Parameter Logistic Model for ELISA to Multianalyte Assay," *Journal of Immunological Methods*, 177, 1–7.

Jones, G., Wortberg, M., Kreissig, S. B., Hammock, B. D., and Rocke, D. M. (1995), "Sources of Experimental Variation in Calibration Curves for Enzyme-Linked Immunosorbent Assay," *Analytica Chimica Acta*, 313, 197–207.

Laurie, D., Manson, A. J., Rowell, F., and Seviour, J. (1989), "A Rapid Qualitative ELISA Test for the Specific Detection of Morphine in Serum or Urine," *Clinica Chimica Acta*, 183, 183–196.

Niewola, Z., Hayward, C., Symington, B. A., and Robson, R. T. (1985), "Quantitative Estimation of Paraquat by an Enzyme Linked Immunosorbent Assay Using a Monoclonal Antibody," *Clinica Chimica Acta*, 148, 149–156.

Raab, G. M. (1981), "Estimation of a Variance Function With Application to Immunoassay," *Applied Statistics*, 30, 32–40.

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

Rajkowski, K. M., Hanquez, C., Bouzoumou, A., and Cittanova, N. (1989), "A Competitive Microtitre Plate Enzyme Immunoassay for

Plasma Testosterone Using Polyclonal Antitestosterone Immunoglobulins," *Clinica Chimica Acta*, 183, 197–206.

Rocke, D. M. (1995), "Optimal Design of Quantitative Immunoassay Protocols," in *New Frontiers in Agricultural Immunoassay*, ed. D. Kurtz, Washington, DC: Association of Official Analytical Chemists, pp. 251–259.

Rocke, D. M., and Lorenzato, S. (1995), "A Two-Component Model for Measurement Error in Analytical Chemistry," *Technometrics*, 37, 176–184.

Rodbard, D. (1981), "Mathematics and Statistics of Ligand Assays: An Illustrated Guide," in *Ligand Assay: Analysis of International Developments on Isotopic and Nonisotopic Immunoassay*, eds. J. Langan and J. J. Clapp, New York: Masson.

Stuart, A., and Ord, J. K. (1987), *Kendall's Advanced Theory of Statistics I* (5th ed.), Oxford, U.K.: Oxford University Press.

Vanderlaan, M., Stanker, L. H., Watkins, B. E., Petrovic, P., and Gorbach, S. (1988), "Improvement and Application of an Immunoassay for Screening Environmental Samples for Dioxin Contamination," *Environmental Toxicology and Chemistry*, 7, 859–870.

Wortberg, M., Jones, G., Kreissig, S. B., Rocke, D. M., and Hammock, B. D. (1995), "An Approach to the Construction of an Immunoarray for Differentiating and Quantitating Cross Reacting Analytes," *Analytica Chimica Acta*, 319, 291–303.

Wortberg, M., Kreissig, S. B., Jones, G., Rocke, D. M., and Hammock, B. D. (1995), "An Immunoarray for the Simultaneous Determination of Multiple Triazine Herbicides," *Analytica Chimica Acta*, 304, 339–352.