Statistics and Computing manuscript No.
(will be inserted by the editor)

Optimal designs for dose-escalation trials and individual allocations in cohorts

Belmiro P.M. Duarte · **Anthony C. Atkinson** · **Nuno M.C. Oliveira**

Received: date / Accepted: date

Abstract Dose escalation trials are crucial in the development of new pharmaceutical products to optimize the amount of drug administered while avoiding undesirable side effects. We adopt the framework established by Bailey (2009) where the individuals are grouped into cohorts, to the subjects in which the placebo or previously defined doses are administered and responses measured. Successive cohorts allow testing higher doses of drug if negative responses have not been observed in earlier cohorts. We propose Mixed Integer Nonlinear Programming formulations for systematically computing optimal experimental designs for dose escalation. We demonstrate its application with i. different optimality criteria; ii. standard and extended designs; and iii. non-constrained (or traditional), strict halving and uniform halving designs. Additionally, we address the allocation of the individuals in a cohort considering previously known prognostic factors. To handle the problem we propose i. an enumerative algorithm; and ii. a Mixed Integer Nonlinear Programming formulation. We demonstrate the application of the enumeration scheme for allocating individuals on an *individual arrival basis*, and of the latter formulation for allocation on a *within cohort basis*.

Keywords Dose escalation experiments · Cohorts · First-in-human trials · Optimal allocation · Prognostic factors

Mathematics Subject Classification (2000) 62K05 · 90C47

Belmiro P.M. Duarte
Instituto Politécnico de Coimbra, Instituto Superior de Engenharia de Coimbra, Department of Chemical and Biological Engineering, Rua Pedro Nunes, Quinta da Nora, 3030−199 Coimbra, Portugal, and CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 Coimbra, Portugal. Tel.: +351-239-790200, Fax: +351-239-790201, E-mail: bduarte@isec.pt

Anthony C. Atkinson
Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom. E-mail: A.C.Atkinson@lse.ac.uk

Nuno M.C. Oliveira
CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 Coimbra, Portugal. E-mail: nuno@eq.uc.pt

1 Motivation

A major focus of Phase I clinical trials is the establishment of the maximum safe dose of a new drug or drug combination. It is customary to use a series of cohorts of subjects, who may be patients or healthy volunteers (Senn, 2007). Some subjects in successive cohorts are allocated a higher dose of the drug than was applied in any previous cohort. A careful discussion of one family of such designs is given by Bailey (2009), who derives designs that can be simply described, and are: i. rational in terms of implementation; and ii. optimal or nearly optimal in terms of the amount of information produced. Unlike Bailey (2009) we use the methods of optimal experimental design to provide designs optimal under three design criteria. We can then establish the efficiency of any proposed design. In addition, we use the methods of sequential optimal design to provide schemes that allow for the presence of any prognostic factors of the individuals, such as age, weight, blood pressure and medical history, that may be available before treatment allocation. The resulting partial balance over prognostic factors leads to increased precision in estimation of treatment differences.

There is a large literature on dose-escalation methods, particularly in cancer trials (Le Tourneau et al., 2009). Algorithm-based methods, often basing decisions only on the results of the current cohort offer negligible toxic death rates (Storer, 1989). Model-based designs such as the continual reassessment method (CRM) (O'Quigley et al., 1990) and its variations (Babb et al., 1998; Cheung, 2005; Neuenschwander et al., 2008) and others, offer significant advantages but assume a generic response model which is sometimes inadequate. Dette et al. (2008) develop designs robust to the specific true model.

We avoid the specification of a relationship between toxicity and dose by assuming that cohorts are homogeneous and are analogous to blocks so that there may be systematic differences between cohorts (Bailey, 2009). The intra-block analysis identifies the difference in responses caused by different doses. The aim is to minimize a specified function of the variances of the estimated differences, which is an optimal design problem maximizing a function of the Fisher Information Matrix (FIM).

Dror et al. (1995) provides a survey of sequential treatment allocation methods in the absence of cohorts and Haines and Clark (2014) and Rosa and Harman (2017) consider designs for dose escalation with cohorts. We are the first to provide an algorithm that works in reasonable time and that extends straight forwardly to sequential allocation using the prognostic factors in the decision. The use of optimal design theory for finding such designs was suggested by Begg and Iglewicz (1980) were the first to suggest the use of optimal design theory for finding such designs, although they used an approximation to the FIM. Atkinson (1982, 2002) proposed methods based on optimal design theory, specifically the D_A-optimal criterion for treatment allocation; treatments are allocate sequentially to one subject at a time in the absence of cohorts. For allocation within cohorts in dose escalation studies, the objective remains the maximization of the information measured by the FIM resulting from optimally allocating the individuals of the same cohort to doses using prognostic information. The final allocation sequence must conform to the dose escalation optimal design.

1.1 Novelty statement and organization

This paper contains four elements of novelty: i. optimization-based formulations to systematically construct optimal designs for dose escalation studies where the individuals are grouped in cohorts for different kinds of exact designs. These include the designs with "strict halving" and "uniform halving" constraints described by Bailey (2009) and designs in which these constraints are absent; ii. optimization-based formulations to systematically allocate the individuals to doses in a cohort on an *individual arrival basis*; iii. optimization-based formulations to systematically allocate the individuals to doses on a *within cohort basis*, and iv. *flexibility* in design construction. Cohorts can be constructed sequentially in response to variations in the number of subjects available. Importantly, allocations in later cohorts can be adjusted to allow for loss of subjects to the trial before responses are measured.

The paper is organized as follows. Section 2 introduces the background and the notation used to formulate the optimal design problem as well as the fundamentals of Mixed Integer Nonlinear Programming (MINLP). Section 3 introduces the formulations used to solve the optimal design problem for standard and extended dose escalation studies, and the modifications required to handle constrained designs, specifically those of the strict halving and uniform halving types. Comparisons for different setups including standard, extended, unconstrained, strict and uniform halving designs are presented in §4 and Appendix A of the Supplementary Material (SM). Section 5 addresses the problem of allocating the individuals to doses using the additional information from prognostic factors and introduces algorithms for: i. sequentially allocating them as they enter the study; and ii. simultaneously allocating all individuals within a cohort. Section 6 demonstrates the application of optimal design-based allocation strategies and analyzes its efficiency relative to random allocation using a measure of the information content for comparison. Section 7 reviews the formulation and offers a summary of the results obtained. Finally, in Appendices A and B of SM, we provide additional results for a more challenging design setups, i.e. setups with more cohorts, doses and individuals to allocate.

2 Optimal designs for cohort-based dose escalation trials

This Section establishes the nomenclature used in the representation of the models. In Section 2.1 we present the experimental design problems outlined above. Then, in Section 2.2, we give an overview of the fundamentals of MINLP which serves to solve the dose escalation design problem as well as the optimal allocation problem in cohorts.

2.1 Optimal experimental design

In our notation bold face lowercase letters represent vectors, bold face capital letters stand for continuous domains, blackboard bold capital letters are used to denote discrete domains and capital letters are adopted for matrices. Finite sets containing

ι elements are compactly represented by $[\![\iota]\!] \equiv \{1, \cdots, \iota\}$. The transpose operation of a matrix or vector is represented by "$^\intercal$". The cardinality of a vector is represented by $\mathrm{card}(\bullet)$, the trace of a matrix by tr, $\mathrm{ldet}(\bullet)$ represents $\ln[\det(\bullet)]$, and $\mathrm{pinv}(\bullet)$ is for the generalized Moore-Penrose inverse of a matrix. $\mathbb{Z}_0$ is the set of non-negative integer numbers.

Exact designs are experimental plans where the relative effort of each experimental condition is a ratio n_i/N satisfying the conditions: i. all n_i's that represent the number of experiments at the i^{th} design point are integer (or null); and ii. the n_i's sum to N. The optimization problem to construct exact designs is not convex, and so finding them is computationally more challenging than finding equivalent approximate optimal designs (Boer and Hendrix, 2000), requiring global optimizers to assure that global optima are attained. This paper focuses on exact designs for dose escalation studies. For simplicity in the remaining sections we use the term designs to indicate exact designs for dose escalation.

Consider the formulation established by Bailey (2009) for design of dose escalation experiments. Suppose that we consider $n-1$ doses order labeled plus the placebo which is given for simplification the number 1; i.e., the vector of treatments is $\mathbf{d} = (d_1,\ d_2, \cdots, d_n)^\intercal \in \mathbb{R}^n$ where $d_1 (= 0) < d_2 < \cdots < d_n$. That is, we have a set of n possible treatments which include the placebo. We also have n_c cohorts, each with $m_k,\ k \in \{1, \cdots, n_c\}$ individuals. A cohort is a group with similar characteristics to which one dose of $\mathbf{d}$ is administered at a previously set time instant. In general we can allow the number of individuals in each cohort to be different but there are practical advantages of having cohorts of equal size. Dose escalation *standard designs* have the number of cohorts equal to $n-1$ (i.e., $n_c = n-1$), whereas in *extended designs* there is an extra cohort so that $n_c = n$. Let $\mathbf{c} = (c_1,\ c_2, \cdots,\ c_{n_c})^\intercal$ be a lexicographic ordered vector containing the designation of the cohorts where c_i refers to cohort i. The number of subjects of cohort k allocated to dose i is $s_{k,i}$, and matrix $S \equiv \{s_{k,i}\}_{k\in\{1,\cdots,n_c\},\ i\in\{1,\cdots,n\}} \in \mathbb{Z}_0^{n_c \times n}$ summarizes the design.

In the first cohort the subjects receive only placebo (dose 1) or dose 2. In cohort k ($2 \leqslant k \leqslant n_c$) the individuals may receive the placebo (d_1) or dose i ($d_i,\ 2 \leqslant i \leqslant k+1$) but no individuals receive higher doses ($i \in \{k+2, \cdots, n\}$). Practically, the design is characterized by having $s_{k,k+1} > 0$ and $s_{k,i} = 0$ for $i \in \{k+2, \cdots, n\}$. The values of $s_{k,i}$ for $i \in \{1, \cdots, k\}$ may be null but the application of Bailey's *diversity principle* minimizes the number of null elements in S (also see Huang and Chapell (2008)). Let the replication of dose i in cohorts 1 to k be $r_{k,i} = \sum_{\kappa=1}^{k} s_{\kappa,i}$, here called the partial replication of size k. Further, the total replication of dose i is denoted by $\varrho_i = \sum_{\kappa=1}^{n_c} s_{\kappa,i}$. The number of subjects in each cohort k is $m_k = \sum_{i=1}^{k} s_{k,i}$, and the total number of individuals in the trial is N. Consequently, $\sum_{k=1}^{n_c} m_k = N$.

For simplicity in establishing the fundamentals we consider standard designs but the extension to extended designs is straightforward. The individual responses are measured on a scale such that the model

$$\mathbb{E}(y_{k,i}) = \tau_i + \beta_k,\ i \in [\![n]\!],\ k \in [\![n_c]\!] \tag{1}$$

holds. Here, $y_{k,i}$ is the response of the individuals of cohort $k \in [\![n_c]\!]$ that received dose $i \in [\![n]\!]$ and $\mathbb{E}(\bullet)$ stands for the expectation. The parameters $\beta_k,\ k \in [\![n_c]\!]$ are

the cohort effects with $\tau_i, i \in [\![n]\!]$ being the dose effects. The aim of the design is to estimate the differences $\tau_i - \tau_j,\ i, j \in [\![n]\!],\ i > j$. The responses are assumed uncorrelated and the observational noise follows a normal distribution with zero mean.

Experimental designs are sought to minimize a function of the variances of the least square estimates of $\tau_i - \tau_j$. For cohort k the model for the differences of response is

$$\mathbb{E}(y_{k,i} - y_{k,j}) = \tau_i - \tau_j,\ i \in \{2, \cdots, n\},\ j \in \{1, \cdots, i-1\}. \tag{2}$$

Throughout the paper we consider that the effects of cohorts are fixed, and constants β_k cancel each other when the differences due to dose effects are of interest. Random cohort effects setups assume each cohort has a different impact on response, and were considered by Haines and Clark (2014); O'Brien (2017) among others.

Let the matrix of parameter differences be $\boldsymbol{\theta} \equiv \{\tau_i - \tau_j :\ i \in \{2, \cdots, n\},\ j \in \{1, \cdots, i-1\}\}$, i.e

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{2,1} & 0 & \cdots & 0 & 0 \\ \theta_{3,1} & \theta_{3,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \theta_{n-1,1} & \theta_{n-1,2} & \cdots & \theta_{n-1,n-2} & 0 \\ \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,n-2} & \theta_{n,n-1} \end{pmatrix} =$$

$$= \begin{pmatrix} \tau_2 - \tau_1 & 0 & \cdots & 0 & 0 \\ \tau_3 - \tau_1 & \tau_3 - \tau_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \tau_{n-1} - \tau_1 & \tau_{n-1} - \tau_2 & \cdots & \tau_{n-1} - \tau_{n-2} & 0 \\ \tau_n - \tau_1 & \tau_n - \tau_2 & \cdots & \tau_n - \tau_{n-2} & \tau_n - \tau_{n-1} \end{pmatrix}.$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{(n-1)^2}$; $\boldsymbol{\Theta}$ is the domain of dose effect differences.

The experimental design ξ is described by matrix S, the Fisher Information Matrix (FIM) of ξ being $\mathcal{M}(\xi)$. For simplicity let $m = m_k,\ k \in [\![n_c]\!]$ (i.e. replication is equal for all cohorts). Then,

$$\mathcal{M}(\xi) = R_n - \frac{S^\intercal\, S}{m}, \tag{3}$$

where R_n is the $n \times n$ diagonal matrix containing the replications of cohorts on its diagonal. The FIM is a circulant matrix, with Moore-Penrose generalized inverse (Anderson, 1972; Pearce, 1983; Searle, 1979)

$$\mathrm{pinv}[\mathcal{M}(\xi)] = \left[\mathcal{M}(\xi) + \frac{J_n}{n}\right]^{-1} - \frac{J_n}{n}, \tag{4}$$

where J_n is the $n \times n$ matrix of ones. The optimal experimental design is obtained by minimizing an appropriate function of the confidence region of the parametric estimates of contrasts $\tau_i - \tau_j$. This is equivalent to minimizing a convex function of $\mathrm{pinv}[\mathcal{M}(\xi)]$ by choice of S or equivalently to maximizing a concave function of its inverse as for $\mathrm{ldet}(\bullet)$. The most commonly used criteria for measuring the amount of information resulting from dose escalation trials are: A–optimality (minimizes the

average variance of the estimated treatment variances); E–optimality (minimizes the maximum of these variances, see Hedayat et al. (1988)) and D–optimality (minimizing the volume of the confidence region of the estimated differences). That is, A–, E– and D–optimal designs each respectively minimize one of the following criteria:

$$\xi_{\mathrm{A}} = \arg\min_{\xi \in \Xi_{\mathrm{A}}} \mathrm{tr}\{\mathrm{pinv}[\mathcal{M}(\xi)]\}, \tag{5a}$$

$$\xi_{\mathrm{E}} = \arg\min_{\xi \in \Xi_{\mathrm{E}}} \lambda_{\max}\{\mathrm{pinv}[\mathcal{M}(\xi)]\}, \tag{5b}$$

$$\xi_{\mathrm{D}} = \arg\max_{\xi \in \Xi_{\mathrm{D}}} \mathrm{ldet}\{[\mathrm{pinv}(\mathcal{M}(\xi))]^{-1}\}, \tag{5c}$$

where Ξ_p is the space of $p-$optimal feasible designs, $p \in \{\text{A–, E–, D–}\}$, and $\lambda_{\max}\{\bullet\}$ represents the maximum eigenvalue of a matrix. The A–, E– and D–optimality efficiencies of a design ξ are defined, respectively, by

$$\mathrm{eff}_{\mathrm{A}}(\xi) = \frac{\mathrm{tr}\{\mathrm{pinv}[\mathcal{M}(\xi^*_{\mathrm{A}})]\}}{\mathrm{tr}\{\mathrm{pinv}[\mathcal{M}(\xi_{\mathrm{A}})]\}} \tag{6a}$$

$$\mathrm{eff}_{\mathrm{E}}(\xi) = \frac{\lambda_{\max}\{\mathrm{pinv}[\mathcal{M}(\xi^*_{\mathrm{E}})]\}}{\lambda_{\max}\{\mathrm{pinv}[\mathcal{M}(\xi_{\mathrm{E}})]} \tag{6b}$$

$$\mathrm{eff}_{\mathrm{D}}(\xi) = \left(\frac{\det[\mathrm{pinv}(\mathcal{M}(\xi_{\mathrm{D}}))^{-1}]}{\det[\mathrm{pinv}(\mathcal{M}(\xi^*_{\mathrm{D}})^{-1}]}\right)^{1/n_\theta}, \tag{6c}$$

where $\xi_p,\ p \in \{\text{A–, E–, D–}\}$ is the $p-$optimal design obtained for the specified criterion, ξ^*_p is the $p-$optimal design used for reference and n_θ the number of parameters to be estimated. Here, we consider $n_\theta = n_c - 1$ which is equal to the number of eigenvalues of $\mathrm{pinv}[\mathcal{M}(\xi)]$ and use it for extended and standard designs; consequently, the D–optimality criterion becomes positively homogeneous (see Pukelsheim 1993, §6.2 and §8.18) which, in turn, allows the comparison of standard and extended designs.

2.2 Mixed-Integer Nonlinear Programming

In this Section we introduce the fundamentals of Mixed-Integer Nonlinear Programming.

MINLP is used for solving the exact design problems introduced in Section 3. MINLP is a class of mathematical programming problems where the objective or some of the constraints are nonlinear and some of the decision variables are constrained to integer values. To optimize a function of n_x continuous variables, $\mathbf{x}$, and n_y discrete variables, $\mathbf{y}$, the general form of a MINLP is

$$\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x},\mathbf{y}) \tag{7a}$$

$$\text{s.t. } \mathbf{g}(\mathbf{x},\mathbf{y}) \leqslant \mathbf{0} \tag{7b}$$

$$\mathbf{h}(\mathbf{x},\mathbf{y}) = \mathbf{0} \tag{7c}$$

$$\mathbf{x} \in \mathbf{X},\ \mathbf{y} \in \mathbf{Y}. \tag{7d}$$

The function (7b) represents a set of r_i inequalities and (7c) a set of equality constraints, $\mathbf{X}$ is a compact set containing continuous variables $\mathbf{x}$, $\mathbf{Y}$ contains the discrete variables $\mathbf{y}$ and (7a) is the objective function. In our design context the vector $\mathbf{x}$ includes the elements of the FIM and its Moore-Penrose generalized inverse and other auxiliary variables used in the Cholesky decomposition which provides fast and accurate calculation of determinants and inverses of information matrices. The vector $\mathbf{y}$ is for the number of individuals allocated to each dose, i.e., the elements $s_{k,i}$ of design matrix S. Finally, the equality constraints are for the FIM computation and its algebraic treatment required to generalize the computation of the information measures outlined above, see Section 3. Here, $\mathbf{Y} \in \mathbb{Z}_0^{n_c \times n}$ is the design domain and is formed by the set of possible non-negative integer values of the elements in the design matrices S.

Exact optimal design problems are typically nonconvex. To guarantee that a global optimum is found, a global solver must be employed. However, global optimization solvers continue to require a long computational time compared to local solvers (Lastusilta et al., 2007), and this may limit their utilization to small and average sized problems. The conditions required for applying deterministic global MINLP solvers are fairly general - typically they only require bounded variables. In our formulations, these assumptions are satisfied by construction as all decision variables are bounded, and we use a global solver.

3 Formulations for computing dose escalation optimal designs of experiments

In this Section we introduce optimization formulations for finding standard and extended optimal designs for dose escalation trials. In Section 3.1 we consider formulations for unconstrained designs, that is those that have no constraints on the allocations other than those required by the optimization problem itself. These are also called *traditional* designs. In Section 3.2 we present specific algebraic operations employed to simplify the numerical computation. In Section 3.3 we consider constrained designs, namely *strict* and *uniform halving* designs, and present the modifications to the optimization formulations required for their computation. Finally, in Section 3.4 we introduce the numerical tools used for solving the optimal design problems. In all the problems we assume that n, n_c and N are known and the cohorts have equal replication; i.e., $m = m_k = N/n_c,\ k \in [\![n_c]\!]$.

3.1 Unconstrained optimal designs for dose escalation trials

This Section introduces the optimization formulations for finding unconstrained designs of both standard and extended type.

First, we consider the formulation for the A–optimality criterion and then we extend it to E– and D–optimality. The problem for finding A–optimal exact designs on Ξ_{A} is defined in (5a). The formulation to handle the optimization problem is as follows:

$$\min_{S,R_n} \ \mathrm{tr}\{\mathrm{pinv}[\mathcal{M}(\xi)]\} \tag{8a}$$

$$\text{s.t. } \mathrm{pinv}[\mathcal{M}(\xi)] = \left[R_n - \frac{S^\intercal\, S}{m} + \frac{J_n}{n}\right]^{-1} - \frac{J_n}{n} \tag{8b}$$

$$m_k = \sum_{i=1}^{k} s_{k,i},\ k \in [\![n_c]\!] \tag{8c}$$

$$\sum_{k=1}^{n_c}\sum_{i=1}^{k} s_{k,i} = N \tag{8d}$$

$$\sum_{k=1}^{n_c} s_{k,i} = R_{n,i,i}, \quad i \in [\![n]\!] \tag{8e}$$

$$R_{n,i,j} = 0, \quad i,j \in [\![n]\!],\ i \neq j \tag{8f}$$

$$s_{k,i} = 0, \quad i > k+1,\ k \in [\![n_c]\!],\ i \in [\![n]\!] \tag{8g}$$

$$s_{k,i} \in \mathbb{Z}_0,\ i \leqslant k+1,\ k \in [\![n_c]\!],\ i \in [\![n]\!] \tag{8h}$$

Equation (8a) is the objective function, (8b) is to compute the Moore-Penrose generalized inverse of the FIM, (8c) is to ensure the replication for each cohort, (8d) ensures that the sum of subjects allocated to the experiment is exactly N, (8e) is to calculate the diagonal elements of R_n corresponding to dose replication, (8f) sets non-diagonal elements to 0, (8g) imposes the constraint that no subjects are allocated to more than one dose not tested before and, finally, (8h) sets the domain of decision variables. The problem (8) falls into the general class of MINLP problems; the complexity of evaluating the objective function is notorious. Further, the problem may have multiple optima.

The inversion of $R_n - S^\intercal\, S/m + J_n/n$ in (8b) is carried out using an algebra based procedure involving three steps: 1. the matrix is decomposed applying the Cholesky decomposition taking advantage of its semidefinite positiveness; 2. the resulting upper diagonal matrix, $\mathcal{U}(\xi)$, is inverted, and produces $\mathcal{U}^{-1}(\xi)$; 3. $\mathcal{U}^{-1}(\xi)$ is then used to compute $[R_n - S^\intercal\, S/m + J_n/n]^{-1} = [\mathcal{U}^{-1}(\xi)]^\intercal\, \mathcal{U}^{-1}(\xi)$; 4. finally, the Moore-Penrose inverse is formed; i.e. $\mathrm{pinv}[\mathcal{M}(\xi)] = [R_n - S^\intercal\, S/m + J_n/n]^{-1} - J_n/n$. This procedure also allows computing the set of eigenvalues of $\mathrm{pinv}[\mathcal{M}(\xi)]$, here represented by $\boldsymbol{\lambda}\{\mathrm{pinv}[\mathcal{M}(\xi)]\}$ as they are the set of diagonal elements of $\mathcal{U}^{-1}(\xi)$. For non-singular FIMs all the eigenvalues of the inverse are nonzero and we designate them as $\lambda_i,\ i \in [\![n]\!]$; for singular FIMs we use instead the positive eigenvalues of the Moore-Penrose generalized inverse. For a detailed analysis of the matrix inversion procedure see Duarte et al. (2021, 2020).

Now, we consider the formulation (8) and adapt it for the E–optimality criterion (5b). The MINLP problem is,

$$\min_{S,R_n} \ t \tag{9a}$$

$$\text{s.t. Equations (8b-8h)} \tag{9b}$$

$$t \geqslant \lambda_i, \quad \lambda_i \in \boldsymbol{\lambda}\{\mathrm{pinv}[\mathcal{M}(\xi)]\},\ i \in [\![n]\!], \tag{9c}$$

where t is the maximum eigenvalue of the Moore-Penrose generalized inverse, and (9c) is to ensure that t is equal or exceeds all eigenvalues of $\mathrm{pinv}[\mathcal{M}(\xi)]$.

Finally, for the D–optimality criterion (5c), the design problem becomes

$$\max_{S,R_n} \ \mathrm{ldet}\{[\mathrm{pinv}(\mathcal{M}(\xi))]^{-1}\} \tag{10a}$$

$$\text{s.t. Equations (8c-8h)}, \tag{10b}$$

where the computation of $\mathrm{ldet}\{[\mathrm{pinv}(\mathcal{M}(\xi))]^{-1}\}$ is also addressed in Duarte et al. (2021, 2020); specifically it is formulated as the sum of the logarithms of the diagonal elements of the inverse of the upper diagonal matrix resulting from the Cholesky decomposition of $\mathcal{M}(\xi)$. Cases where $\mathcal{M}(\xi)$ is singular are addressed in §3.2.

3.2 Algebraic simplifications

This Section presents algebraic strategies used to simplify the solution of optimal design problems (8–10).

The calculation of $\mathrm{pinv}[\mathcal{M}(\xi)]$ in problems (8-10) is challenging, and we adopt specific algebraic simplifications for reducing the complexity. First, we note that the trace of the sum of matrices is equal to the sum of the traces (Kaye and Wilson, 1998) in problem (8):

$$\begin{aligned}\mathrm{tr}\left\{\left[R_n - \frac{S^\intercal\, S}{m} + \frac{J_n}{n}\right]^{-1} - \frac{J_n}{n}\right\} &= \mathrm{tr}\left\{\left[R_n - \frac{S^\intercal\, S}{m} + \frac{J_n}{n}\right]^{-1}\right\} - \mathrm{tr}\left(\frac{J_n}{n}\right)\\ &= \mathrm{tr}\left\{\left[R_n - \frac{S^\intercal\, S}{m} + \frac{J_n}{n}\right]^{-1}\right\} - 1, \end{aligned} \tag{11}$$

Thus, the objective function of (8) is reduced to $\mathrm{tr}\{\left[R_n - \frac{S^\intercal\, S}{m} + \frac{J_n}{n}\right]^{-1}\}$ which replaces (8a).

When all the eigenvalues of $\mathrm{pinv}[\mathcal{M}(\xi)]$ are positive which, for standard designs generally holds because of the semidefinite positiveness of the FIM, $\mathrm{ldet}\{[\mathrm{pinv}(\mathcal{M}(\xi))]^{-1}\}$ is computed as the sum of logarithms of all elements of the diagonal of $\mathcal{U}^{-1}(\xi)$. We note that the designs addressed herein, especially the extended designs, lead to singular FIMs; thus, $\mathrm{rank}\{\mathrm{pinv}[\mathcal{M}(\xi)]\} \leqslant n_c - 1$. When this occurs, the determinant of $\mathcal{M}(\xi)$ is null as is one of the eigenvalues. To overcome this issue we replace the determinant by the pseudo-determinant of the $\mathcal{M}(\xi)$ which is given as the product of the non-zero eigenvalues (Holbrook, 2018). Similarly, the maximum eigenvalue included in the E–optimality problem is chosen in the set of non-zero values of $\boldsymbol{\lambda}\{\mathrm{pinv}[\mathcal{M}(\xi)]\}$. No additional modifications in the procedure are required since the semi-definite positiveness of the FIM guarantees that $\boldsymbol{\lambda}\{\mathrm{pinv}[\mathcal{M}(\xi)]\} \geqslant 0$.

3.3 Constrained optimal designs for dose escalation trials

This Section considers constrained designs and introduces the modifications to optimal design problems (8–10) required for their construction, see Bailey (2009). The constraints result from imposing desirable characteristics on the design in terms of safety, equity and diversity. We consider two constrained designs: i. *strict halving* designs (Senn et al., 2007); and ii. *uniform halving* designs.

We first consider the strict halving design which requires halving the number of individuals allocated to the same dose in two consecutive cohorts ($k-1$ and k). The constraints used for representing this setup are:

$$s_{k,i} = \frac{s_{k-1,i}}{2},\ k \in \{2,\cdots,\log(R_{n,i,i})/\log(2)\},\ i \in \{1\}, \tag{12a}$$

$$s_{k,i} = 1,\ k \in \{\log(R_{n,i,i})/\log(2)+1,\cdots,n_c\},\ i \in \{1\}, \tag{12b}$$

$$s_{k,i} = \frac{s_{k-1,i}}{2},\ k \in \{i,\cdots,\min[i+\log(R_{n,i,i})/\log(2)-2,n_c]\},\ i \in \{2,\cdots,n\}, \tag{12c}$$

$$s_{k,i} = 1,\ k \in \{i,\cdots,\min[i+\log(R_{n,i,i})/\log(2)-1,n_c]\},\ i \in \{2,\cdots,n\}, \tag{12d}$$

Eqs. (12a) and (12c) halve the number of individuals allocated to the same dose in two cohorts ($k-1$ and k) and hold whenever the values of $s_{k,i} \geqslant 1$. Eqs. (12b) and (12d) are for values of $s_{k,i} = 1$ occurring when halving would lead to fractional values below 1. Here, the operator $\min[a,b]$ is used to represent the minimum value between a and b, n_k is the number of doses tested in a cohort with $n_k = k+1$ for all the cohorts except for the last in extended designs where $n_k = k$. The halving procedure cannot be applied to the extra cohort in extended designs since the resulting optimization problem is infeasible; consequently, we limit its application to standard designs and to all the cohorts but the last in extended designs.

We now consider the uniform halving design. Here, the constraints are that the number of individuals allocated to an eligible dose in a cohort is at least one and the partial replication of order k for dose i is larger (or equal) to that for dose $i+1$. The constraints describing these designs are:

$$s_{k,i} \geqslant 1,\ k \in \{2,\cdots,n_c\},\ i \in [\![k+1]\!], \tag{13a}$$

$$r_{k,i} \geqslant r_{k,j},\ k \in \{2,\cdots,n_c\},\ i,j \in [\![k+1]\!],\ i<j, \tag{13b}$$

where (13a) is to have $s_{k,i} > 0$ for $i \leqslant k+1$ and (13b) is to make the partial replication of size k non-decreasing.

In summary, strict halving optimal designs are obtained solving the optimal design problems (8–10) coupled with the constraints (12), and uniform halving designs are obtained solving (8–10) with the constraints (13).

3.4 Numerical strategy and solvers

The formulations in previous sections are coded in the GAMS environment (GAMS Development Corporation, 2013a). GAMS is a general modeling system that supports mathematical programming applications in several areas. Upon execution, the

code describing the mathematical program is automatically compiled, symbolically transcribed into a set of numerical structures; all the information regarding the gradient and matrix Hessian is generated using the automatic differentiation tool and made available to the solver. We provide a sample of such a code in the Supplementary Material, see Appendix C.

Our design problems may have multiple local optima and to guarantee that a global optimum is found, a global solver must be employed. An example of a global solver is BARON. It implements deterministic global optimization algorithms that combine spatial branch-and-bound procedures and bound tightening methods via constraint propagation and interval analysis in a *branch-and-reduce* technique (Tawarlamani and Sahinidis, 2002). However, global optimization solvers require a long computational time compared to local solvers and this may limit their utilization to small and average sized problems. To reduce the CPU required we, instead, used a local MINLP solver – SBB (GAMS Development Corporation, 2013b) – to handle design problems addressed in §4 and §6.2. SBB uses CONOPT as a NLP solver to handle the relaxed nonlinear programs (Drud, 1985) and CPLEX to solve the mixed integer linear programs (GAMS Development Corporation, 2013b).

To reduce the CPU time, we provide an *consistent* initial solution to the MINLP solver; by consistent we mean an initial solution that satisfies all the (equality and inequality) constraints of the design problem. To construct a initial consistent design, we first allocate the individuals to cohorts using equality constraints (12). Next, we compute the design matrix, the FIM and its Moore-Penrose generalized inverse (when required) and let the solver iterate until it converges to the optimum.

All computations in this paper were carried using an AMD 8-Core processor machine running 64 bits Windows 10 operating system with $3.80\,\mathrm{GHz}$. In all problems, the relative and absolute tolerances used to solve the MINLP problems were set to 1×10^{-5}.

4 Optimal designs for dose escalation trials

Here we present A–, E– and D–optimal designs obtained for dose escalation trials. As a first example we consider a standard design with 5 doses (one of them being placebo), 4 cohorts and 32 subjects, i.e. $n = 5$, $n_c = 4$ and $N = 32$. The corresponding extended designs are solved for $n = 5$, $n_c = 5$ and $N = 40$. For simplification we call this scenario Setup 1. It was first used by Haines and Clark (2014) to demonstrate their proposed algorithms. For both standard and extended dose escalation experiments we determine unconstrained, strict and uniform halving optimal designs.

To submit our formulations to a more challenging problem we consider the experiment that includes, for standard designs, 8 doses (including placebo), 7 cohorts and 112 subjects, i.e. $n = 8$, $n_c = 7$ and $N = 112$. The equivalent extended design has $n = 8$, $n_c = 8$ and $N = 128$. This is Setup 2, and we again find traditional, strict and uniform halving optimal designs. To improve the readability of our paper we postpone the presentation of the results for Setup 2 to Appendix A of SM, but incorporate the trends observed in earlier discussions.

The absolute and relative tolerances imposed on the MINLP solver were set to 1×10^{-6} and 1×10^{-7}, respectively, in all design calculations. All computations in this paper were carried out using an AMD 8-Core processor machine running a 64 bits Windows 10 operating system with $3.80\,\mathrm{GHz}$.

Table 1 presents the results obtained for standard traditional designs for Setup 1 obtained with the formulations (8–10) for A–, E– and D–optimality criteria, respectively. Table 4 is for extended traditional designs. For clarification, in Tables "C" is to designate cohort and "T" for treatment (or dose). The integer values forming the design correspond to S matrices that optimize the design criteria. Three aspects are noticeable from comparisons of the designs: i. the results obtained with our formulations are in good agreement with those prescribed by Bailey (2009); ii. they are also in good agreement with those of Haines and Clark (2014, Figure 2); iii. for most of the criteria the standard designs are not equal to extended designs without the last cohort, see Tables 1 and 4. That is, studies with more individuals and an additional cohort may lead to different allocation in the first n_c cohorts. This possibility was not explored by Bailey (2009). However, we note that when standard designs are equal to extended designs without the last cohort flexibility in implementation is increased; the decision to have an additional cohort is postponed. This increment of flexibility is provided by A–optimal designs for 4 and 5 cohorts, as well as by some standard strict halving designs in Tables 2 and 5. These trends extend to Setup 2; compare the results in Tables S1 and S4 in Appendix A of SM.

Tables 2 and 5 present the standard and extended strict halving designs for Setup 1. They show similar trends to those emerging from traditional designs. Tables 3 and 6 are for the standard and extended uniform halving designs. Practically, in most of the cases the trends again coincide with those of the traditional designs. Similar findings are observed for Setup 2, where Tables S1 and S4 present the results for standard and extended traditional designs, respectively; Tables S2 and S5 are for standard and extended strict halving designs, respectively, and Tables S3 and S6 for standard and extended uniform halving designs. The comparison of the designs reveals several noticeable aspects: i. for most of the cases A– and D–optimal designs coincide, whereas the E–optimal designs are different (see Table 1); ii. the standard designs, with the exception of those obtained for the E–optimality criterion (see Table 1), do not include values $s_{k,i} = 0$ in the design matrix; contrarily, extended E–optimality designs of traditional kind include doses not administered in some cohorts (see Table 4); iii. the strict and uniform halving designs avoid the occurrence of zeros in the design matrix by construction, see Tables 5 and 3.

Table 7 displays the optima for each of the design problems solved. The efficiency is determined from optima with Eq. (6); the reference designs, ξ_p^*, $p \in \{\text{A–, E–, D–}\}$, are the traditional designs as they are unconstrained. To simplify the interpretation of the results, the first column is for the number cohorts and the second for the number of treatments, and we recall that standard designs have $n_c = n-1$ and extended designs assume $n_c = n$ with the placebo being one of the treatments. The third column indicates the type of design resulting from the presence of additional constraints. The fourth column lists the number of the table containing the experimental design found. The remaining columns present the optima for A–, E– and D–optimality criteria and the respective efficiencies.

The comparison of efficiencies in Table 7 reveals that i. the traditional designs are systematically more efficient than strict and uniform halving designs, as expected; ii. the uniform halving designs are more efficient than strict halving designs and in some cases they coincide with the traditional designs. Thus, strict halving designs are the most restrictive as they require halving the numbers of individuals allocated in successive cohorts; iii. all the strict and uniform halving designs have efficiencies above $90\,\%$, and for D–optimal designs the efficiencies are above $99\,\%$ which indicates that the loss of information is small if constrained designs are implemented.

The comparison of CPU times in Table 7 shows that: i. D–optimal designs are faster to calculate; ii. constrained designs are easier to find because the constraints (12) and (13) avoid the need of deepening the branch and bound; and iii. consistent initialization of the design problem is crucial to reduce the CPU time.

5 Optimal allocation of individuals in cohorts using prognostic factors

In this Section we address the problem of allocating the individuals to doses in a cohort using prognostic factors, i.e. additional covariates that might influence the response. We note that the optimal design for dose escalation is already available, i.e., the number of individuals to be allocated to each dose in each cohort. The interest now is in maximizing the amount of information gathered from optimally allocating doses to patients taking into account the prognostic factors.

In §5.1 we develop the model and the optimal design problem for allocation of the individuals. Then, in §5.2 we consider the sequential allocation of the individuals of cohorts as they enter the trial. This setup considers that the individuals of the same cohort enter the trial at different time instants but the lags can be infinitesimal. In Section 5.3 we consider the allocation of the complete set of individuals of a cohort; they are considered to enter the trial at the same time instant. The dose escalation optimal designs obtained with the formulations in §3 are assumed to be known and the individuals of each cohort are allocated with the goal of maximizing the amount of information extracted.

5.1 Model and FIM for individual allocations based on optimal design criteria

Here we establish the fundamentals for allocating individuals to doses using prognostic factors and a previously known dose escalation design. The approach relies on the ideas of Atkinson (1982, 2002) for allocating individuals in sequential designs using optimal design criteria.

Let the vector of prognostic factors be $\mathbf{z} \in \mathbb{R}^{n_f}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{n_f}$ be the coefficients quantifying the effect of factors $\mathbf{z}$ on individual responses; n_f is the number of prognostic factors. We first consider the generic linear response model for an individual's response, i.e.,

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\gamma}\,\mathbf{z}^{\intercal} + \boldsymbol{\epsilon}, \tag{14}$$

where $\mathbf{x}$ represents the effect of the dose on the response, the second term of the model is for the effect of the prognostic factors, and $\boldsymbol{\epsilon}$ is the observational noise.

Table 1 Standard traditional optimal designs for dose escalation ($n = 5$, $n_c = 4$ and $N = 32$).

	A–optimality					E–optimality					D–optimality				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
C1	4	4	0	0	0	3	5	0	0	0	4	4	0	0	0
C2	2	3	3	0	0	4	0	4	0	0	3	2	3	0	0
C3	2	1	2	3	0	1	2	1	4	0	1	2	2	3	0
C4	1	1	1	2	3	1	1	1	1	4	1	1	1	2	3

Table 2 Standard strict halving optimal designs for dose escalation ($n = 5$, $n_c = 4$ and $N = 32$).

	A–optimality					E–optimality					D–optimality				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
C1	4	4	0	0	0	4	4	0	0	0	4	4	0	0	0
C2	2	2	4	0	0	2	2	4	0	0	2	2	4	0	0
C3	1	1	2	4	0	1	1	2	4	0	1	1	2	4	0
C4	1	1	1	2	3	1	1	1	2	3	1	1	1	2	3

Table 3 Standard uniform halving optimal designs for dose escalation ($n = 5$, $n_c = 4$ and $N = 32$).

	A–optimality					E–optimality					D–optimality				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
C1	4	4	0	0	0	4	4	0	0	0	4	4	0	0	0
C2	3	2	3	0	0	3	1	4	0	0	3	2	3	0	0
C3	1	2	2	3	0	2	1	1	4	0	1	2	2	3	0
C4	1	1	1	1	4	1	1	1	1	4	1	1	1	1	4

Table 4 Extended traditional optimal designs for dose escalation ($n = 5$, $n_c = 5$ and $N = 40$).

	A–optimality					E–optimality					D–optimality				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
C1	4	4	0	0	0	4	4	0	0	0	4	4	0	0	0
C2	2	3	3	0	0	3	3	2	0	0	2	3	3	0	0
C3	2	1	2	3	0	1	2	1	4	0	2	1	2	3	0
C4	1	1	1	2	3	2	1	0	1	4	1	1	2	2	2
C5	1	1	1	2	3	0	1	2	1	4	1	1	1	2	3

Table 5 Extended strict halving optimal designs for dose escalation ($n = 5$, $n_c = 5$ and $N = 40$).

	A–optimality					E–optimality					D–optimality				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
C1	4	4	0	0	0	4	4	0	0	0	4	4	0	0	0
C2	2	2	4	0	0	2	2	4	0	0	2	2	4	0	0
C3	1	1	2	4	0	1	1	2	4	0	1	1	2	4	0
C4	1	1	1	2	3	1	1	1	2	3	1	1	1	2	3
C5	1	1	1	2	3	1	1	1	1	4	1	1	1	2	3

Table 6 Extended uniform halving optimal designs for dose escalation ($n = 5$, $n_c = 5$ and $N = 40$).

	A–optimality					E–optimality					D–optimality				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
C1	4	4	0	0	0	5	3	0	0	0	4	4	0	0	0
C2	3	2	3	0	0	3	1	4	0	0	3	2	3	0	0
C3	1	2	2	3	0	1	1	1	5	0	1	2	2	3	0
C4	1	1	1	2	3	1	1	1	1	4	1	1	1	2	3
C5	1	1	1	2	3	2	1	1	1	3	1	1	2	2	2

Table 7 Results for standard and extended designs presented in Tables 1-6 and S1-S6 in Appendix A of SM.

				Optimality criteria								
				A–optimality			E–optimality			D–optimality		
n_c	n	Design	Table	Opt	Eff_A (%)	CPU (s)	Opt	Eff_E (%)	CPU (s)	Opt	Eff_D (%)	CPU (s)
4	5	Trad	1	1.9684	100.00	4.56	0.7211	100.00	8.78	-3.0846	100.00	1.20
4	5	StricH	2	1.9747	99.68	3.78	0.7388	97.60	7.82	-3.0462	99.73	0.66
4	5	UnifH	3	1.9781	99.51	1.67	0.7211	100.00	5.42	-3.0374	99.66	0.82
5	5	Trad	4	1.6459	100.00	4.08	0.6000	100.00	12.37	-3.7338	100.00	1.43
5	5	StricH	5	1.6528	99.58	3.10	0.6050	99.17	8.98	-3.6951	99.80	1.05
5	5	UnifH	6	1.6459	100.00	3.21	0.6047	99.22	11.02	-3.7338	100.00	1.51
7	8	Trad	S1	1.7940	100.00	16.21	0.5633	100.00	34.50	-8.1128	100.00	3.08
7	8	StricH	S2	1.8612	96.39	10.89	0.5814	96.89	23.56	-7.9105	99.42	1.35
7	8	UnifH	S3	1.8024	99.53	6.75	0.5633	100.00	26.45	-8.0400	99.79	1.68
8	8	Trad	S4	1.5927	100.00	26.47	0.3991	100.00	35.65	-8.8835	100.00	3.65
8	8	StricH	S5	1.6087	99.01	12.54	0.4201	95.00	16.78	-8.7601	99.71	1.86
8	8	UnifH	S6	1.5931	99.97	13.09	0.4137	96.47	19.32	-8.8822	100.00	3.44

Trad - traditional design; StricH - strict halving design; UnifH - uniform halving design; Opt - optimum.

Without loss of generality, we present the fundamental ideas supposing there is a single prognostic factor, i.e. $n_f = 1$, $\mathbf{z}$ is a single element vector, $\mathbf{z} = (z_1)$, as is $\boldsymbol{\gamma}$, i.e., $\boldsymbol{\gamma} = (\gamma_1)$. Here, we consider the prognostic factor follows a discrete uniform distribution; however, without loss of generalization other setups can be considered, such as continuously distributed factors and factors represented by skewed underlying distributions.

The model for the response of individual i in cohort k is

$$\mathbb{E}(y_{k,i}) = x_{k,i} + \gamma_1 \, z_{1,i}, \qquad k \in [\![n_c]\!], \; i \in [\![m_k]\!], \tag{15}$$

where $y_{k,i}$ is the response, $x_{k,i}$ is the effect of the dose administered to individual i of cohort k, γ_1 is a nuisance parameter and $z_{1,i}$ is the (first) prognostic factor for individual i. In practice $x_{k,i}$ is the effect of a dose chosen from the set that can be administered in cohort k, i.e.

$$x_{k,i} = \sum_{j=1}^{n_k} v_{k,i,j} \, \tau_j, \qquad k \in [\![n_c]\!], \; i \in [\![m_k]\!], \tag{16}$$

where $v_{k,i,j}$ is a binary variable used to represent the allocation of the individual to a given dose and n_k is the number of treatments available for cohort k, with $n_k = k+1$ the number of different treatments available for that cohort. That is, $v_{k,i,j}$ is 1 if individual i is given dose j and 0 otherwise. The allocation must also ensure that $\sum_{j=1}^{n_k} v_{k,i,j} = 1$ for each individual. The variables $v_{k,i,j}$ provide a mathematical representation of the allocation decisions, see Williams (1999). The model (15) becomes

$$\mathbb{E}(y_{k,i}) = \sum_{j=1}^{n_k} v_{k,i,j} \, \tau_j + \gamma_1 \, z_{1,i}, \qquad k \in [\![n_c]\!], \; i \in [\![m_k]\!]. \tag{17}$$

The FIM of model (17) for cohort k is

$$\mathcal{M}_k(\xi) = \left(\begin{array}{c|c} D_k & \mathbf{h}(\mathbf{z}) \\ \hline [\mathbf{h}(\mathbf{z})]^\intercal & \zeta(\mathbf{z}) \end{array} \right), \tag{18}$$

where the $n_k \times n_k$ diagonal matrix D_k has elements $s_{k,i}$ obtained from the optimal dose escalation design. The vector $[\mathbf{h}(\mathbf{z})]^\intercal$ has n_k elements of the form

$$[\mathbf{h}(\mathbf{z})]^\intercal = \underbrace{\left(\textstyle\sum_{j=1}^{n_k} v_{k,i,1} z_{1,i} \; \cdots \; \sum_{j=1}^{n_k} v_{k,i,n_k} z_{1,i} \right)}_{n_k \text{ elements}}.$$

Furthermore, $\zeta(\mathbf{z}) = \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} v_{k,i,1} z_{1,i}^2$ is element (n_k+1, n_k+1) of the FIM and contains the sum of squares of the prognostic factor for all individuals to be allocated within cohort k.

The allocation aims at maximizing the information for estimating the contrasts $\tau_i - \tau_j$ in the presence of the nuisance parameter γ. When interest is in estimation of linear combinations of parameters as we have for (17) the $\mathrm{D_A}$-optimality criterion is appropriate (Atkinson et al., 2007; Sibson, 1974). The experimental design minimizes $\mathrm{ldet}\{L \, [\mathcal{M}_k(\xi)]^{-1} \, L^\intercal\}$ where L is a constant matrix. In our case this matrix contains the coefficients of the combinations of covariates determined by the minimization

and $\mathcal{M}_k(\xi)$ is given by Eq. (18). The generalization for the $\mathrm{A_A}$–optimality criterion is straightforward.

Here, interest is in determining the effects of the contrasts $\tau_i - \tau_j$, ignoring the effects of the nuisance factors $\mathbf{z}$, so that the coefficients of γ in the contrasts are zero. That is, L is a $(n_k - 1) \times (n_k + n_f)$ matrix

$$L = \left(A \middle| 0_{(n_k-1)\times n_f} \right) \tag{19}$$

where A is a $(n_k - 1) \times n_k$ matrix. One form of A has elements of the diagonal equal to -1, the elements of the line above the diagonal being $+1$ and all others zero. From Fedorov (1972, Theorem 2.2.4a) D–optimal designs are invariant with respect to any non-degenerate linear transformation of the parameters to be estimated. Let the contrasts specified by A above be represented as $\theta_i = \tau_i - \tau_{i+1}$; then Fedorov's result applies to $\mathrm{D_A}$–optimal designs under linear transformations of the θ_i. Further, in (19) $0_{(n_k-1)\times n_f}$ designates a $(n_k - 1) \times n_f$ matrix of 0's. The $\mathrm{D_A}$-optimal design satisfies

$$\xi_{\mathrm{D_A}} = \arg \min_{\xi \in \Xi_{\mathrm{D_A}}} \operatorname{ldet}\{L \left[\mathcal{M}(\xi)\right]^{-1} L^\intercal\}. \tag{20}$$

In (20), $\Xi_{\mathrm{D_A}}$ is the set of feasible $\mathrm{D_A}$–optimal designs. The criterion is minimized by choice of the binary variables $v_{k,i,j},\ k \in [\![n_c]\!],\ i \in [\![m_k]\!]$ representing the allocation of individuals to treatments in a cohort.

5.2 Allocation of individuals on a *per arrival basis*

In this Section we formalize the algorithm for allocating the individuals to treatments as they enter the trial. They are allocated sequentially and their treatment may start before the cohort is completed. The goal is to maximize the information obtained from each new entry given knowledge of the prognostic factors. For simplicity we again consider a single prognostic factor; the dose escalation experimental design is already available, and the matrices $D_k,\ k \in [\![n_c]\!]$ can be formed.

To distinguish between prognostic factors of the individual arriving (to be allocated a treatment) and individuals already allocated, we use $z_{1,i},\ \in [\![m_k]\!]$ for the former, and designate the latter by $\omega_{k,i,j}$, where k is the cohort identifier, i the individual identifier, and j the identifier of the dose which they were allocated.

To demonstrate the mechanics of the procedure we consider that cohort k is initially empty, i.e.,

$$[\mathbf{h}(\mathbf{z})]^\intercal = \underbrace{\left(0 \cdots 0\right)}_{n_k}$$

and $\zeta(\mathbf{z}) = 0$. The first individual arrives. Allocation to dose j is represented by setting $v_{k,i,j} = 1$ and all other $v_{k,i,\ell} = 0$ for $\ell \neq j,\ j,\ \ell \in [\![n_k]\!]$. For each possible allocation (i, j), $[\mathbf{h}(\mathbf{z})]^\intercal$ and $\zeta(\mathbf{z})$ are updated with the value of $z_{1,i}$. Then, the respective FIMs are computed (via Eq. (18)), and inverted. Next, the metrics $\varphi = \operatorname{ldet}\{L \left[\mathcal{M}(\xi)\right]^{-1} L^\intercal\}$ for each one are determined and saved. When all possible doses have been considered, a vector with n_k values of $\boldsymbol{\varphi}$ is available. The minimum

value of $\boldsymbol{\varphi}$ is chosen and the individual is allocated to dose $j = \{\ell : \ \varphi_\ell = \min \boldsymbol{\varphi}\}$. Next, the variables $\boldsymbol{\omega}$ containing the sums of prognostic factors of individuals already allocated are updated with $z_{1,i}$. The procedure is iterated for all individuals of the cohort and in each iteration $[\mathbf{h}(\mathbf{z})]^\intercal$ and $\zeta(\mathbf{z})$ are updated to accumulate the prognostic factors of individuals already allocated to each treatment, i.e.

$$[\mathbf{h}(\mathbf{z})]^\intercal = \left(\omega_{k,i,1} \cdots \omega_{k,i,n_k}\right).$$

Similarly, $\zeta(\mathbf{z})$ is updated with the new information and becomes $\sum_{\iota=1}^{i-1} \sum_{j=1}^{n_k} \omega_{k,\iota,j}^2 + z_{1,i}^2$. Then, the FIM is also updated with the new information. Finally, the procedure is iterated for all cohorts. Algorithm 1 presents the pseudo-code for the numerical implementation of this enumeration procedure.

Algorithm 1 Algorithm: allocation on arrival for individuals in a cohort.

procedure ALLOCATEINDIVIDUALINCOHORT($S, \mathbf{z}, \mathbf{m}, n_c, n$)
  for $k \in [\![n_c]\!]$ **do** ▷ Iterate cohorts
    Set elements $(1 : k+1, 1 : k+1)$ of $\mathcal{M}(\xi) \leftarrow \mathrm{diag}(\mathbf{s}_k)$ ▷ Initialize the FIM
    Set L
    for $j \in [\![m_k]\!]$ **do** ▷ Iterate individuals
      for $i \in [\![k+1]\!]$ **do** ▷ Iterate treatments
        Allocate individual j to i^{the} treatment
        Set elements $(k+2, i)$ of $\mathcal{M}(\xi) \leftarrow \sum_j \omega_{k,i,j} + z_{1,i}$
        Set elements $(i, k+2)$ of $\mathcal{M}(\xi) \leftarrow \sum_j \omega_{k,i,j} + z_{1,i}$
        Set element $(k+2, k+2)$ of $\mathcal{M}(\xi) \leftarrow \sum_j \sum_i \omega_{k,j,i}^2 + z_{1,i}^2$
        $\varphi_{k,i,j} \leftarrow \det(L\,[\mathcal{M}(\xi)]^{-1}\,L^\intercal)$
      end for
      Find $j = \{\ell : \ \varphi_\ell = \min \boldsymbol{\varphi}\}$
      Allocate i^{th} individual of k^{th} cohort to j^{th} treatment
      Update $[\mathbf{h}(\mathbf{z})]^\intercal$, $\zeta(\mathbf{z})$ and $\mathcal{M}(\xi)$
    end for
  end for
end procedure

5.3 Allocation of individuals on a *per cohort basis*

Now we address the problem of allocating all the individuals within a cohort, assuming they enter the trial at the same time and the knowledge of their prognostic factors is available. As in §5.2, we consider that the dose escalation design is available, so that the diagonal matrices $D_k, \ k \in [\![n_c]\!]$ in (18) as well as the contrast matrices L can be constructed.

The problem of finding $\mathrm{D_A}$–optimal allocation schemes on $\Xi_{\mathrm{D_A}}$ for each cohort is defined in (20). The formulation of the optimization problem is

$$\min_{\mathbf{v}} \ \mathrm{ldet}\{L\,[\mathcal{M}(\xi)]^{-1}\,L^\intercal\} \tag{21a}$$

$$\text{s.t. } \mathcal{M}_k(\xi) = \left(\begin{array}{ccc|c} s_{k,1} & \cdots & 0 & \sum_{i=1}^{m_k} v_{k,i,1} z_{1,i} \\ 0 & \ddots & 0 & \vdots \\ 0 & \cdots & s_{k,n_k} & \sum_{i=1}^{m_k} v_{k,i,n_k} z_{1,i} \\ \hline \multicolumn{3}{c|}{\sum_{i=1}^{m_k} v_{k,i,1} z_{1,i} \cdots \sum_{i=1}^{m_k} v_{k,i,n_k} z_{1,i}} & \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} v_{k,i,j} z_{1,i}^2 \end{array}\right) \tag{21b}$$

$$\sum_{j=1}^{n_k} v_{k,i,j} = 1, \quad i \in [\![m_k]\!] \tag{21c}$$

$$\sum_{i=1}^{m_k} v_{k,i,j} = s_{k,j}, \quad j \in [\![n_k]\!] \tag{21d}$$

$$v_{k,i,j} \in \{0,1\}. \tag{21e}$$

Equation (21a) is the objective function, (21b) is to form the FIM for each cohort, (21c) ensures that each individual is allocated to a single dose, (21d) that the dose escalation optimal design previously determined is satisfied, and (21e) is to set the decision variables $\mathbf{v}$ as binary variables. The design ξ is formed by the allocation variables $v_{k,i,j}$ that maximize the objective function. The inversion of $\mathcal{M}(\xi)$ is handled algebraically using the approach described in §3.1, see Duarte et al. (2021, 2020). Specifically, the algebraic operations as well as conditions assuring the stability of the Cholesky decomposition are included in the optimal design problem as constraints. The problem falls into the MINLP class and is solved with the global solver used for the construction of dose escalation designs.

6 Optimal allocation of individuals to doses

In this Section we demonstrate the application of the methodologies introduced in §5. First, in Section 6.1 we consider the allocation within cohorts on a *per individual arrival basis*. Then, in Section 6.2 we consider the simultaneous allocation of the complete set of individuals belonging to a cohort. In both scenarios we use the dose escalation traditional D–optimal designs obtained for Setups 1 and 2 as a basis for analysis. These appear in Tables 1, 4, S1, and S4 (the last two are in Appendix A of SM). Tables 1 and S1 display standard designs, and the other two are for extended designs.

For demonstration we consider a single discrete prognostic factor which takes the values 1, 2 or 3, simulated using a uniform integer random number generator.

6.1 Allocation of individuals in cohorts on arrival at different times

In this Section we exploit Algorithm 1 in §5.2 to allocate the individuals as they sequentially enter the trial.

Figure 1 shows the optimal allocation for the D–optimal standard design for Setup 1 (i.e., $n = 5$, $n_c = 4$ and $N = 32$). We notice that the allocation rule based on individual arrivals may not satisfy the dose escalation optimal design as

it aims at maximizing the information available which is partial when building the cohort. An individual with a particular prognostic factor is allocated to a given dose because the amount of information they bring, integrated with that already available, is maximized. However, once more individuals have been enrolled, the amount of information available is higher, which will influence the allocation to doses. In addition, Figure S1 in Appendix B of SM presents the sequential optimal allocation obtained for extended traditional D–optimal design ($n = 8$, $n_c = 8$ and $N = 128$), see Table S4.

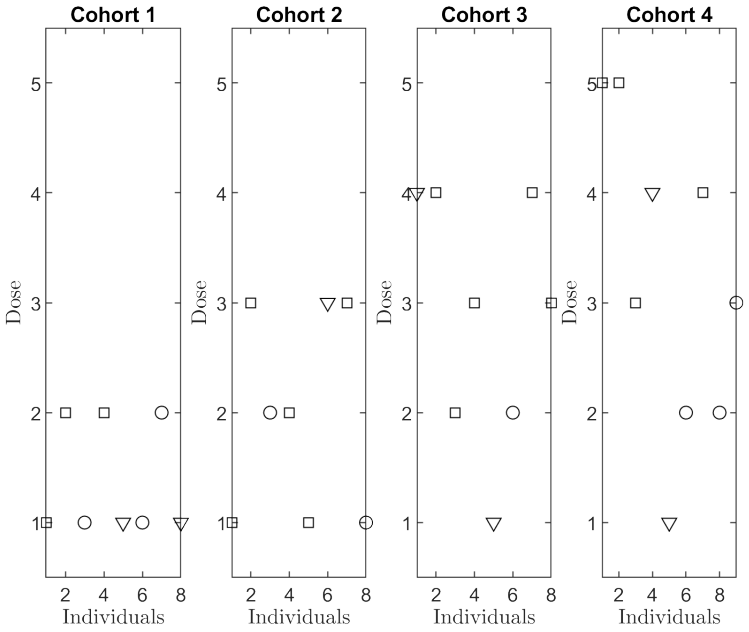

Figure 1 Sequential allocation of individuals to doses on arrival for the standard traditional D–optimal design obtained for $n = 5$, $n_c = 4$ and $N = 32$, see Table 1. Symbols: ○ - prognostic factor=1; ▽ - prognostic factor = 2; and □ - prognostic factor = 3.

To analyze the performance of the optimal sequential allocation strategy proposed we compare its efficiency with randomized allocation, equivalent to allocation in ignorance of the values of the covariate. We consider a set of individuals with the same prognostic factors and randomly vary the order of allocation in each cohort. Then, we determine $\det\{L\ [\mathcal{M}(\xi)]^{-1}\ L^{\intercal}\}$ for each cohort as well as the efficiency using Eq. (6c) and the optimal sequential allocation produced by Algorithm 1 for reference. Finally, we computed the overall efficiency of the randomized allocation as the geometric mean of the efficiencies for all the cohorts. To eliminate the effects of biased randomized allocations we simulate the procedure 1000 times and in each one a different randomly chosen allocation order is used. The results are in Table 8 (columns 6-7), and the optimal sequential allocation (used for reference in efficiency

computations) is advantageous; the *per individual* efficiencies of the randomized allocation schemes are below $100\,\%$.

The measure of efficiency $\text{eff}_{\text{D}}(\xi)$ responds like the variance of the estimate of a single parameter to an increase in sample size, e.g. halving when N is doubled. The effective number of trials for some allocation relative to the optimum design is $N\ \text{eff}_{\text{D}}(\xi)$. As a measure of the loss after N trials due to departure from optimality we therefore use

$$\mathcal{L}_N = N\,[1 - \text{eff}_{\text{D}}(\xi)], \tag{22}$$

a definition which stresses dependence of the loss on N. For randomized designs $\mathcal{L}_N$ is a random variable. The results of Smith (1984a,b) provide asymptotic values $\mathcal{L}_\infty$ for the expected value of the loss for sequential allocation to two treatments. The loss in using random allocation as against the optimal experimental design is the number of covariates n_f. Burman (1996) and Atkinson (2002) used simulation to study small sample properties of the expected value of loss. In the initial stages of a sequential trial, imbalance may be relatively high and the loss may be far from its asymptotic value. It is not clear whether this asymptotic relationship holds for our designs in cohorts when the number of cohorts increases with N. Simulations briefly mentioned in Atkinson (2002) show that, for random allocation to three treatments, the loss is again asymptotically n_f.

For the results in the first line of Table 8 for Setup 1 with $N = 32$ the two efficiencies give losses of 1.40 and 2.14. In line 3 for Setup 2 with $N = 112$ the losses are 1.66 and 2.81. In this example there is a slight increase of loss with N.

6.2 Simultaneous allocation of individuals within cohorts

Now, we consider the allocation of individuals within cohorts assuming they enter the trial simultaneously, and use the formulation (21) in §5.3 to solve the problem. We also limit our analysis to a single prognostic factor, its generation following the same strategy as before. Before the allocation, the prognostic factors of all individuals are known. The allocation should maximize the information produced by the cohort, given the optimal dose escalation design previously obtained.

Figure 2 presents the optimal allocation for standard traditional D–optimal designs obtained for Setup 1, see Table 1. Figure S2 in Appendix B of SM is for the allocation in extended traditional designs obtained for Setup 2, see Table S4. To analyze the performance of the allocation scheme we use the approach described above for sequential allocation. We compared its efficiency with that of a randomized allocation and used 1000 simulated random allocation scenarios for comparison. The results are in Table 8 (columns 8-9) and show the advantages of simultaneous allocation based on optimal design criteria. Another indirect finding is that the simultaneous allocation in cohorts is more efficient than allocation on arrival. This conclusion emerges through use of the randomized allocation as reference in our comparisons.

The CPU time required by both algorithms to carry out the allocation for the various design setups is shown in Table 8. We note that: i. the CPU time required is below $15\,\text{s}$ for both algorithms; ii. the problem of allocating the individuals assuming they enter simultaneously is more sensitive to the number of cohorts and doses. This

Table 8 Comparison of the efficiency and CPU time of optimal allocation schemes based on optimal design criteria vs. randomized allocation (number of simulation scenarios: 1000; reference: optimality-based allocation approaches).

					Allocation scheme							
					per individual basis				*per cohort basis*			
					Optimal		Randomized		Optimal		Randomized	
n_c	n	N	Design	Table	Eff_D (%)	CPU (s)	Eff_D (%)	CPU (s)	Eff_D (%)	CPU (s)	Eff_D (%)	CPU (s)
4	5	32	Trad	1	100.00	10.16	95.63	1.15	100.00	3.41	93.31	1.15
5	5	40	Trad	4	100.00	10.98	98.33	1.32	100.00	5.75	94.34	1.32
7	8	112	Trad	S1	100.00	13.44	98.52	2.27	100.00	12.46	97.49	2.27
8	8	128	Trad	S4	100.00	14.06	98.78	2.55	100.00	16.74	97.78	2.55

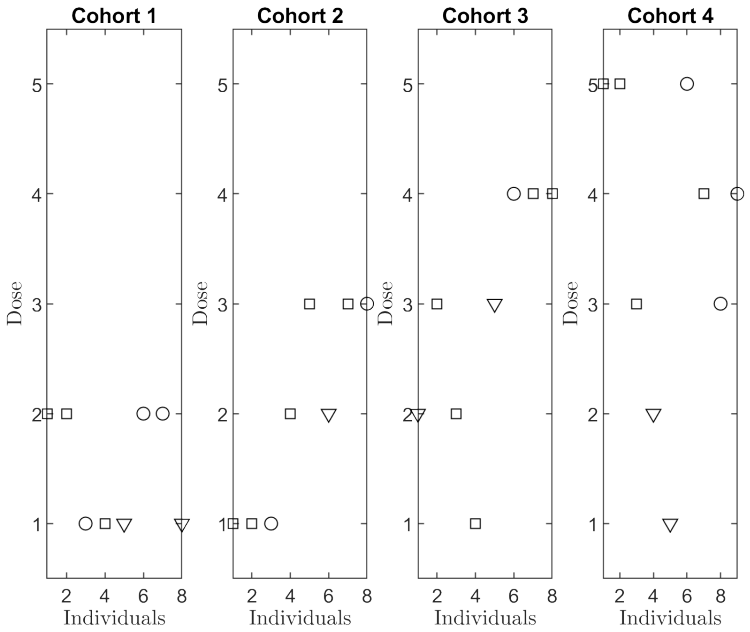

Figure 2 Simultaneous allocation within cohorts of individuals to doses for the standard traditional D–optimal design obtained for $n = 5$, $n_c = 4$ and $N = 32$, see Table 1. Symbols: ∘ - prognostic factor=1; ▽ - prognostic factor = 2; and ▫ - prognostic factor = 3.

trend was expected as the allocation scheme proposed in §6.1 is handled via simulation (see Algorithm 1) using MATLAB® and the simultaneous allocation problems were solved using the GAMS environment, employing the solver SBB as described in §3.4. In the former algorithm the optimization corresponds to the choice of the best allocation for each individual entering the trial in each cohort given the information available at the time. In the latter scheme the numerical solution is obtained by solving (21) to convergence.

For simultaneous optimal allocation problems, the solver easily finds a local optimum, and most of the CPU time is required to prove the global optimality because one may have multiple optima. This trend may occur especially when the prognostic factor(s) is(are) represented by integer values as we have here. Continuous prognostic factors with different values characterizing the individuals lead to problems with a single optimum; consequently global optimality is easier to ensure with reduced CPU time.

We also note that the constrained designs include additional conditions that allow breaking the symmetry of the optimal design problems as in Vo-Thanh et al. (2018) and so reducing CPU time. In general, the imposition of extra constraints simplifies optimization problems and reduces CPU time. The traditional design is typically more efficient but requires more CPU time since the feasibility region to be explored is larger.

Finally, to test the impact of skewed/rare prognostic factors on allocation algorithms, we considered a setup where the occurrence of $z_{1,i}$ in $\{1, 2, 3\}$ is random but the probability of sampling 1 is $50\,\%$, the probability of getting 2 is $35\,\%$, and the probability of 3 is $15\,\%$. This procedure was implemented by sampling from a discrete trinomial function, and this new set of data did not show significant impact on the CPU time required by both algorithms.

7 Conclusions

We have considered the optimal design of experiments for dose escalation in cohort based experiments – a problem conceptualized by Bailey (2009) – and have proposed general formulations for finding exact A–, E– and D–optimal designs, see §3. The problems fall into the MINLP class of problems and we use a global solver to handle them. Our formulations adapt to standard and extended (including an extra cohort) designs as well as non-constrained (traditional) and constrained (of strict and uniform halving type) designs. We demonstrate the application of the formulation with a wide battery of cases, see §4 and Appendix A of SM. Traditional designs are more efficient than constrained designs of strict and uniform halving kind but the difference is small. However, the latter have better properties such as diversity and the absence of individuals tested with a specific treatment in a cohort.

We have also considered the problem of optimally allocating the individuals to doses within the cohorts taking into account prognostic factors and the prescribed optimal dose escalation design. From our knowledge, this problem has never been addressed, and we again formalized it as an MINLP problem, where the goal is to maximize the information obtained from judiciously allocating the individuals. Two distinct strategies are proposed for its solution: i. a sequential allocation scheme in cohorts where the individuals enter the experiment at different time instants; and ii. a simultaneous allocation scheme in cohorts. The former problem was solved with an enumeration algorithm, see §5.2. An MINLP formulation is used to systematically handle the second problem, see Section 5.3. Then, in Section 6 and Appendix B of SM, we demonstrate its application and analyze the advantages of optimal allocation schemes using simulation. Both allocation strategies are more efficient than randomized allocation in terms of the amount of information produced; the allocation on *a per cohort basis* is more efficient than sequential allocation.

The implementation of dose escalation experiments with cohorts of different size may be more complex than equally replicated cohorts. However, numerical tests not reported in this paper show they are slightly more efficient. Typically, these plans require fewer individuals in the first cohorts who are included in cohorts where a larger number of treatments is administered.

Finally, we note that, in a thoughtful discussion at the end of Bailey (2009), there is a justification of the choice of analytical design construction as being more likely to be widely applied than a computer program. We wonder whether the balance has changed over the last 15 years. In particular, due to advances in computing software, we produce exact designs rather than the continuous designs suggested by Bailey. We have answered a further comment by providing designs for two criteria Bailey men-

tions (and one that is not mentioned). It is also suggested that designs found by her methods are (possibly) slightly sub-optimal. We have provided a method of checking their optimality. Indeed, in standard cases, the sub-optimality is slight. We believe that the importance of our computer-based contribution is that it allows the assessment of any proposed design as well as being useful for the kind of non-standard situations listed above. These include forms of allocation allowing for prognostic factors.

Acknowledgements We wish to thank to two anonymous reviewers for their comments and suggestions that helped to improve the quality of the paper.

References

Anderson, D.A.: Designs with partial factorial balance. The Annals of Mathematical Statistics **43**(4), 1333–1341 (1972)

Atkinson, A.C.: Optimum biased coin designs for sequential clinical trials with prognostic factors. Biometrika **69**(1), 61–67 (1982)

Atkinson, A.C.: The comparison of designs for sequential clinical trials with covariate information. Journal of the Royal Statistical Society: Series A (Statistics in Society) **165**(2), 349–373 (2002). DOI https://doi.org/10.1111/1467-985X.00564

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

Babb, J., Rogatko, A., Zacks, S.: Cancer phase I clinical trials: efficient dose escalation with overdose control. Statistics and Medicine **10**(10), 1103–1120 (1998)

Bailey, R.A.: Designs for dose–escalation trials with quantitative responses. Statistics in Medicine **28**(30), 3721–3738 (2009). DOI https://doi.org/10.1002/sim.3646

Begg, C.B., Iglewicz, B.: A treatment allocation procedure for sequential clinical trials. Biometrics **36**(1), 81–90 (1980)

Boer, E., Hendrix, E.: Global optimization problems in optimal design of experiments in regression models. Journal Global Optimization **18**, 385–398 (2000)

Burman, C.F.: On sequential treatment allocations in clinical trials. Ph.D. thesis, Chalmers of University Technology, Göteborg (1996)

Cheung, Y.K.: Coherence principles in dose-finding studies. Biometrika **92**(4), 863–873 (2005)

Dette, H., Bretz, F., Pepelyshev, A., Pinheiro, J.: Optimal designs for dose-finding studies. Journal of the American Statistical Association **103**(483), 1225–1237 (2008)

Dror, S., Faraggi, D., Reiser, B.: Dynamic treatment allocation adjusting for prognostic factors for more than two treatments. Biometrics **51**(4), 1338–1343 (1995)

Drud, A.: `CONOPT`: A GRG code for large sparse dynamic nonlinear optimization problems. Mathematical Programming **31**, 153–191 (1985)

Duarte, B.P.M., Atkinson, A.C., Granjo, J.F.O., Oliveira, N.M.C.: Optimal design of experiments for implicit models. Journal of the American Statistical Association **forthcoming**(0), 1–14 (2021). DOI 10.1080/01621459.2020.1862670

693 Duarte, B.P.M., Granjo, J.F.O., Wong, W.K.: Optimal exact designs of experiments
694 via Mixed Integer Nonlinear Programming. Statistics and Computing **30**, 93–112
695 (2020)
696 Fedorov, V.V.: Theory of Optimal Experiments. Academic Press (1972)
697 GAMS Development Corporation: GAMS - A User's Guide, GAMS Release 24.2.1.
698 GAMS Development Corporation, Washington, DC, USA (2013a)
699 GAMS Development Corporation: GAMS - The Solver Manuals, GAMS Release
700 24.2.1. GAMS Development Corporation, Washington, DC, USA (2013b)
701 Haines, L.M., Clark, A.E.: The construction of optimal designs for dose–escalation
702 studies. Statistics and Computing **24**(1), 101–109 (2014). DOI https://doi.org/10.
703 1007/s11222-012-9356-2
704 Hedayat, A.S., Jacroux, M., Majumdar, D.: Optimal designs for comparing test treat-
705 ments with controls. Statistical Science **3**(4), 462–476 (1988)
706 Holbrook, A.: Differentiating the pseudo determinant. Linear Algebra and its Appli-
707 cations **548**, 293–304 (2018). DOI https://doi.org/10.1016/j.laa.2018.03.018
708 Huang, B., Chapell, R.: Three-dose-cohort designs in cancer phase I trials. Statistics
709 in Medicine **27**(12), 2070–2093 (2008). DOI https://doi.org/10.1016/10.1002/sim.
710 3054
711 Kaye, R., Wilson, R.: Linear Algebra. Oxford Science Publications. Oxford Univer-
712 sity Press (1998)
713 Lastusilta, T., Bussieck, M., Westerlund, T.: Comparison of some high-performance
714 MINLP solvers. Chem Engng Trans. **11**, 125–130 (2007)
715 Le Tourneau, C., Lee, J.J., Siu, L.L.: Dose escalation methods in phase I cancer clini-
716 cal trials. JNCI: Journal of the National Cancer Institute **101**(10), 708–720 (2009).
717 DOI 10.1093/jnci/djp079
718 Neuenschwander, B., Branson, M., Gsponer, T.: Critical aspects of the Bayesian ap-
719 proach to phase I cancer trials. Statistics in Medicine **27**, 2420–39 (2008)
720 O'Brien, R.M.: Mixed models, linear dependency, and identification in age-period-
721 cohort models. Statistics in Medicine **36**(16), 2590–2600 (2017)
722 O'Quigley, J., Pepe, M., Fisher, L.: Continual reassessment method: A practical de-
723 sign for phase 1 clinical trials in cancer. Biometrics **46**(1), 33–48 (1990)
724 Pearce, S.C.: The agricultural field experiment. A statistical examination of theory
725 and practice. John Wiley & Sons Ltd., Chichester (1983)
726 Pukelsheim, F.: Optimal Design of Experiments. SIAM, Philadelphia (1993)
727 Rosa, S., Harman, R.: Optimal approximate designs for comparison with control in
728 dose–escalation studies. TEST **26**, 638–660 (2017). DOI https://doi.org/10.1007/
729 s11749-017-0529-3
730 Searle, S.R.: On inverting circulant matrices. Linear Algebra and its Applications **25**,
731 77–89 (1979). DOI https://doi.org/10.1016/0024-3795(79)90007-7
732 Senn, S.: Statistical Issues in Drug Development. 2nd edn. John Wiley & Sons, Ltd,
733 Chichester, England (2007). DOI 10.1002/9780470723586
734 Senn, S., Amin, D., Bailey, R.A., Bird, S.M., Bogacka, B., Colman, P., Garrett, A.,
735 Grieve, A., Lachmann, P.: Statistical issues in first-in-man studies. Journal of the
736 Royal Statistical Society: Series A (Statistics in Society) **170**(3), 517–579 (2007).
737 DOI https://doi.org/10.1111/j.1467-985X.2007.00481.x

738 Sibson, R.: D_A-optimality and duality. In: Gani, J., Sarkadi, K., Vincze, I. (eds.)
739 Progress in Statistics, Vol.2 – Proc. 9th European Meeting of Statisticians, Bu-
740 dapest, pp. 677–692. North-Holland, Amsterdam (1974)
741 Smith, R.L.: Properties of biased coin designs in sequential clinical trials. The Annals
742 of Statistics **12**, 1018—1034 (1984a)
743 Smith, R.L.: Sequential treatment allocation using biased coin designs. Journal of the
744 Royal Statistical Society, Series B **46**, 519—543 (1984b)
745 Storer, B.E.: Design and analysis of phase I clinical trials. Biometrics **45**(3), 925–937
746 (1989)
747 Tawarlamani, M., Sahinidis, N.: Convexification and Global Optimization in Con-
748 tinuous and Mixed Integer Nonlinear Programming. 1st. edn. Kluwer Academic
749 Pusblishers, Dordrecht (2002)
750 Vo-Thanh, N., Jans, R., Schoen, E.D., Goos, P.: Symmetry breaking in mixed integer
751 linear programming formulations for blocking two-level orthogonal experimental
752 designs. Computers & Operations Research **97**, 96–110 (2018). DOI https://doi.
753 org/10.1016/j.cor.2018.04.001
754 Williams, H.P.: Model Building in Mathematical Programming. 4th. edn. John Wiley
755 & Sons, Ltd., Chichester (1999)