New Developments and Applications in Experimental Design
IMS Lecture Notes – Monograph Series (1998) Volume 34

RESTRICTED OPTIMALITY FOR PHASE I CLINICAL TRIALS

VLADIMIR A. MATS, WILLIAM F. ROSENBERGER[1] AND NANCY FLOURNOY[2]

University of Maryland, University of Maryland and American University

We derive locally c- and D-optimal designs for maximum likelihood estimation of the maximum tolerated dose from a phase I clinical trial. We assume that the response function can be modeled by a distribution function and there is an ethical constraint on the dose space, as patients should not be assigned to potentially highly toxic dose levels. In virtually every case, the c-optimal design will be a one-point design at the unknown quantile of interest. The D-optimal design is two points. We conclude with a discussion of sequential designs for phase I clinical trials and their relevance in the context of the optimality results.

1. Introduction. The objective of a phase I clinical trial is to determine a maximum tolerated dose (MTD), where response is generally dichotomous ("toxicity" or "no toxicity"). The MTD is then passed to further phases of testing; determining efficacy of the therapy is often relegated to these later phases. Correct estimation of the MTD in phase I clinical trials is of critical importance to public health. Assuming that efficacy is monotonically increasing across the dose space, if we underestimate the MTD and pass too low a dose to later testing phases, we risk experimenting with an ineffective treatment. Hence we could sacrifice a potentially life-saving drug by experimenting with the wrong dose. Phase I clinical trials traditionally recruit fairly small numbers of subjects, and hence efficiency of estimation is particularly germane. In addition, experiments are constrained by ethical considerations. Patients enrolled in the study should not be assigned to dose levels far above the MTD. In this paper, we discuss formal experimental design considerations for optimal estimation of the MTD in phase I clinical trials. In particular, we derive the locally c- and D-optimal designs for the maximum likelihood estimator of the MTD under an ethical constraint

---

[1] Research supported by National Institute of Diabetes and Digestive and Kidney Diseases Grant R29-DK51017-03.

[2] Research supported by Environmental Protection Agency Grant CR822636-02.

Received September 1997; revised February 1998.

AMS 1991 subject classifications. Primary 62K05; secondary 62L05.

Key words and phrases. c-optimality, D-optimality, Elfving's method, maximum tolerated dose, quantile estimation, sequential design, toxicity study.

on the dose space and discuss implications in designing studies. By "locally optimal," we assume nominal values for the parameters are available. Ford, Torsney, and Wu (1992) elegantly describe a general framework for constrained optimization, and we refer readers interested in a more general theory to that paper. Our interest is in the very specific application to phase I clinical trials.

2. Estimation of the MTD. In order to discuss estimation of the MTD, we need first to have a rigorous probabilistic definition of the MTD. It is generally accepted that the MTD refers to a quantile of a monotonic curve. If we assume that response is dichotomous and that the probability of toxicity increases as the dose level of a drug increases, then a quantile is the dose level corresponding to a desired probability of toxicity. That is, for a sample space of doses Ω_x, the quantile $\mu \in \Omega_x$ corresponding to a target probability of toxicity Γ is defined by $P\{\text{toxicity is observed given dose level } \mu\} = \Gamma \in (0,1)$. In practice, the experimenter chooses Γ according to the seriousness of the toxicity, and the choice of Γ may vary extensively among diseases studied. Cancer studies, for instance, are often interested in values of Γ ranging from 0.10 to 0.25.

Estimation of μ may be facilitated by assuming the response function follows some probability distribution from a location scale family. If we let the probability of toxicity at dose level x_i be modeled as $F\{(x_i - \alpha)/\beta\}$, where α and β are unknown parameters and F is a distribution function, then

$$\mu = \alpha + F^{-1}(\Gamma)\beta. \tag{2.1}$$

For example, F can be modeled as logistic, probit, or complementary log-log. Under an assumed distribution, one obtains maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β, respectively, and then $\hat{\mu} = \hat{\alpha} + F^{-1}(\Gamma)\hat{\beta}$, the information matrix for a single dose associated with $\hat{\alpha}$ and $\hat{\beta}$ is given by

$$I = \frac{1}{\beta^2} \frac{[f(\lambda)]^2}{F(\lambda)[1 - F(\lambda)]} \begin{bmatrix} 1 & \lambda \\ \lambda & \lambda^2 \end{bmatrix}, \tag{2.2}$$

where f is the density function corresponding to F and $\lambda = (x - \alpha)/\beta$ for $x \in \Omega_x$ [see, for example, Morgan (1992, Section 2.7)]. By summing over doses in (2.2) and inverting the information matrix, we obtain an approximation to the variance-covariance matrix of $(\hat{\alpha},\hat{\beta})$, and hence we can derive an approximation to the asymptotic variance of $\hat{\mu}$ as

$$Var(\hat{\mu}) = Var(\hat{\alpha}) + \left[F^{-1}(\Gamma)\right]^2 Var(\hat{\beta}) + 2[F^{-1}(\Gamma)]cov(\hat{\alpha}, \hat{\beta}), \tag{2.3}$$

which in itself can be estimated.

3. c-Optimality and Elfving's method. Since μ can be estimated as a linear combination of parameters (α, β) by (2.1), a logical choice for an optimality criterion is to minimize a quadratic form of the inverse of the information matrix. The c-optimality criterion [Silvey (1980, p. 13)] is that of minimizing $\boldsymbol{c'I^{-1}c}$ for a fixed vector $\boldsymbol{c}$. For estimating a quantile, the choice for $\boldsymbol{c}$ is $(1, F^{-1}(\Gamma))'$, and the optimality criterion is thus the minimum of (2.3).

Elfving (1952) described a clever graphical technique to determine the optimal design for estimating a linear combination of regression parameters. Using his techniques for the standard linear model

$$y = \alpha x_1 + \beta x_2 + \epsilon \tag{3.1}$$

for (x_1, x_2) in a specified set S where ϵ has mean 0 and variance σ^2. Elfving's method obtains the optimal design points as follows: The design points $(x_{11}, x_{21}), ..., (x_{1n}, x_{2n})$ for estimating $a_1\alpha + a_2\beta$, for constants a_1 and a_2, must be selected from S. See Chernoff (1972, p. 16) for a graphical depiction. Draw S^*, the convex hull generated by S and its reflection through the origin S^-. Then extend a ray $\boldsymbol{z}$ from the origin through the point (a_1, a_2). The location where the ray penetrates S^* yields the solution. If the ray penetrates S^* through S, we have a one-point design at the point (x_1, x_2) of penetration (i.e., take all n design pairs at that point). If it penetrates S^* at a convex combination of points of S and S^-, then we have a two point design at point (x_1, x_2) of S and point $(-x_1, -x_2)$ of S^-. If we let w be the distance from the penetration point to S divided by the distance between S and S^- along the same line segment, we take $n(1-w)$ points at (x_1, x_2) of S and nw points at $(-x_1, -x_2)$ of S^- (assuming nw is an integer, otherwise some rounding scheme is necessary). The variance of the estimator is then $\sigma^2 \, ||\boldsymbol{a}||^2/n \, ||\boldsymbol{z}||^2$, where $\boldsymbol{a}= (a_1, a_2)'$, which gives a direct measure to compare the efficiency of designs. Chernoff (1972) gives a nice proof of Elfving's method; the proof relies on S being closed and bounded. The geometric results concerning optimality criteria of Elfving have been recently generalized in a series of papers by Dette and his colleagues [e.g., Dette (1993a, 1993b), Dette and Studden (1993), and Dette, Heiligers, and Studden (1995)].

Chernoff (1972) recognized that Elfving's method could be used to determine the c-optimal design for the maximum likelihood estimator of a quantile in a nonlinear regression problem. Let a_1=1 and $a_2 = F^{-1}(\Gamma)$, as in (1).

Letting $\beta = 1$, without loss of generality because it is a scaling factor, it is readily seen that setting

$$x_1(\lambda) = f(\lambda)[F(\lambda)(1-F(\lambda)]^{-1/2} \quad \textit{and} \quad x_2(\lambda) = \lambda f(\lambda)[F(\lambda)(1-F(\lambda)]^{-1/2} \tag{3.2}$$

in (3.1) yields the same information matrix as standard linear regression, as given by (2.2). Chernoff (1972) pointed out that since the optimality criterion depends only on the information matrix, Elfving's method can be used to determine c-optimality for the quantile estimation problem. Figure 1 gives the graph of S^* for the logistic distribution. The coordinates are given by (3.2). See Figure 3 of Chernoff (1972, p. 30) for a graph of S^* for the probit model. In all three cases, a ray that penetrates S^* at a convex combination of points of S and S^- corresponds to a two-point c-optimal design.

Recent papers have discussed optimality criteria for binary response [e.g., Wu (1988), Sitter and Wu (1993a, 1993b), and Sitter and Fainaru (1997)]. In particular, Wu (1988) determined the c-optimal design for quantile estimation under various distributions. For each of the distributions he examined, the optimal design was a

one-point design for $\Gamma > 0.18$. For $\Gamma \leq 0.18$, the distributions are highly variable; the minimum value of Γ for which the optimal design is one-point ranges from 0.04 to 0.18. Estimation of smaller quantiles than these requires allocating patients between two points, one low quantile and one highly toxic quantile. This finding leads to some

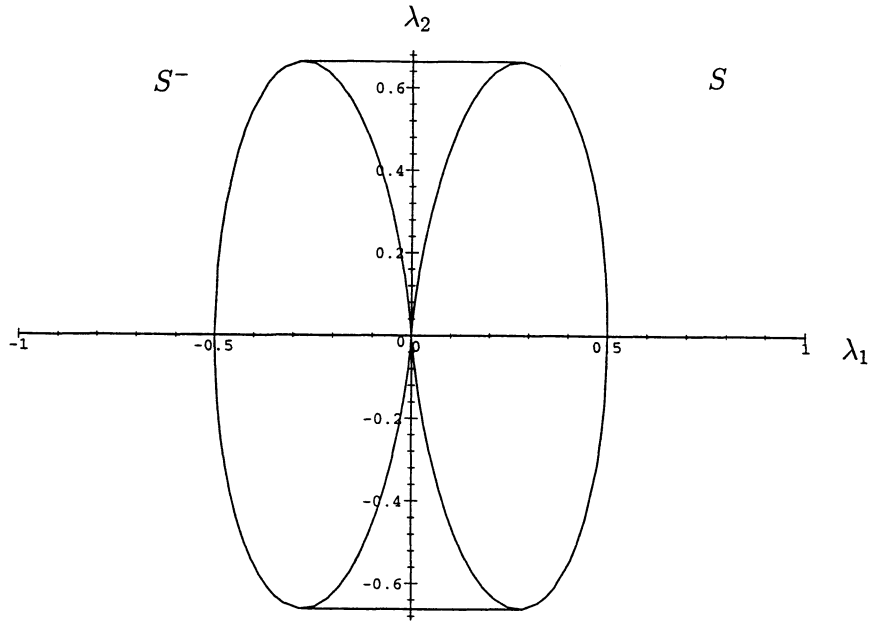

FIG. 1. *S^* for logistic distribution (unrestricted design).*

concern when estimating the quantile corresponding to $\Gamma = 0.10$ or 0.15, sometimes of interest in phase I clinical trials. Depending on the underlying dose-response distribution, the optimal design would be either a one-point design at the target quantile or a two-point design at extreme values. However, Wu's results become altered for phase I clinical trials, because the design space Ω_x is restricted due to ethical considerations. Assume that it is unethical to assign patients above a certain quantile. In the next section, we use Elfving's method to derive the c-optimal design for quantile estimation in a phase I clinical trial with a restricted design space.

4. c-Optimality for restricted dose spaces. Now consider quantile estimation when $x \in \Omega_x$ has the restriction that $x \leq \mu_R$, where μ_R is the quantile corresponding to some $\Gamma_R \in (0, 1)$. For example, we may be interested in estimating μ corresponding to $\Gamma = 0.25$ while imposing the restriction that no patient should be exposed to doses where the probability of toxicity is above $\Gamma_R = 0.33$.

Figure 2 gives the convex hull S^* for the logistic distribution when $\Gamma = 0.15$ and $\Gamma_R = 0.33$. Whereas with an unrestricted dose space, the minimum value of S is the

breakpoint for distinguishing between a one-point and a two-point c-optimal design, shifted left to the tangent point of S with the induced convex hull. This breakpoint corresponds to a quantile which we will denote by Γ_0. Hence, two-point designs arise under only very small values of Γ. We use standard formulas for finding the tangent point corresponding to Γ_0. The following differential equation determines Γ_0 [details are found in Mats (1997)]:

$$\frac{x_2(F^{-1}(\Gamma_R)) + x_2(F^{-1}(\Gamma_0))}{x_1(F^{-1}(\Gamma_R)) + x_1(F^{-1}(\Gamma_0))} = \frac{x_2'(F^{-1}(\Gamma_0))}{x_1'(F^{-1}(\Gamma_0))},$$

where $x_1\{\bullet\}$ and $x_2\{\bullet\}$ are points (3.2) on S^* and $'$ denotes derivative. When $\Gamma < \Gamma_0$, the c-optimal design consists of two-points, given by (x_1^*,x_2^*), the coordinates of the point where $\mathbf{z}$, the ray $[1, F^{-1}(\Gamma)]$, penetrates the convex hull, using the following system:

$$\frac{x_2^* + x_2(F^{-1}(\Gamma_0))}{x_2(F^{-1}(\Gamma_R)) + x_2(F^{-1}(\Gamma_0))} = \frac{x_1^* + x_1(F^{-1}(\Gamma_0))}{x_1(F^{-1}(\Gamma_R)) + x_1(F^{-1}(\Gamma_0))}, \tag{4.1}$$

$$x_2^* = F^{-1}(\Gamma)x_1^*.$$

[Mats (1997)]. Weights are given by the system [Mats (1997)]:

$$w_1 x_1(F^{-1}(\Gamma_0)) = w_2 x_1(F^{-1}(\Gamma_R)), \tag{4.2}$$

$$w_1 + w_2 = 1.$$

In Figure 2, $\Gamma_0 = 0.028$. In determining the c-optimal design for $\Gamma = 0.15$, the ray from the origin through the point $(1, F^{-1}(\Gamma))$ intersects S^* through S. Hence, the c-optimal design consists of one point at μ corresponding to the percentile 0.15. Suppose next we were to estimate the quantile μ corresponding to $\Gamma = 0.025$. According to equations (4.1), the ray passes through S^* at the convex combination of points corresponding to $F^{-1}(0.33)$ and $F^{-1}(0.028)$. So the optimal design for estimating the quantile μ corresponding to $\Gamma = 0.025$ consists of dosages at the levels corresponding to the percentiles 0.33 and 0.028. Equation (4.2) yields that 0.6% of the subjects be assigned to the former and 99.4% assigned to the latter.

In Figure 3, we have the complementary log-log distribution restricted at $\Gamma_R = 0.33$ with $\Gamma = 0.15$. Again we have a one-point design.

Table 1 gives the values of Γ_0 which are the breakpoints distinguishing a one-point and two-point design for a variety of distribution functions. Typically, phase I clinical trials will be designed to target $\Gamma < 0.10$, so one can conclude from this exercise that, typically, a one-point design at the desired quantile μ is c-optimal design for any distribution function and any $\Gamma < \Gamma_R < 0.5$.

Software in MAPLE that draws S^* and gives the c-optimal design for any distribution function, any Γ, and any Γ_R (including unrestricted problems) is available from the first author upon request.

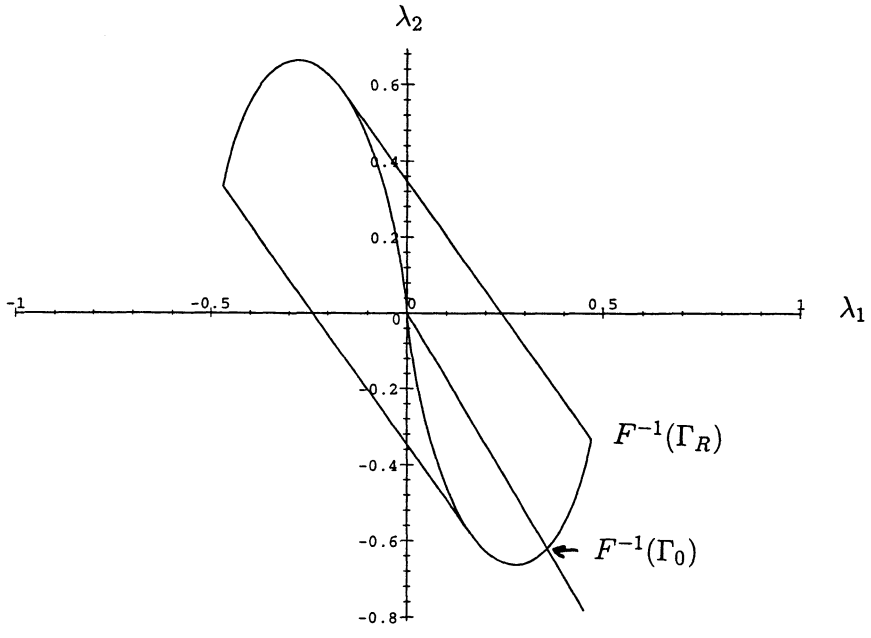

FIG. 2. *S^* for the logistic distribution (unrestricted design) with $\Gamma_R = 0.33$ and the ray passing through the point corresponding to $\Gamma = 0.15$. Here $\Gamma_0 = 0.028$.*

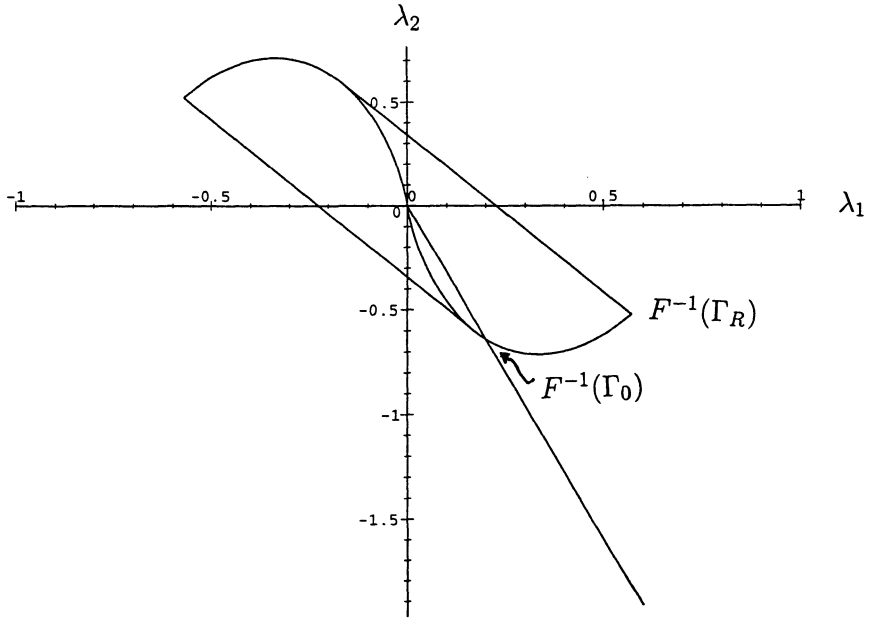

FIG. 3. *S^* for the complementary log-log distribution with $\Gamma_R = 0.33$ and the ray passing through the point corresponding to $\Gamma = 0.15$. Here $\Gamma_0 = 0.029$.*

TABLE 1

Values of Γ_0 on S^* disciminating between one-point and two-point design. Note that $\Gamma_R = 1.0$ is the unrestricted case

Distribution	F(y)	Γ_R=0.25	Γ_R=0.33	Γ_R=0.50	Γ_R=1.0
Logistic	$1 - \frac{1}{1+e^y}$	0.021	0.028	0.044	0.083
Log-log	$1 - e^{-e^y}$	0.021	0.029	0.046	0.109
D. Exp.	$\frac{1}{2}e^y, \quad y < 0$ $1 - \frac{1}{2}e^{-y}, \quad y \geq 0$	0.021	0.029	0.060	0.079
Probit	$\int_{-\infty}^{y} (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}} dt$	0.014	0.019	0.029	0.059

5. *D*-Optimality for restricted dose spaces. The *D*-optimality criteria involves finding a design that maximizes the determinant of the information matrix, given in (2.2). It also has a geometric interpretation similar to that in Elfving's method, where the support points of the *D*-optimal design are the points of contact between S^* and the minimal ellipsoid containing S^* [Silvey (1980)]. However, in practice it is difficult to find these points, and hence numerical optimization methods are more useful.

As before, we use the coordinates given by (3.2). We then maximize the determinant of

$$\mathbf{I}(\lambda) = \sum_{i=1}^{K} w_i x_1^2(\lambda_i) \begin{bmatrix} 1 & \lambda_i \\ \lambda_i & \lambda_i^2 \end{bmatrix},$$

where w_i is the proportion of subjects treated at λ_i. By Fedorov (1972), K is either 2 or 3. Hence, the *D*-optimal design consists of finding the maximum of a function of three variables in the case of a two-point design, or a function of five variables in the case of a three-point design. It is easy to show that, for a two-point design, $w_1 = w_2 = 1/2$. Hence, finding a two-point design reduces to finding the maximum of the function $Q(\lambda_1, \lambda_2) = x_1^2(\lambda_1)x_1^2(\lambda_2)(\lambda_1 - \lambda_2)^2/4$, where $\lambda_i = (x_i - \alpha)/\beta, i = 1, 2$, and $x_i \in \Omega_x$. The surface created by Q is symmetric about the line $\lambda_1 = \lambda_2$, and hence the solution is in the halfplane above the line. It is also symmetric about the origin, so we can look at solutions in the set $\{\lambda_2 \geq \lambda_1\} \cap \{\lambda_2 \geq -\lambda_1\}$. In the restricted case, where $x \leq \mu_R$, we need only search in that region. Define $(\lambda_1^0, \lambda_2^0)$ as the best two-point design.

Generally speaking, the *D*-optimal two point design is not globally *D*-optimal. Conditions for global optimality are given in Ford, Torsney, and Wu (1992) and originate from the equivalence theorem of Kiefer and Wolfowitz (1960). For two-parameter

response functions it is sufficient that

$$\phi(\lambda) := [x_1(\lambda) \quad x_2(\lambda)]\mathbf{I}^{-1}(\lambda_1^0, \lambda_2^0) \begin{bmatrix} x_1(\lambda) \\ x_2(\lambda) \end{bmatrix} \leq 2, \quad \forall \lambda \in \Omega_x. \tag{5.1}$$

If condition (5.1) is not satisfied, then the global D-optimal design is three points, and finding the global three-point design is a more difficult task. Dette and Haines (1994) find the D-optimal three-point design for the double exponential and double reciprocal distributions. Geometrically, the smallest ellipsoid will touch the convex hull in three points for those distributions. However, in the restricted case with $\Gamma_R \leq 0.5$, one can think in geometric terms of the smallest ellipsoid, which will touch the convex hull in only two points. Hence, for our problem, the two-point design satisfies the D-optimality criterion.

We used MATHCAD to find the two-point D-optimal design for each of the distributions of interest and also verifying (5.1). The logistic, probit, and complementary log-log distributions each have a global two-point D-optimal designs. We were not able to find the global D-optimal design for the double exponential distribution. See Dette and Haines (1994) for the global D-optimal design in the unrestricted case. Table 2 gives the D-optimal two-point designs in terms of quantiles of the distribution.

TABLE 2

Two-point D-optimal designs under various restrictions Γ_R *(note that* $\Gamma_R = 1.0$ *is the unrestricted case)*

Distribution	F(y)	Γ_R=0.25	Γ_R=0.33	Γ_R=0.50	Γ_R=1.0
Logistic	$1 - \frac{1}{1+e^y}$	0.037, 0.25	0.051, 0.33	0.083, 0.50	0.083, 0.917
Log-log	$1 - e^{-e^y}$	0.037, 0.25	0.050, 0.33	0.082, 0.50	0.231, 0.930
D. Exp.	$\frac{1}{2}e^y, \quad y < 0$ $1 - \frac{1}{2}e^{-y}, \quad y \geq 0$	0.036, 0.25	0.049, 0.33	0.079, 0.50	0.080, 0.920
Probit	$\int_{-\infty}^{y} (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}} dt$	0.028, 0.25	0.031, 0.33	0.058, 0.50	0.127, 0.873

6. Local optimality and nonlinear models. While Sections 2-4 provide an interesting theoretical perspective on designs for estimating the MTD in a phase I clinical trial, as we have seen, the optimal design for nonlinear problems depends on the unknown parameters. In this case, we have derived optimal designs in terms of the quantiles of the unknown distribution with a restriction placed on the design space

which also corresponds to a quantile of the unknown distribution. These designs are locally optimal in the sense that one can only implement these designs at a selected prior estimate of the unknown parameters. Another approach is to use a Bayesian optimality criterion, which imposes a prior distribution on the parameters and leads to optimal designs on known support points. Prior information is often available in the context of phase I clinical trials, either because the clinicians have a deep understanding of the biological activity of the drug, or there is some information from animal studies. Yet another approach is advocated by Wu (1988), by putting the local optimality results in the context of an adaptive design. The experiment can be started at some preliminary design, and as data accrue, future experiments can be performed at the best point for estimating the parameters. Since, in our findings, local c-optimality considerations will almost always lead to a one-point design for phase I clinical trials whose goal is maximum likelihood estimation of the MTD, sequentially putting points closer to the MTD, or sequentially targeting the MTD, seems like a reasonable approximation to optimality. Ford, Torsney, and Wu (1992) then suggest that the locally optimal design is useful as a reference point in studies of the efficiency of the sequential design; i.e., comparing the relative efficiency of the estimators under the sequential and locally optimal scheme.

There are several proposed methods for targeting a quantile. Stochastic approximation [Robbins and Monro (1951)] and its modifications were proposed to sequentially converge to a target quantile. These are discussed in Morgan (1992, Chapter 8). The very last dose selected is taken as the estimate of μ; as it is a nonparametric procedure, maximum likelihood estimation of μ is not discussed. The Robbins and Monro procedure does not operate on a discrete design lattice, and as such is used in engineering applications where one has a continuous measure on a continuous design space. However, Wetherill and Glazebrook (1986, Chapter 10) give examples of the difficulty of convergence to μ for very large (in our case, very small) values of Γ. Wu (1985) sequentially fits logistic models and updates the maximum likelihood estimator of μ after each patient's response and experiments at the new estimate. McLeish and Tosh (1990) propose a similar technique to sequentially estimate the c- and D-optimal design, and is applicable for sequentially targeting two-point as well as one-point designs. The rather stringent requirements for the existence of maximum likelihood estimators hinder the use of these sequential maximum likelihood methods in small samples.

It is not possible to obtain a sequential design that will ensure that no patients are assigned above μ_R. So one possible design criterion is to compare the efficiency of estimating the target quantile to the best restricted c- and D- optimal designs, but also to analyze the expected number of patients assigned above μ_R.

7. New sequential designs for phase I clinical trials The local optimality criterion we have used may not be appropriate for many phase I clinical trials performed today. First, the optimality criterion is the minimization (in some sense) of an asymptotic approximation to the variance of the maximum likelihood estimator. Most phase I clinical trials recruit small numbers of patients, often less than 25. One would want ensure that the design chosen is optimal for small to moderate sample sizes.

Also, nonparametric estimators of the MTD have been proposed in conjunction with sequential designs that have less variability than the maximum likelihood estimator [see Durham, Flournoy, and Rosenberger (1997)]. Nonparametric estimators may be more attractive since the underlying form of dose-response curve is often unknown.

Secondly, many phase I clinical trials experiment on a finite lattice of dose levels, so that $\Omega_x = \{x_1, ..., x_K\}$ and μ may be spanned by two adjacent dose levels. These doses may be prepackaged and some compounds may not be reducible across the entire spectrum of dose levels. Hence the support points of the c- and D-optimal designs may not even be contained in Ω_x. The optimal selection of a discrete set of doses for experimentation has been discussed elsewhere [e.g., Wong and Lachenbruch (1996)].

Recently, two procedures have been proposed that have promise in phase I clinical trials. The first is the continual reassessment method [O'Quigley, Pepe, and Fisher (1990)]. This is a sequential Bayesian procedure, in the spirit of stochastic approximation. A one-parameter logistic prior is assumed, and after each patient's response, Bayes' formula is used to estimate the dose level corresponding to Γ. Experimentation is performed at the dose level closest to the estimate, so that it operates approximately on a finite lattice. The last dose administered or the maximum likelihood estimator can be used as the estimator of μ. The procedure appears to be robust to distributional assumptions, but there are conditions under which the estimators will be inconsistent [Shen and O'Quigley (1996)].

A second new procedure is random walk rules [Durham and Flournoy (1994)], a variant on up-and-down designs for targeting extreme percentiles. Random walk rules are up-and-down rules, where the patients are sequentially assigned to the next higher, lower, or same dose level as the previous patient, based on the previous patients response and some probability mechanism, say a biased coin. The bias of the coin is determined by Γ, and is selected so that the selected design points will form a unimodal distribution around μ. The design operates on a finite lattice and is nonparametric. Rosenberger, Flournoy, and Durham (1997) show that maximum likelihood estimators have the usual asymptotic properties. Unlike other designs, the finite distribution theory is completely workable [Durham, Flournoy, and Montazer-Haghighi (1995); Durham, Flournoy, and Rosenberger (1997)]. Simulation results on the distribution of design points for the continual reassessment method were nearly identical to exact distributional results for the random walk rules.

8. Conclusion. Sequential designs attempt to target the true MTD, and hence attempt to approximate c-optimality. While these designs do not ensure that all patients are assigned within a restricted dose space, assignment to extremely toxic dose levels can be minimized [Durham, Flournoy, and Rosenberger (1997)]. It will be interesting to compare the variances of estimators from these sequential schemes to the variances obtained from the c and D-optimal designs given in this paper.

We conclude by stating that the correct and efficient estimation of the MTD for further experimentation phases is critical for our public health. Formal design considerations, such as optimality and development of sequential designs to approximate optimality should be considered when planning such a study.

Acknowledgments. The authors thank Professor Herman Chernoff for pointing out the applicability of Elfving's method to this problem and Professor Weng Kee Wong for helpful discussions.

REFERENCES

CHERNOFF, H. (1972). *Sequential Analysis and Optimal Design.* SIAM, Philadelphia.

DETTE, H. (1993a). Elfving's theorem for D-optimality. *Ann. Statist.* **21** 753-766.

DETTE, H. (1993b). A new interpretation of optimality for E-optimal designs in linear regression models. *Metrika* **40** 37-50.

DETTE, H. AND HAINES, L. (1994). E-optimal designs for linear and nonlinear models with two parameters. *Biometrika* **81** 739-754.

DETTE, H., HEILIGERS, B., AND STUDDEN, W. J. (1995). Minimax designs in linear regression models. *Ann. Statist.* **23** 30-40.

DETTE, H. AND STUDDEN, W. J. (1993). Geometry of E-optimality. *Ann. Statist.* **21** 416-433.

DURHAM, S. D. AND FLOURNOY, N. (1994). Random walks for quantile estimation. In *Statistical Decision Theory and Related Topics V* (S. S. Gupta, and J. O. Berger, eds.), 467-476. Springer, New York.

DURHAM, S. D., FLOURNOY, N., AND MONTAZER-HAGHIGHI, A. A. (1995). Up-and-down designs II: exact treatment moments. In *Adaptive Designs* (N. Flournoy and W. F. Rosenberger, eds.), 158-178. Institute of Mathematical Statistics, Hayward.

DURHAM, S. D., FLOURNOY, N., AND ROSENBERGER, W. F. (1997). A random walk rule for phase I clinical trials. *Biometrics* **53** 745-760.

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255-262.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

FORD, I., TORSNEY, B., AND WU, C. F. J. (1992). The use of canonical form in construction of locally optimal designs for nonlinear problems. *J. Roy. Statist. Soc. B* **54** 569-583.

KIEFER, J. AND WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Can. J. Math.* **12** 363-366.

MATS, V. A. (1997). *Design and Likelihood-Based Estimation for Binary Response Experiments Under Ethical Constraints, With Application to Phase I Clinical Trials.* University of Maryland Graduate School, Baltimore (doctoral thesis).

MCLEISH, D. L. AND TOSH, D. H. (1990). Sequential designs in bioassay. *Biometrics* **46** 102-116.

MORGAN, B. J. T. (1972). *Analysis of Quantal Response Data.* Chapman and Hall, London.

O'QUIGLEY, J., PEPE, M., AND FISHER, L. (1990). Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* **46** 33-48.

ROBBINS, H. AND MUNRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **29** 400-407.

ROSENBERGER, W. F., FLOURNOY, N., AND DURHAM, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *J. Statist. Plann. Inf.* **60** 69-76.

SHEN, L. Z. AND O'QUIGLEY, J. (1996). Consistency of continual reassessment method under model misspecification. *Biometrika* **83** 395-405.

SILVEY, S. D. (1980). *Optimal Design.* Chapman and Hall, London.

SITTER, R. R. AND FAINARU, I. (1997). Optimal designs for the logit and probit models for binary data. *Can. J. Statist.* **25** 175-190.

SITTER, R. R. AND WU, C. F. J. (1993a). Optimal designs in binary response experiments: Fieller, D, and A criteria. *Scand. J. Statist.* **20** 329-342.

SITTER, R. R. AND WU, C. F. J. (1993b). On the accuracy of the Fieller intervals for binary response data. *J. Am. Statist. Assoc.* **88** 1021-1025.

WETHERILL, G. B. AND GLAZEBROOK, K. D. (1986). *Sequential Methods in Statistics.* Chapman and Hall, London.

WONG, W. K. AND LACHENBRUCH, P. A. (1996). Tutorial in biostatistics: designing studies for dose response. *Statist. Med.* **15** 343-359.

WU, C. F. J. (1980). Efficient sequential designs with binary data. *J. Am. Statist. Assoc.* **85** 156-162.

VLADIMIR A. MATS
BECTON DICKINSON AND COMPANY
1 BECTON DRIVE
FRANKLIN LAKES, NJ 07417-1880 U.S.A.

WILLIAM F. ROSENBERGER
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MARYLAND, BALTIMORE COUNTY
1000 HILLTOP CIRCLE
BALTIMORE, MD 21250 U.S.A.

NANCY FLOURNOY
DEPARTMENT OF MATHEMATICS AND STATISTICS
AMERICAN UNIVERSITY
WASHINGTON, DC 20016-8050 U.S.A.