Randomizing a Clinical Trial in Neuro-Degenerative Disease

Anthony C. Atkinson[a], Belmiro P.M. Duarte[b,c], David Pedrosa[d,e], Marlena van Munster[d]

[a]*Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom.*
[b]*Polytechnic Institute of Coimbra, ISEC, Department of Chemical & Biological Engineering, Rua Pedro Nunes, 3030–199 Coimbra, Portugal.*
[c]*Univ Coimbra, CIEPQPF, Department of Chemical Engineering, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal.*
[d]*Department of Neurology, University Hospital Marburg, 35043 Marburg, Germany*
[e]*Center for Mind, Brain and Behavior, Philipps-University Marburg, 35032 Marburg, Germany*

Abstract

The paper studies randomization rules for a sequential two-treatment, two-site clinical trial in neuro-degenerative disease. An important feature is that we have values of responses and five potential prognostic factors from a sample of 144 patients similar to those to be enrolled in the trial. Analysis of this sample provides a model for trial analysis. The comparison of allocation rules is made by simulation yielding measures of loss due to imbalance and of potential bias. A major novelty of the paper is use of this sample, via a two-stage algorithm, to provide an empirical distribution of covariates for the simulation: sampling of a correlated multivariate normal distribution is followed by transformation to variables following the empirical marginal distributions. Five allocation rules are evaluated. The paper concludes with some comments on general aspects of the evaluation of such rules and provides a recommendation for two allocation rules, one for each site, depending on the target number of patients to be enrolled.

Keywords: bias, biased-coin design, empirical multivariate distribution, loss, minimization, randomization.

*Corresponding author.
Email addresses: A.C.Atkinson@lse.ac.uk (Anthony C. Atkinson), bduarte@isec.pt (Belmiro P.M. Duarte), david.pedrosa@staff.uni-marburg.de (David Pedrosa), marlena.vanmunster@uni-marburg.de (Marlena van Munster)

1. Introduction

We study methods for randomized treatment allocation for a clinical trial on neuro-degenerative diseases. Two of the best known of such are Alzheimer's and Parkinson's disease (PD). We describe the background to the clinical trial and the forthcoming economic burden of these diseases on advanced societies in the next section. The purpose of this paper is to compare various randomization methods in sequential trials in which patients present with prognostic factors which may be included in the analysis of the data and so should be allowed for in the randomization scheme. We use data from a sample of patients similar to those to be included in the trial

Our focused objective is to provide a scientific basis for the randomization scheme for this particular trial, based on empirical evidence. It is intended that our results will contribute to justify this particular aspect of the trial protocol.

Because we have a clear objective we do not provide a general survey of randomization methods in clinical trials. Such a survey can be found in Rosenberger and Lachin [1]. Several of the methods we compare are derived from forms of randomized treatment allocation introduced by Atkinson [2] using the methods of optimum experimental design. These were extended by Atkinson [3] to include comparisons of the statistical properties of the designs, particularly the loss of efficiency due to randomization and potential bias from the ability to guess the next treatment to be allocated. Both that paper and Rosenberger and Sverdlov [4] contain background material on randomization in sequential clinical trials in the presence of covariates. A recent review of inference after covariate-adaptive randomization is Ma et al. [5]. The review of Sverdlov et al. [6] focuses on the use of the methods of optimum experimental design in clinical trials.

The paper is organized as follows. The medical background and the structure of the proposed two-treatment trial, to be performed at two sites, are described in §2, followed in §3 by the statistical analysis of the sample values of the five covariates (prognostic factors) which may be used in the analysis of the trial results. The analysis of the sample results shows that two variables are important and that a linear regression model should be appropriate for analysis of the clinical trial. The use of randomised forms of the sequential construction of optimum experimental design in sequential clinical trials is introduced in §4. The two important measures of the performance of a trial design, loss and bias, are formalized in §4.2. Protection against the biases that can be introduced by the absence of proper

randomization is especially important in an unblinded trial such as the one we describe. Section 4.3 presents five allocation rules, ranging from deterministic allocation which minimizes the variance of the estimated treatment difference, to random selection of the treatment to be allocated. We also investigate a randomized version of the minimization rule of Pocock and Simon [7].

We use simulation to compare these five rules. A major novelty of our approach is the use of the empirical sample of potential covariates to provide a sampling distribution of covariates, which have some correlations, rather than assuming independent normal distributions for covariate values. The algorithm is in two stages described in §5: sampling of correlated multivariate normal variates is followed by marginal transformation of the sampled normal variates to samples from the empirical marginal distributions of the covariates. The main numerical results on the comparison of allocation rules is in §5. Extensions in §7 explore (i) the effects of designing for either more or fewer covariates than are used in the analysis and (ii) how comparisons of trial designs change if independent normal covariates are sampled instead of those with the empirical distribution. The final section discusses a few more general topics, including a further sampling rule and the use of a concept admissibility in the comparison of trial designs. The paper concludes with recommendations for randomization rules at the two trial sites.

2. Background

2.1. The Trial

PD ranges among the most common neurodegenerative disorders. The disease is mainly characterized by its cardinal motor symptoms bradykinesia, tremor and rigidity, but growing evidence indicates major disability due to non-motor symptoms such as depression, apathy and sleep disturbances] [DOI:https://doi.org/10.1016/S0140-6736(21)00218-X]. The very heterogeneous disease phenotype may account for a great amount of health-related quality of life restrictions but also burden (informal) caregivers severly] [Kalia & Lang, 2015]. With ageing Western societies and the first symptoms of PD usually appearing around the age of 60, current forecasts point to an increasing burden on health systems, but also pose the problem of having to provide universal care for the often very specific problems associated with the progression of PD [Heinzel et al. 2020; Dorsey et al. 2018].

In Germany, care coordination is primarily the responsibility of the resident neurologists and, in some cases, general practitioners. Normally, outpatient care is often only provided once a quarter by the resident neurologists. The coordination of therapies that are tailored to individual needs and the involvement of specially

trained care professionals such as Parkinson Nurses are rarely implemented Prell et al. 2020. In the case of non-mobile patients, these deficits are further aggravated as trips to the doctor's office become even more challenging and home visits are rare in Germany Stangl et al., 2020 - a problem that increases with the degree of illness.

However, this situation is not sustainable. Therefore, modern therapies, especially for neuro-degenerative diseases, are increasingly moving towards a holistic approach to patient care Rajan et al. 2020. Technological advances open up new possibilities in the field of healthcare provision and professional collaboration. The attractiveness of digital technologies lies in their ability to mitigate both mobility-related barriers and economic obstacles. Digital solutions are also suitable for evaluating the disease activity of movement disorders, since tests developed for this purpose can be easily implemented within the framework of eHealth solutions. In Germany, there are mostly regionally implemented digital solutions for PD patients available, but there are no nation-wide healthcare models available van Munster et al. 2020, Stangl et al. 2020. As part of the "ParkProReakt" project, a cross-sectoral, proactive, needs-oriented and technology-supported care model is being developed. The aim of this project is to improve healthcare and achieve a measurably improved quality of life for PD patients. In addition, care givers should be relieved, since the use of digital solutions supports them in assessing the change in the course of the disease. This project is funded under a programme of the federal German government, which, however, is located in the territory of the Ministry of Health Gemeinsamer Bundesausschuss, n.d..

The healthcare model is being evaluated as part of a clinical study where we will look the perception of healthcare professionals working in the model, the impact in the everyday life of people with PD, the economic benefits as well as the effects on patients quality of life.

We will include sequentially a certain number of people at two centres of different sizes. Both centres will accompany and treat a number of patients who are divided 1:1 into controls (receiving only standard care) and an intervention group (with the complex care we developed). Our sample size calculation, in terms of quality of life and according to previous publications REFERENCES, is based on a total of:

Centre 1 = 92/92 and Centre 2 = 54/54, i.e. in total 292 people. There may be some change in the ratio between the centres, but not in the number of cases.

2.2. *Variables*

Results from a representative sample of Marburg's patients. There seem to be two important prognostic factors on which balance is required.

Response variable: Quality of life (QoL), which can be measured with two questionnaires similar: i. the Parkinson's Disease Questionnaire (PDQ) with 39 items ; and ii. the PDQ with 8 items. BD: The 8-item disease questionnaire (PDQ8) [8] is the patient reported outcome measure constructed by taking one question from each of the eight domain of PDQ39 [9]. Of course both pose an oversimplification of what really reduces or represents something a bit abstract such as QoL actually is but it's our primary endpoint and both have been extensively used. BD: The PDQ with 39 items (PDQ39) allows a better and wider characterization of QoL than the PDQ8, but the last is more practical to use and several authors report a strong correlation between the results of both, see Chen et al. [10].

Prognostic Factor 1: Stage of the disease indicated on Hoehn and Yahr scale (H&Y). Increasing values indicate more severe affection on an ordinal scale. I find it somehow arbitrary, but it's very common and easy to assess, and therefore very useful. It is also described here: https://en.wikipedia.org/wiki/Hoehn_and_Yahr_scale.

Prognostic Factor 2: Beck's Depression Inventory (BDI). This indicator is determined from a questionnaire, see Beck et al. [11]. Values from 10 and higher indicate increasing levels of depression.

Other factors seemed not to be important in their data. But here are some results.

Marlena has done a short literature search and there are of course different aspects that negatively correlate with the PDQ39, such as depressive symptoms, disease duration, stage of the disease but also simple things such as age and gender. She has summarised everything in a table that you can find enclosed. But in a nutshell, these are the important factors described in the literature:

Psychological Well-being/ Neuropsychiatric symptoms (Depression, Anxiety)
Disease duration/ Stage of Disease
Demographic (Age, Gender, Area of Living, Income)
Cognitive Impairment

After that we collected a sample of about 150 patients from our outpatient ward, of which many filled out at least some of the questionnaires we tried to replicate this and below you can find the demographics. But if you look at correlations, there are significant correlations with disease stage but especially with

depression but no gender differences and no correlations with age or the cognitive functioning (Moca).

pdq8 - Life quality (QoL) measured on a percentage scale (0-100%). Higher values show less quality of life due to more restrictions on aspects presumably related to QoL.

3. Data Modelling

Because our allocation rules depend on a statistical model of the data, we start with an analysis of the data on the response and five potential prognostic factors taken from patient records (described in §2.2). The analysis leads to building a regression model. In the remaining sections we designate the response variable by $pdq8$, the disease stage expressed on H&Y scale by $h\&y$, the BDI indicator by bdi, the gender of patients by $Gender$, the age by Age and the Montreal Cognitive Assessment (MoCA) test [12] results by $moca$.

There are data from 144 patients on the response $pdq8$ and on the five prognostic factors. Not all variables are available for all patients, but there are an adequate number to establish the correlations between all variables, which are shown in Table 1. The most noticeable features are the correlations of 0.640 and 0.332 between the response $pdq8$ and two of the prognostic factors, i.e. bdi and $h\&y$. These two variables, in turn, have a correlation of 0.325.

[Table 1 about here.]

The regression of $pdq8$ on all five prognostic factors produces the results reported in Table 2. The order of the appearance of the covariates in Table 2 is that of their entry to the linear model obtained via stepwise regression. Surprisingly, in the light of the correlations in Table 1, there is significant regression on bdi but not on $h\&y$. Due to the correlation between the covariates much of the variability of $pdq8$ explained by bdi is already explained by $h\&y$.

[Table 2 about here.]

We checked several models using normal probability plots of the residuals. The left-hand panel of Figure 1 shows the normal QQ plot of the residuals from least squares regression on just bdi and the right-hand panel shows a similar plot from regression on both bdi and $h\&y$. The plot from regression on two variables is appreciably straighter, indicating a more nearly normal distribution of residuals. This plot is also straighter than that of the residuals from regression on all

five variables (not shown). In our exploration of methods for balancing and randomizing treatment allocations we therefore take as our standard allocations those using just two prognostic factors with homoscedastic independent normal errors. However, in §7.1 we also briefly consider the effect of allocations using fewer or more prognostic factors.

[Figure 1 about here.]

As a final introduction to the structure of the data used in our simulations of designs, we give in Figure 2 scatterplots of the response $pdq8$ against $bdi$and $h\&y$, together with histograms of the distributions of the variables. As is to be expected from the analyses given above, there is a stronger relationship between $pdq8$ and bdi. Note also that the marginal distribution of $pdq8$ is not normal. Normality is revealed by the residuals from joint regression on these two prognostic factors.

[Figure 2 about here.]

The plots in Figure 2 also reveal that the distributions of bdi and $h\&y$ are not far from normal. Similar plots for age and $moca$ show no relationship with the response $pdq8$. They also again show covariate distributions that are close to normal. These comments are important when, in §7.2, we explore the properties of designs using normally distributed covariates.

4. Experimental Design

4.1. Sequential Optimum Experimental Design

Patients arrive sequentially. Patient i presents with a vector of $q-1$ prognostic factors z_i and is allocated to one of two treatments, τ_1 or τ_2; the response (here, $pdq8$) for this patient is y_i. The parameter of interest is the treatment difference $\Delta = (\tau_1 - \tau_2)/2$. The regression model for all n observations, in matrix form, is

$$\mathbb{E}(\mathbf{Y}) = \mathbf{a}\Delta + \mathbf{1}\beta_0 + \mathbf{Z}\psi = \mathbf{a}\Delta + F\beta = G\,\omega. \tag{1}$$

In this model $\mathbf{a}$ is the $n \times 1$ vector of allocations with elements $+1$ and -1, depending on whether treatment 1 or treatment 2 is allocated, and $\mathbf{1}$ is the $n \times 1$ vector of ones. The average effect of the two treatments, written as the constant term $\beta_0 = (\tau_1 + \tau_2)/2$, is not of importance. The parameter vector ψ of regression parameters for the prognostic factors is also unimportant, although some balance is required over these variables, which will be included in the analysis of the data.

The constant and covariates are included in the $n \times q$ matrix F. The value of q is important in determining the properties of some allocation rules.

In sequential treatment allocation the covariates and allocations are known for the first n patients, giving a matrix G_n of allocations and explanatory variables in (1). Let patient $n+1$ have a vector z_{n+1} of explanatory variables. If treatment j is allocated, the vector of allocation and explanatory variables for the $(n+1)$st patient is $g_{j,n+1}$, $j = 1, 2$. Results in the sequential construction of optimum experimental designs (see Atkinson [2] and Smith [13, §10]) show that the variance of the estimate $\hat{\Delta}$ after $n+1$ observations is minimized by the choice of that treatment for which the derivative function

$$d_s(j, n, z_{n+1}) = g_{j,n+1}^{\mathrm{T}}(G_n^{\mathrm{T}}G_n)^{-1}g_{j,n+1} - f_{j,n+1}^{\mathrm{T}}(F_n^{\mathrm{T}}F_n)^{-1}f_{j,n+1} \qquad (2)$$

is a maximum. This result is a special case of the use of optimum design theory to minimize the variance of a single parameter estimate in a model with several nuisance parameters, a criterion called $\mathrm{D_s}$-optimality. See Atkinson et al. [14, §10.3] with $s = 1$.

Once the prognostic factors are known for patient $n+1$, treatment allocation in the sequential optimum design of experiments is determined. This procedure leads to a trial in which the variance of $\hat{\Delta}$ is minimized; there is no allowance for randomization. Randomness in the allocations will provide protection against biases and unexpected trends, but at the cost of a slight loss in efficiency, that is an increase of the variance of $\hat{\Delta}$.

4.2. Assessing Rules: Bias and Loss

The loss from randomization is assessed from $\mathrm{Var}(\hat{\Delta})$. Let $b = F^T a$, a "balance" vector which is identically zero when all covariates are balanced across all treatments, which is the goal of the sequential construction of §4.1. Then

$$\mathrm{var}(\hat{\Delta}) = \frac{\sigma^2}{n - b^T(F^TF)^{-1}b} = \frac{\sigma^2}{n - L_n}, \qquad (3)$$

giving an explicit expression for calculation of the loss L_n. The loss is minimized for the balanced design when the estimate of Δ is independent of the estimates of the nuisance parameters. As (3) indicates, the loss quantifies the number of patients on whom information is effectively lost due to imbalance in the trial.

The loss L_n in a specific trial depends on the particular sequence of randomized allocations. In this paper interest is in comparing the properties of various allocation rules, so that the focus is on the expectation $\mathbb{E}(L_n) = \mathcal{L}_n$, approximated

by $\bar{L}_n$, the average over n_{sim} simulations. For some allocation rules theory provides a value for the expected value of the loss $\mathcal{L}_n$ as $n \to \infty$. However, even in such cases, simulation is informative about trials for moderate values of n.

A numerical measure for randomization is selection bias [15] which measures the ability to guess the next treatment to be allocated. Bias depends on the design, the guessing strategy and, for some rules, the value of n. For a particular combination of strategy and design the expected bias $\mathcal{B}_n$ is estimated from n_{sim} simulations as

$$\bar{B}_n = \frac{\text{number of correct guesses of allocation to patient} - \text{number of incorrect guesses}}{n_{\text{sim}}}. \tag{4}$$

This definition is similar to that of (4.2) of Smith [13]. The guessing strategy used in our numerical comparisons is the sensible one of guessing that the treatment for which the allocation probability is higher will be selected.

Amongst many others, Efron [16] and Smith [13] consider that selection bias should not be an issue in double-blind trials with treatment allocation made remotely from the trial, although it may be if there are local attempts towards institutional balance [17]. It is however impossible to blind the trial with which we are concerned. Allocation may be blinded, but the patient and medical staff will know without doubt which treatment has been allocated. For us, then, randomization is particularly important. In general, a trial without randomization appears to lack objectivity. Efron [16] and Smith [13] accordingly study the effect of biased-coin designs on freedom from accidental bias due to omitted factors including time trends and, in the case of Smith [13], correlated errors and outliers. The conclusion of Smith [13] is that biased-coin designs that are not completely random provide good protection against several sources of bias and that selection bias is a good measure of the properties of the design.

Randomization and balance are in conflict. The deterministic rule of sequential optimum design minimizes loss. However, the allocation can always be correctly guessed, so that $\mathcal{B}_n = 1$. The antithesis is the random rule in which the treatment is allocated by the toss of a fair coin. This has the maximum loss of all rules we consider, but it is impossible to have any systematic success in guessing the next allocation, so that $\mathcal{B}_n = 0$. In this paper we study several design strategies intermediate in properties between optimum design and random allocation.

4.3. Five Allocation Rules

We now describe the five rules that we compare in a variety of scenarios for randomizing the experiment. Some of the rules are based on the sequential con-

struction of the optimum design for estimation of Δ. Let the treatment maximizing (2) be $\tau_{[1]}$, which is allocated with probability $\pi([1])$.

Rule D: Deterministic Allocation. This is the sequential construction of the D_s-optimum design; $\pi_D([1]) = 1$. It follows that $\mathcal{L}_\infty = 0$ and, since there is no randomization, $\mathcal{B}_\infty = 1$. The simulations in later sections show that, from very small values of n, $\bar{L}_n \to 0$ and $\bar{B}_n \to 1$.

Rule A: Randomized D_A-optimality. Atkinson [2] introduced a randomized form of the sequential construction of D_A-optimum designs. For two treatments the probability of allocation of treatment j is

$$\pi_A(j) = \frac{d(j, n, z_{n+1})}{d(1, n, z_{n+1}) + d(2, n, z_{n+1})}. \tag{5}$$

Burman [18] showed that for this rule $\mathcal{L}_\infty = q/5$. The values of $d(1, n, z_{n+1})$ and $d(2, n, z_{n+1})$ are not standardized by n. As n increases the difference between the two decreases and as $n \to \infty, \pi_A(j) \to 0.5$. As a consequence, $\mathcal{B}_\infty = 0$.

Rule E: Efron's Biased Coin. Efron [16] introduced a design for the sequential comparison of two treatments in which the under-represented treatment was allocated with probability 2/3. In the presence of covariates, but remaining with two treatments, the under-represented treatment is [1]. Then

$$\pi_E([1]) = 2/3 \tag{6}$$

The loss decreases with n but, from small n, the values of $\mathcal{B}_n$ are close to the asymptotic value of 1/3.

Rule MwC: Minimization with a Coin. The deterministic minimization rule of [7] depends on calculating the total effect on all measures of marginal imbalance when treatment j is allocated. With $q-1$ covariates z, there will be $q-1$ measures to be summed. The individual measures count the number of observations in each category of the covariate. Continuous covariates therefore have to be categorised.

Let the total effect on imbalance be $C(j)$. The allocations are ranked so that $C([1]) \leqslant C([2])$. In this deterministic allocation treatment [1] is allocated, with random allocation if both treatments have the same value of $C(j)$.

We introduce randomization by replacing certain allocation by the 2/3 of Efron's biased coin. Thus

$$\pi_{MwC}([1]) = 2/3, \tag{7}$$

with random allocation if there is a tie, as there may well be, since the prognostic factors are discretized. The deterministic calculations are exemplified by [19] and [3] as well as by [7].

Rule R: Randomized Allocation. $\pi_{\mathrm{R}}([1]) = 0.5$. This is the furthest in properties from deterministic allocation, Rule D. Now since there is complete randomization, $\mathcal{B}_\infty = 0$. A special case of the calculations in Burman [18] is that $\mathcal{L}_\infty = q$, a result that goes back at least to Cox [20].

5. Sampling from the Multivariate Empirical Distribution of Prognostic Factors

Simulation is often used, as here, to find the small sample properties of treatment allocation procedures. Many such investigations, such as Atkinson [3, 21], assume that the prognostic factors are uncorrelated and normally distributed. Here we sample from an approximation to the empirical correlated distribution of the prognostic factors analysed in §3.

In general, it is difficult to sample from multivariate distributions with arbitrary covariances. One possibility is to sample, with replacement, from the $q-1$ dimensional discrete distribution of the observed covariates. An alternative, which gives more sampling points, is to generate a $q-1$ dimensional multivariate normal sample with the desired correlation and then to transform the normal distributions to have the univariate empirical distributions discussed in §3.

Let the $q-1$ prognostic factors of patients entering the trial have the correlation matrix, Γ, constructed from the data in Table 1. Let $\boldsymbol{u}$ be a $q-1$ vector of uncorrelated standard normal variables which we generate using the Box-Muller algorithm. To generate a vector of correlated normal random variables $\mathbf{v}$, we first decompose Γ using the Cholesky decomposition, i.e. $\Gamma = \Lambda\,\Lambda^{\mathrm{T}}$, where Λ is a $(q-1)\times(q-1)$ lower triangular matrix. We then form the elements of the correlated normal q vector for a new patient using the rule

$$v_1 = 1.0 \tag{8a}$$

$$v_i = \sum_{j=1}^{q-1} \Lambda_{i-1,j}\, u_j, \quad i = 2, \cdots, q, \tag{8b}$$

where (8a) is for the constant term and (8b) is for the prognostic factors.

We now further transform the $v_i, i = 2, \cdots, q$ to have the desired empirical distribution. Let the ordered vector of sampled values of the empirical prognostic

factors of §3 for variable i be $\mathbf{s_i}$. Then $\ldots s_{i,k-1} < s_{i,k} < s_{i,k+1} \ldots$ with cdf $F_i(s_{i,k}) = P(S_i \leqslant s_{i,k})$. We sample the distribution of S_i using the cdf of the normal distribution of v_i to provide the probabilities for our correlated sample. That is, let $p_i = \Phi(v_i)$, where Φ is the cdf of the standard normal distribution. Then the values of the simulated covariates z_i are found by numerical search:

$$\text{if} \quad F_i(s_{i,k-1} < p_i \leqslant F_i(s_{i,k}), \quad z_i = s_{i,k}, \quad i = 2, \cdots, q, \tag{9}$$

with $z_1 = 1.0$.

In our analyses we consider: i. $q = 6$, i.e., including all the prognostic factors when samples are extracted via a five-variate normal distribution; ii. $q = 3$, including the variables $h\&y$ and bdi and so sampling from a bivariate normal distribution; and iii. $q = 2$, only the variable bdi as used for prediction of $pdq8$. In this case samples come from a standard normal distribution and $\Gamma = [1.0]$. For non-correlated prognostic factors we consider only the normal case and put $z_i = v_i$ in (8) with $\Gamma = I_{q-1}$ where I_{q-1} is the $q-1$ identity matrix.

6. The Trial Design and Comparison of Allocation Rules: Empirical Prognostic Factors

6.1. The Overall Design of the Sequential Trial

There are two sites for the trial: one is expected to enrol 240 patients and the other 100. It is sensible to randomize separately for the two centres. One reason is that of robustness of the procedure. It will be more straightforward to run two separate schemes, rather than to rely on communication between the centres and the transfer of covariate information. Practically, separate randomization scheme reduce the probability of confusion and errors. The second reason is that it is possible the distribution of covariates at the two site may be different - perhaps due, for example, to socio-economic or demographic factors. In the data analysis we need to be prepared to be able to fit models to well-balanced data from the individual sites, as part of the process that, it is to be hoped, will lead to a single model and analysis for all patients.

The properties of randomization rules depend on the number of patients in the trial. Since it is not certain that the two centres will be able to recruit exactly the specified number of patients, we compare the properties of randomization rules for values of n up to 240. These results are given graphically. Because, however, there are two specific target values of n, we also provide tabulations of the properties of the rules for $n = 100$ and 240.

6.2. Comparison of Allocation Rules: Empirical Prognostic Factors

We start our comparison of the allocation rules taking $q = 3$, that is the intercept and the two prognostic factors bdi and $h\&y$ which are most highly correlated with the response. There were 20,000 simulations in all comparisons.

[Figure 3 about here.]

The results are plotted in Figure 3 and summarized in the central panel of Table 3. The left-hand panel of the figure shows the loss for values of n up to 240. Rule R has a loss of 3 $(= q)$ throughout, in line with the results quoted in §4.3. Reading down in the centre of the plot, the loss for Rule MwC is gradually decreasing, being slightly less than one at $n = 240$. The loss for Rule A settles to a value of $q/5 = 0.6$ just after $n = 100$. The loss for Rule E decreases steadily, becoming less than that for Rule A when n is close to 60. It is however always greater than that for Rule D, for which $\mathcal{L}_\infty = 0$.

The right-hand panel of the figure shows the plot for bias. This has a simpler structure. Rules R, E and D have constant biases of 1, 1/3 and 0, within sampling fluctuation. The bias for MwC is also constant, but lower than that for Rule E because of the occurrence of ties; the value is close to 0.25 rather than 1/3. The bias for Rule A, unlike the others, decreases steadily, in line with the argument of §4.3.

[Table 3 about here.]

We also investigated the properties of the five rules for two further values of q. The two panels of Figure 4 show the plots of loss for $q = 2$ and $q = 6$. Now the losses for Rule R are two and six and those for Rule A tend to 0.4 and 6/5 for large n. For $q = 2$ (the left-hand panel) the losses all proportionately decrease faster than they do for the right-hand panel. This effect is particularly marked for the two rules MwC and E that randomize using Efron's coin. The biases for both values of q are similar in structure to those for $q = 3$ in the right-hand panel of Figure 3 and so are not shown here.

[Figure 4 about here.]

Values of both loss and bias for $q = 6, 3$ and 1 and $n = 100$ and 240 are in Table 3, these being the two values of importance for the trial on neuro-degenerative diseases. The table confirms the suggestion of the figures that Rule A provides a good compromise between loss and bias, low values of both of which are desirable. More generally, the losses for Rules E and MwC in the right-hand panel of Figure 4 show the poor performance of these two rules as q increases

7. Extensions

7.1. Design for an Incorrect Number of Prognostic Factors

It may be that a trial is designed with randomization over q prognostic factors but the final data analysis incorporates r factors, where r may be greater than, or less than q. Results for allocations using the five rules are in Table 4. The top half of the table is when extra covariates are included in the design: balancing is over five covariates, but only two are used in the data analysis. The lower half of the table is for the reverse situation, when two are included in the design and five are used in the analysis. Comparison of these results with those of Table 3 shows very few differences.

There is some recent theoretical work on the properties of deigns when $r > q$, that is "what is the effect of the randomization on the non-randomized covariates?". Unfortunately, this work does not cover our situation as Liu and Hu [22] only consider discretized covariates and Ye et al. [23] develop a model-free approach. Both papers usefully present details of recent work on covariate-adaptive randomization.

[Table 4 about here.]

7.2. Independent Normal Covariates

Many simulation studies of treatment allocation in clinical trials, such as Atkinson [21] have taken the prognostic factors to be independently normally distributed. We now check whether, in our example, the more complicated simulation strategy we have used leads to results distinct from those from the simple assumption of normality.

[Table 5 about here.]

The results for simulations with independent normal prognostic factors when $q = 3$ are in Table 5. Comparison with the central panel of Table 5 again shows only a few slight differences, here between the use of independent normal prognostic factors and the correlated empirical factors coming from the data. The two largest differences are in the reduction in loss for Rule MwC when normal covariates are used.

8. Discussion

The purpose of our paper is to compare the performance of several randomization rules for treatment allocation for a specific clinical trial on treatment of neuro-degenerative diseases. We were fortunate in having available a preliminary set of data from which we were able to estimate the empirical distribution of the prognostic factors. In order to simulate from this empirical distribution, as we describe in §5, we sampled from correlated normal random variables which were then transformed to have the desired marginal distributions. As far as we know this procedure has not previously been used in the context of randomizing treatment allocation in clinical trials. The results of our simulations suggest the use of randomized forms of sequential design construction based on D- or D_s-optimality.

Our results indicate, for our example, that a similar assessment of the relative merits of the different rules is obtained by simulations with independent normally distributed prognostic factors. In retrospect this is not surprising, since most of the correlations between covariates are low and four of the variates, including bdy and $h\&y$ plotted in Figure 2, have distributions close to normal. It is a matter for further exploration as to how general is this result. For methods that allocate according to a function of the information matrix of the design, it is clear that the distribution of the factors will have little effect on the value of loss as $n \rightarrow \infty$, provided the distribution of the prognostic factors has a finite variance. The behaviour of minimization, without randomization, which we did not consider, depends strongly on the distribution and correlation structure of the prognostic factors. Some details are in Figures 2 and Table 2 of Atkinson [3]. However, minimization is not sensitive to a binary covariate, in our case Age. These results also demonstrate the lack of sensitivity of values of loss from Rules R, A and D to the marginal distributions and correlation of the prognostic factors.

There are many other allocation rules that have been studied in the reviews mentioned in §1. One possibility is to use a different function of $d_s(.)$ (2) in the definition of the allocation probability. Atkinson [3] developed ideas on the balance between randomness and information in Ball et al. [24] to replace (5) with the Bayesian form

$$\pi_{\mathrm{B}}(j) = \frac{\{1 + d(j, n, z_{n+1})\}^{1/\gamma}}{\sum_{k=1}^{2} \{1 + d(k, n, z_{n+1})\}^{1/\gamma}}. \tag{10}$$

An advantage of this rule is that initially, for small n, the allocations force balance at the cost of high bias. As n increases the allocation moves towards low bias and a higher loss, although with a proportionately smaller loss for values standardized

by n. . This rule is particularly appropriate if it is not known when the trial is likely to stop. The rate of change of emphasis in the allocation depends on the value of the parameter γ. A suitable value for a specified n can be determined by simulation.

[Figure 5 about here.]

In general, all rules involve a trade-off between bias and loss. Comparisons are helped by the use of the normalized loss, scaled to lie between zero and one:

$$\text{Normalized loss} = \text{Loss}/q.$$

Figure 5 presents the normalized loss vs. bias for all rules for $q = 3$. As we have seen from earlier figures, the comparative properties of the rules depend upon the value of n. We have marked the values for $n = 100$ and 240 on the plot. It is clear that, for all rules except R, increasing n leads to decreasing loss. It is also clear from the closeness of the plotted symbols for $n = 100$ and $n = 240$ that the majority of the change in properties occurs ro small values of n.

The concept of the admissibility of a rule [3] is helpful in interpreting such plots. Small value of both loss and bias are desirable: if Rule 2 has higher levels of bias and loss than Rule 1, then Rule 2 is inadmissable. Rules D and R are always admissible, since they respectively have the minimum values of loss and bias. Figure 5 shows that Rule MwC is inadmissible compared with Rule A, for both values of n of interest. Rule E has lower loss for these values of n than does Rule A. However, the bias is greater; admissibility does not provide a rationale for preferring one design to the other.

[Table 6 about here.]

The situation in the trial of this paper is simple; we require to find good allocation rules for just two values of n; 100 and 240. Table 6 gives results, extracted from Table 3 for Rules A and R, where loss is expressed as a percentage of the number of patients. For Rule A the percentage loss when $n = 100$ is 0.62, rising to 3.0 if Rule R is used. This number may well represent too large a loss of information despite the value of zero for bias. However, when $n = 240$, the percentage loss for Rule R is only 1.52. We therefore suggest that Rule A be used for the centre with 100 patients and Rule R for the centre with 240 patients, the larger sample size leading to a lower potential bias from allocation. Since the centres are randomizing independently, we see no reason why the two sites should follow the same allocation rule.

References

[1] W. F. Rosenberger, J. L. Lachin, Randomization in Clinical Trials: Theory and Practice, 2nd edn, Wiley, New York, 2016.

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

[3] A. C. Atkinson, The comparison of designs for sequential clinical trials with covariate information, Journal of the Royal Statistical Society, Series A 165 (2002) 349–373.

[4] W. F. Rosenberger, O. Sverdlov, Handling covariates in the design of clinical trials, Statistical Science 23 (2008) 404–419.

[5] W. Ma, Y. Qin, Y. Li, F. Hu, Statistical inference for covariate-adaptive randomization procedures, Journal of the American Statistical Association 115 (2020) 1488–1497.

[6] O. Sverdlov, Y. Ryeznik, W. K. Wong, On optimal designs for clinical trials: An updated review, Journal of Statistical Theory and Practice 14 (2020) 1–29.

[7] S. J. Pocock, R. Simon, Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, Biometrics 31 (1975) 103–115.

[8] V. Peto, C. Jenkinson, R. Fitzpatrick, 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 (1998) S10–S14.

[9] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, N. Hyman, The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score, Age and ageing 26 (1997) 353–357.

[10] K. Chen, Y.-J. Yang, F.-T. Liu, D.-K. Li, L.-L. Bu, K. Yang, Y. Wang, B. Shen, R.-Y. Guan, J. Song, et al., Evaluation of PDQ-8 and its relationship with PDQ-39 in China: a three-year longitudinal study, Health and quality of life outcomes 15 (2017) 1–7.

[11] A. T. Beck, R. A. Steer, M. Pompili, BHS, Beck hopelessness scale: manual, Psychological corporation San Antonio, TX, 1988.

[12] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, H. Chertkow, The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment, Journal of the American Geriatrics Society 53 (2005) 695–699.

[13] R. L. Smith, Sequential treatment allocation using biased coin designs, Journal of the Royal Statistical Society, Series B 46 (1984) 519—543.

[14] A. C. Atkinson, A. N. Donev, R. D. Tobias, Optimum Experimental Designs, with SAS, Oxford University Press, Oxford, 2007.

[15] D. Blackwell, J. L. Hodges, Design for the control of selection bias, Annals of Mathematical Statistics 28 (1957) 449–460.

[16] B. Efron, Forcing a sequential experiment to be balanced, Biometrika 58 (1971) 403–417.

[17] S. W. Lagakos, S. J. Pocock, Randomization and stratification in cancer clinical trials: an international survey, in: M. E. Buyse, M. J. Staquet, R. J. Sylvester (Eds.), Cancer Clinical Trials: Methods and Practice, Oxford University Press, Oxford, 1984.

[18] C.-F. Burman, On Sequential Treatment Allocations in Clinical Trials, Department of Mathematics, Göteborg, 1996.

[19] S. Senn, V. V. Anisimov, V. V. Fedorov, Comparisons of minimization and Atkinson's algorithm, Statistics in Medicine 29 (2010) 721–730.

[20] D. R. Cox, Some systematic experimental designs, Biometrika 38 (1951) 312–323.

[21] A. C. Atkinson, Selecting a biased-coin design, Statistical Science 29 (2014). 144–163.

[22] Y. Liu, F. Hu, Balancing unobserved covariates with covariate-adaptive randomized experiments, Journal of the American Statistical Association (2020) 1–12.

[23] T. Ye, Y. Yi, J. Shao, Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods, Biometrika 109 (2021) 33–47. doi:10.1093/biomet/asab015.

[24] F. G. Ball, A. F. M. Smith, I. Verdinelli, Biased coin designs with a Bayesian bias, Journal of Statistical Planning and Inference 34 (1993) 403–421.

9. Acronyms

PD Parkinson's disease

Figure 1: Normal probability plot of residuals from two fitted models: **??** $pdq8$ vs. bdi; **??** $pdq8$ vs. $h\&y$ and bdi.

Figure 2: Scatterplot with histogram for: **??** $pdq8$ vs. bdi; **??** $pdq8$ vs. $h\&y$.

Figure 3: Results for model including 3 parameters (the intercept plus 2 covariates corresponding to $h\&xy$ and bdi – those with the largest correlation with the response): **??** Loss and **??** Bias, as functions of the number of patients.

Figure 4: Losses as a function of n. Left-hand panel $q = 2$ (just bdi) and Right-hand panel $q = 6$ (all prognostic factors).

Figure 5: Normalized loss vs. bias for $q = 3$: empirical correlated covariates. The symbol "$\square$" indicates the performance after 100 patients, and "$\triangledown$" the performance after 240 patients.

Table 1: Correlation matrix between the response and prognostic factors.

	$Gender$	Age	$h\&y$	bdi	$moca$	$pdq8$
$Gender$	1.0000	0.0792	0.0180	-0.1218	-0.0376	-0.0807
Age		1.0000	0.2266	-0.0513	-0.4766	-0.1415
$h\&y$			1.0000	0.3250	-0.6435	0.3318
bdi				1.0000	-0.2689	0.6402
$moca$					1.0000	0.0419
$pdq8$						1.0000

Table 2: Order of addition of covariates to $pdq8$ model (via stepwise regression).

Order of addition	Source	SSE	SSR	d.f.	MSE	MSR	F	Prob>F
1	Intercept			1				
2	bdi	4.4768×10^3	5.5662×10^3	1	172.1852	5.5662×10^3	32.3266	5.560×10^{-6}
3	Age	4.4280×10^3	5.6150×10^3	1	177.1190	48.8340	0.2757	0.6041
4	$Gender$	4.3909×10^3	5.6521×10^3	1	182.9536	37.0898	0.2027	0.6566
5	$h\&y$	4.3756×10^3	5.6674×10^3	1	190.2426	15.3065	0.0805	0.7792
6	$moca$	4.3593×10^3	5.6836×10^3	1	198.1522	16.2311	0.0819	0.7774

SSE - sum of square error; SSR - sum of squares of treatments; d.f. - degrees of freedom; MSE - mean square error (MSE=SSE/(n-d.f.)); MSR - incremental mean of squares of treatments (MSR=SSR_i-SSR_{i-1}); F - F ratio (F=MSR/MSE).

Table 3: Performance of allocation rules after 100 and 240 patients (model with correlated empirical covariates).

		After 100 patients		After 240 patients	
Covariates	Rule	Loss	Bias	Loss	Bias
1	D	0.0160	1.0000	0.0065	1.0000
	R	1.9908	-0.0019	1.9885	0.0148
	A	0.4021	0.0846	0.3997	0.0492
	E	0.1859	0.3334	0.0780	0.3290
	MwC	0.5117	0.2490	0.4237	0.2468
2	D	0.0389	1.0000	0.0161	1.0000
	R	3.0153	-0.0017	3.0135	-0.0054
	A	0.6156	0.1160	0.6051	0.0838
	E	0.3985	0.3257	0.1658	0.3420
	MwC	1.1135	0.2444	0.9239	0.2462
5	D	0.1604	1.0000	0.0644	1.0000
	R	5.9765	-0.0011	6.0126	-0.0076
	A	1.2598	0.1737	1.2277	0.1131
	E	1.4093	0.3414	0.6211	0.3389
	MwC	3.0933	0.2862	2.4296	0.2863

Table 4: Performance of allocation rules after 100 and 240 patients when the design is obtained with q covariates and used in a model including r covariates (models with correlated empirical covariates).

		After 100 patients	After 240 patients
Covariates (q/r)	Rule	Loss	Loss
6/3	D	0.0388	0.0161
	R	3.0151	3.0134
	A	0.6150	0.6049
	E	0.3986	0.1658
	MwC	1.1136	0.9238
3/6	D	0.1594	0.0642
	R	5.9607	5.9545
	A	1.2690	1.2334
	E	1.4152	0.6330
	MwC	3.1035	2.41724

Table 5: Performance of allocation rules after 100 and 240 patients(model with non correlated covariates).

		After 100 patients		After 240 patients	
Covariates	Rule	Loss	Bias	Loss	Bias
2	D	0.0389	1.0000	0.0161	1.0000
	R	3.0153	-0.0017	3.0135	-0.0054
	A	0.6156	0.1205	0.6056	0.0845
	E	0.3975	0.3257	0.1662	0.3420
	MwC	1.0851	0.2447	0.8763	0.2487

Table 6: Performance of allocation rules A and R for 100 and 240 patients - percentage of loss per patient.

	For 100 patients		For 240 patients	
Rule	% Loss	Bias	% Loss	Bias
A	0.62	0.12	0.25	0.08
R	3.02	0.00	1.26	0.00