Anthony C. Atkinson and David J. Pedrosa

Optimum design and sequential treatment allocation in an experiment in deep brain stimulation with sets of treatment combinations

Article (Accepted version) (Refereed)

Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Optimum design and sequential treatment allocation in an experiment in deep brain stimulation with sets of treatment combinations

Anthony C. Atkinson[a]*and David J. Pedrosa[b]

In an experiment including patients who underwent surgery for Deep Brain Stimulation electrode placement, each patient responds to a set of nine treatment combinations. There are sixteen such sets and the design problem is to choose which sets should be administered and in what proportions. Extensions to the methods of non-sequential optimum experimental design lead to identification of an unequally weighted optimum design involving four sets of treatment combinations. In the actual experiment patients arrive sequentially and present with sets of prognostic factors. The idea of loss due to Burman is extended and used to assess designs with varying randomization structures. It is found that a simple sequential design using only two sets of treatments has surprisingly good properties for trials with the proposed number of patients.

Keywords: Burman's loss, extended equivalence theorem, prognostic factors, randomization, sequential design, treatment balance

1. Introduction

The number of readings taken on individual patients may be limited by safety concerns. Our specific example is patients receiving Deep Brain Stimulation (DBS). This relatively novel approach is currently applied on a series of different neurological symptoms such as Parkinson's Disease (PD), pain syndromes or depression [1], [2], [3], especially when conventional treatments are ineffective or cause side effects. For this purpose, electrodes implanted in specific brain areas deliver current pulses either constantly [4] or on demand [5], [6], resulting in dramatic symptom relief. A striking example of successful DBS treatment is tremor. This rhythmic oscillation of one or more extremities is best known as a cardinal symptom of PD. It may, however, also emerge without any other symptoms, being then termed Essential Tremor (ET). As only a minority of patients suffering from ET receives DBS surgery, studies analyzing working mechanisms of this resource-intensive treatment usually contain only moderate sample sizes.

Mechanisms responsible for tremor generation remain a question of debate. Whereas initial investigations presumed a single node responsible for tremor generation, recent findings rather indicate a network of different brain areas interacting and hence producing aberrant signals which are transmitted to the extremities (for review see [7]). Therefore, techniques investigating the entire brain are preferable to promote insight into tremor pathomechanisms while, at the same time, they may also shed light on working mechanisms of DBS. Radioactive imaging and specifically Positron-emission tomography (PET) is an example of such a technique that has already fostered our understanding of many aspects of tremor [8], [9], [10]. PET reproduces the metabolic activity of the brain, that is prominent brain activation during tremor and activation changes due to DBS. Nevertheless, application of radioactive agents is problematic as ET-patients, many of whom are in

[a] *Department of Statistics,*
London School of Economics, London WC2A 2AE, UK
[b] *Klinik und Poliklinik für Neurologie, Universitätsklinik Gießen und Marburg, Baldingerstrasse, D35043 Marburg*
* *Correspondence to: Department of Statistics,*
London School of Economics, London WC2A 2AE, UK. E-mail: a.c.atkinson@lse.ac.uk

middle age, are often reluctant to undergo a potentially harmful investigation. Additionally, dosage limitations per patient leave only a reduced number of conditions which may be ascertained.

In the proposed experiment there are twelve experimental conditions at which measurements can be taken. Although three observations are made on each patient at each stimulation level, it is not possible to observe the response under all twelve conditions and patients are limited to being measured at nine of the conditions. Further restrictions lead to an experimental region consisting of sixteen sets of nine conditions under which patients can be observed. It seems clear that not all sixteen conditions will provide nine observations that are equally informative. The optimum experimental design problem is to determine how many and which of the sixteen sets should be used and in what proportions.

The second design problem is that patients present sequentially with a set of prognostic factors; allocation of treatment combination is then ideally by a form of restricted randomization which aims at some balance over the prognostic factors. An important practical problem is to find a simple sequential design suitable for clinical practice and to measure its efficiency and balancing properties relative to those of the optimum design.

These problems are solved using the theory of optimum experimental design for multivariate observations. In the standard theory for such observations different responses are measured under the same conditions. The technical novelty here is that each multivariate response consists of measurements of the same quantity but under sets of different experimental conditions.

Section 2 introduces the complicated structure of the experiment. Optimum experimental design is introduced in §3 and extended to include experiments with sets of treatment combinations. These more general results are used in §4 to calculate the D-optimum design for the stimulation experiment. It is found that there are 16 distinct D-optimum designs, all with the same value of the design criterion. Sequential treatment allocation is introduced in §5.1 with the sequential construction of designs with prognostic factors in §5.2. The designs are compared using the extension of an idea of loss introduced by [11], which is developed in §5.3. Numerical properties of three sequential designs using this development are in §6. The paper concludes with a brief discussion. The appendix describes the construction of the design matrix.

The paper illustrates the usefulness of the extensions to optimum design theory and to the assessment of sequential designs. An interesting conclusion, in this example, is that a very simple sequential design is virtually as efficient as an appreciably more complicated optimum one.

2. The Structure of the Experiment

Recent literature indicates that while stimulation at high frequencies (HFS) abolishes tremor, it may inversely affect a patient's verbal fluency abilities [12], [13]; ET-patients stimulated at 130Hz may be less shaky while at the same time the ability to name words starting at a specific letter (phonemic verbal fluency) will decay. Otherwise, with pulses delivered at 10Hz, tremor severity increases as does phonemic verbal fluency. Hence, three levels of stimulation were interesting: high frequency stimulation (HFS, 130Hz), low frequency stimulation (LFS, 10Hz) and Off. In the rest of the paper, the three stimulation settings are denoted St_1, St_2 and St_3. The PET recordings are made when the patient is performing one of four activities: resting, lifting the right arm (in order to produce postural tremor), pointing at an object to induce intentional tremor or, in one minute, naming as many words as possible starting with a given letter. These activities will be referred to as a, b, c and d. There are therefore twelve potential treatment combinations. Each patient is measured at each of the three stimulation levels $St_1 - St_3$. However, due to restrictions on the maximum amount of radiation, each patient can only be measured for three of the four activities. There are thus nine observations per patient under distinct treatment combinations. A further constraint on these observations is that in the on or off stimulation levels (St_1, St_3), activities a and d must be included (but not necessarily for St_2). The reason for this constraint is that PET images require normalization because of interindividual differences which may be considerable due to e.g., gender, age or blood flow differences. According to the primary objectives of this study, that is to ascertain differences in activation due to stimulation during either tremor or verbal fluency, the rest condition a with and without stimulation and verbal fluency performance d at both conditions may serve as this reference. Group analyses imply these normalised images in a second step.

Table 1 lists the resulting eight combinations of stimulation level and activity for a single observation. Each patient will receive treatment St_1 and be measured either under the activities in combination A or B, giving rise to three observations. The same patient will receive treatment St_2 and be measured under one of the combinations C, D , E or F. This patient will also receive treatment St_3 and be measured either under the combination G or H. In all, each patient provides nine measurements. To find the design points, that is the allowable values of treatment and activity, we need to select rows from Table 1, one for each treatment. For treatment St_1 there are two combinations, for treatment St_2 there are four and for treatment St_3 two. The resulting 16 = $2 \times 4 \times 2$ sets of experimental conditions are listed in Table 2. In the nomenclature of optimal experimental design, these 16 design points define the design region $\mathcal{X}$. Any experimental design consists of allocating a defined proportion of patients to receive the treatment combinations specified by the members of $\mathcal{X}$. In our

Table 1. The eight allowable combinations of stimulation level and activity for three observations at a single stimulation setting

Combination	Treatment	Activities		
A	St_1	a	b	d
B	St_1	a	c	d
C	St_2	a	b	c
D	St_2	a	b	d
E	St_2	a	c	d
F	St_2	b	c	d
G	St_3	a	b	d
H	St_3	a	c	d

Table 2. The 16 points of the design region $\mathcal{X}$ for nine observations per patient, using the combinations given in Table 1

Point	Combination	Point	Combination
1	ACG	9	ACH
2	BCG	10	BCH
3	ADG	11	ADH
4	BDG	12	BDH
5	AEG	13	AEH
6	BEG	14	BEH
7	AFG	15	AFH
8	BFG	16	BFH

problem not all members of $\mathcal{X}$ are included in the optimum experiment.

3. Optimum Experimental Design

The result of a PET scan is a three-dimensional array of voxels which are analysed by qualitative comparisons to assess the effect of the treatments. Interest is in these effects which translates into good estimation of the parameters of the linear model of §3.1. If there were well-defined numerical responses with known statistical properties, D-optimality, incorporating the statistical model of the errors, would be the appropriate design criterion. Here we do not have numerical responses, but the problem does require the methods of optimum design to provide an efficient and compact design. In order to proceed we make the simplest possible assumption which is D-optimality for homoskedastic errors. We also make the simplest assumption about the nine readings on each patient, namely that they are independent. D-optimum designs for homoskedastic errors are optimum, for example, for responses that are transformed to approximate normality before analysis. They are also close to optimum for generalized linear models with small effects [14], such as would be appropriate if the response were binary, recording whether or not a particular change was observed. Comments on appropriate design criteria when there are numerical responses are in §7. Standard experimental designs, such as the factorial underlying the stimulation experiment, are D-optimum. Our choice is supported by the balanced properties of the designs we found, which are briefly noted in §4.

3.1. Models and Information Matrices

Single readings are taken on each patient under the nine sets of conditions specified in Tables 1 and 2. The model for a single observation is first order in stimulation (three levels) and activity (four levels), so that there are 12 possibilities. If the single observation ℓ comes from a patient receiving stimulation level j and is measured under activity k the linear model for the expected response is

$$E(y_\ell) = \mu + St_j + A_k, \tag{1}$$

where μ is the overall mean. The model includes indicator variables for the mean and the levels of the two factors. The matrix of indicator variables for the twelve single observations is given in Table 6. For this full factorial experiment the

model is over parameterized. As we show in the Appendix, the full rank model has $p = 6$ parameters rather than eight and is conveniently written as

$$E(y_\ell) = \beta^T f_\ell, \quad (\ell = 1, ..., 12). \tag{2}$$

Row ℓ of the 12×6 design matrix F for the full factorial experiment is f_ℓ^T.

With independent errors of observation with constant variance σ^2, estimation of the parameters β is by least squares. The information matrix for an individual observation at conditions ℓ, standardized by taking $\sigma^2 = 1$, is $M_\ell = f_\ell f_\ell^T$.

In the stimulation experiment each patient contributes nine observations. Tables 1 and 2 specify the nine distinct values of ℓ (out of 12) at which observations can be taken on each patient. The last paragraph of the Appendix illustrates calculation of the nine values of ℓ for the first point of $\mathcal{X}$ in Table 2. For patient i let this set be $S(i)$. Then the information matrix for subject i is

$$M_i = \sum_{\ell \in S(i)} M_\ell = \sum_{\ell \in S(i)} f_\ell f_\ell^T. \tag{3}$$

We now consider experimental design. As is standard in the theory of optimum experimental design (for example[15] or [16]), an experimental design ξ allocates a fraction w_i of the patients to condition x_i. A design with m points of support is written as

$$\xi = \left\{ \begin{array}{cc} x_1 & x_2 \dots x_m \\ w_1 & w_2 \dots w_m \end{array} \right\}, \tag{4}$$

where $w_i > 0$ and $\sum_{i=1}^m w_i = 1$. In our example $m \leq 16$. Any realisable experimental design for a total of N patients will require that the weights are ratios of integers, that is $w_i = r_i/N$, where r_i is the number of patients at condition x_i. The mathematics of finding optimal experimental designs and demonstrating their properties is greatly simplified by the consideration of continuous designs in which the integer restriction is ignored. Applicable designs may have to be found by rounding the w_i.

In our experiment the design region $\mathcal{X}$ consists of the 16 design points x_i which each specify one set of the nine values of ℓ at which measurements are taken on a patient. For a patient who is measured at the design point x_i let this set be $S(x_i)$. From (3) the information matrix for the design ξ with m support points can then be written

$$M(\xi) = \sum_{i=1}^m w_i M_i = \sum_{i=1}^m w_i \sum_{\ell \in S(x_i)} M_\ell = \sum_{i=1}^m w_i \sum_{\ell \in S(x_i)} f_\ell f_\ell^T. \tag{5}$$

The 16 points $x_i \in \mathcal{X}$ specify the sets of 9 rows of F which are allowed in the experiment.

D-optimum designs, minimizing the generalized variance of the estimates of β, maximize the determinant $|M(\xi)|$ over the design region $\mathcal{X}$ through choice of the optimum weights w_i^* of the optimum design ξ^*.

3.2. Numerical

The optimum designs in the deep-brain stimulation experiment have to be found numerically. Much of the discussion in the experimental design literature, for example [16, Chapter 3], stresses the desirability of using algorithms that take account of the specific structure of optimum designs. However, the designs for this paper were found using a general purpose numerical algorithm.

The design maximizing $|M(\xi)|$ requires finding the optimum design weights w_i^*. These must be non-negative and sum to one. [15, §9.5] suggest search over an unconstrained space Ψ, using transformation to polar co-ordinates to calculate weights w_i that satisfy the required constraints. Here use was made of a simpler approach taking advantage of the upper and lower constraints on variables in the R function `optim`.

The search variables are ψ_i. Taking

$$w_i = \psi_i / \sum_{j=1}^m \psi_j \tag{6}$$

with $0 \leq \psi_j \leq 1$ provides weights that satisfy the required constraints and are used in (5) to calculate $M(\xi)$. Of course, the w_i are in $m - 1$ dimensions, so that (6) is not unique; the same weights are obtained when all ψ_i are replaced by $a\psi_i, (a \neq 0)$. However the Quasi-Newton BGFS algorithm in the function `optim` did not show any difficulty in converging.

R code for these calculations is available from the first author.

3.3. An Extended General Equivalence Theorem

For experiments with a single observation at each point of $\mathcal{X}$, whether a particular design is D-optimum can be determined from the "general equivalence theorem" for D-optimality due to [17]. In §4 we use an extension of the theorem to elucidate the structure of the optimum design. In §5 this provides a method for sequential treatment allocation with some randomization.

The model for a single response is given in (1). For an experiment with only one reading per patient, it follows from (3) and (5) that the information matrix for a design ξ is

$$M(\xi) = \sum_{i=1}^{m} w_i f_i f_i^T.$$

The equivalence theorem is concerned with the value of sensitivity function

$$d(i, \xi) = f_i^T M^{-1}(\xi) f_i, \tag{7}$$

which, in this case, is the variance of the predicted response at each design point.

For the stimulation experiment the theorem applies, at each design point x_i, to average values over the nine readings per patient. For generality, let there be h such readings. We require the average information matrix for the design ξ; $M_{\text{AVE}}(\xi) = M(\xi)/h$. The average sensitivity function, from (7) is then

$$d_{\text{AVE}}(x, \xi) = \sum_{\ell \in S(x_i)} d(\ell, \xi)/h, \tag{8}$$

where $M(\xi)$ is now given by (5). If $\bar{d}_{\text{AVE}}(\xi)$ is the maximum over $\mathcal{X}$ of $d_{\text{AVE}}(x, \xi)$ and ξ^* is the D-optimum design, the general equivalence theorem for sets of treatment combinations states the equivalence of the following three conditions on ξ^*:

1. The design ξ^* maximizes $|M(\xi)|$;
2. The design ξ^* minimizes $\bar{d}_{\text{AVE}}(\xi)$;
3. The value of $\bar{d}_{\text{AVE}}(\xi^*) = p$, this maximum occurring at the points of support of the design.

As a consequence of 3, we obtain the further condition:

4. For any non-optimum design the value of $\bar{d}_{\text{AVE}}(\xi) > p$.

For $h = 1$ the theorem reduces to that of [17]. The D-optimum design maximizes the determinant $|M(\xi)|$ or equivalently $|M_{\text{AVE}}(\xi)|$. The efficiency of any other design ξ is $\{|M(\xi)|/|M(\xi^*)|^{1/p}$.

The proof of this theorem follows from the additive nature of the information matrix, which itself is a consequence of the independence assumed for the nine measurements on each patient. Standard proofs of the equivalence theorem for univariate observations, such as those in [16, §2.4.2] and [18, §5.2], depend on the directional derivative at a point in $\mathcal{X}$. Here, with the extension to a set of observations, the directional derivative is the sum of the derivatives for the individual observations.

This result is distinct from the equivalence theorem for multivariate observations([15] or [19]) in which several different responses are measured on each unit under the same experimental conditions. Here, the h conditions are specified by $S(i)$ at all of which the same response is measured.

Applications of this extended theorem to response surface designs with sets of treatment combinations are given by [20]. [21] finds designs with treatment combinations for logistic models. Results in [22] for multivariate observations indicate how the results given here might be extended to handle correlation between observations in the same set.

4. D-Optimum Design

The design region $\mathcal{X}$ contains 16 points, but it is unlikely that all will be included in an optimum experiment. Each patient is allocated to one design point and generates $h = 9$ responses from the specified combinations of stimulation level and activity. The first-order model in these two factors contains 6 parameters. For some designs with univariate responses m^*, the number of support points in the optimum design, equals p, but designs with multivariate responses can have fewer than p support points.

Table 3. "Stronger" and "weaker" pairs of design points

Stronger pairs		Weaker pairs	
Point	Point	Point	Point
1	16	3	14
2	15	4	13
7	10	5	12
8	9	6	11
Optimum Weights			
0.36972	0.36972	0.13028	0.13208

To start the numerical exploration of designs, we used weights calculated from (6) to calculate the information matrix $M(\xi)$ in (5). Numerical optimization of the weights over all 16 design points gave several 16-point designs with the same maximized value $|M(\xi^*)| = 12.89389 \times 10^{-5}$. The designs all satisfied the equivalence theorem with $d_{\text{AVE}}(x_i) = 6$ at each design point although the weights depended on the starting point of the numerical search. Such behaviour is evidence of the non-uniqueness of the optimum design. In general, if there are several distinct optimum designs, then any convex linear combination of these component designs will also be optimum.

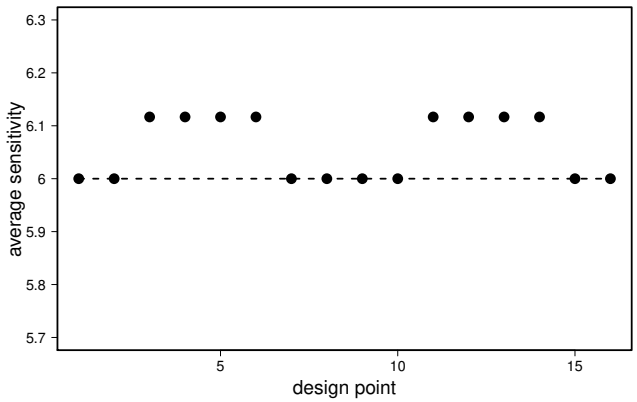

Figure 1. Equivalence theorem. Average sensitivities $d_{\text{AVE}}(x, \xi^*_{(2)})$ for design at a pair of "stronger" design points. The values of 6.1167 at the "weaker" design points show that this design is not D-optimum

To find the component designs searches were made over pairs, triples and so on of design points from $\mathcal{X}$. For each selection of design points, the optimum weights had also to be found numerically. Looking at pairs of design points gives $16 \times 15/2 = 60$ distinct designs. Let these designs be called $\xi_{(2)}$. Searching over these pairs gave four optimum designs for which $|M(\xi^*_{(2)})| = 12.70132 \times 10^{-5}$. For these optimum designs the weights on the two design points were equal. The designs have a very high D-efficiency of $(12.70132/12.89389)^{1/6} = 99.75\%$.

The support points of these four designs were the pairs of points (1,16), (2,15), (7,10) and (8,9), which are called the "stronger pairs" in Table 3. Figure 1 plots the values of the sensitivity function $d_{\text{AVE}}(x, \xi^*_{(2)})$ for the design using the pair (8,9). The sensitivity function is 6 for this pair, as it is for the other stronger pairs. The implication is that designs consisting of convex linear combinations of the four two-point designs will have the same value of $|M(\xi)|$. However, for the other eight support points the value of the sensitivity function is, as Figure 1 shows, 6.1167 at the support points of the weaker pairs. Clearly some of these need to be included in the optimum design.

Numerical search, confirmed by the equivalence theorem, shows that the optimum designs contain 4 support points (so that $m^* < p$). These designs are listed in Table 3. Two of the support points form one of the stronger pairs (left-hand half of table), each with weight 0.36972, and the other two come from one of the "weaker" pairs, each with weight 0.1328. Since any stronger pair can be combined with any weaker pair of points, there are 16 choices of D-optimum design. In addition convex linear combinations of these 16 designs are also D-optimum.

An interesting feature of the designs in Table 3 is that each stronger pair consists of combinations A, B, C, F, G and H. From Table 1 it follows that three activities are measured at each of the two occurrences of the three treatments St_1, St_2 and St_3. The total number of activities in the design are $5a, 4b, 4c$ and $5d$, which is the highest balance possible. The

weaker pairs all consist of combinations A, B, D, E, G and H. Now the total numbers of activities in the design are $6a, 3b, 3c$ and $6d$. These design are less balanced.

Although there is a large class of D-optimum designs, in which all points in $\mathcal{X}$ appear, there are some designs with a low efficiency. The two-point design based on a stronger pair has a D-efficiency of 99.75% whereas that using a weaker pair has an efficiency of 98.01%. Despite these high values, the choice of design points is important; the poorest two-point designs have a determinantal ratio $|M(\xi_{(2)})|/|M(\xi^*)| = 0.554$, so that the D-efficiency is only 90.6%.

5. Sequential Treatment Allocation

5.1. Randomization and Prognostic Factors

In the proposed experiment, patients arrive sequentially and are to be allocated to one of the sixteen treatment combinations in $\mathcal{X}$. There may be several prognostic factors, such as initial severity of the disease, gender, age and body weight, known before allocation. For efficient estimation, some balance is required over such prognostic factors. The purpose is to avoid a disproportionate number of patients with a particular pattern of prognostic factors receiving the same combination. This goal can be met by using the sequential construction of optimum designs to determine the allocation for the next patient.

However, there should also be some randomization to avoid selection bias coming from the doctor's belief that some set of treatment combinations is best for one type of patient and to avoid confounding with unknown factors. See, for example, [23]. Randomization generally leads to a lack of balance in treatment allocation and so to some loss of efficiency if the trial can be stopped after treatment of an arbitrary number of patients. Different allocation rules give rise to different losses of efficiency.

5.2. Sequential Design Construction

The model for the full twelve-trial-factorial was written $E(y) = F\beta$, where y is 12×1 and F is 12×6. An experimental design for n patients chooses n sets of 9 rows from F. If, in addition to these combinations of factors, there are also prognostic factors, the model for n subjects becomes

$$E(\mathbb{Y}_n) = \mathbb{F}_n\beta + \mathbb{G}_n\gamma = \mathbb{H}_n\omega. \tag{9}$$

In our example the response vector $\mathbb{Y}_n$ is $9n \times 1$ and $\mathbb{F}_n$, which defines the set of treatment combinations, is $9n \times 6$. There are q prognostic factors for each patient contained in the $9n \times q$ matrix $\mathbb{G}_n$. All rows of this matrix for each patient are the same. Treatment allocation depends on the prognostic factors, but estimation of γ is not of interest - γ is a nuisance parameter.

In all there are $p + q$ parameters in the linear model (9). With the model terms divided into two groups, the information matrix is partitioned as

$$M(\xi_n) = \begin{bmatrix} \mathbb{F}_n^T\mathbb{F}_n/n & \mathbb{F}_n^T\mathbb{G}_n/n \\ \mathbb{G}_n^T\mathbb{F}_n/n & \mathbb{G}_n^T\mathbb{G}_n/n \end{bmatrix} = \begin{bmatrix} M_{11}(\xi_n) & M_{12}(\xi_n) \\ M_{12}^T(\xi_n) & M_{22}(\xi_n) \end{bmatrix}. \tag{10}$$

The covariance matrix for the least squares estimate of β is $M^{11}(\xi_n)$, the $p \times p$ upper left submatrix of $M^{-1}(\xi_n)$, that is

$$M^{11}(\xi_n) = \{M_{11}(\xi_n) - M_{12}(\xi_n)M_{22}^{-1}(\xi_n)M_{12}^T(\xi_n)\}^{-1}. \tag{11}$$

The D_S-optimum design for the subset of parameters β, with γ treated as a nuisance parameter, maximizes the determinant

$$|M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}^{\mathrm{T}}(\xi)| = |M(\xi)|/|M_{22}(\xi)|. \tag{12}$$

In sequential construction of the design the information matrix incorporates the knowledge about previous allocations and the values of the prognostic factors for all n patients. At the time of treatment allocation the vector of prognostic factors z_{n+1} is known for the new subject, so that the last h rows of $\mathbb{G}_{n+1}$ are known and can be written as copies of the vector $g^T(z_{n+1})$. The last rows of $\mathbb{F}_{n+1}$ are determined by the design point x which, as in (3) determines the values of ℓ. The last rows of $\mathbb{H}_{n+1}$ are thus of the form $h^T(\ell, z_{n+1}) = \{f_\ell^T, \quad g^T(z_{n+1})\}$. The optimum allocation to patient $n + 1$ then, from (12) maximizes the sensitivity function

$$d_s(x, z_{n+1}, \xi_n) = \sum_{\ell \in S(x)} \{h^T(\ell, z_{n+1})M_{\text{AVE}}^{-1}(\xi_n)h(\ell, z_{n+1}) - g^T(z_{n+1})M_{\text{AVE},22}^{-1}(\xi_n)g(z_{n+1})\}. \tag{13}$$

In the next section designs which allow for prognostic factors are compared with those in which there are no such factors. Then $M(\xi_n) = \mathbb{F}_n^T\mathbb{F}_n/n$ and, from (13), these designs are found by allocating to maximize the sensitivity function

$$d(x, \xi_n) = \sum_{\ell \in S(x)} f_\ell^T M_{\text{AVE}}^{-1}(\xi_n) f\ell.$$

5.3. Designs and their Assessment

Sequential construction of the D-optimum designs provides designs with highest efficiency, but they lack randomization; given all the relevant information, the treatment to be allocated to patient $n+1$ can be calculated with certainty. These designs are compared in the next section with two randomized designs. The D-optimum design of §4 has unequal weights at four treatment combinations. In one design treatments are randomized to these conditions with probabilities proportional to the design weights, the "correctly randomized" design. In the other, the "completely randomized" design, there is equal probability of allocating a patient to one of the four treatment combinations.

To compare designs we use an extension of the loss introduced by Burman [11] for two-treatment trials. The D-efficiency of a design ξ_n relative to the optimum design ξ^* is

$$E_n = \{|M(\xi_n)|/|M(\xi^*)|\}^{1/p}.$$

If there are prognostic factors the efficiency is instead defined in terms of the information matrices $M^{11}(\xi_n)$ and $M^{11}(\xi^*)$ from (11). For sequential design construction E_n rapidly tends to one. Randomized allocation is used to give reduced bias at some cost in efficiency. For randomized designs, two conditions have to be satisfied for E_n to tend to one:

(i). The randomized allocation must target the optimum design;

(ii). The prognostic factors must be well-behaved. If Z_n is the $n \times q$ matrix of prognostic factors, there must exist a limit for $Z_n^T Z_n/n$ as $n \to \infty$.

For a design with homoskedastic univariate responses and two treatments α_1 and α_2 interest is often in the treatment difference $\Delta\alpha = \alpha_1 - \alpha_2$. Then $\text{Var}(\Delta\hat{\alpha})$ is minimized with $n/2$ patients receiving each treatment when $\text{Var}(\Delta\hat{\alpha}) = 4\sigma^2/n$. For any other design the variance will be greater and can be written as $4\sigma^2/(n - L_n)$. Burman calls L_n the loss - it measures on how many patients information is lost compared to the optimum design.

The efficiency of the design is the ratio of the variances of $\Delta\hat{\alpha}$ under the two designs so that

$$E_n = 1 - L_n/n \qquad L_n = n(1 - E_n). \tag{14}$$

Both E_n and L_n are random variables, depending on the particular allocation, which may depend on the prognostic factors, current and past and the allocation history. The least efficient reasonable rule is random allocation, for which E_n rapidly converges to one. However (14) shows that the loss L_n does not therefore have to go to zero as n increases. Results on the asymptotic behaviour of L_n are given by [24] for a number of randomization rules. For random allocation $E(L_\infty) = q$. For deterministic allocation the asymptotic value is zero - all other reasonable rules have expectations between these two. See [25] for extensive comparisons for univariate responses. For the complicated treatment structure and multivariate response of the current design, calculation of the efficiency leads to the value of loss in (14). Apart from the behaviour under deterministic allocation, little is known even about asymptotic values. Simulation is used to determine the expectation of loss for finite sample sizes.

6. Numerical Comparison of Designs

Since the D-optimum design is known, the problem of sequential treatment allocation can be reduced from searching over the 16 treatment combinations in $\mathcal{X}$ to four points of one of the optimum designs of Table 3. Table 4 contains average values of L_n for three randomization rules for sequential treatment allocation:

(i). Sequential Design Construction

(ii). Correct Randomized Allocation (with the probabilities in Table 3).

(iii). Complete randomization over the four treatment combinations of the optimum design; each is then selected with probability one quarter.

In the argument in §5.3 leading to the relationship between loss and efficiency (14) each patient yielded one observation, so that the number of patients on whom information was conceptually lost due to imbalance was the same as the number of lost observations. With sets of treatment combinations each patient in general gives rise to h observations. Now loss can be defined in terms of lost observations, that is $L_n = nh(1 - E_n)$ with, in this experiment, $h = 9$.

Numerical results are in Table 4 for 1,000 simulations and for n up to 200, starting from a four-patient design with one patient allocated to each treatment combination. The results in the left-hand half of the table are for loss in the absence of prognostic factors. The first column show that sequential design construction does indeed produce designs with virtually zero loss. Correct randomization of the treatments with the frequencies of the optimum design produces a loss of around 0.6, whereas complete randomization produces values of loss that have increased to 4.585 on average when $n = 200$. This randomization is targeting a design with equal weights at the four support points for which $|M(\xi)| = 12.73439 \times 10^{-5}$ giving a D-efficiency of 99.79276%. From (14) the expected loss due to targeting a non-optimum design increases linearly

with n, here reaching 3.73 when $n = 200$. The values in the table are slightly higher than this due to the effect of randomization.

The results in the right-hand half of the table are for simulations including five prognostic factors, independently distributed with standard normal distributions. All losses increase due to the unbalancing effect of the covariates. Initially all designs are unbalanced with average losses in the range 12-13. With increasing n the losses for the sequential design are around 7.8. Those for the correctly randomized design are around one higher than those for the sequential design. The losses for the completely randomized design first decrease as the design becomes more balanced and then increase again due to the effect of targeting a non-optimum design. The increases in loss for the completely randomized designs compared to those targeting the optimum design are comparable in the presence or absence of prognostic factors.

These results depend on the distributions assumed for the prognostic factors. For two treatment designs with univariate responses, simulations in [26] compare the properties of designs when the factors have a lognormal distribution with those when the prognostic factors are normally distributed. The effect on loss is surprisingly small. A poor starting design is also rapidly discounted by the sequential procedures for the construction of designs. Here starting from the points of the optimum design, even if they are equally weighted, has a negligible effect, given the high D-efficiency of this design.

Table 4. Average loss for randomized and non-randomized sequential optimum designs for the six model parameters. Left-hand columns, no prognostic factors, right-hand columns, five prognostic factors

	No prognostic factors			Five prognostic factors		
n	Sequential design	Correctly randomized	Completely randomized	Sequential design	Correctly randomized	Completely randomized
10	6.34E-02	0.437	0.724	12.45	12.41	13.24
50	6.20E-03	0.618	1.726	8.16	9.41	11.10
100	4.92E-05	0.631	2.638	7.79	8.92	11.66
200	9.84E-05	0.604	4.585	7.71	8.60	13.28

The designs in Table 4 use four treatment combinations. In §4 designs on the stronger pair of design points were shown to have a very high D-efficiency of 99.74952%. We therefore conclude by examining the average loss for this simpler design.

Simulation results for average loss for the design with two treatment combinations are in Table 3. With only two treatment combinations, the optimum design puts half the observations at each combination. Consequently the correct randomization is complete randomization with probability one half. In both the presence and absence of prognostic factors, the losses for the sequential designs are close to those for the randomized designs due to the slight effect of randomizing over only two treatment combinations. Further, in both halves of the table, the losses for the randomized designs are similar to those for the completely randomized designs with four treatment combinations. This is caused by the similar efficiencies of the two designs: 99.79276% for equal replication of the four treatment combinations and 99.74952% for the design with two combinations. For this design the expected loss when $n = 200$ is 4.51, agreeing closely with the result for the sequential design in the absence of prognostic factors.

Table 5. Two treatment combinations. Average loss for randomized and non-randomized sequential designs for the six model parameters. Left-hand columns, no prognostic factors, right-hand columns, five prognostic factors

	No prognostic factors		Five prognostic factors	
n	Sequential design	Randomized	Sequential design	Randomized
10	0.255	0.581	11.65	11.61
50	1.127	1.530	8.85	9.57
100	2.254	2.670	9.84	10.51
200	4.509	4.897	12.14	12.35

7. Discussion

These result demonstrate the power of the extension of the general equivalence theorem in finding the structure of the D-optimum design for the experiment in deep brain stimulation. Further, the results show that the loss for the simpler sequential design randomizing over two treatment combinations is hardly less than that for the more complicated design requiring unequal randomization over four treatment combinations. Since it was expected that about 35 patients would be recruited overall, Tables 4 and 5 show there is virtually no loss in using the simpler design, which has appreciable operational advantages.

For the data on deep brain stimulation, the assumption of independent errors might with advantage be replaced by a linear mixed model, allowing for correlation between observations from individual patients. The generalized equivalence theorem extends to such designs in which the responses within a set are correlated. An introduction and references are given by [16, §2.11]. A further extension is that the design also has a split-plot component since stimulation levels are relatively hard to change. Readings under all three conditions are taken before the stimulation level is altered [27, Chapters 10 and 11]. The ordering of the conditions could also be considered in case there are carry-over effects from one activity to another.

Appendix: Models for the Full Factorial Experiment

The linear model for a single observation was written in (1) as

$$E(y_\ell) = \mu + St_j + A_k.$$

For the underlying full factorial experiment there are 12 combinations of treatment and activity. A parametrization of the design matrix in terms of indicator variables for the two factors is given in Table 6.

Table 6. Design matrix for the full 3×4 factorial experiment. There are eight columns of indicator variables and the model is overparameterized

Index	Mean	Treatment			Activity			
ℓ	μ	St_1	St_2	St_3	a	b	c	d
1	1	1	0	0	1	0	0	0
2	1	0	1	0	1	0	0	0
3	1	0	0	1	1	0	0	0
4	1	1	0	0	0	1	0	0
5	1	0	1	0	0	1	0	0
6	1	0	0	1	0	1	0	0
7	1	1	0	0	0	0	1	0
8	1	0	1	0	0	0	1	0
9	1	0	0	1	0	0	1	0
10	1	1	0	0	0	0	0	1
11	1	0	1	0	0	0	0	1
12	1	0	0	1	0	0	0	1

This model is overparameterized. The sum of the columns for the three treatments is the same as the column for the mean, as is the sum for the four columns of activity. A full-rank model, with $p = 6$, is conveniently obtained by deleting

the columns in Table 6 for the first treatment and the first activity to give the design matrix

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

The 16 design points in $\mathcal{X}$ correspond to various selections of nine rows from F. In Table 1 Combination A picks rows 1, 4 and 10. For Combination C $\ell = 2, 5, 8$ and for G, 3, 6 and 12. Thus, from Table 2 $S(x_1) = (1, 2, 3, 4, 5, 6, 8, 10, 12)$. These values of ℓ of course depend on the order of the rows of F.

Acknowledgement

We are grateful to the referees for comments which helped us clarify the argument of our paper, as did comments of the associate editor. We also appreciate suggestions which led to the discussion of the balance of optimum designs at the end of §4.

D.P. was a recipient of a post-doctoral fellowship PE2291-1 from the Deutsche Forschungsgemeinschaft held at the Medical Research Council Brain Network Dynamics Unit and Nuffield Department of Clinical Neurosciences of the University of Oxford, OX3 9DU, Oxford, U.K..

References

1. Benabid AL, Chabardes S, Mitrofanis J, Pollak P. Deep brain stimulation of the subthalamic nucleus for the treatment of parkinson's disease. *The Lancet Neurology* 2009; **8**(1):67–81.
2. Levy RM, Lamb S, Adams JE. Treatment of chronic pain by deep brain stimulation: long term follow-up and review of the literature. *Neurosurgery* 1987; **21**(6):885–893.
3. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C, Schwalb JM, Kennedy SH. Deep brain stimulation for treatment-resistant depression. *Neuron* 2005; **45**(5):651–660.
4. Bergman H, Wichmann T, DeLong MR. Reversal of experimental Parkinsonism by lesions of the subthalamic nucleus. *Science* 1990; **249**(4975):1436–1438.
5. Little S, Pogosyan A, Neal S, Zavala B, Zrinzo L, Hariz M, Foltynie T, Limousin P, Ashkan K, FitzGerald J, *et al.*. Adaptive deep brain stimulation in advanced parkinson disease. *Annals of Neurology* 2013; **74**(3):449–457.
6. Cagnan H, Pedrosa D, Little S, Pogosyan A, Cheeran B, Aziz T, Green A, Fitzgerald J, Foltynie T, Limousin P, *et al.*. Stimulating at the right time: phase-specific deep brain stimulation. *Brain* 2017; **140**(1):132–145.
7. Raethjen J, Deuschl G. The oscillating central network of essential tremor. *Clinical Neurophysiology* 2012; **123**(1):61–64.
8. Boecker H, Wills AJ, Ceballos-Baumann A, Samuel M, Thompson PD, Findley LJ, Brooks DJ. The effect of ethanol on alcohol-responsive essential tremor: a positron emission tomography study. *Annals of Neurology* 1996; **39**(5):650–658.
9. Wills A, Jenkins I, Thompson P, Findley L, Brooks D. Red nuclear and cerebellar but no olivary activation associated with essential tremor: a positron emission tomoraphic study. *Annals of Neurology* 1994; **36**(4):636–642.
10. Jenkins I, Bain P, Colebatch J, Thompson P, Findley L, Frackowiak R, Marsden C, Brooks D. A positron emission tomography study of essential tremor: evidence for overactivity of cerebellar connections. *Annals of Neurology* 1993; **34**(1):82–90.
11. Burman CF. *On Sequential Treatment Allocations in Clinical Trials*. Department of Mathematics: Göteborg, 1996.
12. Pedrosa DJ, Auth M, Eggers C, Timmermann L. Effects of low-frequency thalamic deep brain stimulation in essential tremor patients. *Experimental Neurology* 2013; **248**:205–212.
13. Pedrosa DJ, Auth M, Pauls KAM, Runge M, Maarouf M, Fink GR, Timmermann L. Verbal fluency in essential tremor patients: the effects of deep brain stimulation. *Brain Stimulation* 2014; **7**(3):359–364.
14. Cox DR. A note on design when response has an exponential family distribution. *Biometrika* 1988; **75**:161–164.
15. Atkinson AC, Donev AN, Tobias RD. *Optimum Experimental Designs, with SAS*. Oxford University Press: Oxford, 2007.
16. Fedorov VV, Leonov SL. *Optimal Design for Nonlinear Response Models*. Chapman and Hall/ CRC Press: Boca Raton, 2014.
17. Kiefer J, Wolfowitz J. The equivalence of two extremum problems. *Canadian Journal of Mathematics* 1960; **12**:363–366.
18. Pronzato L, Pázman A. *Design of Experiments in Nonlinear Models*. Springer: New York, 2013.
19. Fedorov VV. *Theory of Optimal Experiments*. Academic Press: New York, 1972.

20. Atkinson AC. Optimum experiments with sets of treatment combinations. *mODa 11 – Advances in Model-Oriented Design and Analysis*, Müller CH, Kunert J, Atkinson AC (eds.). Springer: Heidelberg, 2016; 17–24.
21. Atkinson AC. Optimum experiments for logistic models with sets of treatment combinations. *A Festschrift in Honor of Hans Nyquist on the Occasion of His 65th Birthday*, Fackle-Fornius E (ed.). Department of Statistics, Stockholm University: Stockholm, 2015; 44–58.
22. Atkinson AC, Fedorov VV, Herzberg AM, Zhang R. Elemental information matrices and optimal experimental design for generalized regression models. *Journal of Statistical Planning and Inference* 2014; **144**:81–91.
23. Rosenberger WF, Lachin JL. *Randomization in Clinical Trials: Theory and Practice, 2nd edn*. Wiley: New York, 2002.
24. Smith RL. Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society, Series B* 1984; **46**:519—543.
25. Atkinson AC. Optimal model-based covariate-adaptive randomization designs. *Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects*, Sverdlov O (ed.). Chapman and Hall/ CRC Press: Boca Raton, 2015; 131–154.
26. Atkinson AC. The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Society, Series A* 2002; **165**:349–373.
27. Goos P, Jones B. *Optimal Design of Experiments: A Case Study Approach*. Wiley: New York, 2011.