The Construction of Optimal Designs for Dose-escalation Studies

Linda M. Haines and Allan E. Clark
Department of Statistical Sciences, University of Cape Town, Rondebosch 7700, South Africa.

Abstract:

Methods for the construction of A-,MV-,D- and E-optimal designs for dose-escalation studies are presented. Algebraic results proved elusive and explicit expressions for the requisite optimal designs are only given for a restricted class of traditional designs. Recourse to numerical procedures and heuristics is therefore made. Complete enumeration of all possible designs is discussed but is, as expected, highly computer intensive. Two exchange algorithms, one based on block exchanges and termed the Block Exchange Algorithm and the other a candidate-set-free algorithm based on individual exchanges and termed the Best Move Algorithm, are therefore introduced. Of these the latter is the most computationally effective. The methodology is illustrated by means of a range of carefully selected examples.

Keywords: Dose-escalation studies; A-, MV-, D- and E-optimal designs; Complete enumeration; Exchange Algorithms.

1. Introduction

Dose-escalation trials are commonly used in first-in-man studies of a new drug. However many issues relating to the planning and execution of such trials were brought into stark relief by the Te Genero disaster. The Royal Statistical Society working party, drawn together to examine the statistical issues relating to the Te Genero trial, was particularly critical of the design used in the trial, that is of the allocation of individuals to cohorts and to treatments within cohorts (Senn et al, 2007). In a subsequent paper Bailey (2009) examined a broad range of issues relating to the design of dose-escalation trials, formulated some key desiderata in this regard and recommended a suite of easy-to-use designs.

The aim of the present study is to devise efficient methods for the construction of optimal designs for dose-escalation studies and to critically appraise the form and usefulness of such designs. The paper is organized as follows. The necessary notation and the criteria of interest are introduced in Section 2. The ensuing sections then relate to methods of construction of the requisite optimal designs, with an algebraic approach discussed in Section 3, complete enumeration of all possible designs in Section 4 and the use of heuristic algorithms based on exchange procedures in Section 5. Illustrative examples are provided in each section. Some broad conclusions and pointers for future research are given in Section 6.

2. Preliminaries

Consider a dose-escalation study comprising n doses and a placebo, and thus $t = n + 1$ treatments, and c cohorts. Following Bailey (2009), the placebo is labelled 0 and increasing doses are labelled from 1 to n. The number of cohorts c is usually taken to be n which specifies a standard design or $n + 1$ which specifies an extended design. Dose-escalation is incorporated into the trial by ensuring that for cohort k, where $k = 1, \ldots, n$, at least one individual is allocated to dose k and further individuals are allocated to the treatments $0, \ldots, k$ but not to doses $k + 1, \ldots, n$. For extended designs there is no restriction on the allocation of individuals to treatments in cohort $n+1$. The setting can thus be summarized by the $c \times (n + 1)$ incidence matrix S with entry s_{ki} equal to the number of individuals allocated to treatment i in cohort k where $i = 0, 1, \ldots, n$ and $k = 1, \ldots, c$. For any given allocation the rows of S sum to $\underline{m} = (m_1, \ldots, m_c)$, where m_k represents the number of individuals allocated to cohort k, $k = 1, \ldots, c$, and the columns of S sum to $\underline{r} = (r_0, \ldots, r_n)$, where r_i is the number of times treatment i is replicated, $i = 0, \ldots, n$. In addition the constraints $s_{kk} > 0$, $s_{ki} \geq 0$ for $i \leq k-1$ and $s_{ki} = 0$ for $i > k$ where $k = 1, \ldots, n$ necessarily hold. Note that it is common to take the numbers of individuals allocated to each cohort to be the same, that is to set $\underline{m} = m\underline{1}$ where $\underline{1}$ is a vector of 1's.

This dose-escalation setting can be modelled as an incomplete block design with cohorts corresponding to blocks. Specifically, suppose that responses are continuous and that cohort and treatment effects are fixed. Then the response y_{ki} to the ith treatment in the kth cohort can be modelled as

$$y_{ki} = \mu + \beta_k + \tau_i + e_{ki}, \quad k = 1, \ldots, c, \ i = 0, \ldots, n,$$

where μ denotes the overall mean effect, β_k the effect of the kth cohort, τ_i the effect of the ith treatment and the error term e_{ki} is normally distributed with mean 0 and variance σ^2 independently of all other such error terms. It then follows immediately from standard results pertaining to block designs (John and Williams, 1995, p. 12) that the information matrix for the treatment effects adjusted for blocks is given by

$$L = R - S^T K^{-1} S$$

where R is a diagonal matrix of order t with diagonal entries $r_0, \ldots, r_n$ and K is a diagonal matrix of order c with diagonal entries $m_1, \ldots, m_c$. Note that $L\underline{1} = \underline{0}$ where $\underline{0}$ is a vector of 0's and thus that L is singular with $\text{rank}(L) < n+1$. If the block design is connected however, then $\text{rank}(L) = n$ and all treatment contrasts are estimable. Only connected designs are considered here.

The essential design problem in a dose-escalation study is to allocate individuals to cohorts and to treatments within cohorts in a manner that is in some sense optimal or near-optimal. In the present study, following Bailey (2009), optimality criteria based on the variances of the least squares estimators of the pairwise treatment differences are considered. Specifically, the variance of the estimator of $\tau_i - \tau_j$ is given, up to the constant of proportionality σ^2, by $w_{ij} = \underline{c}_{ij}^T G \underline{c}_{ij}$ where G is a generalized inverse of L and the vector $\underline{c}_{ij}$ has elements 1 and -1 in rows i and j respectively and zeros elsewhere for i and j integers such that $0 \leq i < j \leq n$. Thus the variance scaled by the efficiency factor is given by $v_{ij} = \dfrac{N}{2t} w_{ij}$, where $N = \sum_{k=1}^{c} m_k$ is the total number of individuals. The A-optimality criterion can then be introduced as the average and the MV-optimality criterion as the maximum of the scaled pairwise variances and these criteria can be written formally as $\bar{v} = \frac{1}{\binom{t}{2}} \sum\limits_{0 \leq i < j \leq n} v_{ij}$ and $v_{max} = \max\limits_{0 \leq i < j \leq n} v_{ij}$ respectively.

It is convenient, in particular computationally, to introduce the set of pairwise treatment differences as $C\underline{\tau}$ where C is the $\dfrac{t(t-1)}{2} \times t$ matrix with the row corresponding to $\tau_i - \tau_j$ given by $\underline{c}_{ij}^T$ for $0 \leq i < j \leq n$ and $\underline{\tau}$ is the vector $(\tau_0, \ldots, \tau_n)$. Then the variance matrix of the least squares estimator of $C\underline{\tau}$ is given by CGC^T with diagonal entries corresponding to the pairwise treatment variances. Thus A-optimal designs are those designs which minimize $tr(CGC^T) = tr(GC^TC)$ and, since $C^TC = tI - J$, where I is the identity matrix of order t and J is the $t \times t$ matrix of 1's, and also since $G\underline{1}$ is a constant multiple of $\underline{1}$ for any g-inverse G of L, those designs which minimize $tr(G)$. Note that if G is taken to be the Moore-Penrose inverse, that is $G = (L + \frac{1}{t}J)^{-1} - \frac{1}{t}J$, then $tr(CGC^T) = tr(G)$ (Bailey, 2009). In addition MV-optimal designs are those designs which minimize the maximum of the diagonal entries of CGC^T. Other criteria based on the non-zero eigenvalues of L, written $\lambda_1, \ldots, \lambda_n$ where $n = t - 1$, have been formulated within the context of block designs but are less commonly used than A- and MV-optimality (John and Williams, 1995, p. 32). Of these criteria, D- and E-optimality are arguably the better known and in a sense most meaningful. Thus, in the present study, D-optimal designs which maximize the geometric mean of the non-zero eigenvalues of L, namely $\prod\limits_{i=1}^{n} \lambda_i$, and E-optimal designs which maximize the minimum eigenvalue of L, namely $\min\limits_{i=1,\ldots,n} \lambda_i$, are also introduced.

Finally, if block effects are taken to be random, the model for the response y_{ki} to the ith treatment in the kth cohort in the dose escalation setting can be written as

$$y_{ki} = \tau_i + \eta_k + e_{ki}, \quad k = 1, \ldots, c, \; i = 0, \ldots, n,$$

where η_k is the random effect of the kth cohort with mean 0 and variance σ_C^2 independently

of all other error terms (Bailey, 2009). Then, provided the number of individuals allocated to each cohort is the same, that is $m_k = m$ for $k = 1, \ldots, c$, it follows that the pairwise variances of the treatments can be obtained by replacing the matrix L introduced above by

$$L = R - \frac{1-\theta}{m} S^T S - \frac{\theta}{cm} \underline{r}\underline{r}^T$$

where $\theta = \dfrac{\sigma^2}{\sigma^2 + m\sigma_C^2}$ (Goos, 2002, p. 105; Bailey, 2009). Note that if θ is taken to be 0, results for the fixed block effects model are obtained and that if θ is taken to be 1 those for a completely random design accrue (John and Williams, 1995, p. 143; Bailey, 2009).

3. Algebraic Results

The most elegant approach to constructing optimal designs in the dose-escalation setting is to derive explicit expressions for the pairwise treatment variances for broad classes of designs and to use these to formulate and optimize appropriate criteria with respect to the allocation. Bailey (2009) provides some guidelines. Indeed a very general strategy would be to find a tractable algebraic expression for the matrix L but unfortunately this would only seem to be feasible in the case of certain traditional designs.

To be specific, consider a traditional design with m individuals per cohort, with a individuals allocated to the placebo and $b = m - a$ to dose k in cohorts $k = 1, \ldots, n$ and, in the extended design case, with x_1 individuals allocated to the placebo and an equi-distribution of $x_2 = \dfrac{m - x_1}{n}$ individuals to each of the remaining doses. Suppose further that the block effects are random. Then for an extended design the $t \times t$ matrix L is specified by taking

$$S = \begin{bmatrix} a\underline{1} & bI \\ x_1 & x_2\underline{1}^T \end{bmatrix}, \underline{r} = \begin{bmatrix} na + x_1 \\ (b + x_2)\underline{1}^T \end{bmatrix}$$

and

$$R = \begin{bmatrix} na + x_1 & \underline{0}^T \\ \underline{0} & (b + x_2)I \end{bmatrix}.$$

Now L can be partitioned in the form

$$L = \begin{bmatrix} \ell_{11} & \ell_{12}\underline{1}^T \\ \ell_{21}\underline{1} & \ell_{22,I}I + \ell_{22,J}J \end{bmatrix}$$

where $\ell_{11}, \ell_{12} = \ell_{21}, \ell_{22,I}$ and $\ell_{22,J}$ are scalars and it is a straightforward matter to derive explicit expressions for the non-zero eigenvalues of L, and thus for the Moore-Penrose inverse of L (Pringle and Rayner, 1971, p. 23), as

$$\lambda_1 = \frac{b(a+b\theta)+mx_2}{m} \quad \text{with multiplicity } n-1$$

and

$$\lambda_n = \frac{(n+1)\left[cb(a+b\theta)+cmx_2-nc(1-\theta)x_2^2-n\theta(b+x_2)^2\right]}{mc}.$$

Full details are presented in the Appendix. These non-zero eigenvalues can now be used directly to specify the $A-, D$- and E-optimality criteria (John and Williams, 1995, Section 2.4). Furthermore the pairwise variances can be expressed in terms of the eigenvalues λ_1 and λ_n as

$$w_{0i} = \frac{1}{n}\left(\frac{n-1}{\lambda_1}+\frac{n+1}{\lambda_n}\right) \text{ for } i = 1, \ldots, n$$

and

$$w_{ij} = \frac{2}{\lambda_1} \text{ for } 1 \leq i < j \leq n,$$

thereby specifying the MV-optimality criterion. Full details are again given in the Appendix. Note that results for the standard designs can be recovered by setting $c = n$ and $x_2 = 0$ and for fixed block effects by setting $\theta = 0$. The use of the above formulae in constructing optimal traditional designs is now illustrated by means of the following example.

Example 3.1: Consider a standard traditional design with random cohort effects, as specified above. It then follows that the non-zero eigenvalues of L are given by $\frac{b(a+b\theta)}{m}$ with multiplicity $n-1$ and $\frac{(n+1)ab}{m}$ with multiplicity 1, the scaled pairwise variances by

$$v_{0i} = \frac{(a+b)^2(an+b\theta)}{2(n+1)ab(a+b\theta)} \text{ for } i = 1, \ldots, n,$$

$$v_{ij} = \frac{n(a+b)^2}{(n+1)b(a+b\theta)} \text{ for } i < j \leq n, i = 1, \ldots, n-1$$

and their mean by

$$\bar{v} = \frac{(a+b)^2(an^2+b\theta)}{ab(n+1)^2(a+b\theta)}.$$

These expressions can be invoked to construct optimal designs over the permitted range of θ, namely $0 \leq \theta \leq 1$. To be specific, suppose that $n = 4$ and $m = 16$. Then the requisite A-optimal designs depend on the values of θ and b and this dependence is illustrated by

the plot of optimal b against θ shown in Figure 1. Note that values of θ corresponding to transitions from b to $b+1$ are given by feasible solutions to quadratics in θ obtained by equating values of $\bar{v}$ for b and $b+1$, $b = 8, \ldots, 12$. These A-optimal designs are necessarily more efficient than the Senn designs but are very much less efficient than the corresponding halving designs, at least for $\theta < 0.849$ (Bailey, 2009). The MV-optimal designs are more interesting in that the changes in optimal b with θ, as shown in Figure 1, reflect switches between the maximum of the scaled pairwise variances v_{0i} and v_{ij} and in that they are more efficient than the recommended halving design for θ values greater than 0.540.

4. Complete Enumeration

It is worth beginning the numerical search for optimal designs for a dose-escalation study by investigating the feasibility of a complete enumeration of all possible designs. In particular, suppose that a_k individuals are free to be allocated to p_k treatments in the kth cohort with the remaining number of individuals in that cohort, $m_k - a_k$, preallocated to a specified set of treatments. Then the total number of ways in which the allocation over all cohorts can be attained is given by

$$N_A = \prod_{k=1}^{c} \binom{a_k + p_k - 1}{p_k - 1}.$$

The resultant compositions for each cohort specify allocations to be added to the prespecified allocation and are termed block additions. The number N_A increases exponentially with increasing numbers of doses and individuals in the study and in addition depends sensitively on the values of a_k and p_k in the kth cohort, $k = 1, \ldots, c$. As an illustration, values of N_A for standard and extended designs with $n = 2, 3, 4$ and 5, $m = 2n$, at least one individual allocated to each permitted treatment within the cohorts $k = 1, \ldots, n$ and a free allocation of individuals to treatments in the additional cohort of the extended designs are presented in Table 1.

Complete enumeration proceeds broadly as follows:

Enumeration Algorithm

Step 1: Specify the fixed allocation of numbers of individuals to treatments in the design.

Step 2: Generate the candidate block additions for each cohort.

Step 3: Cycle through all possible designs recursively, building each design by taking a block addition from the list for each cohort and recording the values of the criteria of interest.

Table 1: Total numbers of standard and extended designs for $n = 2, 3, 4, 5$, $m = 2n$, $s_{ki} \geq 1, k = 1, \ldots, n, i = 0, 1, \ldots, k$ and $s_{n+1,i} \geq 0, i = 0, 1, \ldots, n$

n	N_A: standard design	N_A: extended design
2	9	135
3	500	42,000
4	180,075	89,137,125
5	432,081,216	1,297,539,891,648

Step 4: Order the enumerated designs according to each of the A-, MV-, D- and E-optimality criteria and select the best designs.

Step 2 involves listing all block additions, that is compositions or ordered partitions, for the kth cohort, $k = 1, \ldots, c$. A command for generating compositions is available in Mathematica but not in other computer packages such as Gauss or R. However routines to implement the NW Algorithm of Nijenhaus and Wilf (1978, Chapter 5) for generating compositions are readily devised by following closely the original code. Note that the NW Algorithm was also used by Fisher and Hall (1992) to enumerate bootstrap samples. Step 3 in the enumeration procedure can be achieved by recursive looping through the block additions in each cohort. Finally note that the enumeration procedure can be readily adapted to find (M, S)-optimal designs, that is designs for which the sum of the non-zero eigenvalues of L, $\sum_{i=1}^{n} \lambda_i$, is a maximum and the sum of squares, $\sum_{i=1}^{n} \lambda_i^2$, is a minimum (John and Williams, 1995, p. 32). The complete enumeration of optimal designs for a dose-escalation setting is now illustrated by means of the following example.

Example 4.1: The 89,137,125 extended designs for the setting $n = 4, m = 8$, $s_{ki} \geq 1$ for $k = 1, \ldots, 4, i = 0, 1, \ldots, k$ and $s_{5,i} \geq 0$ for $i = 0, 1, \ldots, 4$ were completely enumerated using routines written in Gauss. The optimal designs are not unique but this is due, at least in part, to the fact that designs in which the allocations for the placebo and for dose 1 are switched are equivalent. Selected A-, MV-, D-, E- and (M, S)-optimal designs, which are deemed to be best across all criteria, are presented in Figure 2 and their key properties are summarized in Table 2.

Note that the recommended halving design given in Bailey (2009) is MV-optimal and in addition is highly A-, D- and E-efficient. Note also that there are in fact 300 M-optimal

Table 2: The number of optimal designs, N_{opt}, and the A-, MV-, D- and E-criterion values for the designs in Figure 2

Optimal		Criterion Values			
Design	N_{opt}	A	MV	D	E
A	2	1.2919	1.6054	2.3491	4.3255
MV	10	1.3231	1.5123	2.7123	4.6201
D	4	1.3055	1.6691	2.3402	4.0320
E	14	1.3778	1.6614	3.1254	4.6398
(M, S)	2	1.3506	1.8213	2.4019	3.6233

designs, that is designs maximizing $\sum_{i=1}^{n} \lambda_i$, but of these only 2 attain the minimum $\sum_{i=1}^{n} \lambda_i^2$ and are thus (M, S)-optimal.

5. Algorithmic Construction

Complete enumeration rapidly becomes infeasible as the numbers of doses in the study and the numbers of individuals free to be allocated to treatments increases. Recourse must then be made to heuristic algorithms which do not guarantee to find the the globally optimal design (Atkinson, Donev and Tobias, 2007, Chapter 12). In the present case two heuristics are proposed, the one involving block exchanges and the other individual moves.

5.1 Block Exchange

Arguably the most widely used heuristic for the construction of block designs involves compiling a candidate list of feasible blocks and embedding this into a Fedorov-type exchange procedure. In the present case the algorithm can be implemented broadly as follows.

Block Exchange Algorithm

Step 1: Specify the fixed allocation of numbers of individuals to treatments. Generate the candidate block additions for each cohort.

Step 2: Generate a starting design by choosing a block addition at random for each cohort. Repeat until the design is connected.

Step 3: For each cohort in turn, evaluate the criterion for exchanges of the current block addition with all the remaining candidate block additions and perform the exchange which effects the largest favourable change in criterion value.

Step 4: Repeat Step 3 until no further favourable exchanges can be effected. Stop.

This search procedure can be run many times and the best overall design selected as being near-optimal.

In fact Step 3 can be modified in the following way.

Step 3.1: For each cohort, evaluate the criterion for exchanges of the current block addition with all the remaining candidate block additions. Perform the exchange ONLY in the cohort for which the change in criterion value is most favourable, that is perform the best exchange for the design overall.

This modification proved to be ineffective over a wide range of examples however and is not discussed further.

An immediate drawback to the Block Exchange Algorithm is that the numbers of candidate compositions, that is block additions,

$$N_X = \sum_{k=1}^{c} \binom{a_k + p_k - 1}{p_k - 1},$$

can become prohibitively large as the numbers of doses and individuals in the study increases. To illustrate this fact, values for N_X for standard and extended designs of the form specified in Table 1 but with $n = 6, \ldots, 12$ are summarized in Table 3. The ideas underpinning the Block Exchange Algorithm are now illustrated by means of the following example.

Example 5.1: Suppose that a dose-escalation setting with 8 doses, 8 cohorts, 16 individuals per cohort and $s_{ki} \geq 1$ for $k = 1, \ldots, 8$ and $i = 0, 1, \ldots, k$ is of interest and suppose further that A-optimal designs for $\theta = 0, 0.25, 0.5$ and 0.75 are to be considered. The requisite optimal, or more precisely near-optimal, designs were constructed by invoking the block exchange procedure outlined above $1,000$ times and these designs are summarized in Figure 3. The difference between the designs with changing θ is striking, with more individuals being allocated to the highest dose in each cohort as θ increases. This trend can be attributed to the fact that designs tend towards an equi-replication of treatments as the between-cohort variance σ_C^2 decreases.

Table 3: Total numbers of candidate block additions for standard and extended designs with $n = 6, \ldots, 12$, $m = 2n$, $s_{ki} \geq 1, k = 1, \ldots, n, i = 0, 1, \ldots, k$ and $s_{n+1,i} \geq 0, i = 0, 1, \ldots, n$

n	N_X: standard design	N_X: extended design
6	1,485	20,049
7	5,811	122,091
8	22,818	758,289
9	89,845	4,776,670
10	354,521	30,399,536
11	1,401,291	194,938,011
12	5,546,381	1,257,224,081

5.2 Best Individual Moves

An alternative heuristic to that of block exchange, based on moving single individuals between treatments and termed the Best Move algorithm, was devised and can be summarized as follows.

Best Move Algorithm

Step 1: Specify the fixed allocation of numbers of individuals to treatments. Generate a starting design by allocating the balance of individuals to treatments randomly within each cohort, preserving the structure of the required design setting.

Step 2: For each cohort in turn, evaluate the criterion for all permissible and distinct moves of individuals from one treatment to another and perform that move which effects the largest favourable change in criterion value.

Step 3: Repeat Step 2 until no further moves give an improvement in criterion value. Stop.

Again this search procedure can be run many times and the best overall design selected as being near-optimal. Step 2 can be modified so that the algorithm is less greedy as follows.

Step 2.1: For each cohort, evaluate the criterion for all permissible and distinct moves of individuals to the treatments. Perform the move ONLY in the cohort for which the change in criterion value is most favourable, that is take the best move for the design overall.

The appeal of the best move algorithm is that it is candidate-set-free, that is it does not require block additions for each of the cohorts to be generated. Indeed the algorithm, and specifically the variant which incorporates Step 2, is in the spirit of candidate-set-free coordinate-exchange, an approach which is proving to be highly effective both in the context of the construction of exact optimal designs (Meyer and Nachtsheim, 1995; Jones and Goos, 2007) and in other areas as well (Hastie, Tibshirani and Friedman, 2009, Section 3.8.6). The algorithm can also be regarded, in essence, as an interchange algorithm, with the structure of the design problem allowing for a particularly straightforward interchange procedure as compared with those invoked for more general block and row-and-column designs (John and Williams, 1995, p. 100).

Best move algorithms were invoked to construct near-A-optimal designs for the setting of Example 5.1 and the results are presented below. Application of the algorithms to large-scale dose-escalation studies follows immediately and more details, specifically computational, are presented elsewhere (Clark and Haines, 2012).

Example 5.2: The best move algorithms incorporating Step 2 and Step 2.1 were run $1,000$ times for the dose-escalation settings of Example 5.1 and the near-A-optimal designs summarized in Table 5 were again obtained. Computationally, the block exchange algorithm took approximately 80 times and the best move algorithm based on Step 2.1 approximately 5 times as long as the best move algorithm based on Step 2. The space requirements for the best move algorithms were necessarily minimal as compared with those for the block exchange algorithm. There was however no clear pattern across the block exchange and the best move algorithms in terms of the percentage of runs that were successful in providing the near-A-optimal designs.

The best move algorithm can also be invoked to construct designs for somewhat unconventional dose-escalation settings, as for example those in which a total of M individuals are to be allocated to the treatments with minimal constraints on the numbers of individuals in each cohort. The latter application is illustrated by means of the following example.

Example 5.3: Suppose that $M = 40$ individuals are to be allocated to a dose-escalation study with $n = 4$ doses, $c = 4$ cohorts and θ taken to be 0, that is with no random effects. The only constraint placed on the designs for this setting is that at least one individual must be assigned to the placebo and one to the highest permitted dose in each cohort. A-, MV-, D- and E-near-optimal designs were constructed by running the best move algorithm with Step 2 $10,000$ times and these designs are summarized in Figure 4. The percentage of times the near-optimal designs were obtained varied widely across the criteria, ranging from 0.09%

for the MV-criterion to 75.40% for the D-criterion. The designs are all somewhat extreme and reflect the fact that, with a minimal constraint on numbers allocated to each cohort, a tendency towards equi-replication prevails.

6. Conclusions

The main aim of the present study has been to develop methods for constructing A-, MV-, D- and E-optimal designs for dose-escalation trials. Algebraic results proved elusive and explicit expressions for the requisite optimal designs could only be found for a restricted class of traditional designs. Recourse was therefore made to numerical procedures which provided considerable flexibility in terms of choice of criteria and allocation of individuals across treatments within the cohorts. Of the methods investigated numerically, complete enumeration and the best move algorithm are to be recommended. Complete enumeration guarantees to find the globally optimal design for a particular setting but its use is limited in that it is computationally highly expensive. In contrast the best move algorithm is a candidate-set-free coordinate-exchange algorithm so that, while not guaranteeing to find optimal designs, it is efficient in terms of computer-time and -space. The optimal designs presented in the examples tend to display unusual patterns which may be difficult to introduce to the practitioner. In addition the results indicate that Bailey's recommended halving designs perform well in terms of criteria-based efficiencies. On balance therefore it would seem that numerically-constructed optimal designs for dose-escalation studies provide valuable benchmarks and that the attendant routines offer considerable flexibility in terms of choice of criterion and allocation. A suite of programs to construct optimal designs for dose-escalation studies by invoking complete enumeration and the best move algorithm is being developed in R and will be reported elsewhere (Clark and Haines, 2012).

Acknowledgments

This work was supported by funds from the University of Cape Town and the National Research Foundation, South Africa. The study represents an extension of the first author's contribution to the discussion of Bailey (2009). A part of the paper was completed while the first author was a Visiting Fellow with the Design and Analysis of Experiments Programme at the Isaac Newton Institute of Mathematical Sciences, Cambridge, UK, in 2011 and she would like to thank Rosemary Bailey for some insightful discussions during that visit. Sayi Toutou and Mark Steinhaus provided help with the computer programming and their input is appreciated.

Appendix

Consider an extended traditional design with information matrix L as specified in Section 3. Then it is straightforward to show that the $n \times n$ matrix

$$L_{22} = \ell_{22,I} I + \ell_{22,J} J = \left[\frac{b(a+b\theta)+mx_2}{m}\right] I - \left[\frac{c(1-\theta)x_2^2 + \theta(b+x_2)^2}{cm}\right] J$$

has eigenvalues $\lambda_1 = \dfrac{b(a+b\theta)+mx_2}{m}$ with multiplicity $n-1$ and attendant eigenvectors $(1,-1,0,\ldots,0)$, $(1,0,-1,\ldots,0)$ through to $(1,0,0,\ldots,-1)$, or any basis thereof, and

$$\lambda_{n,22} = \frac{cb(a+b\theta) + cmx_2 - nc(1-\theta)x_2^2 - n\theta(b+x_2)^2}{mc}$$

with multiplicity 1 and eigenvector $\underline{1} = (1,1,1,\ldots,1)^T$. It is thus clear that the matrix L itself has $n-1$ eigenvalues λ_1 with eigenvectors $(0,1,-1,0,\ldots,0)$, $(0,1,0,-1,\ldots,0)$ through to $(0,1,0,0,\ldots,-1)$, or any basis thereof. Furthermore it is readily seen that the remaining eigenvalues of L are $\lambda_n = (n+1)\lambda_{n,22}$ with eigenvector $(-n,1,\ldots,1)$ and 0 with eigenvector $\underline{1}$. Thus the non-zero eigenvalues of the Moore-Penrose inverse of L, namely L^g, are $\dfrac{1}{\lambda_1}$ with multiplicity $n-1$ and $\dfrac{1}{\lambda_n}$ with multiplicity 1 and the corresponding eigenvectors are the same as those of L, that is for λ_1 and λ_n respectively (Pringle and Rayner, 1971, pp. 23-24).

Consider now a pairwise comparison of treatment effects, written $\underline{c}^T\underline{\tau}$, with variance

$$Var(\underline{c}^T\hat{\underline{\tau}}) = \underline{c}^T L^g \underline{c} = \underline{c}^T H \Lambda_g H^T \underline{c}$$

where Λ_g is a diagonal matrix with diagonal elements the eigenvalues of L^g and H is an orthogonal matrix with columns the corresponding normalized eigenvectors. Specifically suppose, without loss of generality, that the matrices Λ_g and H can be expressed in partitioned form as

$$\Lambda_g = \begin{bmatrix} \dfrac{1}{\lambda_n} & \underline{0}^T & 0 \\ \underline{0} & \dfrac{1}{\lambda_1} I & \underline{0} \\ 0 & \underline{0}^T & 0 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} -\sqrt{\dfrac{n}{n+1}} & \underline{0}^T & \dfrac{1}{\sqrt{n+1}} \\ \dfrac{1}{\sqrt{n(n+1)}}\underline{1} & B & \dfrac{1}{\sqrt{n+1}}\underline{1} \end{bmatrix}$$

where B is an $n \times (n-1)$ matrix with columns any normalized orthogonal basis of the $n-1$ vectors $(1,-1,0,\ldots,0)$, $(1,0,-1,\ldots,0)$ through to $(1,0,0,\ldots,-1)$. Then for $\underline{c}^T = \underline{c}_{01}^T = (1,-1,0,\ldots,0)$, the first column of B can be taken to be $\dfrac{1}{\sqrt{n(n-1)}} \begin{bmatrix} -(n-1) \\ \underline{1}^T \end{bmatrix}$

and all other columns of B are orthogonal to this vector and have first element 0. Thus it follows immediately that $\underline{c}_{01}^T L^g \underline{c}_{01} = \dfrac{1}{n}\left(\dfrac{n+1}{\lambda_n} + \dfrac{n-1}{\lambda_1}\right)$. In addition, for $\underline{c}^T = \underline{c}_{12}^T = (0, 1, -1, 0, \ldots, 0)$, the first column of B can be taken to be $\dfrac{1}{\sqrt{2}}\,\underline{c}_{12}$ so that all other columns of B are orthogonal to $\underline{c}_{12}$ and thus $\underline{c}_{12}^T L^g \underline{c}_{12} = \dfrac{2}{\lambda_1}$. Similar arguments to those used for $\underline{c}_{01}$ and $\underline{c}_{12}$ hold for all contrast vectors of the form $\underline{c}_{0i}$ for $i = 2, \ldots, n$and $\underline{c}_{ij}$ for $1 \leq i < j \leq n$ respectively.

References

Atkinson, A.C., Donev, A.N. and Tobias R.D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press: Oxford.

Bailey R.A. (2009). Designs for dose-escalation trials with quantitative responses. *Statistics in Medicine*, 28, 3721–3738.

Clark A.E. and Haines L.M. (2012). Work in progress.

Fisher N.I. and Hall P. (1991). Bootstrap algorithms for small samples. *Journal of Statistical Planning and Inference*, 27, 157–169.

Goos P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments.* Springer-Verlag:New York.

Hastie T.J., Tibshirani R.J. and Friedman J. (2009). *The Elements of Statistical Learning.* Second Edition. Springer: New York.

John J.A. and Williams E.R. (1995). *Cyclic and Computer Generated Designs.* Second Edition. Chapman & Hall: London.

Jones B. and Goos P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs. *Applied Statistics*, 56, 347–364.

Meyer R.K. and Nachtsheim C.J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37, 60–69.

Nijenhuis A. and Wilf H.S. (1978). *Combinatorial Algorithms for Computers and Calculators.* Academic Press: New York.

Pringle R.M. and Rayner A.A. (1971). *Generalized Inverse Matrices with Applications to Statistics*. Griffin: London.

Senn S., Amin D., Bailey R.A., Bird S.M., Bogacka B., Colman P., Garrett A., Grieve A. and Lachmann P. (2007). Statistical issues in first-in-man studies. *Journal of the Royal Statistical Society, Series A*, 170, 517–579.

Figure 1: Plots of optimal b against θ for standard traditional designs with $n = 4$, $m = 16$ and random cohort effects. Solid lines correspond to A-optimal and dashed lines to MV-optimal designs.

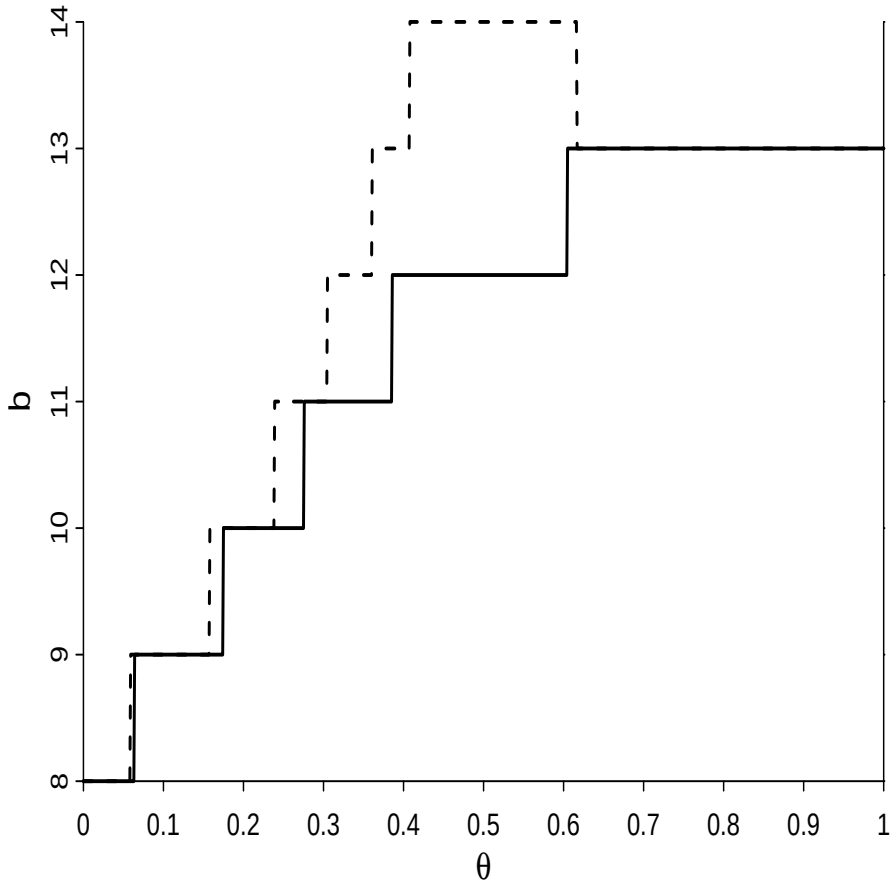

Figure 2: Selected A-, MV-, D-, E- and (M,S)-optimal extended designs for $n = 4$ and $m = 8$ with $s_{ki} \geq 1, i = 0, \ldots, k, k = 1, \ldots, 4$ and $s_{5i} \geq 0, i = 0, \ldots, 4$.

A-optimal

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	2	3	3	0	0
Cohort 3	2	1	2	3	0
Cohort 4	1	1	1	2	3
Cohort 5	1	1	1	2	3

MV-optimal

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	2	2	4	0	0
Cohort 3	1	1	2	4	0
Cohort 4	1	1	1	1	4
Cohort 5	1	1	1	2	3

D-optimal

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	2	3	3	0	0
Cohort 3	2	1	2	3	0
Cohort 4	1	1	2	2	2
Cohort 5	1	1	1	2	3

E-optimal

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	2	2	4	0	0
Cohort 3	2	1	1	4	0
Cohort 4	1	1	1	1	4
Cohort 5	0	1	2	1	3

(M,S)-optimal

Dose	0	1	2	3	4
Cohort 1	4	4	0	0	0
Cohort 2	3	2	3	0	0
Cohort 3	2	2	2	2	0
Cohort 4	1	1	2	2	2
Cohort 5	1	1	2	2	2

Figure 3: Near-A-optimal designs for $n = 8$ doses, $c = 8$ cohorts and $m = 16$ individuals per cohort with $s_{ki} \geq 1, i = 0, \ldots, k, k = 1, \ldots, 8$.

$\theta = 0$

Dose	0	1	2	3	4	5	6	7	8
Cohort 1	8	8	0	0	0	0	0	0	0
Cohort 2	5	5	6	0	0	0	0	0	0
Cohort 3	3	3	4	6	0	0	0	0	0
Cohort 4	2	2	3	3	6	0	0	0	0
Cohort 5	2	1	2	2	3	6	0	0	0
Cohort 6	1	1	1	2	2	3	6	0	0
Cohort 7	1	1	1	1	1	2	3	6	0
Cohort 8	1	1	1	1	1	1	1	3	6

$\theta = 0.25$

Dose	0	1	2	3	4	5	6	7	8
Cohort 1	8	8	0	0	0	0	0	0	0
Cohort 2	5	4	7	0	0	0	0	0	0
Cohort 3	3	3	4	6	0	0	0	0	0
Cohort 4	1	2	3	4	6	0	0	0	0
Cohort 5	1	1	1	3	4	6	0	0	0
Cohort 6	1	1	1	1	2	4	6	0	0
Cohort 7	1	1	1	1	1	1	3	7	0
Cohort 8	1	1	1	1	1	1	1	2	7

$\theta = 0.5$

Dose	0	1	2	3	4	5	6	7	8
Cohort 1	8	8	0	0	0	0	0	0	0
Cohort 2	4	5	7	0	0	0	0	0	0
Cohort 3	3	2	4	7	0	0	0	0	0
Cohort 4	1	1	3	4	7	0	0	0	0
Cohort 5	1	1	1	2	4	7	0	0	0
Cohort 6	1	1	1	1	1	4	7	0	0
Cohort 7	1	1	1	1	1	1	3	7	0
Cohort 8	1	1	1	1	1	1	1	2	7

$\theta = 0.75$

Dose	0	1	2	3	4	5	6	7	8
Cohort 1	8	8	0	0	0	0	0	0	0
Cohort 2	4	5	7	0	0	0	0	0	0
Cohort 3	2	1	5	8	0	0	0	0	0
Cohort 4	1	1	1	4	9	0	0	0	0
Cohort 5	1	1	1	1	3	9	0	0	0
Cohort 6	1	1	1	1	1	2	9	0	0
Cohort 7	1	1	1	1	1	1	1	9	0
Cohort 8	1	1	1	1	1	1	1	1	8

Figure 4: Standard A-, MV-, D-, E-optimal designs for $n = 4$ doses and $M = 40$ individuals assigned to the trial with $s_{k0}, s_{kk} \geq 1$ and $s_{ki} \geq 0, i = 1, \ldots, k-1, k = 1, \ldots, 4$.

A-optimal

Dose	0	1	2	3	4
Cohort 1	1	1	0	0	0
Cohort 2	1	1	1	0	0
Cohort 3	1	1	1	1	0
Cohort 4	5	5	6	7	8

MV-optimal

Dose	0	1	2	3	4
Cohort 1	1	1	0	0	0
Cohort 2	1	1	1	0	0
Cohort 3	1	1	2	2	0
Cohort 4	5	5	5	6	8

D-optimal

Dose	0	1	2	3	4
Cohort 1	1	1	0	0	0
Cohort 2	1	1	1	0	0
Cohort 3	1	1	1	1	0
Cohort 4	6	6	6	6	7

E-optimal

Dose	0	1	2	3	4
Cohort 1	1	1	0	0	0
Cohort 2	1	0	1	0	0
Cohort 3	1	1	1	2	0
Cohort 4	5	6	6	6	8