RESEARCH ARTICLE

Optimum Designs for Clinical Trials in Personalized Medicine when Response Variance Depends on Treatment

Belmiro P.M. Duarte[a,b,c], Anthony C. Atkinson[d]

[a]Polytechnic Institute of Coimbra, ISEC, Rua Pedro Nunes, 3030–199 Coimbra, Portugal;
[b]INESC Coimbra, Universidade de Coimbra, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal;
[c]CERES, Universidade de Coimbra, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal;
[d]Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom.

ARTICLE HISTORY
Compiled June 28, 2024

Abstract
We study optimal designs for clinical trials when the value of the response and its variance depend on treatment and covariates are included in the response model. Such designs are generalizations of Neyman allocation, commonly used in personalized medicine when external factors may have differing effects on the response depending on subgroups of patients. We develop theoretical results for D-, A-, E- and D_A-optimal designs and construct Semidefinite Programming (SDP) formulations that support their numerical computation. D-, A-, and E-optimal designs are appropriate for efficient estimation of distinct properties of the parameters of the response models. Our formulation allows finding optimal allocation schemes for a general number of treatments and of covariates. Finally, we study frequentist sequential clinical trial allocation within contexts where response parameters and their respective variances remain unknown. We illustrate, with a simulated example and with a redesigned clinical trial on the treatment of neuro-degenerative disease, that both theoretical and SDP results, derived under the assumption of known variances, converge asymptotically to allocations obtained through the sequential scheme. Procedures to use static and sequential allocation are proposed.

KEYWORDS
Optimal designs; Neyman allocation; Covariates; Semidefinite programming; Sequential allocation.

1. Motivation

In the customary model for clinical trials, for example Rosenberger and Lachin (2016); Rosenberger and Sverdlov (2008), the expected response to treatment includes a linear model to allow for the effect of prognostic factors, or covariates, and an additive constant for the effect of the treatment. There is a single model for all patients. With normal response, with which we are chiefly concerned, the errors are usually assumed to be independent and normally distributed, with constant variance. The test with the highest power for the difference of two treatments then has equal allocation to the treatments.

The simple linear model does not hold for the forms of personalized medicine with which we are concerned. One departure is due to patient-treatment interaction. Yang *et al.* (2024) discuss a sequential Phase III trial to compare *Anastrozole* versus *Tamoxifen* for cancer treat-

Contact: Belmiro P.M. Duarte. Email: bduarte@isec.pt

ment in postmenopausal women. The patients were randomly assigned (1:1) to receive either oral *Tamoxifen* or *Anastrozole* per day for 5 years. Margolese *et al.* (2016) reported the primary results from this study. A total of 3104 patients were enrolled; *Anastrozole* was found superior to *Tamoxifen* in the younger than 60-year-old group, but in the 60 and older group *Tamoxifen* was the preferred treatment. Different models of response are thus required for the younger and older patients.

A second set of examples for Phase III trials is mentioned by Hyun *et al.* (2016). They recall the observed heterogeneity of the results of patients' response to treatments in a stroke prevention trial (The Stroke Prevention in Atrial Fibrillation Investigators, 1990); a difference between the *Aspirin* treatment group and the placebo group in the number of strokes was found in patients receiving anticoagulation, but not among patients without anticoagulation therapy. If this patient-treatment interaction had not been noticed, *Aspirin* would have been recommended for the general population. Here again a different response-treatment relationship is required for the two groups. There may then be sets of covariates for which one treatment is to be preferred and other sets of covariates for which another treatment is preferred. When the covariates of the patient are known, the case in sequential clinical trials, it is then possible to choose the best treatment for that particular patient. Even if these effects are strong, there may also be regions in which there is little to choose between treatments and some other criterion, such as cost, will be the major consideration.

At present, numerous trials incorporating personalized medicine practices are being utilized for a variety of purposes, including optimizing the selection of the most effective procedures to implement (Burnett *et al.*, 2020).

It is reasonable to expect that not only will the response relationship vary between subgroups, but that there may also be a difference in the variance of the response between subgroups. In the major simulations in their papers, both Yang *et al.* (2024) and Hyun *et al.* (2016) consider logistic responses, so that there is no independent variance parameter to be varied between subgroups. However, both write their regression models for patient-treatment interaction with a variance independent of subgroup. In our paper, both the model for the response and the variance of the observations may depend upon treatment allocation.

Throughout we are concerned with designs for sequential clinical trials. To this end, we use the methods of optimal experimental design to obtain procedures providing efficient parameter estimates and so powerful tests of hypotheses. In order to understand the properties of these designs, we start by exploring the properties of non-sequential designs for treatment-interaction models. For example, even if only the variances differ in the two (or more) treatment groups, generalizations of the results of Neyman (1934) to regression show that, in general, the treatment allocation will be skewed, rather than 1:1.

There is an appreciable computational load, both in finding the optimal designs and in simulating the clinical trial. We provide details of our algorithms in the hope that they will be of use in the development and evaluation of designs for this problem. Furthermore, our findings demonstrate that non-sequential optimal designs, obtained for scenarios with known response variances, provide the limits for stable allocation ratios obtained from sequential designs.

This paper contains three elements of novelty: (i) the development of theoretical results for D-, A-, E- and D_A-optimal designs for Neyman allocation when covariates are considered and the treatments have unequal (known) variances of the response; (ii) the development of Semidefinite Programming (SDP) formulations to generalize numerical computation of designs to larger trials (both more treatments and more covariates) for scenarios where the variances are known; and (iii) the development of sequential designs to simultaneously (and optimally) allocate the patients and update the model response parameters and variances.

The paper is organized as follows. Section 2 provides the background and the notation

used for problem formulation. Section 3 presents the theoretical tools developed to help in the computation of optimal designs for allocation. Section 4 presents SDP-based formulations to handle the problem systematically and Section 5 exhibits the properties of sequential designs. While Sections 3 and 4 delve into the static allocation of individuals to competing treatments – defined as the determination of the proportion of individuals allocated to each model prior to the trial's initiation – Section 5 addresses the sequential optimal allocation problem utilizing information criteria. The findings from the preceding sections serve as asymptotic results for those presented in Section 5 and can offer valuable insights into delineating sequential optimal allocation schemes, particularly when error variances are either available or estimable before the trial is initiated. In Section 6, we present the results from the application of the developed tools for both static and sequential allocation to a problem concerning Parkinson's disease patients, where they are assigned to two alternative treatments. Although the focus in our paper is on optimum experimental designs, we conclude with a brief discussion of the application of our results to sequential clinical trials.

2. Background and Notation

2.1. Generalized Neyman Allocation

Neyman allocation (Neyman, 1934) allocates patients with a weight proportional to the standard deviation of the observations receiving each treatment. There are, however, no covariates. Atkinson (2015) presents optimum experimental designs for first-order regression models, that is, models with covariates but no interaction or higher-order terms, when the responses to the two treatments have different variances. Wang and Ai (2016) extended these results to an arbitrary number of treatments and a general regression model. Although these papers provide designs when response variance depends on treatment, the model for the expected response is the same for all treatments. We find designs when both the expected response and its variance depend upon treatment.

2.2. Purposes of Optimal Experimental Design

Here we describe some advantages of optimal experimental design and introduce some of the procedures. Berger and Wong (2009) give an introduction to optimal design in social and biomedical research. A survey of the use of optimal experimental in clinical trials is Sverdlov *et al.* (2020).

Optimal experimental designs are model-based. They provide a structured data collection plan aimed at maximizing the amount of information gathered, the way in which the experiment is designed depending upon the information sought. This paper focuses specifically on estimation of the parameters of the model. Our aim is to find optimal experimental designs to minimize the variance of the estimates of the model parameters in specific models for clinical trials. These models establish a relationship between the response and the covariates as elucidated by Kiefer (1961). Various measures can be employed to minimize the confidence ellipsoid for the parameters, including: (i) D-optimality criterion, corresponding to the ellipsoid's volume; (ii) A-optimality, corresponding to the sum of diagonals of the ellipsoid in each dimension; (iii) E-optimality, corresponding to the minimum of the diagonals of the ellipsoid; and (iv) D_A-optimality, representing the volume concerning a specific combination of parameters denoted by matrix A. Each of these criteria is formulated as a convex (concave) function of the Fisher Information Matrix (FIM), the inverse of which is a lower bound for the parametric covariance matrix (Cramér, 1999; Rao, 1945). Thus, optimal experimen-

tal designs emerge through the minimization of a convex function of the inverse of the FIM. The first three criteria (i.e., (i)-(iii)) belong to Kiefer's class of optimality criteria, denoted by Φ_δ, where $\delta \in (-\infty, 0)$ represents the coefficient in the Kiefer general class of criteria. A-optimality corresponds to $\delta = -1$ and is formulated as the minimization of the trace of the inverse of the FIM, E-optimality to $\delta \to -\infty$ and corresponds to the maximization of the minimum of the eigenvalues of the FIM, and D-optimality to $\delta \to 0$ where it corresponds to the maximization of the determinant of the FIM.

2.3. Fundamentals of Optimal Experimental Design

This Section establishes the nomenclature and the fundamental background used in the subsequent sections.

In our notation bold face lowercase letters represent vectors, bold face capital letters continuous domains, blackboard bold capital letters discrete domains and capital letters are 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 "$^\mathsf{T}$". The cardinality of a vector is represented by $\mathrm{card}(\bullet)$ and the trace of a matrix by $\mathrm{tr}(\bullet)$.

Let $k \in [\![K]\!]$ represent categorical covariates in the linear response model, $i \in [\![I]\!]$ represent treatments, and $j \in [\![J]\!]$ represent the levels each factor can have in the experiment. Here, K, I, and J denote the number of categorical covariates, treatments, and factor levels in the response model, respectively, while k, i, and j are counters used for covariates, treatments, and levels. In this simplified configuration, we consider ategorical covariates with only two levels: one designated as -1 (lower) and the other as $+1$ (upper). According to the bound on the determinant of a Hadamard matrix (Hadamard, 1893; Brenner, 1972), extreme observations are the most informative for the D-optimality criterion in first-order models.

We stress that these limited points of support of the designs reduce the computational burden in demonstrating the properties of our designs in the non-sequential setting. In the application to the construction of sequential clinical trials in §5, the values of the covariates are sampled from continuous distributions.

When $i \in \{1, 2\}$ (two treatments) and we have k covariates, the model is as follows

$$\eta_1(\mathbf{x}, \boldsymbol{\theta}) = \alpha_1 + \beta_{1,1}\, x_1 + \beta_{1,2}\, x_2 + \cdots + \beta_{1,k}\, x_k, \quad \mathrm{var}(\eta_1) = \sigma_1^2 \tag{1a}$$

$$\eta_2(\mathbf{x}, \boldsymbol{\theta}) = \alpha_2 + \beta_{2,1}\, x_1 + \beta_{2,2}\, x_2 + \cdots + \beta_{2,k}\, x_k, \quad \mathrm{var}(\eta_2) = \sigma_2^2, \tag{1b}$$

where $\eta_i(\mathbf{x}, \boldsymbol{\theta})$ represents the response when the individuals are submitted to treatment i, $\mathbf{x} = (x_1, x_2, \cdots, x_K)$ is the vector of covariates, $\mathbf{X}$ the design space, $\boldsymbol{\theta}$ the set of parameters to be estimated, which contains all α's and β's, and $\boldsymbol{\Theta} \subset \mathbb{R}^{n_\theta}$ the domain of the parameters to be estimated, n_θ being the number of parameters involved in all treatment models. Here, each model in (1) is associated with a treatment i, featuring unique parameters and observational error. Practically, a total of $2\,(K+1)$ parameters needs to be estimated, although the number is reduced if, for some treatments $k, \beta_{1,k} = \beta_{2,k}$. In general, all $i \in [\![I]\!]$ treatments have a normally distributed error with expectation 0 and standard deviation σ_i, this being distinct for each treatment.

The allocation is optimized to maximize the information extracted from the patients' responses while considering the predictions from the model. Consequently, our objective is to determine the allocation of individuals to treatment 1, enabling the estimation of parameters $(\alpha_1, \beta_{1,1}, \cdots \beta_{1,K})^\mathsf{T}$, and to allocate individuals to treatment 2 for estimation of the parameters of the second treatment model. In the context of approximate optimal designs, such as we have here, w represents the fraction of individuals allocated to treatment 1, while $1 - w$

denotes the fraction assigned to treatment 2. This yields the design matrix:

$$\xi = \begin{pmatrix} w & 1-w \\ 1 & 2 \end{pmatrix}.$$

Here, the upper row of ξ signifies the fraction allocated to each treatment, while the lower row denotes the treatment counter itself. In instances where the number of individuals is small enough to necessitate an exact design, rounding procedures can be employed to determine the allocation, as outlined in Pukelsheim and Rieder (1992).

Let $\tau = \sigma_2^2/\sigma_1^2$ and, without loss of generality (wlog), let $\sigma_1^2 = 1$. For the sake of simplicity, we consider that the fraction of individuals allocated to group 1 is w and the fraction allocated to group 2 is $1-w$. For a single model when allocation is to either $+1$ or -1 the Fisher Information Matrix (FIM) for the experimental design ξ is the $(K+1) \times (K+1)$ diagonal matrix. Here σ^2 is a multiplicative constant, which does not affect the properties of the design. For two treatments the FIM is the $2 \cdot (K+1) \times 2 \cdot (K+1)$ diagonal matrix. Consequently,

$$M(\xi) = \begin{pmatrix} w & 0 & \cdots & 0 & 0 \\ 0 & \frac{1-w}{\tau} & \cdots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \cdots & w & 0 \\ 0 & 0 & \cdots & 0 & \frac{1-w}{\tau} \end{pmatrix}. \tag{2}$$

It is clear that, now, the value of τ will , in general, affect the design.

3. Theoretical results

Here, we present some new theoretical results for Neyman allocation developed with the help of symbolic algebraic tools. The goal is computing continuous experimental designs (i.e., the vector of weights) to minimize the uncertainty of all estimates of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of interest in both models (1).

First, we consider model (1) and develop theorems for optimal allocation to treatments. For demonstration purposes we assume models including two treatments ($I = 2$), two covariates ($K = 2$) and factors with two levels ($J = 2$), i.e. $\mathbf{X} \equiv \otimes_{i=1}^{J} \{-1, +1\}$ is a finite set of points that result from the combination of levels allowed for each covariate; here, $x_i \in \{-1, +1\}$, $i \in [\![I]\!]$. We note that when x_i is continuous in $[-1, +1]$ and the models are linear with respect to the parameters, the maximum information is obtained by locating the experiments at the bounds of the domain. Thus, by considering only the extremes of the design space, we are maximizing the information obtained from each group through choice over $\mathbf{X}$.

The goal is to construct continuous optimal designs where all the parameters of interest are to be estimated and the error variances of the responses to the various treatments are previously known. The premise that error variances are known, though uncommon, has been explored in several studies (Atkinson, 2015; Bai *et al.*, 2002; Rosenberger, 1993; Hu *et al.*, 2006), indicating its relevance in certain contexts. In our case, this assumption is adopted due to our focus on deriving asymptotic results for Neyman allocation, particularly when covariates are considered. This analysis is pivotal for validating the sequential designs discussed in Section 5, where we no longer need to make this assumption, the values being estimated from

the data.

The total number of individuals to be allocated to treatments is $N \to +\infty$, i.e., we focus on infinite sized experiments. Thus, we let w_i be the fraction of individuals allocated to treatment i, so that the number of individuals allocated is $n_i = N \times w_i$. In this Section, in which we only consider two treatments, we write w_1 as w and w_2 as $1 - w$. Further, w^* is the optimal allocation. Equivalent exact designs can be obtained from continuous designs assuming a given number N of individuals in the design, using a rounding procedure (Pukelsheim and Rieder, 1992) or Mixed Integer Nonlinear Programming formulations (Duarte *et al.*, 2020). We start with the most used information-theoretical criteria (D-, A- and E-: see Kiefer (1959); Atkinson *et al.* (2007)) which measure different aspects of the efficiency of the allocation schemes for estimation of the parameters of the response models for each group. E-optimal designs can be hard to implement in practice. However, they correspond to $\delta \to +\infty$ in Kiefer's generalized criterion; we consider them since they provide a bound in the allocation schemes based on the Neyman rule. In Theorems 1–4 we develop theoretical results for the optimal allocations.

Theorem 1. *The D-optimal experimental design for model* (1) *has weights* $w^* = 1/2$ *independently of the values of* τ *and* k.

Proof. Let $\det[M(\xi)] = w^{K+1}\ (1-w)^{K+1}/\tau^{K+1}$ and $\nabla_\xi \det[M(\xi)] = (1-2\,w)\ (K+1)\ w^K\ (1-w)^K/\tau^K$. The first order condition for the optimality of the design is that $\nabla_\xi \det[M(\xi)] = 0$, which leads to $w = 0.5$. The second order condition is that $x\ \nabla^2_\xi \det[M(\xi^*)]\ x^\mathsf{T} \leq 0,\ \forall x \in \mathbb{R}$ and ξ^* is the design that satisfies the first order condition. By substitution, $\nabla^2_\xi \det[M(\xi^*)] = -(K+1)/(2^{2K-1}\ \tau^{K+1})$ which validates the second order optimality condition. □

Theorem 2. *The A-optimal experimental design for model* (1) *allocates* $w^* = 1/(1+\sqrt{\tau})$.

Proof. Let $\mathrm{tr}[M^{-1}(\xi)] = (K+1)/w + (K+1)\ \tau/(1-w)$ and $\nabla_\xi \mathrm{tr}[M^{-1}(\xi)] = \tau\ (K+1)/(1-w)^2 - (K+1)/w^2$. The first order condition for optimality leads to $w = 1/(1+\sqrt{\tau})$. Further, $\nabla^2_\xi \mathrm{tr}[M^{-1}(\xi^*)] = 2\ (K+1)\ (1+\sqrt{\tau})^4/\sqrt{\tau}$ which validates the second order optimality condition $x\ \nabla^2_\xi \mathrm{tr}[M^{-1}(\xi^*)]\ x^\mathsf{T} \geq 0,\ \forall x \in \mathbb{R}$. □

With $w_1^* = 1/(1+\sqrt{\tau})$ and, therefore $w_2^* = \sqrt{\tau}/(1+\sqrt{\tau})$, the A-optimal allocation is proportional to the standard deviation of the responses to the two treatments, that is the original Neyman allocation.

Theorem 3. *The E-optimal experimental design for model* (1) *allocates* $w^* = 1/(1+\tau)$.

Proof. Let $\min \lambda_{\max}[M^{-1}(\xi)] = \min\{1/w,\ \tau/(1-w)\}$ which minimization requires that $1/w = \tau/(1-w)$. Consequently, the solution is obtained for $w^* = 1/(1+\tau)$. □

These results depend on all parameters in the two models having different values. Atkinson (2015) gives results when some of the parameters in the two response models are the same. Here, we illustrate the point when, in models (1), $\alpha_1 = \alpha_2$. The D-optimal design leads to

$$w^* = \frac{3\,\tau - 7 + \sqrt{9\,\tau^2 - 2\,\tau + 9}}{10\,(\tau - 1)}.$$

For the E-optimal design $w^* = 1/(1+\tau)$. Finally, the A-optimal design problem leads to

an algebraic expression that requires numerical solution for each τ:

$$\frac{2\,\tau}{(w^*-1)^2} - \frac{2}{(w^*)^2} - \frac{\tau\,(\tau-1)}{(\tau\,w^* - w^* + 1)^2} = 0.$$

Now, we consider estimation of (some of) the differences $\alpha_1 - \alpha_2$ and $\beta_{1,k} - \beta_{2,k}$ in (1). For demonstration we consider that all the $k+1$ parameter differences are to be estimated and find the optimal experimental design for such a purpose. This would be appropriate when interest is in whether there is a difference between patients' responses to treatments 1 and 2 that is a function of the combination of covariate values in $\mathbf{X}$. D_A-optimal designs (Silvey, 1980) allow efficient determination of linear combinations, such as differences, of the parameters.

Theorem 4. *The D_A-optimal experimental design for model (1) when*

$$A \in \mathbb{R}^{2(K+1)\times m} : A^\mathsf{T} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix},$$

and $1 \leq m \leq K+1$ allocates $w^* = 1/(1+\sqrt{\tau})$.

Proof. Let the objective be $\min_\xi \det[A^\mathsf{T}\, M^{-1}(\xi)\, A]$. Matrix A encapsulates the linear relationships between the parameters to be estimated, with m representing the total number of such relations. Considering the $2\cdot(K+1)\times m$ size matrix A, $\det[A^\mathsf{T}\, M^{-1}(\xi)\, A] = [1/w + \tau/(1-w)]^m$. Then, the first-order optimality condition leads to $w = 1/(1+\sqrt{\tau})$. The computation of $\nabla^2_\xi \det[A^\mathsf{T}\, M^{-1}(\xi^*)\, A]$ leads to $2m\,(1+\sqrt{\tau})^{2m+2}/\sqrt{\tau}$ which validates the second order optimality condition. □

Suitable modification of the matrix of contrasts A allows the generalization to problems where only a subset of parameter differences are to be tested. Together, these theorems (1-4) allow the construction of tables presenting w^* as a function of $\tau = \sigma_2^2/\sigma_1^2$.

4. Results obtained with Semidefinite Programming formulations

Here, we present SDP formulations to automate the computation of continuous optimal designs for Neyman allocation in cases where more challenging setups are considered.

For demonstration we consider the model (1) including K covariates, i.e. $\mathbf{x} \equiv (x_1, x_2, \cdots, x_K)$. The levels of the covariates are also $\{-1, +1\}$, and the variances of each group are σ_1^2 and σ_2^2, respectively. Here, we consider two groups (i.e., $I = 2$) but the formulations can be easily extended to more than two treatments and more covariate levels. The goal is estimating all the $2\cdot(K+1)$ parameters of both models as in §3. The formulations for D-, A-, E-optimality criteria are currently state of the art and appear in Appendix A. In all formulations, the optimum reached upon convergence is denoted as "Opt" and serves as the reference for comparison. By "optimum at convergence", we specifically refer to the value attained by the objective function following the resolution of the SDP problem under consideration. The formulation can be adapted to handle a different number of factors, factors having a different number of levels and ranges and even continuous factors, although these have to be discretized, see Duarte *et al.* (2016).

The D_A-optimality criterion can also be formalized as an SDP problem. Let the setup

be the same as in §3 before Theorem 4. To derive the SDP formulation for D_A-optimality we use the equivalence of the linear matrix inequalities $[A^\mathsf{T} M^{-1}(\xi)A]^{-1} - C \succeq 0_m$ and $M(\xi) - ACA^\mathsf{T} \succeq 0_{2\cdot(K+1)}$ where C is a semidefinite positive matrix of size $m \times m$, see Harman and Sagnol (2015, Lema 2.1). Then, we consider the equivalence relation $\min_\xi \det[A^\mathsf{T} M^{-1}(\xi)A] \equiv \max_\xi \det\{[A^\mathsf{T} M^{-1}(\xi)A]^{-1}\}$ to express the problem as a maximization program since the operator $\det(\bullet)$ of a semidefinite positive matrix is a concave function (Ben-Tal and Nemirovski, 2001). That is,

$$
\begin{aligned}
\text{Opt} = \max_{\xi, C, B} \quad & t && \text{(3a)}\\
\text{s.t.} \quad & M(\xi) - A\, C\, A^\mathsf{T} \succeq 0_{2\ (K+1)} && \text{(3b)}\\
& \begin{pmatrix} C & B^\mathsf{T} \\ B & \operatorname{diag}(B) \end{pmatrix} \succeq 0_{2m} && \text{(3c)}\\
& t \leq \prod_{i=1}^{m} B_{i,i}^{1/m} && \text{(3d)}\\
& \sum_{i=1}^{I} w_i = 1 && \text{(3e)}\\
& 0 \leq w_i \leq 1, \quad i \in [\![I]\!]. && \text{(3f)}
\end{aligned}
$$

Equation (3) represents the Semidefinite Programming (SDP) problem aimed at determining the D_A-optimal designs for each value of τ. Here, the matrix B represents an upper triangular matrix to be determined, C is a positive semidefinite matrix to be determined, and t corresponds to the hypograph (maximum value) of the convex set defined by the concave function that we seek to maximize.

Table 1 presents D-, A- and E- optimal designs obtained for $K = \{2, 3\}$, $I = 2$ and $x_k \in \{-1, +1\}$, $k \in [\![K]\!]$ as the value of τ varies. The designs obtained are represented by w_1, the relative fraction of individuals to be allocated to the first group, and the optimum of the objective function at convergence (Opt). The fraction of subjects allocated to group 2 is given by $w_2 = 1 - w_1$. In general, the fraction of observations in the optimal design allocated, at the points of the 2^K factorial, to treatment i when there are K covariates is $w_i/2^K$. Because the optimal allocations (i.e., w_1) for both values of K are equal and only differences on the optima are noted, we use Table 1 to compactly present all results where the optima obtained for both values of K are in adjacent columns. The results indicate that the optimal values for D- and E-optimal designs decrease as τ and K increase. In contrast, for the A-optimality criterion, they increase with both τ and K.

In the table, the optimal designs, although not the values of the optima, are independent of K. The D-optimal designs are also independent of $\tau = \sigma_2^2/\sigma_1^2$. This result is consistent with the theorems in §3.

The D_A-optimal designs obtained via SDP for $K = 2$ by solving (3) are in Table 2. Together with the results in Tables 1-3 they verify Theorems 1-4.

Now we consider the model with three groups ($I = 3$) and two covariates ($K = 2$) when each covariate has two levels, $x_k \in \{-1, +1\}$, $k \in [\![K]\!]$, i.e.

$$
\begin{aligned}
\eta_1(\mathbf{x}, \boldsymbol{\theta}) &= \alpha_1 + \beta_{1,1}\, x_1 + \beta_{1,2}\, x_2, \quad \text{var}(\eta_1) = \sigma_1^2 && \text{(4a)}\\
\eta_2(\mathbf{x}, \boldsymbol{\theta}) &= \alpha_2 + \beta_{2,1}\, x_1 + \beta_{2,2}\, x_2, \quad \text{var}(\eta_2) = \sigma_2^2 && \text{(4b)}\\
\eta_3(\mathbf{x}, \boldsymbol{\theta}) &= \alpha_3 + \beta_{3,1}\, x_1 + \beta_{3,2}\, x_2, \quad \text{var}(\eta_3) = \sigma_3^2. && \text{(4c)}
\end{aligned}
$$

Table 1. D-, A- and E-optimal designs for model (1) (including two treatments, two covariates, $x_k \in \{-1, +1\},\ k \in [\![K]\!]$ and $w_2 = 1 - w_1$).

		D-opt. design			A-opt. design			E-opt. design		
			Opt			Opt			Opt	
σ_1^2	σ_2^2	w_1	$K=2$	$K=3$	w_1	$K=2$	$K=3$	w_1	$K=2$	$K=3$
1.0000	0.2000	0.5000	0.2795	0.1398	0.6910	25.1331	67.0217	0.8333	0.2083	0.1042
1.0000	0.4000	0.5000	0.1976	0.0988	0.6126	31.9789	85.2722	0.7143	0.1786	0.0893
1.0000	0.6000	0.5000	0.1614	0.0807	0.5635	37.7903	100.7742	0.6250	0.1562	0.0781
1.0000	0.8000	0.5000	0.1398	0.0699	0.5279	43.0663	114.8433	0.5556	0.1389	0.0694
1.0000	1.0000	0.5000	0.1250	0.0625	0.5000	48.0000	128.0000	0.5000	0.1250	0.0625
1.0000	1.2500	0.5000	0.1118	0.0559	0.4721	53.8328	143.5542	0.4444	0.1111	0.0556
1.0000	1.6667	0.5000	0.0968	0.0484	0.4365	62.9839	167.9580	0.3750	0.0937	0.0469
1.0000	2.5000	0.5000	0.0791	0.0395	0.3874	79.9473	213.1929	0.2857	0.0714	0.0357
1.0000	5.0000	0.5000	0.0559	0.0280	0.3090	125.6656	335.1084	0.1667	0.0417	0.0208

Table 2. D_A-optimal designs for model (1) (including two treatments, two covariates and $x_k \in \{-1, +1\},\ k \in [\![K]\!]$ and $w_2 = 1 - w_1$).

σ_1^2	σ_2^2	w_1	Opt
1.0000	0.2000	0.6910	0.4775
1.0000	0.4000	0.6126	0.3752
1.0000	0.6000	0.5635	0.3175
1.0000	0.8000	0.5279	0.2786
1.0000	1.0000	0.5000	0.2500
1.0000	1.2500	0.4721	0.2229
1.0000	1.6667	0.4365	0.1905
1.0000	2.5000	0.3874	0.1501
1.0000	5.0000	0.3090	0.0955

For simplicity we assumed $\sigma_1^2 = 1.0$ and varied σ_2^2 and σ_3^2; the latter was set to $2 \cdot \sigma_2^2$, i.e. $\tau_1 = \sigma_2^2/\sigma_1^2$ and $\tau_2 = 2 \times \tau_1$. The optimal designs are in Table 3.

Table 3. D-, A- and E-optimal designs for response model (4) (including three treatments, two covariates and $x_k \in \{-1, +1\},\ k \in [\![K]\!]$, $w_3 = 1 - w_1 - w_2$, $\sigma_2^2 = 2 \times \sigma_1^2$ and $\sigma_3^2 = 2 \times \sigma_2^2$).

		D-opt. design			A-opt. design			E-opt. design		
σ_1^2	σ_2^2	w_1	w_2	Opt	w_1	w_2	Opt	w_1	w_2	Opt
1.0000	0.2000	0.3333	0.3333	0.1934	0.4808	0.2150	51.9003	0.6250	0.1250	0.1562
1.0000	0.4000	0.3333	0.3333	0.1218	0.3958	0.2503	76.6216	0.4545	0.1818	0.1136
1.0000	0.6000	0.3333	0.3333	0.0930	0.3484	0.2699	98.8457	0.3571	0.2143	0.0893
1.0000	0.8000	0.3333	0.3333	0.0768	0.3165	0.2831	119.7770	0.2941	0.2353	0.0735
1.0000	1.0000	0.3333	0.3333	0.0661	0.2929	0.2929	139.8823	0.2500	0.2500	0.0625
1.0000	1.2500	0.3333	0.3333	0.0570	0.2703	0.3022	164.2066	0.2105	0.2632	0.0526
1.0000	1.6667	0.3333	0.3333	0.0471	0.2429	0.3136	203.3702	0.1667	0.2778	0.0417
1.0000	2.5000	0.3333	0.3333	0.0359	0.2076	0.3282	278.4658	0.1176	0.2941	0.0294
1.0000	5.0000	0.3333	0.3333	0.0226	0.1563	0.3495	491.2659	0.0625	0.3125	0.0156

To utilize the results presented in Tables 1-3 effectively in practical scenarios, researchers should follow these steps:

(i) Use given values of σ_1^2 and σ_2^2 to compute τ;
(ii) based on the number of covariates in the response model (K), select the appropriate Table of results;
(iii) set the optimality criterion of interest;
(iv) refer to the chosen Table from step (ii) and find the corresponding value of w for the chosen optimality criterion and the computed value of τ from step (i);

(v) allocate $\lceil N \cdot w_i \rceil$ individuals to each treatment i.

For example, if $\tau = 0.4$ (as indicated in the second row of Table 1), $K = 2$, and the research objective is the implementation of A-optimal designs, then $w = 0.6126$ (that is, allocating $61.26\,\%$ of individuals to the first treatment and the remaining $38.74\,\%$ to the second). These explicit designs can be used for the comparison of power. If there is uncertainty about the parameter values, designs for several sets of numbers can be compared.

In our work, we solved all the Semidefinite Programming problems in `Matlab®` using the `cvx` environment combined with the solver `Mosek` that uses an efficient Interior Point algorithm (Ye, 1997). The relative and absolute tolerances used to solve the SDP problem were set to 1×10^{-5}.

5. Sequential Designs

When the response models are linear, as in (1), inference for all parameters, or for linear combinations of parameters, of the linear models does not depend on the values of the parameters. However, as we have seen above, the optimal design does depend on $K-1$ ratios of variances.

The designs found in §4 depend on the value of τ or, in the case of Table 3, two variance ratios. To overcome the lack of knowledge of the true value of τ we use a sequential design, a solution appropriate for clinical trials in which patients arrive sequentially and, given the values of their prognostic factors, are allocated to a specific treatment. In frequentist analysis of sequential experiments, the information from observations is used to update the parameter estimates. Experiments are designed using the available information, with one or more observations being obtained using this design. The parameter estimates are updated using all observations providing a new optimum experimental design. The process terminates when sufficient accuracy has been obtained in the parameter estimates, or when all resources are exhausted. Atkinson *et al.* (2007, §17.3) describe sequential and non-sequential methods for optimal design when the information matrix depends on unknown parameters.

We consider frequentist sequential designs for Neyman allocation where K prognostic factors are considered, see the model (1). Also, let $I = 2$ (i.e., two treatments) where each treatment has an unknown variance, unlike the known variances of the classical Neyman allocation. Our goal is to demonstrate that sequential optimal designs, where the variances of the treatments are iteratively updated from analysis of the responses, tend to the optimal designs obtained theoretically, or with SDP-based formulations, assuming the variances are previously known.

The procedure starts by allocating the first arriving n_f individuals randomly to either treatment. For simplicity we allocate $n_f/2$ to each treatment, with n_f even. The prognostic factors of each individual are used to construct the respective FIM so that the complete set of six parameters (α_1, α_2, $\beta_{1,1}$, $\beta_{2,1}$, $\beta_{1,2}$ and $\beta_{2,2}$) is to be estimated. The FIM of an individual allocated to treatment i ($i \in \{1, 2\}$) is represented by

$$M_1(\xi) = \frac{1}{\hat{\sigma}_1^2} \begin{pmatrix} 1 & 0 & x_1 & 0 & x_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ x_1 & 0 & x_1^2 & 0 & x_1\,x_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ x_2 & 0 & x_1\,x_2 & 0 & x_2^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and}$$

$$M_2(\xi) = \frac{1}{\hat{\sigma}_2^2} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & x_1 & 0 & x_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_1 & 0 & x_1^2 & 0 & x_1\, x_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_2 & 0 & x_1\, x_2 & 0 & x_2^2 \end{pmatrix}. \tag{5}$$

$M_i(\xi)$ is the FIM used for allocating the individual to treatment i and $\hat{\sigma}_i^2$ is the estimated variance of the observational errors of the responses of individuals allocated to that treatment. These estimates come from Least Squares (LS) fitting of the two models $\eta_i(\mathbf{x}, \boldsymbol{\theta})$ using all available responses. The fraction of individuals allocated to each group in the initialization is therefore $w_i^{(n_f)} = 1/I$. In the remainder of the paper the value in the superscript is used to designate the the number of individuals previously enrolled. This nomenclature is also adopted for global and treatment FIM's. For clarification, we use the term *global FIM* to represent the FIM accounting for all individuals allocated previously, $\mathcal{M}^{(\ell-1)}(\xi)$, the term *treatment FIM* to represent the FIM for all the individuals previously allocated to a given treatment, $\mathbb{M}_i^{(\ell-1)}(\xi)$, and the term *individual FIM* to the FIM for a single individual entering the plan which is given by (5). Further, let the vector $\mathbf{p}$ represent the allocations of the individuals; each element of $\mathbf{p}$ is 1 if the individual is allocated to the first treatment and 2 if allocation is to treatment 2.

For simulating the procedure, we take the covariates as uniformly distributed in $[-1, +1]$, distinct from the design region in previous sections where $x_i \in \{-1, +1\}$. For the prognostic factors we generate variables from the uniform distribution in $[-1, +1]$; i.e. $x_i \sim \mathcal{U}(-1, +1),\ i \in [\![I]\!]$. The global FIM obtained after enrolling evenly the n_f individuals is calculated by summing the FIM's of all the individuals allocated to each treatment.

Now, we start allocating the individuals using information-based criteria. That is, every individual arriving (corresponding to an iteration) is enrolled to the treatment that maximizes the amount of information measured by a given Φ-optimality criterion. Consider the arrival of individual ℓ with known prognostic factors represented by the extended vector $\mathbf{z}_\ell = (1.0, \mathbf{x}_\ell)$. Here, $\mathbf{x}_\ell$ denotes the covariates specific to the ℓ^{th} individual, while in Sections 2-4, $\mathbf{x}$ denotes the candidate experimental points.

The global FIM obtained considering all the $\ell - 1$ individuals previously allocated is given by

$$\mathcal{M}^{(\ell-1)}(\xi) = \sum_{i=1}^{I} w_i^{(\ell-1)}\, \mathbb{M}_i^{(\ell-1)}(\xi), \tag{6}$$

where $w_i^{(\ell-1)},\ i \in [\![I]\!]$ is the proportion of individuals allocated to each treatment and $\mathbb{M}_i^{(\ell-1)}(\xi)$ is the treatment FIM for treatment i:

$$\mathbb{M}_i^{(\ell-1)}(\xi) = \sum_{j=1}^{\ell-1} 1_{p_j=i}\, M_i^{(j)}(\xi). \tag{7}$$

Here $1_{p_j=i}$ is the indicator function, with value 1 if the j^{th} individual ($j \in [\![\ell - 1]\!]$) was allocated to treatment i and 0 otherwise. The $M_i^{(j)}(\xi)$ are the individual FIM's; see Eq. (5).

We choose the treatment that maximizes the overall information for the allocation of individual ℓ. That is, we calculate the sensitivity for each treatment to $\mathbf{z}_\ell$ using the directional

derivative of the criterion under analysis and choose the maximum. When the D-, A-, E- or $\mathrm{D_A}$-optimality criteria are used, the respective sensitivity functions (Atkinson *et al.*, 2007, §10.1) are:

$$d_{\mathrm{D},i} = \mathbf{z}_\ell \left[\mathcal{M}^{(\ell-1)}(\xi)\right]^{-1} \mathbf{z}_\ell^\mathsf{T}/\hat{\sigma}_i^2, \quad i \in [\![I]\!] \tag{8a}$$

$$d_{\mathrm{A},i} = \mathbf{z}_\ell \left[\mathcal{M}^{(\ell-1)}(\xi)\right]^{-2} \mathbf{z}_\ell^\mathsf{T}/\hat{\sigma}_i^2, \quad i \in [\![I]\!] \tag{8b}$$

$$d_{\mathrm{E},i} = \mathbf{z}_\ell \left(\mathbf{v}_{\min}\, \mathbf{v}_{\min}^\mathsf{T}\right) \mathbf{z}_\ell^\mathsf{T}/\hat{\sigma}_i^2, \quad i \in [\![I]\!] \tag{8c}$$

$$d_{\mathrm{D_A},i} = \mathbf{z}_\ell \left[\mathcal{M}^{(\ell-1)}(\xi)\right]^{-1} A^\mathsf{T} \left(A \left[\mathcal{M}^{(\ell-1)}(\xi)\right]^{-1} A^\mathsf{T}\right)^{-1} A \left[\mathcal{M}^{(\ell-1)}(\xi)^{-1}\right] \mathbf{z}_\ell^\mathsf{T}/\hat{\sigma}_i^2 \quad i \in [\![I]\!] \tag{8d}$$

where $d_{\Phi,i}$ is the Φ-optimality criterion sensitivity to treatment i, $\Phi = \{\mathrm{D}, \mathrm{A}, \mathrm{E}, \mathrm{D_A}\}$ and $\mathbf{v}_{\min} \in \mathbb{R}^{n_\theta}$ is the eigenvector associated to $\lambda_{\min}[\mathcal{M}^{(\ell-1)}(\xi)]$.

We consider the D-optimal criterion for analysis, the next steps of the allocation being common to all criteria. After computing the sensitivities of both treatments to the arriving individual, treatment 1 is allocated if $d_{D,1} \geq d_{D,2}$ and treatment 2 otherwise. This allocation rule may be written as

$$\begin{cases} \text{the individual is allocated to treatment 1} & \text{if} \quad d_{D,1} \geq d_{D,2} \implies p_\ell = 1 \\ \text{the individual is allocated to treatment 2} & \text{if} \quad d_{D,1} < d_{D,2} \implies p_\ell = 2 \end{cases}.$$

We note that this allocation policy does not involve randomization rules; instead it allocates the individual to the treatment that maximizes the information expected given their prognostic factors.

Let individual ℓ be allocated to treatment $i^{\max} = \{i : \; d_{D,i} = \max(\mathbf{d}_D), \; i \in [\![I]\!]\}$. Then, we update the allocation vector $\mathbf{p}$; that is, we set $p_\ell = i^{\max}$. We also use the updated set of the individuals allocated to treatment $i^{\max}$ to re-estimate the parameters of $\eta_{i^{\max}}(\mathbf{x}, \boldsymbol{\theta})$ using LS and update the error variance of the model, $\hat{\sigma}^2_{i^{\max}}$. The treatment FIM of treatment $i^{\max}$ is also updated, as well as the weights, which become

$$w_i^{(\ell)} = \frac{\sum_{j=1}^{\ell} 1_{p_j=i}}{\ell}, \quad i \in [\![I]\!]. \tag{9}$$

Finally, the treatment FIM's and the global FIM are updated using (7) and (6), respectively. The procedure is then iterated for the remaining $n_p - \ell$ individuals until we have enrolled n_p. To avoid the influence of the randomness in the construction of the covariates and of the response we repeated the procedure n_s times. Then, we computed averages of the weights for each individual allocated and of the variances of the errors of both response models.

Algorithm 1 systematically outlines the procedure for sequential allocation based on information criteria. To demonstrate its mechanics we consider the allocation of the ℓ^{th} individual. Its implementation necessitates the knowledge of the following quantitities:

(i) the optimality criterion considered for allocation, denoted as C;
(ii) the number of competing treatments, denoted as I;
(iii) the number of covariates, denoted as K;
(iv) the vector of covariates for the ℓ^{th} individual, denoted as $\mathbf{x}_\ell$;
(v) the Fisher Information Matrices (FIMs) for each treatment and the global FIM after the

allocation of $(\ell-1)^{\text{th}}$ individual, denoted as $\mathcal{M}^{(\ell-1)}(\xi)$; and

(vi) the estimates of error variances for all treatments, denoted as $\hat{\sigma}_i^2$ obtained after allocating $(\ell-1)$ individuals.

This iterative procedure is applied to all individuals, $\ell \in [\![n_p]\!]$, as they enter the trial.

Algorithm 1 Allocation of the ℓ^{th} individual.

procedure ALLOCATEINDIVIDUAL(Input: Optimality Criterion (C), I, K, $\mathbf{x}_\ell$, $\mathcal{M}^{(\ell-1)}(\xi)$, $\hat{\sigma}_i^2$)
  Construct the vector $\mathbf{z}_\ell \equiv (1, \mathbf{x}_\ell)$
  for i in $[\![I]\!]$ **do**
    Compute the sensitivity functions $d_{\text{C},i}$ for $\mathbf{x}_\ell$ using Eq. 8
  end for
  Find the maximum sensitivity value, $\max_i d_{\text{C},i}$, among all treatments
  Allocate the individual to the treatment with the maximum sensitivity value
  Update $\mathbf{w}^{(\ell)}$, $\mathcal{M}^{(\ell)}(\xi)$ and $\hat{\sigma}_i^2$ (Eqs. (6-7,9))
end procedure

To test the algorithm described and check whether the sequential designs can be used to overcome the requirement that the variances are known, we consider two treatments with responses

$$\eta_1(\mathbf{x}, \boldsymbol{\theta}) = 0.5 + 0.2\, x_1 + 0.4\, x_2 + \epsilon_1 \tag{10a}$$

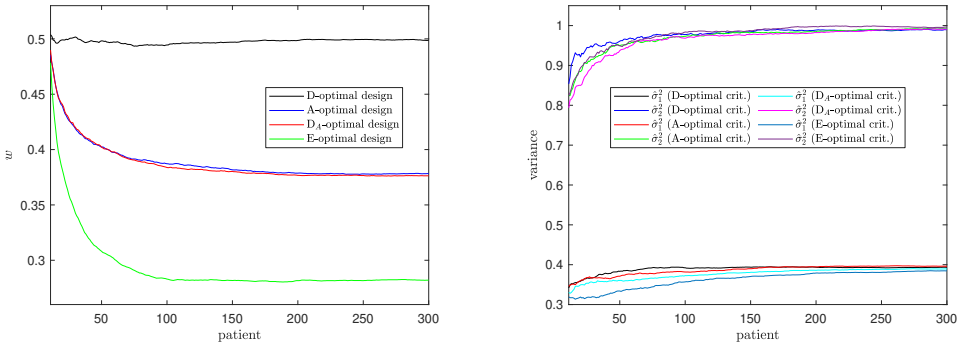
$$\eta_2(\mathbf{x}, \boldsymbol{\theta}) = 0.6 + 0.4\, x_1 + 0.5\, x_2 + \epsilon_2, \tag{10b}$$

respectively, both covariates $x_i \in [-1, +1]$, $i \in [\![I]\!]$ were sampled using a uniform random number generator. The observational error of the response to each treatment is normally distributed with zero mean and different standard deviations, i.e. $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, $i \in [\![I]\!]$ where σ_i is the standard deviation for treatment i. We notice that the error variances are unknown in this context and are to be iteratively estimated using LS as individuals accumulate to treatments. To simulate the responses we consider that $\sigma_1^2 = 0.4$ and $\sigma_2^2 = 1.0$. That is, $\sigma_1 = 0.6325$, $\sigma_2 = 1.0$, so that $\tau = 2.5$. Practically, the values of ϵ for each individual are obtained by sampling from zero mean normal distributions with different standard deviations using a normal random generator.

We consider the $\{\text{D,A,E,D}_\text{A}\}$-optimal design criteria previously introduced for testing the procedure. The number of individuals initially randomly allocated to each treatment is 5. Further, we set n_p to 300 and n_s to 500.

The results are in Figure 1; the plots at the left present the evolution of w_1 for all the optimality criteria and the plots at the right the estimated error variances for both treatments. We notice that for the D-optimality criterion $w_1 \to 0.5$, see Figure 1(a), corroborating the results obtained in §4 for $\tau = 2.5$ and the theorems in §3. Additional numerical tests, not shown in the paper, illustrate the lack of dependence of w_1 on the value of τ for D-optimal designs. For the A- and D_A-optimality criteria, w_1 tends to 0.38, and the values obtained using both SDP and the theoretical results are $1/(\sqrt{2.5}+1) \approx 0.3874$. The weight for the E-optimal allocation is close to 0.28 and the values obtained with previous tools are $1/(2.5+1) \approx 0.2857$. The numerical results for other values of τ from the sequential optimal design algorithm also show strong agreement with the optimal allocations previously obtained. Finally, we note in Figure 1(b), the remarkably fast convergence of the estimates of the variances of the observational errors for both treatments (and of the ratio τ), independently of the criterion.

One very interesting point is that the sequential optimal designs in $[-1, +1]$ tend to the optimal designs in $\{-1, +1\}$ as we claimed in Section 3. The weights for the sequential designs tend to the optimal, but the designs can't since the sequential x_i are given, not chosen. Further, they also allow capturing the estimates of the parameters of each response model and



(a) (b)

Figure 1. Sequential optimal designs for response models (10) $x_i \in [-1, +1]$, $i \in [\![I]\!]$, $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, $\sigma_1^2 = 0.4$, $\sigma_2^2 = 1.0$ and $\tau = 2.5$: (a) w_1 for D-, A-, E- and $\mathrm{D_A}$-optimality criteria; (b) estimated observational error variances for D-, A-, E- and $\mathrm{D_A}$-optimality criteria.

their respective variances.

6. Application: A Clinical Trial In Neuro-Degenerative Disease

In this section, we apply the methodologies outlined in Sections 3 and 5 to a real-world case study with two prognostic factors, one non-normal. Specifically, we examine a clinical trial outlined in Atkinson *et al.* (2023), focusing on the randomization process for Parkinson's disease patients. The primary aim of this study was to devise an effective randomization protocol for allocating patients to one of two treatment groups:

(i) treatment 1: Involving a pioneering procedure centered around digital technologies combined with some human intervention; and
(ii) treatment 2: Encompassing a conventional treatment and monitoring approach.

For assessing treatment efficacy, Atkinson *et al.* (2023) employed Quality of Life (QoL) as the outcome measure, assessed using the Parkinson's Disease Questionnaire with 8 items (PDQ-8) (Peto *et al.*, 1998), which we represent as $pdq8$; higher scores denote a lower quality of life. Out of six potential explanatory variables, data analysis showed that the two important ones were: (i) disease duration and stage, denoted as $h\&q$, assessed on the Hoehn and Yahr scale (Hoehn and Yahr, 1967, 1998); and (ii) psychological well-being and neuropsychiatric symptoms, such as depression and anxiety, labeled as bdi, based on questionnaire responses (Beck *et al.*, 1988).

We take the models for the effect of treatment to be those in Equation (1), x_1 being $h\&q$, while the second factor x_2 represents bdi. In the dataset employed for modeling the response to both treatments, $h\&q$ and bdi are confined to integer values. Precisely, $h\&q$ ranges from 1 to 5 ($h\&q \in [1, 5]$), and bdi varies from 1 to 35 ($bdi \in [1, 35]$). Consequently, the design space considered in this context is therefore $\mathbf{X} \equiv [1, 5] \times [1, 35]$, with the covariate values randomly sampled from $\mathbf{X}$. The variable $h\&y$ was fitted to a normal distribution with a mean of 2.3837 and a standard deviation of 0.2857, whereas bdi was fitted to a Gamma distribution with a shape parameter of 1.7678 and a scale parameter of 6.8145. These specific parameters were determined through fitting to past data. In order to sample only at the points of $\mathbf{X}$ these distributions were discretized. In general, the sequential procedure is the same whether the covariates are discrete or continuous. The model parameters for each treatment are as follows:

(i) For treatment 1: $(\alpha_1, \beta_{1,1}, \beta_{1,2}) = (2.0941, 3.9057, 1.1089)$, and $\sigma_1^2 = 207.9267$;
(ii) For treatment 2: $(\alpha_2, \beta_{2,1}, \beta_{2,2}) = (-0.1015, 4.1312, 1.5117)$, and $\sigma_2^2 = 107.8993$.

Although the uncertainty of treatment 1 it emerges as the preferable option due to its lower expected Quality of Life (QoL). Upon visual examination of the graphical representations of both models forecasts, we discern that within a specific region ($\Omega \equiv (h\&q, bdi) \in [1, 5] \times [1, 35] \cap 0.2251 \cdot h\&q + 0.4028 \cdot bdi \leq 2.1956$) of the design space where both bdi and $h\&q$ are low, treatment 2 gains preference.

In this specific case, the value of $\tau = \sigma_2^2/\sigma_1^2$ is calculated to be 0.5189. Consequently, applying the tools outlined in §3-§4, the static optimal allocation lead to $w_1 = 0.5$ ($w_2 = 0.5$) if D-optimality criterion is used, $w_1 = 0.5813$ ($w_2 = 0.4187$) if A- or D_A-optimality criteria are considered and $w_1 = 0.6584$ ($w_2 = 0.3416$) if E-optimality is employed. These allocations provide validation for the observed trends in the sequential allocation scheme.

We now apply the algorithm outlined in §5 for assigning treatments to individuals using all the optimality criteria considered (A-, D-, E- and D_A-optimality). The D_A-optimality criterion is used to find the parameter differences $\alpha_1 - \alpha_2$, $\beta_{1,1} - \beta_{2,1}$ and $\beta_{1,2} - \beta_{2,2}$, i.e., the matrix A have a structure similar to that considered in Theorem 4. For simulation purposes, we fix $n_s = 1000$ and $n_p = 300$ (equivalent to 300 individuals). The values of $h\&q$ and bdi are randomly sampled from a normal distribution with a mean of 2.3837 and a standard deviation of 0.2857, and from a Gamma distribution with a shape parameter of 1.7678 and a scale parameter of 6.8145, respectively, after which they are rounded to the nearest integer. The prediction error ϵ_i for each treatment is simulated sampling from a normal distribution $\mathcal{N}(0, \sigma_i^2)$, σ_i^2 listed above.

The results in Figure 2, exhibit a display akin to the example illustrated in §5. We note that

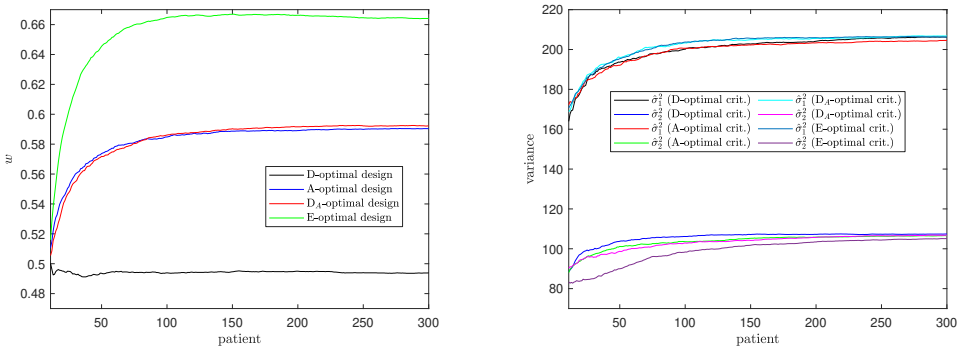
(i) For the D-optimality criterion, $w_1 \to 0.5$, as shown in Figure 2(a). Here, the value is systematically below 0.5 but the difference is below 0.05.
(ii) For both the A-optimality and D_A-optimality criteria, $w_1 \to 0.58$.
(iii) Under the E-optimality criterion, $w_1 \to 0.65$.

In all scenarios, the convergence values closely match those derived from asymptotic designs utilizing the theoretical results of §3. This exhibits the suitability of these results for establishing optimal *a priori* allocation policies, particularly when good estimates of the response variances are available before the start of the trial, or when sequential allocation is not possible. Additionally, as depicted in Figure 2(b), we observe the convergence of the estimates of observational error variance for both treatments towards their respective "true" values.

7. Discussion

One assumption in this paper is that it is known that both the response model and the variance of the response depend upon the treatment allocated. However, at the start of the Phase III trial, there may be no clear indication of the existence of a subgroup of patients with a distinct response pattern. Yang *et al.* (2024) describe enrichment designs in which the treatment allocations are changed once a significant subgroup has been identified.

We have assumed that, within each group, the variance of the response is constant. A referee has commented that the variance may depend not only on treatment, but also on other covariates. In this case there is an additional, often linear, model for the response variance, perhaps also depending on treatment. Atkinson and Cook (1995) derive optimum designs when there is a single response model with Fedorov and Leonov (2014, Chapter 6 and 7) providing pharmaceutical examples.



(a) (b)

Figure 2. Sequential optimal designs for response models in representing the QoL in Parkinson's disease treatment comparison. Two covariates; Model parameters: $(\alpha_1, \beta_{1,1}, \beta_{1,2}) = (2.0941, 3.9057, 1.1089)$, $(\alpha_2, \beta_{2,1}, \beta_{2,2}) = (-0.1015, 4.1312, 1.5117)$; Variances: $\sigma_1^2 = 207.9267$, $\sigma_2^2 = 107.8993$; Covariates domain: $x_1 \in [1, 5]$, $x_1 \sim \mathcal{N}(2.3837, 0.2857)$, $x_2 \in [1, 35]$, $x_2 \sim \Gamma(1.7678, 6.8145)$ (a) w_1 for D-, A-, E- and D_A-optimality criteria; (b) estimated observational error variances for D-, A-, E- and D_A-optimality criteria.

The sequential design procedures described in this paper are deterministic; given the history of covariates and allocations, conveniently summarisable as sufficient and ancillary statistics, the allocation to the next individual is certain. Our interest in the designs and algorithms described in this paper is for their use in clinical trials, when some randomization should be introduced into the allocation to avoid biases. For sequential trials following the standard model of a constant treatment effect, Atkinson (1982) suggested an allocation rule based on the sensitivity function of D_A-optimality with a biased-coin randomization rule calculated from the sensitivity functions of the treatments. A negative effect of randomization is a slight loss in the precision of parameter estimates. Several randomization rules, some based on sequential optimum design for the standard model, and the associated losses are described and compared in Atkinson (2014). A recent survey of the balance between statistical efficiency and randomness in clinical trails is Sverdlov and Ryeznik (2023).

As demonstrated, the allocation of individuals in trials based on information criteria serves as a natural and statistically robust method for maximizing information. This method relies on a transparent metric to gauge the information contained within each response. Moreover, it opens up numerous avenues for further research. One such avenue involves exploring models of increasing complexity, including nonlinear and Gaussian process models dependent on covariates. Additionally, a common and intriguing area of investigation involves considering the dependence of response variance on prognostic factors, mentioned above. By incorporating the relationships between predictor variables and variance into the model, we relax the assumption of homoscedastic variance errors (Snijders and Bosker, 2012). Such an approach would also necessitate the utilization of techniques tailored to nonlinear model responses.

However, design criteria from the theory of optimum experimental design focus on information maximization. Ethical considerations, on the other hand, require that as many individuals as possible receive the best treatment; in the case of personalized medicine, the best treatment for the individual (Palmer and Rosenberger, 1999). This leads to minimum regret designs, which seek to maximize the proportion of individuals who receive the best treatment. Of course, if the parameters are poorly estimated many allocations will be incorrect. One way forward is to combine information maximization with minimum regret-based designs. Here, ethics and information can be combined by the use of a compound design criterion which contains a flexible parametric combination of measures of the achievement

of the two goals. These designs offer an interesting avenue to explore in the future.

Acknowledgements. We are most grateful to the referees whose careful and thoughtful comments led to much improvement of our paper.

References

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

Atkinson, A.C., 2014. Selecting a biased-coin design, *Statistical Science*, 29, 144–163.

Atkinson, A.C., 2015. Optimum designs for two treatments with unequal variances in the presence of covariates, *Biometrika*, 102, 494–499.

Atkinson, A.C. and Cook, R.D., 1995. D-optimum designs for heteroscedastic linear models, *Journal of the American Statistical Association*, 90, 204–212.

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

Atkinson, A.C., Duarte, B.P., Pedrosa, D.J., and van Munster, M., 2023. Randomizing a clinical trial in neuro-degenerative disease, *Contemporary Clinical Trials Communications*, 33, 101140.

Bai, Z.D., Hu, F., and Rosenberger, W.F., 2002. Asymptotic properties of adaptive designs for clinical trials with delayed response, *Annals of Statistics*, 30, 122–139.

Beck, A.T., Steer, R.A., and Pompili, M., 1988. *BHS, Beck hopelessness scale: Manual*, Psychological corporation San Antonio, TX.

Ben-Tal, A. and Nemirovski, A.S., 2001. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, Philadelphia: Society for Industrial and Applied Mathematics.

Berger, M. and Wong, W.K., 2009. *An Introduction to Optimal Designs for Social and Biomedical Research*, New York: Wiley.

Brenner, J., 1972. The Hadamard maximum determinant problem, *The American Mathematical Monthly*, 79 (6), 626–630.

Burnett, T., Mozgunov, P., Pallmann, P., Villar, S.S., Wheeler, G.M., and Jaki, T., 2020. Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs, *BMC Medicine*, 18, 1–21.

Cramér, H., 1999. *Mathematical Methods of Statistics*, vol. 43, Princeton University Press.

Duarte, B.P.M., Granjo, J.F.O., and Wong, W.K., 2020. Optimal exact designs of experiments via Mixed Integer Nonlinear Programming, *Statistics and Computing*, 223, 505–560.

Duarte, B.P.M., Wong, W.K., and Oliveira, N.M.C., 2016. Model-based optimal design of experiments - Semidefinite and nonlinear programming formulations, *Chemometrics and Intelligent Laboratory Systems*, 151, 153 – 163.

Fedorov, V.V. and Leonov, S.L., 2014. *Optimal Design for Nonlinear Response Models*, Boca Raton: Chapman and Hall/ CRC Press.

Hadamard, J., 1893. Résolution d'une question relative aux déterminants, *Bull. des Sciences Math.*, 2, 240–246.

Harman, R. and Sagnol, G., 2015. Computing D-optimal experimental designs for estimating treatment contrasts under the presence of a nuisance time trend, *in:* A. Steland, E. Rafajłowicz, and K. Szajowski, eds., *Stochastic models, Statistics and their Applications*, Springer, vol. 122, 83–91.

Hoehn, M.M. and Yahr, M.D., 1967. Parkinsonism, *Neurology*, 17 (5), 427–427.

Hoehn, M.M. and Yahr, M.D., 1998. Parkinsonism: onset, progression, and mortality, *Neurology*, 50 (2), 318–318.

Hu, F., Rosenberger, W.F., and Zhang, L.X., 2006. Asymptotically best response-adaptive randomization procedures, *Journal of Statistical Planning and Inference*, 136 (6), 1911–1922.

Hyun, S.W., Huang, T., and Zhu, H., 2016. Efficient and ethical adaptive clinical trial designs to detect treatment-covariate interaction, *in:* O. Sverdlov, ed., *Modern Adaptive Randomized Clinical Trials*,

Boca Raton FL: CRC Press, 309 – 326.
Kiefer, J., 1959. Optimum experimental designs (with discussion), *J. Roy. Statist. Soc. B*, 21, 272–319.
Kiefer, J., 1961. Optimum experimental designs V, with applications to systematic and rotatable designs, *in: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Univ of California Press, vol. 1, 381–405.
Margolese, R.G., Cecchini, R.S., Julian, T.B., Ganz, P.A., Costantino, J.P., Vallow, L.A., Albain, K.S., Whitworth, P.W., Cianfrocca, M.E., Brufsky, A.M., and et al, 2016. Anastrozole versus tamoxifen in postmenopausal women with ductal carcinoma in situ undergoing lumpectomy plus radiotherapy (NSABP B-35): a randomised, double-blind, phase 3 clinical trial, *The Lancet*, 387 (10021), 849–856.
Neyman, J., 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection, *Journal of the Royal Statistical Society*, 97 (4), 558–625.
Palmer, C.R. and Rosenberger, W.F., 1999. Ethics and practice: Alternative designs for phase III randomized clinical trials, *Controlled Clinical Trials*, 20, 172–186.
Peto, V., Jenkinson, C., and Fitzpatrick, R., 1998. PDQ-39: a review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures, *Journal of Neurology*, 245 (1), S10–S14.
Pukelsheim, F. and Rieder, S., 1992. Efficient rounding of approximate designs, *Biometrika*, 79, 763–770.
Rao, C.R., 1945. Information and the accuracy attainable in the estimation of statistical parameters, *Reson. J. Sci. Educ*, 20, 78–90.
Rosenberger, W.F., 1993. Asymptotic inference with response-adaptive treatment allocation designs, *Annals of Statistics*, 21, 2098–2107.
Rosenberger, W.F. and Lachin, J.L., 2016. *Randomization in Clinical Trials: Theory and Practice, 2nd edn*, New York: Wiley.
Rosenberger, W.F. and Sverdlov, O., 2008. Handling covariates in the design of clinical trials, *Statistical Science*, 23, 404–419.
Silvey, S.D., 1980. *Optimum Design*, London: Chapman and Hall.
Snijders, T.A.B. and Bosker, R.J., 2012. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*, Sage, 2nd ed.
Sverdlov, O. and Ryeznik, Y., 2023. Balancing the objectives of statistical efficiency and allocation randomness in randomized controlled trials, *Statistics in Biopharmaceutical Research*.
Sverdlov, O., Ryeznik, Y., and Wong, W.K., 2020. On optimal designs for clinical trials: An updated review, *Journal of Statistical Theory and Practice*, 14 (1), 1–29.
The Stroke Prevention in Atrial Fibrillation Investigators, 1990. Design of a multicenter randomized trial for the Stroke Prevention in the Atrial Fibrillation Study, *Stroke*, 21, 538–545.
Vandenberghe, L. and Boyd, S., 1996. Semidefinite programming, *SIAM Review*, 8, 49–95.
Vandenberghe, L. and Boyd, S., 1999. Applications of semidefinite programming, *Applied Numerical Mathematics*, 29, 283–299.
Wang, Y. and Ai, M., 2016. Optimal designs for multiple treatments with unequal variances, *Journal of Statistical Planning and Inference*, 171, 175–183.
Yang, L., Diao, G., and Rosenberger, W.F., 2024. A two-stage covariate-adjusted response-adaptive enrichment design, *Statistics in Biopharmaceutical Research*, dOI: 10.1080/19466315.2024.2308877.
Ye, Y., 1997. *Interior Point Algorithms: Theory and Analysis*, New York: John Wiley & Sons.

Appendix A. Formulations to determine the optimal allocation via SDP

Here we list SDP formulations for the D-, A-, E- and D_A-optimality criteria. The first three were introduced in Vandenberghe and Boyd (1996, 1999); Ben-Tal and Nemirovski (2001).

We start with the formulation for D-optimal designs.

$$\text{Opt} = \max_{\xi, B} t \tag{A.1a}$$

$$\text{s.t.} \begin{pmatrix} M(\xi) & B^{\mathsf{T}} \\ B & \operatorname{diag}(B) \end{pmatrix} \succeq 0_{4(k+1)} \tag{A.1b}$$

$$t \leq \prod_{i=1}^{2(k+1)} B_{i,i}^{1/[2(k+1)]} \tag{A.1c}$$

$$\sum_{i=1}^{I} w_i = 1 \tag{A.1d}$$

$$0 \leq w_i \leq 1, \quad i \in [\![I]\!]. \tag{A.1e}$$

Now, the formulation for computing A-optimal designs

$$\text{Opt} = \min_{\xi, B} t \tag{A.2a}$$

$$\text{s.t.} \begin{pmatrix} M(\xi) & I_{2(k+1)} \\ I_{2(k+1)} & B \end{pmatrix} \succeq 0_{4(k+1)} \tag{A.2b}$$

$$t \geq \sum_{i=1}^{2(k+1)} B_{i,i} \tag{A.2c}$$

$$\sum_{i=1}^{I} w_i = 1 \tag{A.2d}$$

$$0 \leq w_i \leq 1, \quad i \in [\![I]\!], \tag{A.2e}$$

and E-optimal designs.

$$\text{Opt} = \max_{\xi, t} t \tag{A.3a}$$

$$\text{s.t. } M(\xi) - t\, I_{2(k+1)} \succeq 0_{2(k+1)} \tag{A.3b}$$

$$\sum_{i=1}^{I} w_i = 1 \tag{A.3c}$$

$$0 \leq w_i \leq 1, \quad i \in [\![I]\!]. \tag{A.3d}$$