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

OPTIMUM EXPERIMENTAL DESIGNS FOR CHEMICAL KINETICS AND CLINICAL TRIALS

BY ANTHONY C. ATKINSON

London School of Economics

Optimum designs are found for estimating the order, as well as the rate, of chemical reactions. Compound design criteria provide designs easily specified by experimenters. A second application is to the design of sequential clinical trials when prognostic factors need to be considered. Comparisons are given of several design strategies, including some of "biased-coin" type.

1. Introduction. The paper describes two applications of the methods of optimum experimental design. The general theory is described in Section 2 with an emphasis on compound D_A-optimality. Models arising in chemical kinetics are described in Section 3: the emphasis is on designs for determining the *orders*, rather than just the rates, of reactions. Compound criteria provide flexible families of designs easily specified by experimenters. Section 4 briefly mentions design construction when the kinetic differential equations do not have analytical solutions. The last section uses a randomized form of D_A-optimality to provide "biased-coin" designs for treatment allocation in sequential clinical trials with prognostic factors. Asymptotic comparisons with other procedures are augmented by simulations for finite sample sizes.

2. Optimum design theory.

2.1. *Experimental designs.* The experiment consists of measuring the response y at conditions specified by the values of k factors or explanatory variables represented by the $k \times 1$ vector x. One experimental run yields one observation y_i and the experimental design is a list of the n sets of conditions $x_i, i = 1, \ldots, n$, not necessarily distinct, at which measurements are to be made. We assume that the response, y, for a multiple regression model has mean $E(Y) = F\beta$ and dispersion matrix $D(y) = \sigma^2 I, \sigma^2 > 0$ with I the $n \times n$ identity matrix. The ith row of the $n \times p$ extended design matrix F is

Received September 1997.
AMS 1991 subject classifications. 62K05.
Key words and phrases. Biased-coin design; compound design criterion; D_A-optimum design; order of reaction; parameter sensitivities.

the $1 \times p$ vector $f^T(x_i)$, a function, usually a low order polynomial, of the explanatory variables.

We use a general definition of an experimental design in which ξ is a continuous design specifying a set of m distinct points in a design region $\mathcal{X}$ and the proportions, w_i, of observations taken at these points

$$\xi = \left\{ \begin{array}{c} x_1, \ldots, x_m \\ w_1, \ldots, w_m \end{array} \right\}.$$

The x_i are the points of support of ξ and w_i the design weights. In practice, when only n observations can be taken, an exact design will be required. Often the optimum exact design is approximated by a design with the number of trials at x_i the integer closest to nw_i. The information matrix of a design ξ for the p parameters β is thus given by

$$M(\xi) = F^T W F, \tag{2.1}$$

where

$$W = \text{diag}\{w_1, \ldots, w_m\}$$

and now F is $m \times p$.

2.2. *Compound D-optimality and D_s-optimality.* D-optimum designs maximize the logarithm of the determinant of the information matrix, $\log |M(\xi)|$ or, equivalently, minimize the asymptotic generalized variance of the parameter estimators. For the various applications in this paper we need the extension to generalized D-optimality in which the function to be minimized is

$$\sum_{j=1}^{H} \alpha_j \log |A_j^T M_j^{-1}(\xi) A_j|. \tag{2.2}$$

The criterion permits designs for H different models which may be fitted to the data, for the jth of which the information matrix is $M_j(\xi)$. The matrix A_j defines s_j linear combinations of the p_j parameters in model j which are of experimental importance and the non-negative weights α_j express the relative importance of the different aspects of the design. Examples of compound D-optimum designs for linear models are given by Atkinson and Donev (1992, Chapter 21), and by Cook and Wong (1994). Here they are used in Section 3.3 for nonlinear models.

The General Equivalence Theorem of Kiefer and Wolfowitz (1960) makes it possible to check the optimality of a candidate continuous D-optimum design. With the standardized variance of the prediction at x from one model defined by

$$d(x, \xi) = f^T(x) M^{-1}(\xi) f(x), \tag{2.3}$$

the relevant part of the Equivalence Theorem states that, for the optimum design, ξ^*, the maximum value of $d(x, \xi^*)$ over the design region, $\mathcal{X}$, is p, the number of parameters in the model, and further that this maximum value is attained at the support points ξ^*. The theorem also provides a basis for algorithmic construction of D-optimum designs [Wynn (1970)].

For the compound criterion in (2.2) the equivalence theorem becomes

$$(2.4)\quad d(x,\xi^*) = \sum_{j=1}^{H} \alpha_j f_{(j)}^T(x) M_j^{-1}(\xi^*) A_j \{A_j^T M_j^{-1}(\xi^*) A_j\}^{-1} A_j^T M_j^{-1}(\xi^*) f_{(j)}(x)$$
$$\leq \sum_{j=1}^{H} \alpha_j s_j.$$

If only an s subset of the parameters β_2 is of interest, let the parameters be partitioned as $\beta^T = (\beta_1^T \quad \beta_2^T)$ with the information matrix $M(\xi)$ partitioned so that the information for β_1 is $M_{11}(\xi)$. Then the D_s-optimum design for β_2 minimizes the special case of (2.2) with $H = 1$ and $A^T = (0 \quad I_s)$. If $M^{-1}(\xi^*)$ is partitioned so that the covariance matrix for $\hat{\beta}_2$ is $M^{22}(\xi^*)$, it follows that the design minimizes $\log|M^{22}(\xi^*)|$, which is identical to maximizing $\log\{|M(\xi)|/|M_{11}(\xi)|\}$. The Equivalence Theorem for D_s-optimum designs states that, for the optimum measure ξ^*

$$(2.5)\qquad d(x,\xi) = f^T(x) M^{-1}(\xi) f(x) - f_{(1)}^T(x) M_{11}^{-1}(\xi) f_{(1)}(x) \leq s,$$

where $f_{(1)}^T(x)$ relates to the information matrix $M_{11}(\xi)$.

2.3. *Nonlinear models.* In the next section we are concerned with nonlinear examples in which the single explanatory variable is the time t at which the measurement is taken. The response for the nonlinear regression model has mean $E(Y) = \eta(t,\psi)$, with dispersion matrix as before. Here η is a function nonlinear in at least one of the p parameters ψ and t is defined on the time interval $\mathcal{T}$. The information matrix of a design ξ for the parameters ψ is given by

$$M(\xi,\psi) = F^T W F,$$

where

$$F = \begin{pmatrix} f^T(t_1,\psi) \\ \vdots \\ f^T(t_p,\psi) \end{pmatrix},$$

and the vector $f^T(t_i,\psi)$ has j-th element

$$f_j(t_i,\psi) = \frac{\partial \eta(t_i,\psi)}{\partial \psi_j}, \quad \text{for} \quad j = 1,\ldots,m.$$

The information matrix, although formally similar to that for the linear model, now depends on the unknown parameter, ψ. A natural way of accommodating the obvious problems which follow from this dependence, is to adopt a best guess for the parameters, say ψ^o, and to consider designs which maximize an appropriate function of $M(\xi,\psi)$ evaluated at $\psi = \psi^o$ [Chernoff (1953)]. Such designs are termed locally optimum. An alternative, if prior information on the distribution of ψ is available, is to incorporate the information into the design criterion.

2.4. *Bayesian designs.* In Bayesian D-optimality, we find the maximum of the expectation, with respect to the prior parameter distribution $p(\psi)$, of the logarithm of the determinant of the information matrix, that is

$$E_\psi \log |M(\xi, \psi)| = \int \log |M(\xi, \psi)| p(\psi) d\psi, \tag{2.6}$$

is maximized. A discussion of this form of Bayesian experimental design is given by Chaloner and Larntz (1989). Chaloner and Verdinelli (1995) give a full review. Several examples are given by Atkinson and Donev (1992, Chapter 19).

The mathematical relationship between this Bayesian criterion and the criterion for compound D-optimality (2.2) becomes clear if the integral in (2.6) is calculated numerically, when the criterion reduces to a weighted sum of the logarithms of determinants for the various values of ψ.

An Equivalence Theorem also applies here. For the D-optimum design ξ^* we must have

$$d(t, \xi^*) = \int f^T(t, \psi) M^{-1}(\xi^*, \psi) f(t, \psi) p(\psi) d\psi \leq p, \tag{2.7}$$

where p is still the number of parameters in the model. The maxima of (2.7) are once more at the points of support of the design, a feature which is useful in constructing and checking optimum designs.

The locally optimum designs of the previous subsection for nonlinear models almost invariably put trials at p design points. A feature of Bayesian designs is that, as prior uncertainty about ψ increases, the number of design points also increases.

3. A nonlinear model: general decay.

3.1. *Models from chemical kinetics.* Locally D-optimum designs for nonlinear models were introduced by Box and Lucas (1959). The models came from chemical kinetics. Although the chemistry of the reactions was not stressed, it was assumed that the orders of the reactions were known and that the rates of the reactions were the parameters of interest. Atkinson and Bogacka (1997) extended this work to include the determination of orders of reaction, using the compound optimum design criterion described in Section 2.2 to provide methods for aiding experimenters in balancing different objectives of the experiment. A summary of the problem follows.

Reviews of optimum experimental design for nonlinear models in general include Ford, Titterington, and Kitsos (1989), Ford, Torsney, and Wu (1992) and Atkinson and Haines (1996). For models arising specifically in chemical kinetics, both Box and Lucas (1959) and Atkinson and Donev (1992) find locally D-optimum designs for the nonlinear response model resulting from first-order decay

$$A \xrightarrow{\theta} B$$

in which the concentration of chemical A at time t is given by the nonlinear function

$$[A] = \eta(t, \theta) = e^{-\theta t} \qquad (\theta, t \geq 0), \tag{3.1}$$

if it is assumed that the initial concentration of A is 1. The model comes from solution of the differential equation

$$\frac{d[A]}{dt} = -\theta[A], \tag{3.2}$$

which assumes first-order kinetics. More generally, the order of the reaction may be represented by the parameter λ, when the differential equation becomes

$$\frac{d[A]}{dt} = -\theta[A]^{\lambda}, \tag{3.3}$$

so that, for (3.1) and (3.2) $\lambda = 1$.

In a simple case such as (3.3) the more general differential equation can also be integrated to yield a new model in which the expected value of concentration at time t is

$$[A] = \eta(t, \psi) = \{1 - (1 - \lambda)\theta t\}^{1/(1-\lambda)} \quad (\lambda, \theta, t \geq 0; \lambda \neq 1), \tag{3.4}$$

where $\psi = [\theta, \lambda]^T$. As $\lambda \to 1$, (3.4) reduces to (3.1). For $\lambda < 1, [A] = 0$ for $t \geq 1/\{(1 - \lambda)\theta\}$. However, for many kinetic models, it is only possible to obtain explicit expressions like (3.4) for a few values of λ. The implications for experimental design of the lack of analytical solutions are discussed in Section 4.

Calculation of locally D- or D_s-optimum designs for the general decay model (3.4) requires values of the derivatives

$$\begin{aligned} f_1(t, \psi) &= \partial\eta(t, \psi)/\partial\theta \qquad \text{and} \\ f_2(t, \psi) &= \partial\eta(t, \psi)/\partial\lambda, \end{aligned} \tag{3.5}$$

which can be found analytically to be:

$$f_1(t, \psi) = -t\{1 - (1 - \lambda)\theta t\}^{\lambda/(1-\lambda)} \tag{3.6}$$

and

$$\begin{aligned} f_2(t, \psi) = \{1/(1 - \lambda)^2\}[\log\{1 - (1 - \lambda)\theta t\} + (1 - \lambda)\theta t/\{1 - (1 - \lambda)\theta t\}] \\ \times\{1 - (1 - \lambda)\theta t\}^{1/(1-\lambda)}, \text{ for } t \leq 1/\{(1 - \lambda)\theta\}. \end{aligned} \tag{3.7}$$

The case of exponential decay, $\lambda = 1$, requires special attention. Differentiation of the response (3.1) yields

$$f_1(t, \psi) = -t\exp(-\theta t), \tag{3.8}$$

which can also be found as the limit of (3.6) as $\lambda \to 1$. The same limiting operation on (3.7) yields

$$\lim_{\lambda \to 1} f_2(t, \psi) = \frac{1}{2}(\theta t)^2 \exp(-\theta t). \tag{3.9}$$

Although (3.8) is found directly with $\lambda = 1$, calculation of the derivative (3.9) requires the derivative (3.7) for general λ. So, even to find designs for testing whether $\lambda = 1$, we need (3.7).

3.2. *Some designs.*

3.2.1. *D-optimum designs for the rate θ.* We partition the parameter as $\psi = (\theta \quad \lambda)^T$, with $M_{11}(\xi, \psi) = M_{11}(\xi, \theta)$ the information for θ. If the order of reaction, λ, is known the locally D-optimum design for θ puts all the trials where $\log|M_{11}(\xi, \theta^o)|$ is a maximum. In the case of one-point designs $|M_{11}(\xi, \theta^o)| = |f_1^2(\xi, \theta^o)|$, so the absolute value $|f_1(\xi, \theta^o)|$ is maximized. Differentiation of (3.6) with respect to t, followed by equating the derivative to zero, shows that the optimum design concentrates all trials at $t = 1/\theta^o$, a result which is well known for the special case $\lambda = 1$. Atkinson and Bogacka (1997) provide plots of the variance (2.3) which demonstrate that the one-point designs are optimum.

3.2.2. *D-optimum designs for both parameters.* If both parameters are of interest, the D-optimum design is found which maximizes $\log|M(\xi, \psi^o)|$. For the two-parameter model the D-optimum design has two points of support, with weight $1/2$ at each, the times of the design points depending on the prior values of the parameters. For the D-optimum design for θ of the previous subsection, $t = 2$ when $\theta^o = 0.5$. The D-optimum design for both parameters has time points either side of this value. As λ^o increases from 0.5, t_1 remains sensibly unchanged, decreasing from 1.27 to 1.24 at $\lambda^o = 2$. Over the same range t_2 increases rapidly from 3.09 reaching 11.03 for $\lambda^o = 2$. This emphasis on high values of t when λ^o is large is in line with the increase in response at large t as λ increases.

3.2.3. *D_s-optimum designs for the order of reaction λ.* If the main purpose of the experimenter is to determine the order of the reaction, with the actual value of the rate constant θ of secondary importance, the D_s-optimum design for λ is appropriate. The required design maximizes $\log(|M(\xi, \psi)|/M_{11}(\xi, \theta))$.

Like the D-optimum designs above, the D_s-optimum designs have two points of support. Unlike them, however, the design weights are not equal: w_2, the weight for the upper time point, ranges from 0.570 when $\lambda^o = 0.5$ to 0.693 when $\lambda^o = 2$. Over the same set of values of λ, t_1 decreases from 0.980 to 0.849 while t_2 increases from 3.33 to 17.38. These designs require more extreme values of time than do the D-optimum designs.

3.3. *Compound optimum designs.* Three different designs were described in the previous subsection: the D-optimum design for θ with λ known, maximizing $\log|M_{11}(\xi, \theta^o)|$, the D-optimum design for θ and λ maximizing $\log|M(\xi, \psi^o)|$ and the D_s-optimum design for the order λ maximizing $\log\{|M(\xi, \psi^o)|/|M_{11}(\xi, \theta^o)|\}$. We now find compound optimum designs which provide a practically useful balance between the objectives of these individual designs.

For determining the order of the general decay model, the most extreme designs are the D-optimum design for θ alone, which is completely uninformative about λ, and the D_s-optimum design for λ, which is most informative. Combining these two as in

(2.2) yields the compound design criterion

$$\Phi(\xi, \psi) = (1 - \alpha) \log M_{11}(\xi, \theta) + \alpha \log\{|M(\xi, \psi)|/M_{11}(\xi, \theta)\}. \tag{3.10}$$

In (3.10) α, $(0 \leq \alpha \leq 1)$ expresses the experimenter's relative interest in determination of the order of the reaction, with $\alpha = 1$ corresponding to interest solely in order determination. When $\alpha = 0.5$, the criterion reduces to a scalar multiple of that for D-optimality. Since multiplication of the design criterion does not affect the optimum design, we have a precise interpretation of D-optimality as intermediate between the other two criteria.

Despite the precise meaning to be given to $\alpha = 0.5$, it is unlikely that an experimenter will be able to specify a numerical value of α. To obtain a suitable design we find, by numerical optimization, the designs maximizing the compound criterion (3.10) for a series of values of α. For each we calculate the efficiencies of the design relative to the D-optimum design for the estimation of rate, the D_S-optimum design for estimation of order and relative to the D-optimum design for both. A plot of these efficiencies against α makes it possible to choose a design with a balance of efficiencies for all aspects of the problem.

Plots of these efficiencies are given in Figure 1. The compound designs themselves are not given. However for $0.5 \leq \alpha \leq 1$ the designs consist of experiments at two values of time with unequal weights, both the weights and times being between those for the D- and D_S-optimum designs of Sections 3.2.2 and 3.2.3. As $\alpha \to 0$, the design problem approaches that of D-optimality for θ when λ is known, yielding the one-point design at $t = 2$ of Section 3.2.1. The compound design reflects this since $t_1 \to 2$, as $\alpha \to 0$, while w_2 decreases towards zero.

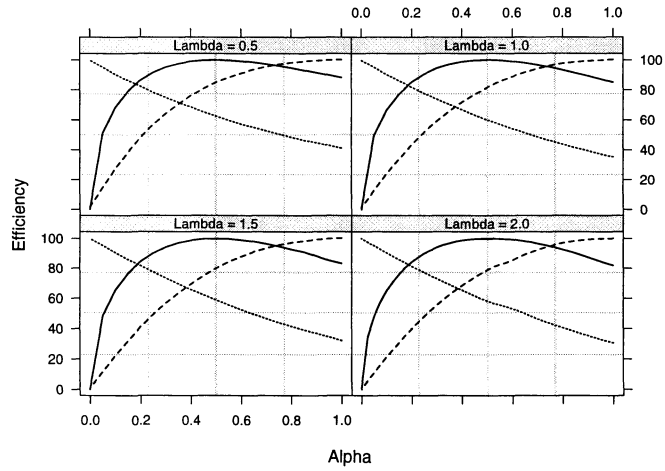

FIG. 1. *Efficiencies of compound optimum designs: D-optimum, E_D, continuous line; D_s-optimum for λ, E_λ, dashed line; D-optimum for θ, E_θ, dotted line.*

The efficiencies are calculated using the optimum designs for the particular aspect of interest. Let the optimum compound design be ξ_c^* and the D-optimum design for

estimating θ be ξ_θ^*. Then the percentage efficiency of the compound design if only θ is of interest is

$$E_\theta = 100 M_{11}(\xi_c^*, \theta^o)/M_{11}(\xi_\theta^*, \theta^o). \tag{3.11}$$

Likewise, if the D_s-optimum design for estimating λ is ξ_λ^*, the relevant efficiency is

$$E_\lambda = 100 \frac{|M(\xi_c^*, \psi^o)|/M_{11}(\xi_c^*, \theta^o)}{|M(\xi_\lambda^*, \psi^o)|/M_{11}(\xi_\lambda^*, \theta^o)}. \tag{3.12}$$

However, if the D-optimum design for θ and λ is ξ_D^*, the efficiency is

$$E_D = 100\{|M(\xi_c^*, \psi^o)|/|M(\xi_D^*, \psi^o)|\}^{1/2}, \tag{3.13}$$

the square root being required as the determinant is for a model with two parameters.

As an example, suppose $\lambda = 2$. The compound design for $\alpha = 0.5$ has, of course, $E_D = 100\%$. The other efficiencies are $E_\theta = 58.2\%$ and $E_\lambda = 79.0\%$. For $\alpha = 0.375$ the values are $E_D = 97.9\%$, $E_\theta = 67.2\%$ and $E_\lambda = 65.5\%$, high efficiencies for any way in which the data may be analyzed.

4. Other kinetic models. Atkinson and Bogacka (1997) give fuller details of the designs sketched above and also find Bayesian designs using the results of Section 2.4. An important extension of the general method is to models in which the kinetic differential equations have to be solved numerically. Analytical expressions are then not available for the derivatives of the function η with respect to the parameters θ and λ. In the chemical literature these derivatives are known as first-order sensitivities.

The method used here for determining the order of a kinetic model depends on embedding the kinetic equations in more general ones of unknown order. Unfortunately introduction of the general order parameter λ into kinetic equations renders an analytical solution most unlikely. As an example, for the two successive reactions

$$A \stackrel{\theta_1}{\rightarrow} B \stackrel{\theta_2}{\rightarrow} C, \tag{4.1}$$

in which the concentration of B is measured, if B is formed at rate $\theta_1[A]^{\lambda_1}$, given by (3.3), and itself reacts with order λ_2, the concentration of B at time t is governed by the differential equation

$$\frac{d[B]}{dt} = \theta_1[A]^{\lambda_1} - \theta_2[B]^{\lambda_2}. \tag{4.2}$$

Analytical expressions for $[B]$ can only be found for a few values of λ_1 and λ_2, the case of both parameters equal to one being given by Box and Lucas (1959). Numerical values of the sensitivities for such models can be found using what Valko and Vajda (1984) call the "direct method" in which differentiation of the differential equation defining the response, here (4.2), with respect to the parameters yields a set of simultaneous differential equations for the sensitivities. Solution of the equations provides a grid of numerical values of the sensitivities, which have to be interpolated in the construction of the design. Similar methods have been employed in the chemical literature

to find designs for the rate constants in complicated models, for example Nathanson and Saidel (1985), where the sets of equations to be solved are linear. Although the theory of the construction and properties of these designs is similar to that for the general decay model, which is the subject of the previous section, the computational requirement is much heavier.

5. Clinical trials.

5.1. *General design considerations.* The second use of the theory of Section 2 is in the sequential design of clinical trials. In order to achieve some balance as well as randomness Atkinson (1982) suggested a randomized form of the sequential construction of D_A-optimum designs maximizing (2.2) with $H = 1$. The method, together with related procedures, has been thoroughly investigated by Smith (1984a), Smith (1984b), Wei, Smythe, and Smith (1986) and by Burman (1996). Here these comparisons are summarized and extended.

Patients arrive sequentially and are to be given one of t treatments. In order to avoid the suspicion of conscious or unconscious bias, treatments should be allocated at random. But, because it is not known when the trial will terminate, there needs to be a balance of the number of patients receiving each treatment. If there are just two treatments and no measurements of prognostic factors on the patients, one possibility is the biased-coin design of Efron (1971) in which the under-represented treatment is assigned with probability 2/3. To extend this idea to t treatments and prognostic factors requires modelling the response.

Let

$$E(Y) = G\omega = E\alpha + Z\gamma, \tag{5.1}$$

where E is the $n \times t$ matrix of indicator variables for the treatments with one non-zero entry per row, and Z is the $n \times (q-1)$ matrix of prognostic factors, including interactions and other terms if required. The γ are nuisance parameters. However interest is only in contrasts between the α, not in the mean level of response which is an additional nuisance parameter, making q in all. For example, if there are only two treatments, experiments should be designed to estimate $\alpha_1 - \alpha_2$ with minimum variance. With more than two treatments a set of $t-1$ contrasts orthogonal to the mean is required, for example

$$L^T = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ & & \vdots & & \\ 1 & 0 & 0 & \dots & -1 \end{pmatrix}. \tag{5.2}$$

If a D_A-optimum design is found using these contrasts, the exact form is unimportant, provided the contrasts span the $t-1$ dimensional space orthogonal to the overall mean. Since the γ in (5.1) are nuisance parameters, the contrasts in (5.2) need augmenting by a matrix of zeroes

(5.3) $$A^T = (L^T \quad 0)$$

to reflect the required interest in the parameters.

In the iterative construction of D_A-optimum designs the next trial would be added where the variance (2.4) was a maximum over the design region. For clinical trials the design region is the set of t treatments which can be allocated. But the design will depend not only on the previous allocations and the matrix Z, but also on z_{n+1}, the vector of prognostic factors for the new patient. To emphasize this dependence the variances can be written $d_A(j, \xi_n, z_{n+1})$, $j = 1, \ldots, t$. Asymptotically all treatments will be allocated equally often and the variances will tend to equality. To provide a randomized form of this iterative construction, Atkinson (1982) suggests allocating treatment j with probability

(5.4) $$\pi_A(j|z_{n+1}) = \frac{d_A(j, \xi_n, z_{n+1})}{\sum_{j=1}^{t} d_A(j, \xi_n, z_{n+1})}.$$

In (5.4) the variances $d_A(.)$ could be replaced by any monotone function $\phi\{d_A(.)\}$. We do not explore this, but compare the performance of (5.4) with other suggestions in the literature. The emphasis in the comparisons is on the variance of parameter estimates.

5.2. *Two treatments.* With two treatments the parameter of interest is $\Delta = \alpha_1 - \alpha_2$. The model (5.1) can be written

(5.5) $$E(Y) = a\Delta + 1\beta_o + Z\gamma = a\Delta + F\beta,$$

where a is the $n \times 1$ vector of allocations with elements +1 and -1, and the constant term and covariates are included in the $n \times q$ matrix F. Then

(5.6) $$\mathrm{var}(\hat{\Delta}) = \sigma^2\{a^T a - a^T F(F^T F)^{-1} F^T a\}^{-1}.$$

In (5.6) it is meaningful to let $b = F^T a$, a "balance" vector which is identically zero when all covariates are balanced across all treatments. Also $a^T a = n$, so that (5.6) can be written in the revealing form

(5.7) $$\mathrm{var}(\hat{\Delta}) = \frac{\sigma^2}{n - b^T(F^T F)^{-1} b} = \frac{\sigma^2}{n - \mathcal{L}_n},$$

where $\mathcal{L}_n$ is the loss after n trials. This important measure expresses the loss of information due to imbalance. If the design is exactly balanced, $\mathcal{L}_n$ is zero. Otherwise the loss of information is expressed in terms of number of patients. For the randomized designs studied here $\mathcal{L}_n$ is a random variable. The results of Smith (1984a) and of Smith (1984b) provide asymptotic values $\mathcal{L}_\infty$. As in Burman (1996), simulation is used here to study the progress of the loss towards its asymptotic value. In the initial stages of the trial imbalance may be relatively high and the loss $\mathcal{L}_n$ may be far from $\mathcal{L}_\infty$.

With one exception, all allocation rules considered here depend on the quantity

(5.8) $$R_{n+1} = f_{n+1}^T (F^T F)^{-1} b,$$

with $f_{n+1}^T = (1 \;\; z_{n+1}^T)$ and b the balance vector after n trials. The rules are expressed in terms of probabilities

$$\pi(1) = \text{prob}(a_{n+1} = 1 | R_{n+1} < 0).$$

DETERMINISTIC (SEQUENTIAL DESIGN CONSTRUCTION).

$$\pi_D(1) = 1.$$

The treatment with larger variance $d_A(j, \xi_n, z_{n+1})$ is always selected. Asymptotically there is no loss, $\mathcal{L}_\infty = 0$.

COMPLETELY RANDOMIZED.

$$\pi_R(1) = 0.5,$$

with $\mathcal{L}_\infty = q$, the number of nuisance parameters, including the constant.

These two rules represent the extremes. The loss of any other rule will be bounded by these values.

EFRON'S BIASED-COIN.

$$\pi_E(1) = 2/3,$$

with again $\mathcal{L}_\infty = 0$. Other values than 2/3 will give a different rate of convergence to $\mathcal{L}_\infty$ and a different probability that the clinician can guess correctly which treatment will be allocated next.

ATKINSON'S D_A-OPTIMALITY.

Results from the inverse of partitioned matrices show that allocation according to (5.4) can be rewritten as

$$\pi_A(1) = 0.5 - R_{n+1}/(1 + R_{n+1}^2),$$

with $\mathcal{L}_\infty = q/5$.

BALANCED COVARIATES.

This rule does not depend on R_{n+1}. The values of the $q-1$ covariates are dichotomised about their individual medians, giving 2^{q-1} possible cells in which the value of z_{n+1} could lie. The under-represented treatment in the cell is then allocated, the probability being 0.5 if the numbers of the two treatments are equal. A randomized version could have a biased coin within each cell. If q is not small, the large number of cells may be sparsely filled, leading to lack of balance over the margins of the table.

5.3. *Simulations.* Figure 2 shows the average results of one thousand simulations of sequential designs for n up to 200. There are four prognostic factors ($q = 5$) simulated independently from the standard normal distribution. The results support the theoretical conclusions above. For the deterministic allocation, D, the loss rapidly declines to almost zero, whereas for the completely random allocation, R, the value is near 5. The randomized D_A-optimum design, A, quickly achieves a loss close to $q/5 = 1$. Efron's rule, E, has a loss less than that of A for n greater than about 100, a loss which continues to decline as n increases. At 200 trials $\mathcal{L}$ for the covariate balanced strategy is about 1.63, the second largest value.

The designs for $q = 10$ in Figure 3 show similar features for rules R, A and D. Efron's biased-coin strategy, E, now does not become comparable in loss with A until n is near 200 - for small n and large q the probabilities given by (5.4) can be much greater than 2/3, reflecting large values of the variance $d_A(j, \xi_n, z_{n+1})$ for one of the treatments. However, the most striking difference between $q = 5$ and $q = 10$ is in the behaviour of the covariate balancing designs C. With $2^9 = 512$ cells to be filled, a large amount of imbalance is possible, reflected in the slow decline of the loss.

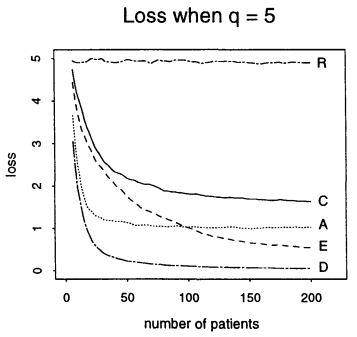

FIG. 2. *Loss $\mathcal{L}_n$ for five strategies for sequential allocation of treatments with $q = 5$: A, D_A-optimality; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin and R, Random. Means of 1,000 simulations.*

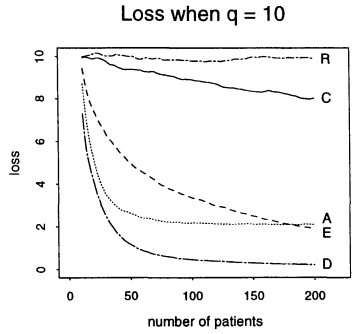

FIG. 3. *Loss $\mathcal{L}_n$ for five strategies for sequential allocation of treatments with $q = 10$: A, D_A-optimality; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin and R, Random. Means of 1,000 simulations.*

A general point about the simulations is that smoothed average values have been plotted. The curves, especially those for R and C, show evidence of random fluctuation. It would be interesting to study the change of the distribution of $\mathcal{L}_n$ with n , not just the expected value. A second point is that (5.6) shows that $\text{var}(\hat{\Delta})$ does not depend on the scaling of the prognostic factors z. A small simulation suggests that replacing normally distributed variables with those from a t distribution has little effect. The effect of correlation among the prognostic factors remains to be investigated.

5.4. *Three or more treatments.* With more than two treatments it is not possible to rewrite the model in a form such as (5.5) and so to define the loss in terms of the variance of the estimator of a single parameter. But an expression can still be found for the loss $\mathcal{L}_n$.

The D_A-optimum design for the model (5.1) requires minimization of

$$|A^T(G^TG)^{-1}A| = |L^T\{E^TE - E^TZ(Z^TZ)^{-1}Z^TE\}^{-1}L|, \tag{5.9}$$

similar in form to (5.6). For the optimum design, in which an equal number of patients is allocated to each treatment, and $E^TZ = 0$, so that there is balance over all prognostic factors,

$$|A^T(G^TG)^{-1}A| = t^t/(n^{t-1}).$$

The efficiency of any other design is then

$$\mathcal{E}_n = \left(\frac{t^t/(n^{t-1})}{|A^T(G^TG)^{-1}A|}\right)^{1/(t-1)},$$

so that the effective number of trials is $n\mathcal{E}_n$. The generalization of (5.7) is

$$\mathcal{L}_n = n(1 - \mathcal{E}_n).$$

The theoretical results referenced above cover designs for more than two treatments. The small sample properties of $\mathcal{L}_n$ can again be found by simulation.

5.5. *Discussion.* The focus in this section has been on the variance of parameter estimates. It is clear that the smaller the amount of randomization the smaller the loss $\mathcal{L}_n$, but the larger the chance of bias. However, if patients are algorithmically allocated a treatment in order of arrival, the opportunities for biasing the trial would seem to be extremely limited. If this is so, the reduction of variance is an appropriate design criterion.

Acknowledgments. I am very grateful to the Committee for the invitation to attend the conference and for financial support, to Bill Rosenberger for encouraging comments and to Weng Kee Wong for additional references.

REFERENCES

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67.

ATKINSON, A. C. AND B. BOGACKA (1997). Compound, D- and D_s-optimum designs for determining the order of a chemical reaction. *Technometrics* **39** 347–356.

ATKINSON, A. C. AND A. N. DONEV (1992). *Optimum Experimental Designs.* Oxford University Press, Oxford.

ATKINSON, A. C. AND L. M. HAINES (1996). Designs for nonlinear and generalized linear models. In *Handbook of Statistics, Vol. 13*, (S. Ghosh and C. R. Rao, eds.), 437–475. Elsevier, Amsterdam.

BOX, G. E. P. AND H. L. LUCAS (1959). Design of experiments in nonlinear situations. *Biometrika* **46** 77–90.

BURMAN, C.-F. (1996). *On Sequential Treatment Allocations in Clinical Trials.* Department of Mathematics, Göteborg (doctoral thesis).

CHALONER, K. AND K. LARNTZ (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inf.* **21** 191–208.

CHALONER, K. AND I. VERDINELLI (1995). Bayesian experimental design: a review. *Statist. Sci.* **10** 273–304.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. of Math. Statist.* **24** 586–602.

COOK, R. D. AND W. K. WONG (1994). On the equivalence between constrained and compound optimal designs. *J. Amer. Statist. Assoc.* **89** 687–692.

EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417.

FORD, I., D. M. TITTERINGTON, AND C. P. KITSOS (1989). Recent advances in nonlinear experimental design. *Technometrics* **31** 49–60.

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

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

NATHANSON, M. H. AND G. S. SAIDEL (1985). Multiple-objective criteria for optimal experimental design: application to ferrokinetics. *Amer. J. Physiology* **248** R378–R386.

SMITH, R. L. S. (1984a). Properties of biased coin designs in sequential clinical trials. *Ann. Statist.* **12** 1018—1034.

SMITH, R. L. S. (1984b). Sequential treatment allocation using biased coin designs. *J. Roy. Statist. Soc. B* **46** 519—543.

VALKO, P. AND S. VAJDA (1984). An extended ODE solver for sensitivity calculations. *Computers and Chemistry* **8** 255–271.

WEI, L. J., R. T. SMYTHE, AND R. L. S. SMITH (1986). K-treatment comparisons with restricted randomization rules in clinical trials. *Ann. Statist.* **14** 265—274.

WYNN, H. P. (1970). The sequential generation of *D*-optimal experimental designs. *Ann. Math. Statist.* **41** 1055–1064.

DEPARTMENT OF STATISTICS
LONDON SCHOOL OF ECONOMICS
LONDON WC2A 2AE
UNITED KINGDOM