Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

AN ADAPTIVE DESIGN FOR MAXIMIZATION OF A CONTINGENT BINARY RESPONSE

BY ZHENGQING LI, STEPHEN D. DURHAM* AND NANCY FLOURNOY

The University of Wisconsin, The University of South Carolina and The American University

Abstract

Treatment at dose x may be toxic or non-toxic, and if it is non-toxic it may or may not result in cure. We wish to maximize the probability of a cure. A class of adaptive sequential designs for a family of parametric models is proposed. The designs are constructed so that the information in previous trials is used to determine the dose level for the next trial. Criteria for the existence of a maximum for the probability of a cure are given. After calcuating the maximum likelihood estimates for the model parameters, the next dose level is chosen to be the level for which the estimated probability of a cure is maximized. Necessary and sufficient conditions for the existence of maximum likelihood estimators are given. The sequential dose levels turn out to be consistent and asymptotically optimal under certain conditions.

Received September 1992; revised August 1994.

*Work supported in part by NSF Grant EPSCoR#OSR-9108772 and by the Westinghouse Savannah River Company, through the South Carolina University Research and Education Foundation.

AMS 1991 subject classification. Primary 62L15; secondary 62L05.

Key words and phrases. Dose-response, opposing hazard functions, sequential maximum likelihood, toxicity study, survival.

1. Introduction. Treatment at dose x may be toxic or it may be non-toxic. If it is non-toxic, it may or may not result in a cure. This situation arises in many fields of research. In cancer therapy, it is current practice to select the dosages of a new drug by first considering its toxicity. If treatment is non-toxic, it may or may not result in a cure. In testing the compressive strength of engineered fibers, a fiber may or may not fail after it is mounted to a support tab and stressed under tension to a predetermined level. If it holds under this initial tension, the recoil test [see Hayes, Edie and Durham (1992)] is initiated by cutting the fiber at the midpoint. Since the fiber is cut in half to initiate the test, each mounted fiber produces two observations. If the initial stress level is sufficiently high (but not high enough to lead to a failure), each half of the fiber will fail due to compressive stresses generated as the stored strain energy is recovered. At lower initial stress levels the fiber halves may both survive the test or only one half may survive and the other fail under recoil after separation. The main goal is to find the dose that maximizes the probability of a cure in the medical study and to find the stress level that maximizes the probability of a recoil failure without tensile failure.

The toxicities considered in this paper are assumed to be fatal, that is, if toxicity is observed, then cure or non-cure cannot be observed. Therefore, there are three kinds of responses possible at each treatment level: toxicity, non-toxicity and non-cure, and non-toxicity and cure. Suppose the set of feasible dose levels is D_l. The objective is, by the choice of x from D_l, to maximize

$$\begin{aligned} &P\{\text{non-toxicity and cure} \\ &\quad = P\{\text{non-toxicity }\}P\{\text{cure}|\text{non-toxicity}\}. \end{aligned}$$

It is reasonable to assume that the probability of toxicity and the probability of a cure given non-toxicity can be modeled as distribution functions of dose levels x though they may possibly be defective. Consequently, we assume the following parametric models for the response curves:

$$\begin{aligned} F(x) &= H(x;\Theta) = P\{\text{toxicity is fatal at } x\}; \\ G(x) &= K(x;\Psi) = P\{\text{cure} \mid \text{toxicity is non-fatal at } x\}, \end{aligned}$$

where Θ and Ψ are parameter vectors. Then

$$P\{\text{there is a cure at } x\} = (1 - F(x))G(x).$$

Our goal is to find values of x in D_l that maximize $(1-F(x))G(x)$ over D_l. Hereafter, when we mention maximum values, we mean the global maximum over D_l. If $(1-F(x))G(x)$ has the same maximum value over D_l at more than one dose level, we choose the smallest dose level as the optimal design point, which is called x^*. The reason we choose the smallest as the optimal design point is that we generally prefer the response of non-toxicity and non-cure to that of toxicity. Theoretically, we expect a small dose level to result in fewer toxicities than a relatively large dose level when they have the same probability of cure.

For the design we consider, we assume that the dose level x is bounded, i.e., $|x| \leq D$, where D is some finite dose level. D is determined by previous knowledge and from other experiments. The dose level x could be negative since we may take transformations of actual dosages (for example, the log-transformation). Hereafter, we always assume that x is the dose level after an appropriate transformation.

In this paper, we present two adaptive sequential designs that eventually maximize the probability of a cure within the domain of dose levels prescribed. Sequential designs are constructed such that all of the information in the previous trials is used to determine the dose level for the next trial. That is, estimated response curves are constructed using current maximum likelihood estimators based on all of the data available in the previous trials, and the next dose is determined from the current estimated response curves through their connection with the parameters. We propose two different procedures. One is applicable to a design space of continuous dose levels (for example, $D_l = (-\infty, +\infty)$ or $[a, b]$), and one is applicable to a space of discrete dose levels $D_l = \{d_1, d_2, ..., d_m\}$. In order to ensure statistical consistency for the first design, a grouped design is used in which the sequence of group sizes is required to diverge to infinity (although the group sizes are not necessarily nondecreasing) as the experiments continue. Detailed design procedures are given in Sections 2.2. In Section 2.1, we discuss the existence of x^* and the dependence of x^* on F and G. Necessary and sufficient conditions for the existence and uniqueness of maximum likelihood estimators are given in Section 2.3. Under certain conditions, the consistency of the maximum likelihood estimators and the asymptotic optimality is proved for both procedures in Section 3.

2. A class of adaptive sequential designs based on the estimated response curves. In order to find an optimal design point,

x^*, that maximizes the probability of a cure over D_l, we assume the following parametric models:

$$
\begin{aligned}
& F(x) = H(x;\Theta) = P\{\text{toxicity at } x\ \}; \\
& G(x) = K(x;\Psi) = P\{\text{cure} \mid \text{non-toxicity at } x\}; \\
& \lim_{x\to-\infty} H(x;\Theta) = 0; \quad \lim_{x\to+\infty} H(x;\Theta) = 1; \\
& \lim_{x\to-\infty} G(x;\Psi) = 0; \quad \lim_{x\to+\infty} G(x;\Psi) = 1.
\end{aligned} \tag{2.1}
$$

For simplicity, we take F and G to be proper distribution functions; H and K are continuous functions of x, and Θ and Ψ are parameter vectors.

When $D_l = \{d_1, d_2, ..., d_m\}$, the existence of x^* under assumptions (2.1) is trivial. When $D_l = (-\infty, +\infty)$, the maximum value of the probability of a cure (i.e. the maximum value of $(1-F(x))G(x)$) always exists over D_l under assumptions (2.1) as we will prove in Theorem 1, and in many cases the maximum can be found by solving the equation

$$g(x)\overline{F}(x) - f(x)G(x) = 0, \tag{2.2}$$

where $g(x) = G'(x), f(x) = F'(x), \overline{F}(x) = 1 - F(x)$. For certain distributions, F and G, x^* can be expressed as a function of Θ and Ψ explicitly.

For example, suppose $\overline{F}(x)$ and $G(x)$ are negative and positive extreme value distributions, respectively, i.e.

$$
\begin{aligned}
\overline{F}(x) = \overline{H}(x;\theta_1) &= exp\{-exp(\alpha_1 + \beta_1 x)\}, \quad \text{where } \beta_1 > 0; \\
G(x) = K(x;\theta_2) &= exp\{-exp(-\alpha_2 - \beta_2 x)\}, \quad \text{where } \beta_2 > 0.
\end{aligned}
$$

when $D_l = (-\infty, +\infty)$, the maximum probability of cure is obtained by solving (2.2),

$$x^* = \frac{log(\beta_2/\beta_1) - \alpha_1 - \alpha_2}{\beta_1 + \beta_2}.$$

If the probability of toxicity follows a logistic function and the probability of cure given non-toxicity follows an exponential function, namely,

$$
\begin{aligned}
\overline{F}(x) = \overline{H}(x;\theta_1) &= \frac{1}{1 + exp(\alpha_1 + \beta_1 x)}, \\
G(x) = K(x;\theta_2) &= exp\{\beta_2(x - \alpha_2)\},
\end{aligned}
$$

where $\beta_1 > 0, \beta_2 > 0$, and if $D_l = (-\infty, \alpha_2]$, then for $\beta_1 > \beta_2$, we have

$$x^* = \frac{log\{\beta_2/(\beta_1 - \beta_2)\} - \alpha_1}{\beta_1}.$$

Sometimes x^* cannot be obtained from (2.2). For example, if

$$\overline{F}(x) = \overline{H}(x;\theta_1) = exp\left(-\frac{x-\alpha_1}{\beta_1}\right), \quad \text{where} \quad \beta_1 > 0, \quad x \geq \alpha_1,$$

$$G(x) = K(x;\theta_2) = exp\left(\frac{x-\alpha_2}{\beta_2}\right), \quad \text{where} \quad \beta_2 > 0, \quad x \leq \alpha_2,$$

then the probability of cure is

$$\overline{F}(x)G(x) = exp\left\{\left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)x + \frac{\alpha_1}{\beta_1} - \frac{\alpha_2}{\beta_2}\right\}.$$

Since $\alpha_1 \leq x \leq \alpha_2$, x^* equals α_1 or α_2. If an explicit expression for x^* does not exist, we may still be able to find x^* from (2.2) using numerical methods. The existence of x^* is ensured by the following theorem.

THEOREM 1. *Let $F(x)$ and $G(x)$ be continuous distribution functions and let $D_l = (-\infty, +\infty)$. If $\overline{F}(x)G(x) \neq 0$ for some x, then x^* exists.*

PROOF. Without loss of generality, we assume $\overline{F}(0)G(0) > 0$. Since

$$\lim_{|x| \to +\infty} \overline{F}(x)G(x) = 0,$$

then there exists a $M > 0$ such that $\overline{F}(x)G(x) < \overline{F}(0)G(0)$ for $|x| \geq M$. $\overline{F}(x)G(x)$ is continuous, and therefore, it achieves its nonzero maximum value at $x^* \in [-M, +M]$, where $\overline{F}(x^*)G(x^*) \geq \overline{F}(0)G(0) > 0$. Consequently, x^* exists by its definition. □

Theorem 2 gives sufficient conditions for the uniqueness of the maximum value of $(1 - F(x))G(x)$ over $D_l = (-\infty, +\infty)$.

THEOREM 2. *If $F(x)$ and $G(x)$ are distribution functions with second derivatives, and $-log(1-F)$ and $-log(G)$ are strictly convex functions for all finite x. If $0 < \overline{F}(x)G(x) < 1$ for some finite x, then there exists a unique $x^* \in (-\infty, +\infty)$ maximizing $\overline{F}(x)G(x)$.*

PROOF. Without loss of generality, we assume $0 < \overline{F}(0)G(0) < 1$. For convenience, define

$$t(x) =: -log(G) - log(1 - F).$$

Then we have $\lim_{|x|\to+\infty} t(x) = +\infty$ and $t(0) > 0$. There exists $N > 0$ such that $t(x) > t(0)$ for $|x| \geq N$. In $[-N, N]$, $t(x)$ is continuous. Therefore $t(x)$ achieves its minimum value $x^* \in (-N, N)$ where $t'(x^*) = 0$. Since $t(x)$ is a strictly convex function, $t''(x) > 0$ and $t'(x)$ is strictly increasing. Therefore $t'(x) = 0$ has a unique root x^*. □

If F and G are multivariate distributions so that x is multidimensional vector, the additional complexity of the response functions prevent us giving a simple criterion for the existence of x^* [see Rockefeller (1972)].

2.1. The dependence of x^* on the underlying response distributions. Suppose that x^* exists and is a root of (2.2). In order to see how x^* depends on the underlying distributions F and G, let X be a random variable having distribution F and let Y be a random variable having distribution G. We define

$$\lambda_F(x) =: \frac{f(x)}{1 - F(x)} \quad \text{and} \quad \mu_G(x) =: \frac{g(x)}{G(x)}.$$

To interpret the failure rate function $\lambda_F(x)$, it is known [see, for example, Ross (1983)] that

$$P\{X \in (x, x + dx) | X > x\} \approx \lambda_F(x) \cdot dx.$$

Therefore, suppose that patient survives dose level x, $\lambda_F(x)dx$ represents the probability that an additional dose dx would have been toxic.

To interpret the reverse failure rate $\mu_G(x)$, note that

$$\begin{aligned} P\{Y \in (x - dx, x) | Y < x\} &= \frac{P\{Y \in (x - dx, x), Y < x\}}{P\{Y < x\}} \\ &= \frac{P\{Y \in (x - dx, x)\}}{P\{Y < x\}} \\ &\approx \frac{g(x) \cdot dx}{G(x)} \\ &= \mu_G(x) \cdot dx. \end{aligned}$$

That is, $\mu_G(x)dx$ represents the probability of a non-cure for a small dose reduction dx given that a cure occurs below dose level x and x is not toxic.

Based on this interpretation for $\lambda_F(x)$ and $\mu_G(x)$, it is reasonable to assume that $-\lambda_F(x)$ and $\mu_G(x)$ are decreasing functions of x. Let

$$K_{(F,G)}(x) = -\lambda_F(x) + \mu_G(x) = \frac{g(x)\overline{F}(x) - f(x)G(x)}{\overline{F}(x)G(x)}.$$

Then, if $-\lambda_F(x)$ and $\mu_G(x)$ are decreasing functions of x, $K_{(F,G)}(x)$ is a decreasing function of x under these assumptions. Furthermore, if x^* exists and is a root of (2.2) such that $\overline{F}(x^*)G(x^*) \neq 0$, then from (2.2),

$$K_{(F,G)}(x^*) = 0.$$

We write $x^* = x^*(F, G)$. The dependence of x^* on F and G is given by the following theorem and it turns out to be the following: the larger λ_F is, the smaller x^* is; the larger μ_G is, the larger x^* is.

THEOREM 3. *Let A represent the class of pairs of distribution functions $F(x)$ and $G(x)$ that satisfy the following conditions:* (1) *x^* exists and is a root of the equation* (2.2) *and $\overline{F}(x^*)G(x^*) \neq 0$; and* (2) *$-\lambda_F(x)$ and $\mu_G(x)$ are decreasing functions of x. Then for $(F_1, G_1), (F_2, G_2) \in A$, we have*

(i) *For fixed G, $\lambda_{F_1}(x) \geq \lambda_{F_2}(x)$ for all $x \Rightarrow x^*(F_1, G) \leq x^*(F_2, G)$;*

(ii) *For fixed F, $\mu_{G_1}(x) \geq \mu_{G_2}(x)$ for all $x \Rightarrow x^*(F, G_1) \geq x^*(F, G_2)$;*

(iii) *$K_{(F_2,G_2)}(x) \geq K_{(F_1,G_1)}(x)$ for all $x \Rightarrow x^*(F_2, G_2) \geq x^*(F_1, G_1)$.*

PROOF. We only prove (*iii*), because (*i*) and (*ii*) are the special cases of (*iii*). If $\overline{F}(x)G(x) \neq 0$ at x^*, then

$$K_{(F,G)}(x) = 0 \Longleftrightarrow g(x)\overline{F}(x) - f(x)G(x) = 0.$$

For convenience, let $x_1^* = x^*(F_1, G_1)$, and $x_2^* = x^*(F_2, G_2)$. Since $K_{(F_1,G_1)}(x) \leq K_{(F_2,G_2)}(x)$ for all x,

$$K_{(F_2,G_2)}(x_2^*) = 0 = K_{(F_1,G_1)}(x_1^*) \leq K_{(F_2,G_2)}(x_1^*).$$

Now $x_2^* \geq x_1^*$ is implied by the decreasing nature of $K_{(F_2,G_2)}(x)$. □

2.2. A class of adaptive designs. For a continuous interval of dose levels, our goal is to estimate x^* after each trial or group of trials and treat the next group at the level of this estimator. To see that x^* is a function of Θ and Ψ requires us to consider sequential design procedures for estimating x^* through its connection with Θ and Ψ.

(2.3)

CONTINUOUS INTERVAL SEQUENTIAL PROCEDURE:

1. Define the set of feasible dose levels D_l.

2. Find good estimators

$$\hat{\Theta}_n = \hat{\Theta}\{(y_{ij}, x_i) : i = 1, 2, ..., n, j = 1, 2, ..., k_i\}$$

for Θ and

$$\hat{\Psi}_n = \hat{\Psi}\{(y_{ij}, x_i) : i = 1, 2, ..., n, j = 1, 2, ..., k_i\}$$

for Ψ, where y_{ij} is the binary responses of the jth treatment at the ith trial and k_i is the number of experiments at the ith trial.

3. Define $\hat{F}_n(x) = H(x; \hat{\Theta}_n)$ and $\hat{G}_n(x) = G(x; \hat{\Psi}_n)$, and choose the next dose x^*_{n+1} such that x^*_{n+1} maximizes $(1 - \hat{F}_n(x))\hat{G}_n(x)$ in such a way that the group size, k_{n+1}, at x^*_{n+1} diverges, i.e., $\lim_{n\to\infty} k_n = \infty$. If $(1 - \hat{F}_n(x))\hat{G}_n(x)$ has the same maximum value at more than one dose level in D_l, choose the smallest as the next dose.

The sample sizes, k_n, in design (2.3) are not necessarily strictly increasing, but we must eventually put more and more experiments at doses close to the optimal point as trials continue. The sequence of dose levels x^*_n in design (2.3) may take any value in a continuous interval, and we do not decide the dose of the next group of trials before we complete the current group.

In practice, we may only be able to do the experiments at certain designated dose levels due to constraints inherent in the experiments or for the experimenter's convenience. Design (2.3) may not be advisable in this situation. Design (2.4) which follows maybe more appropriate for a discrete set of dose levels. It has the additional advantage that it can proceed in a strictly sequential manner.

(2.4)

DISCRETE LEVEL SEQUENTIAL PROCEDURE:

1. Define the set of feasible dose levels $D_l = \{d_1, d_2, ..., d_m\}, d_i \neq d_j$ for $i \neq j$.

2. Find good estimators

$$\hat{\Theta}_n = \hat{\Theta}\{(y_{ij}^k, d_i) : i = 1, 2, \cdots, m, j = 1, 2, ..., n_i^k, k = 1, 2, ..., n\}$$

and

$$\hat{\Psi}_n = \hat{\Psi}\{(y_{ij}^k, d_i) : i = 1, 2, ..., m, j = 1, 2, ..., n_i^k, k = 1, 2, ..., n\},$$

where n_i^k is the number of the experiments at d_i at the kth trial and $n_i^k \geq 0, y_{ij}^k$ is the binary response of the jth experiment at d_i at the kth trial. Assume that $\sum_{i=1}^m n_i^k \geq 1$ for all k.

3. Define $\hat{F}_n(x) = H(x; \hat{\Theta}_n)$ and $\hat{G}_n(x) = K(x; \hat{\Psi}_n)$, and choose the next design point, x_{n+1}^*, from D_l to maximize $(1 - \hat{F}_n(x))\hat{G}_n(x)$ over D_l. If $(1 - \hat{F}_n(x))\hat{G}_n(x)$ has the same maximum value at more than one point in D_l, choose the smallest as the next design level.

The sequence of doses x_n^* are given by maximizing $(1 - \hat{F}_n(x))\hat{G}_n(x)$ through its connection with the parameters. Some of the n_i^k in step 2 above can be zero, but we require that at least one experiment be done at some d_i at each trial, i.e., $\sum_{i=1}^m n_i^k \geq 1$ for all k.

The next issue is the choice of good estimators $\hat{\Theta}_n$ and $\hat{\Psi}_n$ in designs (2.3) and (2.4). We use the maximum likelihood estimators (MLEs) for Θ and Ψ based on all of the data in the previous trials. That way all of the information in the previous trials can be used in suggesting how the next trial should be conducted [cf. Wu (1985)]. For the computation of the maximum likelihood estimators for the parameters, the NLIN and CATMOD procedures in SAS, and GLIM are available.

It is easy to see that x_n^* in designs (2.3) and (2.4) is the optimal choice of the next design point in D_l based on estimated response curves, $\hat{F}_n(x)$ and $\hat{G}_n(x)$. We will prove its consistency through its connection with the parameter estimators (i.e., the MLEs). To start procedures (2.3), we need an initial dose and an interval, and to start

procedure (2.4) we need a discrete set of doses and an initial dose belonging to the set. Both proceduresrequire that the experiments be performed on at least two distinct dose levels in order that the MLEs exist. Whether the MLEs are computed using design (2.3) or (2.4), it is critical not to start the iterations until the conditions for the existence and uniqueness of the MLEs are satisfied. A premature start may lead to inconsistent estimators [see Wu (1985)].

2.3. The maximum likelihood estimators for the parameters. If, for $i = 1, 2, ..., k$, x_i is a dose level and n_i experiments are run at x_i, the likelihood function is

$$L = \prod_{i=1}^{k} \binom{n_i}{l_i \quad m_i} q_i^{l_i} r_i^{m_i} p_i^{n_i - l_i - m_i},$$

where l_i is the number of toxic responses at x_i, m_i is the number of cures with non-toxic responses at x_i, and $n_i - l_i - m_i$ is the number of non-cure with non-toxic responses at x_i. The q_i, r_i and p_i are given by $q_i = F(x_i)$, $r_i = \overline{F}(x_i)G(x_i)$ and $p_i = \overline{F}(x_i)\overline{G}(x_i)$.

The log-likelihood function is

(2.5) $\log(L)$

$$= C + \sum_{i=1}^{k} l_i log(q_i)$$

$$+ \sum_{i=1}^{k} m_i log(r_i) + \sum_{i=1}^{k} (n_i - l_i - m_i) log(p_i),$$

where C is a constant.

The following lemma is a basic consequence of (2.5).

LEMMA 1. *Under the assumption* (2.1), *if the first derivatives of* $H(x; \Theta)$ *and* $K(x; \Psi)$ *with respect to each component of* Θ *and* Ψ, *respectively, exist, then the maximum likelihood equations for* Θ *are functionally independent of those for* Ψ.

PROOF. From (2.5) and (2.1), we have

(2.6) $\log(L)$

$$
\begin{aligned}
= \; C \; &+ \; \textstyle\sum_{i=1}^{k} l_i log H(x_i;\Theta) \\
&+ \; \textstyle\sum_{i=1}^{k} m_i[log\overline{H}(x_i;\Theta) + log K(x_i;\Psi)] \\
&+ \; \textstyle\sum_{i=1}^{k}(n_i - l_i - m_i)[log\overline{H}(x_i;\Theta) + log\overline{K}(x_i;\Psi)] \\
= \; C \; &+ \; \textstyle\sum_{i=1}^{k} l_i log H(x_i;\Theta) + \sum_{i=1}^{k}(n_i - l_i) log\overline{H}(x_i;\Theta) \\
&+ \; \textstyle\sum_{i=1}^{k} m_i log K(x_i;\Psi) \\
&+ \; \textstyle\sum_{i=1}^{k}(n_i - l_i - m_i) log\overline{K}(x_i;\Psi).
\end{aligned}
$$

It is obvious that the first derivatives of $\log(L)$ with respect Θ are independent of those with respect to Ψ, so long as Θ is not a function of Ψ.

By Lemma 1, the MLEs for Θ and Ψ can be computed separately. For the implementation of designs (2.3) and (2.4), the existence and uniqueness of the MLEs is very important. Assuming that there are at least two distinct x_i's, $F(x) = H(\theta_1 + \theta_2 x)$ and $G(x) = K(\beta_1 + \beta_2 x)$, the following theorem is a consequence of Lemma 1 and Theorem(iii) in Silvapulle (1981).

THEOREM 4. *Suppose F and G are continuous distribution functions and $-logF$, $-log(1-F)$, $-logG$ and $-log(1-G)$ are convex functions and that $F(x) = H(\theta_1 + \theta_2 x)$ and $G(x) = K(\beta_1 + \beta_2 x)$. Then the MLEs for $\Theta = (\theta_1, \theta_2)$ and $\Psi = (\beta_1, \beta_2)$ exist and are unique iff the following conditions are satisfied*

$$(x^+_{1min}, x^+_{1max}) \bigcap (x^-_{1min}, x^-_{1max}) \neq \phi \tag{2.7}$$

and

$$(x^+_{2min}, x^+_{2max}) \bigcap (x^-_{2min}, x^-_{2max}) \neq \phi, \tag{2.8}$$

where

$$
\begin{aligned}
x^+_{1max(min)} &= max(min)\{x_i\text{: fatal toxicity at } x_i, i = 1, ..., k\}, \\
x^-_{1max(min)} &= max(min)\{x_i\text{: non-fatal toxicity at } x_i, i = 1, ..., k\}, \\
x^+_{2max(min)} &= max(min)\{x_i\text{: there is no cure at } x_i, i = 1, ..., k\}, \\
x^-_{2max(min)} &= max(min)\{x_i\text{: there is a cure at } x_i, i = 1, ..., k\}.
\end{aligned}
$$

PROOF. From Lemma 1 and (2.6), it is easy to see the MLEs for Θ

$= (\theta_1, \theta_2)$ and $\Psi = (\beta_1, \beta_2)$ are special cases of the theorem in Silvapulle (1981). The interlocking set of conditions (2.7) and (2.8) are simplified according to condition Π in that theorem. □

The conditions (2.7) and (2.8), once satisfied, are always satisfied and cannot be violated by the addition of more observations. When dose level is a multiple dimensional vector and it appears in the inner product form $X\Theta$ in the distribution functions, we still can give conditions similar to (2.7) and (2.8) [see Silvapulle (1981)]. The following corollary is a direct result of Theorem 4.

COROLLARY. *Under the conditions of Theorem 4, if there exist* $x_i \neq x_j$ *and* $x_{i'} \neq x_{j'}$ *such that* $0 < l_i < n_i, 0 < l_j < n_j, 0 < m_{i'} < n_{i'} - m_{i'}$ *and* $0 < m_{j'} < n_{j'} - l_{j'}$, *then the MLEs for* Θ *and* Ψ *exist and are unique.*

3. Statistical consistency and convergence to optimality. Through its connection with parameters, we have from (2.2), that x_n^* in (2.3) and (2.4) will converge to the optimal value x^* if the sequential MLEs for the parameters converge to the true parameters. It is known that [see Lehmann (1983)] the MLEs are consistent under suitable regularity conditions for independent identically distributed observations. In estimating the percentiles of the response curve, Wu (1985) proved the consistency under the rather restrictive conditions that the sequential MLEs converge uniformly to constants. The following theorem concerning designs (2.3) and (2.4) proves the consistency of the MLEs for the parameters under rather broad conditions.

THEOREM 5. *Let* $F(x) = H(x; \Theta)$ *and* $G(x) = K(x; \Psi)$ *be probability distribution functions, where*

$$\Theta = (\theta^1, \theta^2, ..., \theta^s) \text{ and } \Psi = (\beta^1, \beta^2, ..., \beta^t).$$

Let D_l *be the set of doses. Suppose* F *and* G *satisfy the following conditions:*

(i) *The parameter spaces for each component of* Θ *and* Ψ *are open intervals.*

(ii) *Both distribution functions* $H(x; \Theta)$ *and* $K(x; \Psi)$ *have their common supports, respectively.*

(iii) *For all* i *and* j, $\partial F/\partial\theta^i$, $\partial G/\partial\beta^j$, $\partial^2 F/\partial(\theta^i)^2$ *and* $\partial^2 G/\partial(\beta^j)^2$ *exist and are bounded for* $x \in D_l$.

(iv) *There is a constant* $b > 0$ *such that* $|\partial F/\partial\theta^i| > b$, $|\partial G/\partial\beta^j| > b$, $F\overline{F} > b$ *and* $G\overline{G} > b$ *for* $x \in D_l$.

Then no matter how the experimental doses are taken, given that the MLEs exist and are unique we have

(a) *If* D_l *is finite closed interval and* $\lim_{k\to\infty} n_k = \infty$, *then* $\hat{\theta}_k^i \overset{a.s.}{\to} \theta^i$ *and* $\hat{\beta}_k^j \overset{a.s.}{\to} \beta^j$ *as* $k \to \infty$, *where* $\hat{\theta}_k^i$ *and* $\hat{\beta}_k^j$ *are the unique MLEs for* θ^i *and* β^j *based on experiments at* $x_0, x_1, ..., x_k$. $x_i \in D_l$ *for all* i, *and* n_i *is the group size at* x_i.

(b) *If* $D_l = \{d_1, d_2, ..., d_m\}$ *with* $m \geq s + t$ *and each* $n_{ik} \to \infty$ *as* $k \to \infty$, *where* n_{ik} *is the group size at* d_i *by the* k*th trial; then* $\hat{\theta}_k^i \overset{a.s.}{\to} \theta^i$ *and* $\hat{\beta}_k^j \overset{a.s.}{\to} \beta^j$ *as* $k \to \infty$, *where* $\hat{\theta}_k^i$ *and* $\hat{\beta}_k^j$ *are the unique MLEs for* θ^i *and* β^j *based on the previous* k *trials at* D_l.

PROOF. For simplicity of notation, we only prove the theorem for the case in which Θ and Ψ are one-dimensional, i.e., $\Theta = \theta$ and $\Psi = \beta$.

(a) From (2.6), the log-likelihood function is

$$L_k = C + \sum_{i=1}^{k} l_i log F(x_i) + \sum_{i=1}^{k} (n_i - l_i) log\overline{F}(x_i) + \sum_{i=1}^{k} m_i log G(x_i) + \sum_{i=1}^{k} (n_i - l_i - m_i) log\overline{G}(x_i).$$

The likelihood equation is

$$\frac{\partial L_k}{\partial\theta} = \sum_{i=1}^{k} \frac{l_i - n_i F(x_i)}{F(x_i)\overline{F}(x_i)} \cdot \frac{\partial F}{\partial\theta} = \sum_{i=1}^{k} h(x_i;\theta) = 0,$$

where

$$h(x_i;\theta) =: \frac{l_i - n_i F(x_i)}{F(x_i)\overline{F}(x_i)} \cdot \frac{\partial F}{\partial\theta}.$$

Taking a Taylor series expansion of $h(x_i;\hat{\theta}_k)$ as a function of $\hat{\theta}_k$ about θ (considering x_i to be constant), we have

$$h(x_i;\hat{\theta}_k) = h(x_i;\theta) + (\hat{\theta}_k - \theta)k(x_i;\xi_k), \tag{3.1}$$

where $k(x_i;\theta) =: \partial h(x_i;\xi_k)/\partial\xi_k$, and ξ_k lies between θ and $\hat{\theta}_k$. Summing (3.1) over i and using $\sum_{i=1}^{k} h(x_i;\hat{\theta}_k) = 0$ yields

$$\frac{\partial L_k}{\partial\theta} = \sum_{i=1}^{k} h(x_i;\theta) = -(\hat{\theta}_k - \theta)\sum_{i=1}^{k} k(x_i;\xi_k). \tag{3.2}$$

If $\sum_{i=1}^{k} k(x_i;\xi_k) = 0$, then $\sum_{i=1}^{k} h(x_i;\theta) = 0$ from (3.1). By the uniqueness of MLE, $\hat{\theta}_k = \theta$. Without loss of generality, we assume

$$\sum_{i=1}^{k} k(x_i;\xi_k) \neq 0.$$

Hence from (3.2) we have

$$\hat{\theta}_k - \theta = -\frac{\sum_{i=1}^{k} h(x_i\,\theta)}{\sum_{i=1}^{k} k(x_i;\xi_k)} = \frac{a_k}{b_k}, \tag{3.3}$$

where

$$a_k = -\frac{\sum_{i=1}^{k} h(x_i;\theta)}{\sum_{i=1}^{k} n_i}, \quad b_k = \frac{\sum_{i=1}^{k} k(x_i;\xi_k)}{\sum_{i=1}^{k} n_i}.$$

By the strong law of large numbers, $l_i/n_i - F(x_i) \to 0$ a.s. as $n_i \to \infty$. Furthermore,

$$\frac{l_i/n_i - F(x_i)}{F\overline{F}} \cdot \frac{\partial F}{\partial\theta} \to 0 \quad \text{a.s. as } n_i \to \infty$$

since $F\overline{F} > b$ and $\partial F/\partial\theta$ is bounded. Noting that $x_n \to 0$ a.s. implies $\sum_{i=1}^{n} x_i/n \to 0$ a.s. as $n \to \infty$, we have

$$\begin{aligned} a_k &= -\frac{\sum_{i=1}^{k} h(x_i;\theta)}{\sum_{i=1}^{k} n_i} \\ &= -\frac{\sum_{i=1}^{k} n_i(l_i/n_i - F(X_i))(\partial F/\partial\theta)/(F\overline{F})}{\sum_{i=1}^{k} n_i} \longrightarrow 0 \end{aligned} \tag{3.4}$$

almost surely as $k \to \infty$.

For the denominator in (3.3),

$$(3.5) \qquad b_k = \frac{\sum_{i=1}^{k} k(x_i;\xi_k)}{\sum_{i=1}^{k} n_i}$$

$$= \frac{\sum_{i=1}^{k} n_i(l_i/n_i - F(x_i))(\partial^2 F/\partial\xi_k^2)/(F\overline{F})}{\sum_{i=1}^{k} n_i}$$

$$- \frac{\sum_{i=1}^{k} n_i(l_i/n_i - F(x_i))(\partial F/\partial\xi_k)^2(1-2F)/(F\overline{F})^2}{\sum_{i=1}^{k} n_i}$$

$$- \frac{\sum_{i=1}^{k} n_i(\partial F/\partial\xi_k)^2/(F\overline{F})}{\sum_{i=1}^{k} n_i}.$$

By the same reasoning as used to obtain (3.4), the first two terms in the right side of (3.5) converge to zero a.s. as $k \to \infty$. As for the third term, since $1/F\overline{F} \geq 4, |\partial F/\partial\theta|^2 \geq b^2$, we have

$$\left| \frac{\sum_{i=1}^{k} n_i(\partial F/\partial\theta)^2/(F\overline{F})}{\sum_{i=1}^{k} n_i} \right| \geq 4b^2 \quad \text{for all} \quad k.$$

Thus, from (3.5), we have $|b_k| \geq 2b^2$ if k is large enough. Consequently, $\hat{\theta}_k \to \theta$ a.s. as $k \to \infty$ follows from (3.3) and (3.4).

Since $l_i/n_i \to F(x_i) \neq 1$ a.s. as $n_i \to \infty$, we have $n_i - l_i = n_i(1 - l_i/n_i) \to \infty$ as $n_i \to \infty$. By Lemma 1 and the same methods as above, we have $\hat{\beta}_k \to \beta$ a.s. as $k \to \infty$.

(b) Let $n_{ik} = \sum_{j=1}^{k} n_i^j, l_{ik} = \sum_{j=1}^{k} l_i^j$, and $m_{ik} = \sum_{j=1}^{k} m_i^j$, where

- l_i^j is the number of toxicities at d_i at the jth trial,
- m_i^j is the number of cure with non-toxic responses at d_i at the jth trial,
- n_i^j is the number of experiments at d_i at the jth trial,
- l_{ik} is the total number of toxicities at d_i by the kth trial,
- m_{ik} is the total number of cures at d_i by the kth trial,
- n_{ik} is the total number of experiments at d_i by the kth trial.

The likelihood function is

$$L_k = \prod_{i=1}^{m} \binom{n_{ik}}{l_{ik} \quad m_{ik}} q_i^{l_{ik}} r_i^{m_{ik}} p_i^{n_{ik}-l_{ik}-m_{ik}},$$

where q_i, r_i and p_i are defined in Section 2.3. The log-likelihood function is

$$log(L_k) = C + \sum_{i=1}^{m} l_{ik} log F(x_i) + \sum_{i=1}^{m} (n_{ik} - l_{ik}) log \overline{F}(x_i)$$
$$+ \sum_{i=1}^{m} m_{ik} log G(x_i) + \sum_{i=1}^{m} (n_{ik} - l_{ik} - m_{ik}) log \overline{G}(x_i).$$

Therefore,

$$\frac{\partial log(L_k)}{\partial \theta} = \sum_{i=1}^{m} \frac{n_{ik}(l_{ik}/n_{ik} - F(x_i))}{F\overline{F}} \cdot \frac{\partial F}{\partial \theta} = \sum_{i=1}^{m} h_k(x_i; \theta),$$

where

$$h_k(x_i; \theta) = \frac{n_{ik}(l_{ik}/n_{ik} - F(x_i))}{F\overline{F}} \cdot \frac{\partial F}{\partial \theta}.$$

By the Taylor series expansion of $h_k(x_i; \hat{\theta}_k)$ as a function of $\hat{\theta}_k$ about θ, we have

$$h_k(x_i; \hat{\theta}_k) = h_k(x_i; \theta) + (\hat{\theta}_k - \theta) q_k(x_i; \xi_k),$$

where $q_k(x_i; \xi_k) = \partial h_k(x_i; \xi_k)/\partial \xi_k$, and ξ_k lies between θ and $\hat{\theta}_k$. Since $\sum_{i=1}^{m} h_k(x_i; \hat{\theta}_k) = 0$, by the same reasoning as (3.3),

$$\hat{\theta}_k - \theta = -\frac{\sum_{i=1}^{m} h_k(x_i; \theta)}{\sum_{i=1}^{m} q_k(x_i; \xi_k)}. \tag{3.6}$$

If $n_{ik} \to \infty$, by the strong law of large numbers,

$$\left| \frac{h_k(x_i; \theta)}{\sum_{i=1}^{m} n_{ik}} \right| \leq \frac{|l_{ik}/n_{ik} - F(x_i)|}{F\overline{F}} \frac{\partial F}{\partial \theta} \to 0 \quad \text{a.s.}$$

as $k \to \infty$. Therefore, we have

$$\frac{\sum_{i=1}^{m} h_k(x_i; \theta)}{\sum_{i=1}^{m} n_{ik}} \to 0 \quad \text{a.s. as} \quad k \to \infty.$$

The way of dealing with the denominator in (3.6) is similar to that for part (a). We complete the proof following methods analogous to those used for (a). □

From Theorem 5, we know x_n^* in designs (2.3) and (2.4) converges to x^* through its connection with the parameters. The key idea for ensuring consistency of design (2.3) is to put more experiments at doses

close to x^*. The conditions (iii) and (iv) over D_l for the underlying distributions in the Theorem 5 are not hard to satisfy under (i) and (ii).

For example, if

$$F(x) = 1 - \frac{1}{1 + exp(\alpha + \beta x)}, \quad \text{where} \quad \beta > 0, \quad -\infty < \alpha < +\infty,$$

it is obvious that (i) and (ii) are satisfied. Since

$$\left|\frac{\partial F}{\partial \alpha}\right| = \left|\frac{\partial^2 F}{\partial \alpha^2}\right| = F(x), \quad \left|\frac{\partial F}{\partial \beta}\right| = xF(x), \quad \text{and} \quad \left|\frac{\partial^2 F}{\partial \beta^2}\right| = x^2 F(x),$$

(iii) and (iv) are satisfied for $x \in D_l$.

The following corollary, which gives the asymptotic optimality of x_n^* for designs (2.3) and (2.4), is a natural consequence of Theorem 5.

COROLLARY. *Under assumption (2.1) and the conditions in Theorem 5, the sequentially estimated optimal design points $x_n^* \in D_l$, defined in designs (2.3) and (2.4) converges to $x^* \in D_l$ a.s.*

4. Discussion. We have proved the asymptotic optimality for the designs proposed in Section 2.2 by first proving the consistency of the sequential MLEs for the parameters in the response curves. However, we need to define the set of feasible dose levels D_l and give the initial designs such that the MLEs based on the initial designs exist and are unique before the treatments begin in a sequential manner. The conditions for the existence and uniqueness of MLEs, once satisfied, are never violated by adding more observations. Ways to choose the initial designs needs further study. Theoretically we only require that the conditions for the existence and uniqueness of MLEs be satisfied before the sequential procedures begin. Defining the set of feasible dose levels D_l is mainly based on previous knowledge about the treatments. It is not always easy to estimator the variance of x_n^* theoretically due to the sequential character of x_n^*. Further study is needed to evaluate the variance of x_n^*.

References

HAYES, G.J., EDIE, D.D. and DURHAM, S.D. (1992). Statistical analysis of recoil test data for carbon fibers. *Twentieth Biennial Conference on Carbon*, 326-327.

LEHMANN, E.L. (1983). *Theory of Point Estimation.* New York: Wiley.

ROCKEFELLER, R.T. (1972). *Convex Analysis.* Princeton, N. J.

ROSS, S.M. (1983). *Stochastic Processes.* John Wiley and Sons.

SILVAPULLE, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response model. *Journal of the Royal Statistical Society B* **43** 310-313.

WU, C.F.J. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association* **80** 974-984.

ZHENGQING LI
DEPARTMENT OF STATISTICS
THE UNIVERSITY OF WISCONSIN
1210 DAYTON STREET
MADISON, WI 53706-1693

STEPHEN D. DURHAM
DEPARTMENT OF STATISTICS
THE UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208

NANCY FLOURNOY
DEPARTMENT OF MATHEMATICS AND STATISTICS
THE AMERICAN UNIVERSITY
WASHINGTON, DC 20016-8050