DESIGN METHODS FOR SOME DOSE–RESPONSE MODELS

LEO W. G. STRIJBOSCH
Computer Applications Group, Department of Business and Economics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

RONALD J. M. M. DOES
Centre for Quantitative Methods, Nederlandse Philips Bedrijven B.V., HCM-711, P.O. Box 218, 5600 MD Eindhoven, The Netherlands

AND

WILLEM ALBERS
Department of Applied Mathematics, Twente University, P.O. Box 217, 7500 AE Enschede, The Netherlands

SUMMARY
A recently described design method for one-parameter biomedical models such as limiting or serial dilution assays is generalized to two-parameter models for which the dose–response relationship can be expressed as a linear regression model with parameters \(a\) (intercept) and \(b\) (slope). Design formulae are proposed for three different cases in which prior information about the unknown regression parameters \(a\) and \(b\) is available (a known, known and neither known, respectively). A suitable transformation of the two-parameter model enables the direct application of the one-parameter design method to the first two cases, while the third needs more advanced considerations. Two experimental designs, taken from the literature, are reproduced as closely as possible using the methods described, thus showing under which circumstances these designs are suitable.

1. INTRODUCTION
Good experimental design is crucial for effective statistical analysis. This paper deals with the experimental design for quantal two-parameter biomedical models, for which the dose–response relationship can be expressed as a linear regression model \(y = a + bx\), where \(x\) is a function of the dose \(z\), and \(y\) is a function of the probability \(P(x)\) of a positive response. A well-known example of such a model is the logit regression model, resulting from the logistic function

\[ P(x) = \frac{1}{1 + \exp(- (a + bx))}, \]

where \(x = \log z\) and \(y\) is the logit of \(P(x)\). This model can be used to study the toxicity of a drug, or, for example, the relationship between dose of an insecticide and mortality of a target insect species.

As an example of the application of the logistic model, consider the study of the relationship between the dose of a drug and the mortality of a target group of mice. A dose \(x\) (actually the logarithm of the true dose) of a drug is given to each of \(n\) mice. To formulate the model it is useful
to assume that each mouse possesses its own tolerance against a particular drug and dies when the dose level exceeds this tolerance. It is not unusual to suppose that the tolerance $T_j$ of the $j$th mouse is a random observation drawn from a logistic distribution which is identical for all mice. Then $P(T_j < x) = 1/[1 + \exp(-\alpha + \beta x)]$.

The one-parameter biomedical models ($\alpha = 0$), such as limiting or serial dilution assays (LDAs, SDAs), where $P(x) = 1 - \exp(-\beta x), x = z$ and $y = -\log(1 - P(x))$, form a special case. These models are used widely in many areas, including public hygiene, bacteriology, virology and immunology. The aim of a dilution assay is to estimate the frequency of one particular cell type among other similar cells (LDA), or to estimate the density of particular cells or micro-organisms per unit volume of solution (SDA). Experimental design for both dilution assays requires choice of the number of doses, the number of replicate cultures at each dose and the (mean) number of cells per replicate at each dose. Dose groups for which $\beta x$ is much greater than 1 will most surely respond positively (the value of $P$ is near 1), dose groups for which $\beta x$ is much less than 1 are likely to respond negatively (the value of $P$ is near 0). It is clear that efficient determination of the unknown parameter in dilution analysis requires a setup whereby $\beta x \sim 1$ for at least some dose groups (the value of $P$ is intermediate). Since some dose groups should be informative for values of $\beta$ in a certain range, information about $\beta$ must be included in the design. Recently, Strijbosch et al. have developed a simple design method for the one-parameter model. A summary of this method is given in Section 2. An experimental design for the two-parameter model is more complicated, since prior knowledge of two different parameters has to be incorporated. Several authors have suggested appropriate designs, but these methods do not provide formulae for design characteristics, based on prior estimates of both parameters. In this paper the design method for the one-parameter models is generalized to two parameters. Three special cases are dealt with and each is illustrated with an example (Section 3). The examples clarify the circumstances under which a design, taken from the literature, is suitable, since this design is reproduced as closely as possible by carefully choosing the corresponding design parameters.

2. A DESIGN METHOD FOR THE ONE-PARAMETER BIOMEDICAL MODEL $y = \beta x$

Strijbosch et al. propose design formulae for dilution assays which are based on the experimenter's criteria. They can be viewed as more elaborate than the suggestions of Finney (pp. 428, 435 and 437) and yield a setup with known and suitable properties. The following criteria have to be chosen by the experimenter:

(i) lower ($\beta_1$) and upper ($\beta_2$) values for the unknown parameter $\beta$ ($0 < \beta_1 \leq \beta \leq \beta_2$);
(ii) lower ($P_1$) and upper ($P_2$) values for a fraction of positive replicates at a dose considered sufficiently informative ($0 < P_1 < P_2 < 1$) (in the sense of the Fisher information for individual doses);
(iii) an integral number ($d$) of doses which will yield sufficiently informative data as specified under (ii), ($d \geq 1$).

Let $y_1 = -\log(1 - P_1)$, and $y_2 = -\log(1 - P_2)$, so that $0 < y_1 < y_2$ (log denotes the natural logarithm). On the basis of the criteria chosen, the number of doses, $m$, and the number of cells, $x_j$, per replicate at dose $j, j = 1, \ldots, m$, can be calculated as follows:

\begin{align*}
    m &= d + \lceil \log(\beta_2/\beta_1)/\log(c) \rceil, \\
    x_1 &= \sqrt[3]{y_1 y_2 c^{-m}/(\beta_1 \beta_2)}, \\
    x_j &= x_1 c^{j-1}, \quad j = 2, \ldots, m,
\end{align*}

(1) (2) (3)
DESIGN METHODS FOR SOME DOSE–RESPONSE MODELS

Figure 1. Construction of a design $x_1, \ldots, x_m$ assuming that for $\beta_1 \leq \beta \leq \beta_2$ exactly 3 log-doses $x$ produce responses $y = \beta x$ in the interval from $y_1$ to $y_2$.

where

$$c = \frac{y_2}{y_1}^{1/d},$$

and $\lfloor \cdot \rfloor$ denotes the ‘integral part of’.

These design formulae can be explained heuristically with the assistance of Figure 1 which shows the lines $y = y_1, y = y_2, y = \beta_1 x$ and $y = \beta_2 x$. Suppose that $d = 3$, and $x_1, x_2$ and $x_3$ are given values satisfying the condition $y_1 < \beta_2 x_1 < \beta_2 x_2 < \beta_2 x_3 \leq y_2$. The value of $x_4$ can then be found by letting $\beta$ decrease from $\beta_2$ to $\beta_0$ which is determined by the point $(x_1, y_1)$. For $\beta_0 < \beta \leq \beta_2$, it follows that $y_1 < \beta x_1 < \beta x_2 < \beta x_3 \leq y_2$, while $\beta x_1 \leq y_1$ for $\beta \leq \beta_0$. Thus a new value $x_4$ has to be added to the series to satisfy the design requirements (ii) and (iii). In general it follows that $x_{j+d} = x_j y_2/y_1$, justifying the geometrical series (3) and (4). Formula (2) follows from a condition of symmetry: $x_1 x_m = y_1 y_2/(\beta_1 \beta_2)$. As $x_1 x_m = x_1^2 c^{m-1}$, (2) is obtained. Finally, (1) is explained by observing that typically $x_1 \approx y_1/\beta_2$, while $x_{m-d+1} \approx y_1/\beta_1$, and hence $c^{m-d} = x_{m-d+1}/x_1 \approx \beta_2/\beta_1$.

The resulting assay design has the property that for each $\beta$ such that $\beta_1 \leq \beta \leq \beta_2$ exactly $d$ values of $x_1, \ldots, x_m$ satisfy the following condition:

$$y_1 < \beta x_j \leq y_2, \quad j = 1, \ldots, m.$$  

In this design method the interval from $\beta_1$ to $\beta_2$ can be interpreted as the interval corresponding with a uniform prior distribution for $\beta$. The rationale of the method is that the statistical properties of the maximum likelihood estimator of $\beta$ are as far as possible independent of the true value of $\beta$. These statistical properties can be influenced further by the choices for $P_1, P_2$ and $d$. Attempting to minimize the maximum asymptotic value of the coefficient of variation in the interval $\beta_1 \leq \beta \leq \beta_2$, with an equal number of replicates in each dose group, leads to a heuristic formula for the determination of $P_1$ and $P_2$ when $\beta_2/\beta_1$, $d$ and $m$ are given.\(^{10,11}\) This coefficient of variation is defined by $1/(\beta \sqrt{[I(\beta)]})$, where $I(\beta) = \sum_{j=1}^{n_j}(n_j x_j^2/(\exp(\beta x_j) - 1))$ is the Fisher information. The number of replicates $n_j$, $j = 1, \ldots, m$, are preferably equal and can be determined on the basis of a user-supplied coefficient of variation.\(^{10}\)

This section is concluded with a numerical example based on a design taken from the literature. We investigate parameter values for our design method which will approximately reproduce this
3. DESIGN METHODS FOR THE TWO-PARAMETER MODELS \( y = \alpha + \beta x \)

After a linearizing transformation many dose–response relationships can be expressed as a linear regression model with two unknown parameters: \( y = \alpha + \beta x \), where \( x = \log z \), \( z \) is the dose, and 
\[
y = f(P(x)),
\]
where \( P(x) \) is the probability of a positive response.\(^1\)\(^2\) For example, the logistic model
\[
P(x) = \frac{1}{1 + \exp[-(\alpha + \beta x)]}
\]
can be written as
\[
y = \log\left\{ P(x)/(1 - P(x)) \right\} = \alpha + \beta x.
\]

It is important to establish that the model changes when \( z \) is expressed in other units. Therefore, 
the use of a model \( y = \alpha + \beta x \) presupposes a specific unit for \( z \). By analogy with the one-

DESIGN METHODS FOR SOME DOSE-RESPONSE MODELS

1357 parameter case (formulae (1) to (4)), design methods for the two-parameter model have been
developed when a prior point estimate for $\alpha$ and an interval estimate for $\beta$ are available (Section
3.1), or when a point estimate for $\beta$ and an interval estimate for $\alpha$ are available (Section 3.2), or
when interval estimates for both $\alpha$ and $\beta$ are available (Section 3.3). The choice of $P_1$ and $P_2$
(or equivalently $y_1$ and $y_2$ based on equation (7)) in the two-parameter cases is more complicated
than in the one-parameter case and is dependent on the goal of the analysis. When applying the
Spearman–Kärber method to estimate an ED50 (the median effective dose which on average
produces a response in 50 per cent of subjects), for example, one of the requirements is that the
proportion of positively responding subjects varies from 0 to 1. This leads to a choice for $P_1$
near 0 and for $P_2$ near 1. In estimating the ED$y$ where $y$ is close to 0 or 1, it is considered important12
to locate the doses around the ED$y$. This can easily be achieved by choosing $P_1$ and $P_2$ such that
$P_1 < y < P_2$ and applying the most appropriate design method of Section 3.

3.1. A design method for the case $\alpha = \alpha_0$ and $\beta_1 \leq \beta \leq \beta_2$

Suppose that $\alpha_0$ is a point estimate for $\alpha$ and that an interval estimate for $\beta$ is from $\beta_1$ to $\beta_2$. Just
as in Section 2, assume that $0 < \beta_1 \leq \beta_2$. We shall concentrate on the more interesting, and at the
same time more complicated, case where $\alpha_0 < y_1 < y_2$. Then the design formulae (1) to (4) can be
applied directly:

$$m = d + \lfloor \log(\beta_2/\beta_1)/\log(c) \rfloor,$$

$$x_1 = \sqrt{[(y_1 - \alpha_0)(y_2 - \alpha_0)c^{1-m}/(\beta_1 \beta_2)]},$$

$$x_j = x_1 c^{j-1}, \quad j = 2, \ldots, m,$$

where

$$c = \begin{cases} y_2 - \alpha_0, \\ y_1 - \alpha_0 \end{cases}^{1/d}.$$

(11)

The resulting design has the property that for each $\beta$ in the interval from $\beta_1$ to $\beta_2$ exactly $d$ values
of $x_1, \ldots, x_m$ satisfy the condition $y_1 < \alpha_0 + \beta x_j < y_2$. Note that for the doses in this design we
may write:

$$z_j = z_{j-1}^{1/d},$$

(12)

which means that the original doses should be chosen equidistantly on a log-log scale. This is
remarkable, since in most designs in the literature, doses equidistant on the log-scale
are recommended or used in Monte Carlo experiments. Finally, in the remaining case
$y_1 \leq \alpha_0 \leq y_2$, it is easy to see that $m = d$, and $x_j, j = 1, \ldots, d$ merely need to satisfy
$y_1 < \alpha_0 + \beta_2 x_j \leq y_2, j = 1, \ldots, d$.

Again we conclude with a numerical example based on a design taken from the literature and
investigate which parameter choices in our design method will approximately reproduce this
design. Berkson\textsuperscript{15} presents in his Table I an experiment with doses of 40, 60, 80, 100, 120, 140, 160,
180, 200, 250 and 300 units. The recorded experiment yields, assuming the logistic model (6),
estimates $\hat{\delta} = -10.33$ and $\hat{\beta} = 5.66$. Since he uses the transformation $x = \log_{10}(z)$, we must take
here $\beta = 5.66/\log(10) = 2.46$. Assuming $\alpha_0 = -10.3$ and $P_1 = 1 - P_2$, we find that each of the
following combinations, applying (8) to (11), produce a design with 11 doses and approximately
40 and 300 units for the first and last dose, respectively: $d = 1, \beta_1 = 1.80, \beta_2 = 2.79$ and $P_1 = 0.444$; $d = 2, \beta_1 = 1.84, \beta_2 = 2.73$ and $P_1 = 0.390$; $d = 3, \beta_1 = 1.88, \beta_2 = 2.67$ and $P_1 = 0.338$; $d = 4, \beta_1 = 1.92, \beta_2 = 2.61$ and $P_1 = 0.290$. It is clear, however, that the spacing
can not be reproduced exactly.
Figure 3. Illustration of the conditions for the construction of the design method in Section 3.3. The values $\hat{y}$, $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$, and $\hat{x}_4$ refer to the choice for the weights of translation and rotation in the construction of the design formulae (22) to (24).

3.2. A design method for the case $\alpha_1 \leq \alpha \leq \alpha_2$ and $\beta = \beta_0$

Suppose that an interval estimate for $\alpha$ is from $\alpha_1$ to $\alpha_2$ and that $\beta_0$ is a point estimate for $\beta$. Then the following design formulae can be deduced from (1) to (4):

$$m = d + [(\alpha_2 - \alpha_1)/e],$$

$$x_1 = (y_1 + y_2 - \alpha_1 - \alpha_2 + e(1 - m))/(2\beta_0),$$

$$x_j = x_1 + (j - 1)e/\beta_0, \quad j = 2, \ldots, m,$$

where

$$e = (y_2 - y_1)/d.$$  (16)

This follows from the transformations $y^* = e^y$, $\beta^* = e^x$, $x^* = e^{\beta_0 x}$, and applying (1) to (4) to the model $y^* = \beta^* x^*$. The property of the resulting design is that for each $\alpha$ in the interval from $\alpha_1$ to $\alpha_2$ exactly $d$ values of $x_1, \ldots, x_m$ satisfy the condition $y_1 < \alpha + \beta_0 x_j \leq y_2$. Note that in this situation doses are equidistant on the log-scale.

We will now consider the same example as in Section 3.1 using a fixed value for $\beta$ and an interval estimate for $\alpha$. Assuming $\beta_0 = 2.45$ and $P_1 = 1 - P_2$, we find that each of the following combinations, applying (13) to (16), produces a design with 11 doses starting with 40 and ending with 300 units: $d = 1$, $\alpha_1 = -14.0$, $\alpha_2 = -9.04$ and $P_1 = 0.439$; $d = 2$, $\alpha_1 = -13.73$, $\alpha_2 = -9.28$ and $P_1 = 0.379$; $d = 3$, $\alpha_1 = -13.5$, $\alpha_2 = -9.53$ and $P_1 = 0.323$; $d = 4$, $\alpha_1 = -13.23$, $\alpha_2 = -9.76$ and $P_1 = 0.271$.

3.3. A design method for the case $\alpha_1 \leq \alpha \leq \alpha_2$ and $\beta_1 \leq \beta \leq \beta_2$

Figure 3 illustrates construction of the design method in this section. Suppose that interval estimates for $\alpha$ and $\beta$ are from $\alpha_1$ to $\alpha_2$ and from $\beta_1$ to $\beta_2$, respectively, and that $\alpha_1 < \alpha_2 < y_1$, $0 < \beta_1 < \beta_2$ and $\beta_1/\beta_2 \leq (y_2 - \alpha_2)/(y_2 - \alpha_1)$. This configuration corresponds to the situation which is of primary interest. By analogy with Section 3.1 we shall concentrate on this case. The last condition means that the slope ($\beta_0$) of the line through the points $(0, \alpha_2)$ and
((y_2 - \alpha_1)/\beta_2, y_2) is greater than or equal to \beta_1. Then, combining the results of the two preceding sections, a method can be invented for this case. In the first place, the approximate number of doses is determined. Applying (8) to (11) with \(x_0 = \alpha_1\) yields \(m_1\) and \(c_1\). To cover all remaining combinations of \(x\) and \(\beta\), it is sufficient to extend the design applying (8) to (11) once more to \(y = \alpha_2 + \beta x\), with \(\beta^0 \leq \beta \leq \beta_2\), where \(\beta^0 = \beta_2(y_2 - \alpha_2)/(y_2 - \alpha_1)\). This yields \(m_2\) and \(c_2\). Since the \(m_1\) and \(m_2\) doses overlap an interval corresponding with \(d\) doses, \(m = m_1 + m_2 - d\) doses are needed for a design with the property that at least \(d\) values satisfy \(y_1 < x + \beta x_j \leq y_2\) (for any \(\alpha_1 \leq x \leq \alpha_2\) and \(\beta_1 \leq \beta \leq \beta_2\)). It is not possible to construct a design with the property that exactly \(d\) values satisfy this condition. It follows that

\[ m = d + \lfloor \log(\beta_2/\beta_1)/\log(c_1) \rfloor + \lfloor \log((y_2 - \alpha_1)/(y_2 - \alpha_2))/\log(c_2) \rfloor, \]

where

\[ c_1 = \left\lfloor \left(\frac{y_2 - \alpha_1}{y_1 - \alpha_1}\right)^{1/d} \right\rfloor, \]

and

\[ c_2 = \left\lfloor \left(\frac{y_2 - \alpha_2}{y_1 - \alpha_2}\right)^{1/d} \right\rfloor. \]

Obvious choices for \(x_1\) and \(x_m\) are

\[ x_1 = (y_2 - \alpha_2)c_1^{-d}/\beta_2, \]

\[ x_m = (y_1 - \alpha_1)c_1^{1/d}/\beta_1. \]

When it is important that for all considered parameter values \(x\) and \(\beta\), the proportions of positive responders in an experiment vary from \(P_1\) to \(P_2\), better choices for \(x_1\) and \(x_m\) are \(x_1 = (y_1 - \alpha_2)/\beta_2\) and \(x_m = (y_2 - \alpha_1)/\beta_1\). This is the case, for example when the Spearman–Kärber method is applied to estimate an ED50.

Now, it follows from the results of the Sections 3.1 and 3.2, that the choice of \(x_j\), \(j = 2, \ldots, m - 1\), should depend on the ratio of the lengths of the intervals from \(\alpha_1\) to \(\alpha_2\) and from \(\beta_1\) to \(\beta_2\). When the first is large and the second is small, equidistant values for \(x_j\), \(j = 1, \ldots, m\), are advisable, whereas in the opposite case a geometrical series is preferable. Combining these ideas, the following design formulae are proposed:

\[ x_j = \{T[x_1 + (x_m - x_1)(j - 1)/(m - 1)] + WRx_1(x_m/x_1)^{(j-1)(m-1)}/(T + WR), \]

\[ j = 2, \ldots, m - 1, \]

where

\[ T = (\alpha_2 - \alpha_1)/(y_1 + y_2 - \alpha_1 - \alpha_2), \]

and

\[ R = (\beta_2 - \beta_1)/(\beta_1 + \beta_2). \]

Motivation for the choice of the weights \(T\) (for translation) and \(R\) (for rotation) is as follows (see Figure 3). The design formulae of Section 3.1 (\(\alpha_1 = \alpha_2 = \alpha_0\)) could be found by studying the effect of rotation of the line \(y = \alpha_0 + \beta x\) around the point \((0, \alpha_0)\), on the log-doses. Accordingly, the design formulae of Section 3.2 (\(\beta_1 = \beta_2 = \beta_0\)) could be found by studying the effect of translation of the line \(y = \alpha + \beta_0 x\) from \(y = \alpha_1 + \beta_0 x\) to \(y = \alpha_2 + \beta_0 x\) on the log-doses. These effects must be quantified somehow to determine the weights \(R\) and \(T\). Since the effect of rotation is different for \(y_1\) and \(y_2\), all effects are considered for \(\bar{y} = (y_1 + y_2)/2\) only. Furthermore, the
Figure 4. Graph of the functions $f_j(\alpha, \beta) = 1/[1 - \exp\{- (\alpha + \beta x_j)\}]$ at each extreme of either parameter interval. The $x_j$'s, $j = 1, \ldots, m$, follow from formulae (17) to (24) with design parameters $\alpha_1 = -10.7, \alpha_2 = -9.9, \beta_1 = 2.03, \beta_2 = 2.65, P_1 = 0.38, P_2 = 0.80$ and $d = 4$ and $W = 1$ resulting in $m = 11$. 
Figure 4. (Continued)
effects of rotation for \( \alpha = \alpha_1 \) and \( \alpha = \alpha_2 \), respectively, are accumulated, as are the effects of translation for \( \beta = \beta_1 \) and \( \beta = \beta_2 \), respectively. Thus, the effect of rotation can be expressed as \((\bar{x}_3 - \bar{x}_1) + (\bar{x}_4 - \bar{x}_2)\), whereas the effect of translation is \((\bar{x}_4 - \bar{x}_3) + (\bar{x}_2 - \bar{x}_1)\).

Since \( \bar{x}_1 = (\bar{y} - \alpha_2)/\beta_2 \), \( \bar{x}_2 = (\bar{y} - \alpha_1)/\beta_2 \), \( \bar{x}_3 = (\bar{y} - \alpha_2)/\beta_1 \) and \( \bar{x}_4 = (\bar{y} - \alpha_1)/\beta_1 \), the quotient of \( R \) and \( T \) is:

\[
\frac{R}{T} = \frac{(\beta_2 - \beta_1)(y_1 + y_2 - \alpha_1 - \alpha_2)}{(\alpha_2 - \alpha_1)(\beta_1 + \beta_2)}.
\]

Hence, (23) and (24) represent possible choices for \( R \) and \( T \). \( W \) is a weighting factor of which the optimal value is dependent on the purpose of the analysis. Under certain restricted conditions, Ashton\(^1\) (p. 30) gives the asymptotic variance of an ED50. Limited Monte Carlo simulations comparing the ratio of the square root of this variance and ED50 indicate that \( W \) should be larger than 1. Further investigation is needed on this subject.

We will now consider the example used in the two previous sections using interval estimates for both \( \alpha \) and \( \beta \). Note that it is much more difficult now to find combinations of parameters reproducing the design. We present only one combination. Applying (17) to (24) with \( \alpha_1 = -10.7 \), \( \alpha_2 = -9.9 \), \( \beta_1 = 2.03 \), \( \beta_2 = 2.65 \), \( d = 4 \), \( W = 1 \), \( P_1 = 0.38 \) and \( P_2 = 0.80 \), results in the doses 41, 49, 58, 70, 84, 102, 125, 154, 191, 239 and 301 units. Figures 4(a)-(d) present a graph of this design at each extreme of both parameter intervals.

If both conditions \( \alpha_1 < \alpha_2 < y_1 \) and \( \beta_1/\beta_2 \leq (y_2 - \alpha_2)/(y_2 - \alpha_1) \) are not satisfied, alternatives for formulae (17) to (24) are possible applying the same concepts.

**ACKNOWLEDGEMENT**

The authors thank Professor D. J. Finney for encouragement and stimulating discussions.

**REFERENCES**

