Adaptive Biased-Coin Designs for Clinical Trials with Several Treatments

Anthony C. Atkinson, Department of Statistics,
London School of Economics, London WC2A 2AE, UK

October 29, 2003

Abstract

Adaptive designs are used in phase III clinical trials for skewing the allocation pattern towards the better treatments. We use optimum design theory to provide a skewed biased-coin procedure for sequential designs with continuous responses. The skewed designs are used to provide adaptive designs, the performance of which is studied numerically for designs with three treatments. Important properties are loss and the proportion of allocation to inferior treatments. Regularisation to provide consistent parameter estimates greatly improves both these properties.

Keywords: c-optimal design; limiting allocation proportion; minimization; randomization; regularisation.

1 Introduction

Patients arrive sequentially, perhaps for a phase III clinical trial, and are to be allocated one of t treatments. Adaptive designs skew the allocation proportion to the eventually best treatment by using earlier responses to determine the next allocation. For efficient estimation of the treatment effect the allocation needs to be approximately balanced over the prognostic factors and covariates of the individual patients. There needs also to be some randomization in the allocation. The paper uses optimum design theory to provide balance, augmented by a biased coin (Atkinson 1982) for randomization. We describe optimum designs that give a skewed allocation and show how they can be used to provide adaptive designs that favour the best treatment. These skewed designs will also be of importance in their own right.

For example, in the comparison of a treatment with a control, it might be desired to give a specified majority of patients the new treatment.

There is an extensive literature on randomization and adaptive designs in clinical trials. Recent book length treatments include Matthews (2000) and Rosenberger and Lachin (2002). Virtually all of the reported work on adaptive designs is for binary responses in the absence of prognostic factors, with designs generated from urn models. In contrast we use optimum design theory for regression models with responses that are either normal or can be made so by transformation. Our results also cover generalized linear models where the treatment effects are sufficiently small that the effect on the design of the iterative weights used in parameter estimation can be ignored (Cox 1988).

The combination of optimum design theory for linear models and biased-coin designs (Atkinson 1982) provides expressions for the probabilities of allocation of the treatments; the resulting randomization depends on the allocation history and the covariates. The lack of balance from restricted randomization results in increased variance of parameter estimates that can be expressed as a "loss", that is the number of patients on whom information is not available compared with the balanced design. Comparisons of several rules for non-adaptive biased-coin designs are in Atkinson (2002). Adaptive designs were combined with biased-coin designs by Atkinson and Biswas (2004b) whose development was mostly for the comparison of two treatments. Here we extend the theory to any number of treatments and give examples of designs when $t = 3$. Our objective is to present adaptive designs for which we can quantify the tradeoff between efficiency of estimation of treatment differences, partially randomized allocation and reduction of the number of patients receiving inferior treatments.

The model and parameter estimates are presented in Section 2, together with a method for skewing treatment allocations. Section 3 describes a variety of biased-coin rules and compares the average losses from the resulting skewed designs. Adaptive designs are introduced and their average properties explored in Section 4. In the early stages of the trial adaptive designs can diverge far from the optimum due to poor parameter estimates. Section 5 describes a regularisation that leads to improved average properties for the designs. The distribution of loss and of the proportions of treatments allocated for individual clinical trials are the subject of §6. Section 7 concludes with some comments on the properties of adaptive designs. Mathematical results on optimum design are in an appendix. Although the theory is for any number of treatments, the numerical examples are for exactly three.

2 Models, Variances and Loss

2.1 Models

In general let there be t treatments, one of which may be the control. The vector of unknown treatment effects is α and the patient presents with a vector x_i of covariates. The results of the trial will be analysed using the regression model

$$E(y_i) = g_i^T \omega = h_i^T \alpha + z_i^T \theta. \tag{1}$$

Here h_i is a vector of t indicator variables, the one non-zero element indicating which treatment the patient received. One linear combination of the α is of interest and the $(q - t + 1) \times 1$ vector z_i contains those covariates, including any powers or interactions of the elements of x_i, which will be used to adjust the responses when estimating α.

2.2 Skewed Allocations

We consider designs to estimate particular linear combinations of the elements of ω. We write the linear combination as

$$l^T \omega = l_1^T \alpha + l_2^T \theta. \tag{2}$$

We are not interested in designing to estimate θ, so the $q - t + 1$ elements of θ are nuisance parameters and all elements of l_2 are zero.

We initially assume that the elements of l_1 are known. We then use adaptive designs in which they are estimated from the observations already to hand. To develop adaptive designs, Atkinson and Biswas (2004b) introduced a procedure that yielded a skewed allocation in which a proportion p of the patients received treatment 1. They designed to estimate the linear combination

$$l_1^T \alpha = p\alpha_1 - (1-p)\alpha_2 \qquad (0 \le p \le 1), \tag{3}$$

with minimum variance, so that $l_1 = \{p \ -(1-p)\}^T$. Simulations similar to those for unskewed designs in Atkinson (2002) show that use of biased-coin designs to estimate this linear combination in the presence of covariates produces the desired skewed allocation for a fixed value of p. If $r_1 n$ patients receive treatment 1 and $r_2 n$ treatment 2, r_1 converges to p and r_2 to $1 - p$.

We now extend this procedure to designs for t treatments by use of the linear combination

$$l_1^T \alpha = \pm p_1 \alpha_1 \pm \ldots \pm p_t \alpha_t, \tag{4}$$

with the proportions $p_j, j = 1, \ldots, t$ such that $0 < p_j < 1$ and $\sum p_j = 1$. In the absence of covariates the variance of $l_1^T \hat{\alpha}$ is minimised when the proportion of patients receiving treatment j is $r_j = p_j$, which is the desired allocation. Our simulations show how the rate at which r_j converges to p_j depends upon the design criterion.

We are interested in only one linear combination of the α. There are therefore a further $t - 1$ linear combinations spanning the space of the α that are not of interest. Together with the $q - t + 1$ elements of θ there are therefore in all q nuisance parameters. As we shall see, the properties of the designs depend on q.

2.3 Loss

With more than two treatments interest may be in estimation of two or more linear combinations of the treatment parameters. Expressions for the variance of a general set A of linear combinations of the estimated treatment effects is given by Atkinson (1982) and by Atkinson (2002). The special case for a single linear combination is in (A4).

Let the estimate of the linear combination from the optimum skewed design with $r_j = p_j$ be $l^T \widehat{\alpha_*}$. Then, in the absence of covariates,

$$\text{var}\{l^T \widehat{\alpha_*}\} = \sigma^2/n, \tag{5}$$

which is also the variance for an optimum skewed design with balance over the covariates.

For other designs we find the variance from (A4) with the linear combination given by (A5). We can compare these designs using either the ratio of variances, that is the efficiency E_n, or we can use the loss (Burman 1996), calculated by Atkinson (2002) for eleven rules for unskewed treatment allocation. From (A4) the efficiency of any design is

$$E_n = 1/\left\{ n l^T (G_n^T G_n)^{-1} l \right\}. \tag{6}$$

The loss L_n is defined by writing the variance (A4) as

$$\text{var}\{l^T \hat{\alpha}\} = \frac{\sigma^2}{n - L_n}, \tag{7}$$

so that

$$L_n = n(1 - E_n). \tag{8}$$

With a random element in treatment allocation, the loss L_n is a random variable, depending upon the particular trial and pattern of covariates. Let

$E(L_n) = \mathcal{L}_n$. The results of Atkinson (2002) show that, for random allocation of two treatments in the unskewed case, $\mathcal{L}_n \to q$, the number of covariates. For the randomised c-optimum rule (A9), the asymptotic value $\mathcal{L}_\infty$ when $t = 2$ is $q/5$ and for non-random allocation zero. The loss can be interpreted as the number of patients on whom information is lost due to the lack of optimality of the design.

One advantage of loss as a measure of design performance is that it approaches the informative asymptotic value relatively quickly. For the schemes considered in Atkinson (2002) interpretation of the values of loss using (8) shows that the efficiency of all designs asymptotically tends to one: loss is a more sensitive measure of design performance.

3 Sequential Designs for Skewed Allocation

3.1 Optimum Designs and Biased-coin Designs

The variance of the linear combination of estimated treatment effects in (A4) does not depend upon any unknown parameters. Given the vector z_{n+1} of covariate values for the $n+$ 1st patient, we can calculate the decrease in variance due to allocating each of the t treatments singly to this patient. The sequential construction of optimum designs described in §A.3 allocates that treatment for which the decrease in variance is greatest. When, as here, the variance is that of a linear combination of parameter estimates, the design criterion is that of c-optimality.

A disadvantage of sequentially constructed optimum designs is that it possible to guess correctly the treatment that each patient will receive. Atkinson (2002) compares several rules for the allocation of two treatments in which biased-coin designs are used. These allocate the treatment indicated by the sequential construction of designs with a probability greater than $1/t$. The rules were compared for loss and allocation bias, related to the probability of correctly guessing the allocation. We use four of these rules here, extending them, where necessary, to the allocation of several treatments.

3.2 Classical Allocation Rules

We denote by $\pi(j|x_{n+1})$ the conditional probability that the $(n+1)$st patient, with prognostic factors x_{n+1}, receives treatment j. In some cases these probabilities depend upon the ordering of the treatments by the variances $d_c(j, n, z_{n+1})$. We use $\pi([j]|x_{n+1})$ to represent the probability of allocating the treatment with the jth largest value of the variance.

D: Deterministic (Sequential Design Construction)

In order to achieve balance that treatment should be allocated for which $d_c(j, n, z_{n+1})$, $j = (1, \ldots, t)$ is largest

$$\pi_D([1]|x_{n+1}) = 1.$$

Asymptotically, for any reasonable distribution over time of prognostic factors, the design will be balanced over the factors, when allowance is made for the skewing induced by the linear combinations l, and there will be no loss: $\mathcal{L}_\infty = 0$.

R: Completely Randomized

For skewed designs

$$\pi_R(j|x_{n+1}) = p_j$$

and $\mathcal{L}_\infty = q$, the value for unskewed designs.

These two rules represent the extremes of rules which aim for skewing and balance over both the short and long term. The losses of the other rules considered here are bounded by these values.

A: $\mathbf{D_A}$-Optimality

We consider only the special case when interest is in one linear combination, so the D_A-optimum criterion of Atkinson (1982) reduces to c-optimality. From (A9)

$$\pi_A(j|x_{n+1}) = \frac{p_j d_c(j, n, x_{n+1})}{\sum_{k=1}^{t} p_k d_c(k, n, x_{n+1})}. \tag{9}$$

For two treatments Burman (1996) shows that $\mathcal{L}_\infty = q/5$.

E: Efron's Biased-Coin

In Efron's original biased-coin design (Efron 1971), with two treatments and no prognostic factors, the probability of allocation of the under-represented treatment is $\pi_E([1]|x_{n+1}) = b_{[1]} = 2/3$. As for the deterministic rule, $\mathcal{L}_\infty = 0$.

When there are covariates the allocation depends upon the ordering of the treatments by the variances $d_c(j, n, z_{n+1})$. Let the rank of treatment j by this ordering be $R(j)$. For unskewed allocations we can take

$$\pi_E(j|x_{n+1}) = b_j = 2\{t + 1 - R(j)\}/\{t(t + 1)\}, \tag{10}$$

when the b_j sum to one. For skewed allocation we need to weight the b_j by the skewing proportions p_j to obtain

$$\pi_E(j|x_{n+1}) = b_j p_j / \sum_{k=1}^{t} b_k p_k. \tag{11}$$

3.3 Bayesian Biased-Coin Designs

The comparisons of Atkinson (2002) for two treatments and unskewed allocation showed that Bayesian biased-coin designs derived from the general approach of Ball, Smith, and Verdinelli (1993) have good properties in terms of bias and loss. Balance is forced at the start of the trial but, as n increases, the allocation becomes increasingly random and so safer from allocation bias. The extension to skewed allocation, with examples for two treatments, is given by Atkinson and Biswas (2004a). For t treatments the probabilities of allocation are

$$\pi_B(j|x_{n+1}) = \frac{p_j\{1 + d_c(j, n, x_{n+1})\}^{1/\gamma}}{\sum_{k=1}^{t} p_k\{1 + d_c(k, n, x_{n+1})\}^{1/\gamma}}, \tag{12}$$

where γ is a parameter, to be chosen by the clinician, that controls the rate of change of the criterion from balance to randomness. We leave to another paper the exploration of the properties of these designs.

3.4 Numerical Results for Skewed Designs

In this section the four non-Bayesian allocation rules are compared for the linear combination of three treatments with $l_1 = (0.8 \quad -0.15 \quad 0.05)^T$. The proportion of treatments allocated is therefore expected to be 0.8, 0.15 and 0.05, the minus sign in the definition of l serving to avoid the generation of singular designs. The small value of 0.05 for treatment 3 was deliberately chosen to exhibit any instabilities that might exist in the procedure for generating adaptive designs. As we see in §6 the proportion of patients allocated to treatment 3 is sensitive to the design criterion. In the comparisons of this section we find the loss for $q = 5$ and 10. The results shown are the averages of 10,000 simulations of 800 patient trials with the elements of the prognostic factors z_i independently normally distributed with variance one.

The plots of Figure 1 show the losses, as functions of patient number: the left-hand panel is for $q = 5$ and the left-hand panel for $q = 10$. The losses are similar to those in Figures 4 and 5 of Atkinson (1999) which were for unskewed D_A-optimality when two contrasts orthogonal to the treatment mean were of interest. They are also similar to the plots in Atkinson (2002)

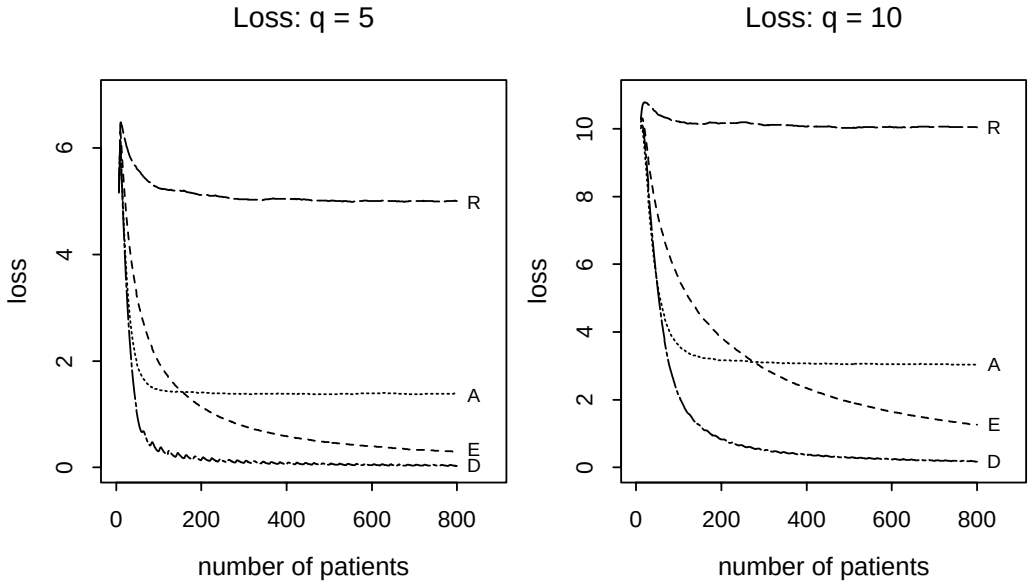

Figure 1: Designs for $l_1 = (0.8 \quad -0.15 \quad 0.05)^T$. Average losses L_n for four allocation rules: A, D_A-optimality; D, deterministic; E, Efron's biased coin and R, random. Averages of 10,000 simulations: Left-hand panel $q = 5$; Right-hand panel $q = 10$

for unskewed designs for two treatments. The main differences are that we have extended the simulations to 800 patients and that there is a slight increase in all losses for the skewed allocation. This arises because, with 16:3:1 ratios of allocation, the skewed designs are on average slightly less well balanced than those for unskewed allocation. In both panels the loss for rule R is close to q and that for D decreases to zero, faster for $q = 5$ than for $q = 10$. The losses for rule E also decrease to zero, becoming less than those for rule A, which are stable after $n = 100$. For two treatments the asymptotic losses for A are $q/5$. Here $3q/10$ seems closer to the value. The numbers for L_{800} are in Table 1.

There is also some fine structure for the loss for rule D that is particularly evident in the left-hand panel of Figure 1. The slight saw-tooth pattern arises because, on average, only one in every twenty patients receives treatment 3. In the sequential construction of an optimum design without any randomness this treatment would be allocated regularly at steps of 20 in n. Here the randomness introduced by sampling the prognostic factors is not sufficient to completely destroy this pattern. For the other rules the randomness in the allocation does destroy this pattern arising from balancing the design.

Rules R and D are the two most extreme in Figure 1. Figure 2 is a plot for

Table 1: Average loss L_{800} for skewed allocations from 10,000 simulations, with target proportions 0.8, 0.15 and 0.05

Rule	$q = 5$	$q = 10$
A	1.39	3.04
D	0.03	0.17
E	0.30	1.26
R	5.01	10.04

these two rules of the average values of the proportions r_j receiving the three treatments. In both panels the proportions start at 1/3 since the algorithm initially allocates three patients to each treatment. Thereafter the values of r_1 and r_2 approach 0.8 and 0.15 in a similar manner. The difference comes in the plot of the values of r_3 which approaches 0.05 more rapidly for rule D than for rule R. This is to be expected since rule D is forcing the r_j to mimic the p_j as quickly as possible.

4 Adaptive Designs

4.1 Link Function Based Adaptive Design for Two Treatments

So far we have assumed that the values of the p_j are known. For two treatments Atkinson and Biswas (2004b) extend the link-function based adaptive design of Bandyopadhyay and Biswas (2001) to provide a design with some randomness in which p_1 and p_2 are estimated from the data.

The purpose was to skew allocation towards the better treatment. Assume that large values of the response y are desired. Let $\hat{\Delta} = \hat{\alpha}_1 - \hat{\alpha}_2$. Then they suggest that the adaptive probability of allocating treatments should be calculated using the estimate

$$\hat{p}_1 = \Phi(\hat{\Delta}/T),$$

where $\Phi(x/T)$ is the distribution function of a $N(0, T^2)$ random variable and $\hat{p}_2 = 1 - \hat{p}_1$. If $\hat{\Delta}$ is positive, that is if $\hat{\alpha}_1 > \hat{\alpha}_2$, then $\hat{p}_1$ is > 0.5. These estimated values are used in calculating the variance $d_c(j, n, z_{n+1})$ for the $(n+1)$st patient. They then apply the allocation rules of §3.2. As it becomes clearer that treatment 1 is superior to treatment 2, the allocation proportion converges to $J\{(\alpha_1 - \alpha_2)/T\}$ with the speed of convergence depending on the allocation rule.

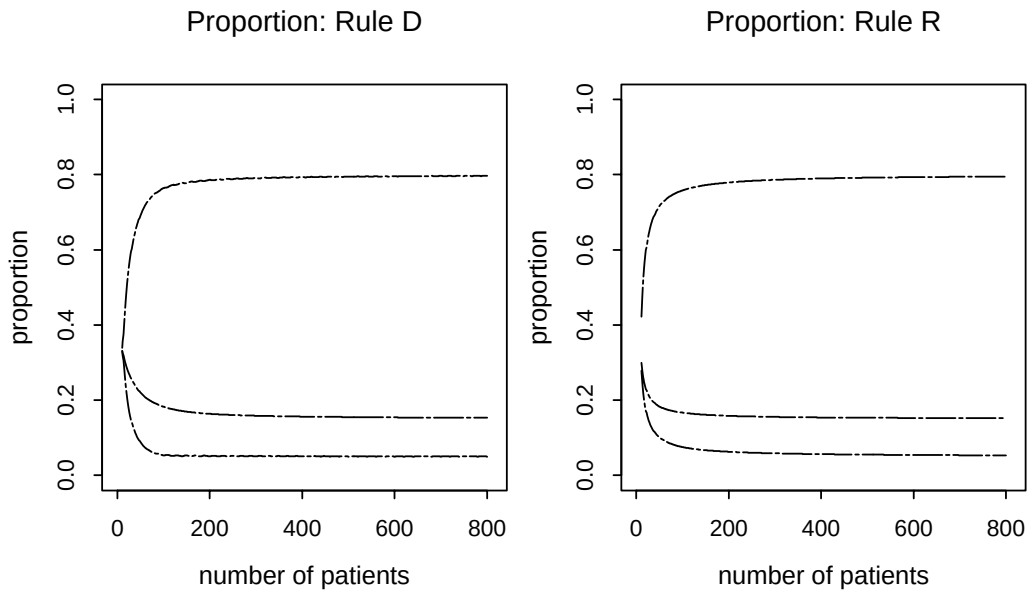

Figure 2: Designs for $l_1 = (0.8 \quad -0.15 \quad 0.05)^T$. Average ratios r_j of treatments allocated for two allocation rules when $q = 5$. Averages of 10,000 simulations: Left-hand panel rule D; Right-hand panel rule R. In particular r_3 converges more slowly to 0.05 for rule R than it does for rule D

4.2 Adaptive Design for Several Treatments

Suppose that after n patients have been treated the estimated treatment parameters are $\hat{\alpha}_j$. To extend the adaptive design criterion to more than two treatments we need to preserve the invariance of the procedure to the overall treatment mean. Accordingly let

$$\bar{\alpha} = \sum_{j=1}^{t} \hat{\alpha}_j / t \quad \text{and} \quad \hat{\Delta}_j = \hat{\alpha}_j - \bar{\alpha}.$$

As before we use the cumulative normal distribution to obtain estimated coefficients $\hat{p}_j$ by setting

$$p'_j = \Phi(\hat{\Delta}_j / T) \quad \text{and} \quad \hat{p}_j = p'_j / \sum_{k=1}^{t} p'_k. \tag{13}$$

For $t = 2$ this reduces to the design procedure of §4.1 except that the standard deviation T is replaced by $2T$

4.3 Numerical Results for Adaptive Designs

In this section the four allocation rules are compared under the same conditions as before, but with the weights $\hat{p}_j$ for the linear combination of treatments estimated from the previous n observations, so that it is assumed that the responses from all patients are available before allocation is made for patient $n+1$. The three treatment means were taken as $\alpha_1 = 3.968$, $\alpha_2 = 0.645$ and $\alpha_3 = 0$. Together with $T = 1$ the algorithm of §4.2 yields the values of the weights p_j as 0.8, 0.15 and 0.05, the values that were used in the earlier calculations. The values of the responses are simulated by adding an independent standard normal error to the value of α for the allocated treatment.

The resulting average losses are plotted in Figure 3. Comparison with

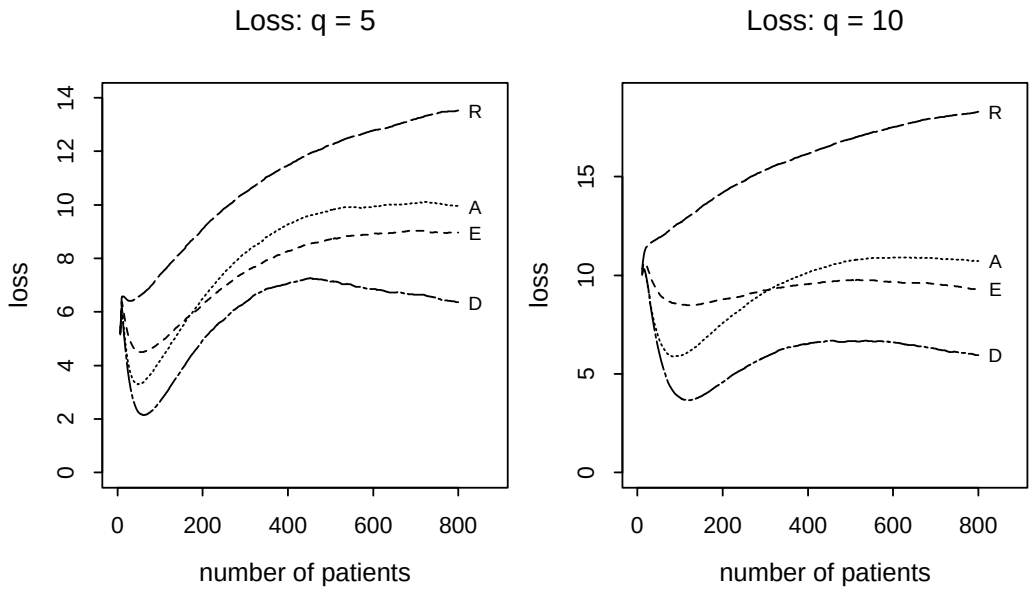

Figure 3: Adaptive designs with $\sigma = 1$. Average losses L_n for the four allocation rules: A, D_A-optimality; D, deterministic; E, Efron's biased coin and R, random. Averages of 10,000 simulations: Left-hand panel $q = 5$; Right-hand panel $q = 10$. To be compared with Figure 1

the designs for known skewing in Figure 1 shows that the values of L_{800} have increased by approximately 8 for all rules except that for sequential design construction D, where they have increased by around 6. The numbers for L_{800} are in Table 2. The effect of the increased uncertainty due to the use of the adaptive designs is not only to increase the average loss but to change the shape of the curves. When the observation error is small or, equivalently

Table 2: Average loss L_{800} for unregularised and regularised allocations from 10,000 simulations

	Unegularised		Regularised	
Rule	$q = 5$	$q = 10$	$q = 5$	$q = 10$
A	9.96	10.72	4.57	6.33
D	6.36	5.95	3.34	3.50
E	8.96	9.28	3.72	5.15
R	13.52	18.28	9.05	14.24

as n becomes large, the values of the losses will approach those in Figure 1. The curves in Figure 3 for rules A, E and D show that the highest values of average loss have been passed and that the values are starting to decrease. However, the loss for the completely random rule R is still increasing at $n = 800$.

5 Regularisation

A few of the simulated adaptive designs gave rise to exceptionally high losses. These usually occurred because, due to observational error, treatment 3 seemed even worse than it was. Consequently, the proportion r_3 for some trials was sensibly less than the optimum value of 0.05. To avoid such problems the adaptive designs were regularised to ensure that each treatment continued to be allocated throughout the trial, yielding consistent estimates of the α_j. Three of the first nine patients are allocated to each treatment. Thereafter, if the number allocated to any treatment was below $\sqrt{n}$, that treatment is allocated when n is an integer squared. For our 800 trial design with 3 patients allocated initially to each treatment, the first regularisation could occur when $n = 16$ and the last when $n = 784$.

The effect of the regularisation should be that r_3 is forced to have values above 0.05 until $n = 400$. However, the implementation used in the simulations only checks for balance when n is exactly an integer squared. One minor effect is that if two proportions are less than $\sqrt{n}$, only one will be increased. The other is that the proportion will be below $1/\sqrt{n}$ until the correction is made.

The resulting average losses for the regularised designs are plotted in Figure 4. Comparison with the designs for known skewing in Figure 1 shows that the increases in the average loss L_{800} are around three to four, that

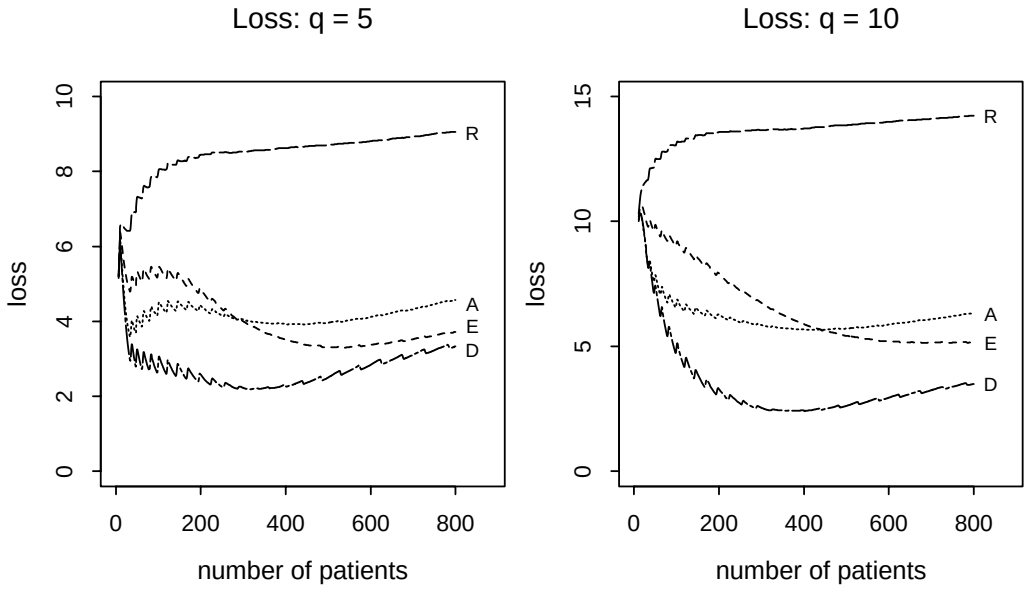

Figure 4: Regularised adaptive designs with $\sigma = 1$. Average losses L_n for the four allocation rules: A, D_A-optimality; D, deterministic; E, Efron's biased coin and R, random. Averages of 10,000 simulations: Left-hand panel $q = 5$; Right-hand panel $q = 10$. To be compared with Figures 1 and 3

is around half the values of the increases for the unregularised designs in Figure 3. The numbers are again in Table 2.

The effect of the regularisation is therefore appreciably to reduce the average loss. A further effect is also clear in both panels of Figure 4, especially, but not only, in the traces for rule D. The sawtooth pattern of increases in loss for $n < 400$ occurs at each point when n is an integer squared and some designs are being forced to move away from the optimum since we should have $r_3 > 0.05$. Around $n = 400$ the effect of the regularisation does not show. Above that the effect is a decrease in loss as designs for which r_3 is too low are being forced towards the optimum. These effects are small compared with the overall effect of the regularisation which is, for these values of p_j, to prevent designs from having extremely small values of r_3.

The effect of the regularisation on the proportion r_3 can be seen in the plots of average proportions for two allocation rules with $q = 5$ shown in Figure 5. In the left hand panel, for rule D, the average values of r_3, and, to a lesser extent, of r_2, increase at each regularisation point until n is around 200. For rule R, shown in the right-hand panel, the average value of r_3 decreases more slowly and the regularisation has a less obvious effect. The overall

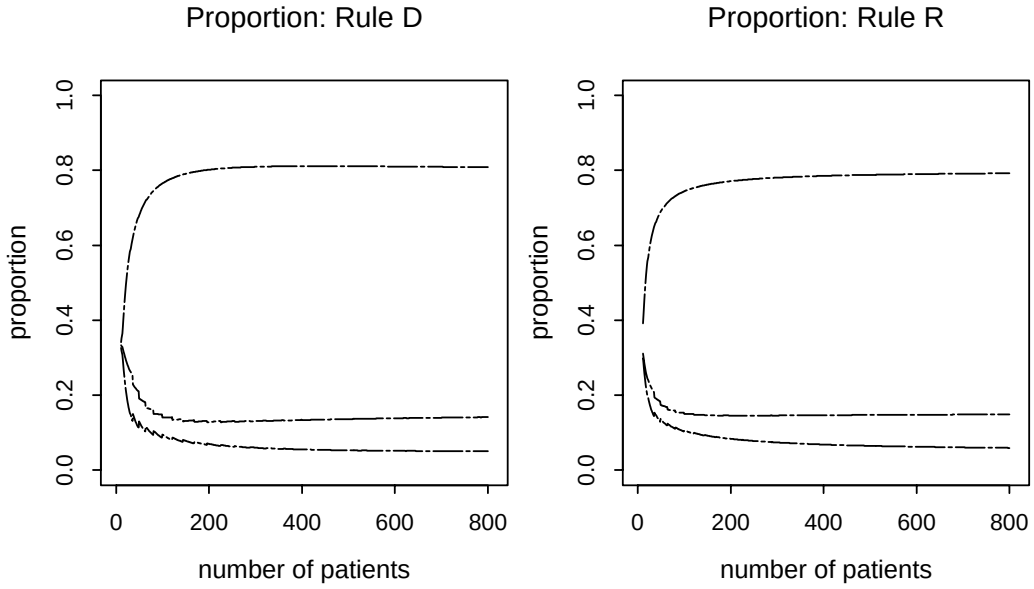

Figure 5: Regularised adaptive designs with $\sigma = 1$. Average ratios r_j of treatments allocated for two allocation rules when $q = 5$. Averages of 10,000 simulations: Left-hand panel rule D; Right-hand panel rule R. Again r_3 converges more slowly to 0.05 for rule R than it does for rule D. To be compared with Figure 2

effect of the regularisation is that, by ensuring occasional measurements from patients receiving treatment 3, consistent estimates of, in particular, α_3 are obtained; in consequence, under-estimates of the parameters at the start of the trial do not cause continuing serious departures from the optimum allocation.

6 The Distribution of Loss

Although the average properties of the design are of interest, it is important and helpful to look at the individual trials: it is little consolation for a clinician with a poorly balanced trial to be assured that the average properties of trials produced by the randomization scheme are excellent. The properties of the individual trials reinforce the discussion about the effect of regularisation in §5. We compare the properties of regularised and unregularised designs using rule A when $q = 5$ and $\sigma = 1$.

The left-hand panel of Figure 6 shows boxplots of the distribution of 1,000 values of L_n for n from 100 to 800. The distribution has its longest upper tail at $n = 400$. From n from 600 onwards the largest loss is around 35. The plot for average loss in Figure 4 indicates that the average loss for this regularised design increases slightly from $n = 400$ and this is evident from the trend of

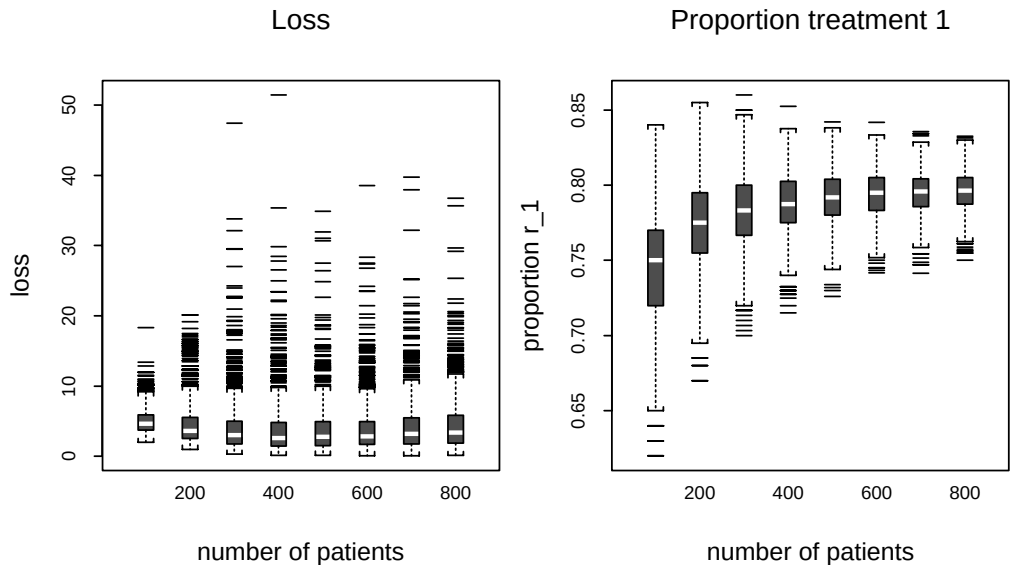

Figure 6: 1,000 individual regularised adaptive designs for rule A when $q = 5$ and $\sigma = 1$. Left-hand panel: boxplots of loss L_n. Right-hand panel: proportion r_1 of patients receiving the first treatment

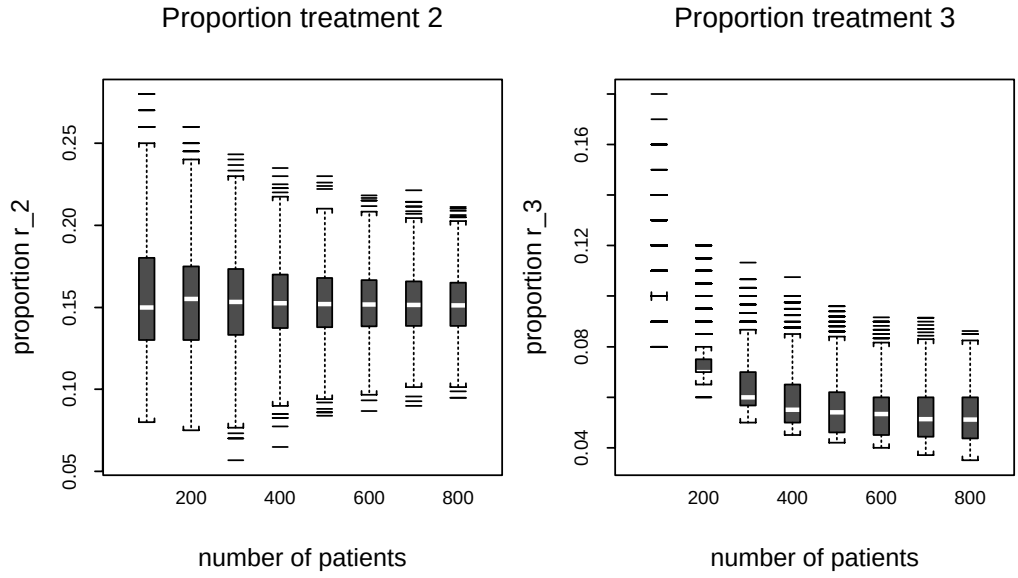

Figure 7: 1,000 individual regularised adaptive designs for rule A when $q = 5$ and $\sigma = 1$. Left-hand panel: proportion r_2 of patients receiving the second treatment. Right-hand panel: proportion r_3 of patients receiving the third treatment

the centres of the boxplots. The right-hand panel of Figure 8 shows boxplots for r_1, the proportion of patients receiving treatment 1. These converge

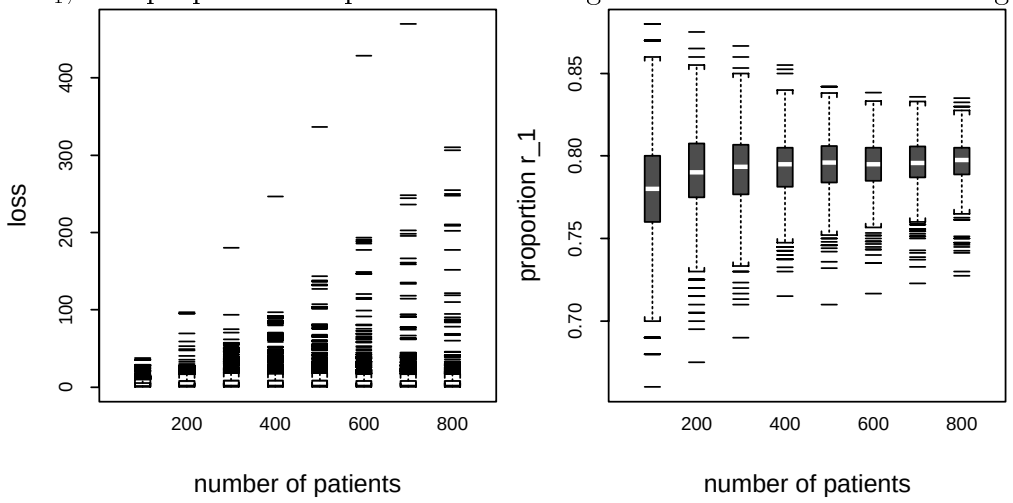

Figure 8: 1,000 individual unregularised adaptive designs for rule A when $q = 5$ and $\sigma = 1$. Left-hand panel: boxplots of loss L_n. Right-hand panel: proportion r_1 of patients receiving the first treatment

steadily from below towards 0.8 - even as early as $n = 100$ the effect of the starting value of 1/3 has been virtually eradicated. At this point both r_2 and r_3 should have minimum values of 0.1, so that the maximum value of r_1 is 0.8. In fact, for the reasons described in the second paragraph of this section, one trial has a value of 0.84 and several have $r_1 = 0.81$. The boxplot is thus slightly misleading for this value of n, as it is for $n = 200$, because its symmetry does not reflect the underlying distribution.

The plots for the proportions r_2 and r_3 are in Figure 7. The value of r_2 converges steadily towards 0.15. However, the plot for r_3 is very different. Initially the regularisation forces virtually all trials to have a value of 0.1 at $n = 100$: the minimum is 0.08. Likewise, for $n = 200$ there is little spread in the distribution of values. Then, as n increases, the mean value of r_3 decreases, as does the minimum value, constrained by the regularisation. Some trials have higher value of r_3, above 0.08, but the number is decreasing.

The corresponding plots for unregularised designs are very different. We know from Table 2 that the average loss at $n = 800$ more than doubles from 4.57 for the regularised design to 9.96 for the unregularised design. The left-hand panel of Figure 7 shows that this is not only an increase in the average, but that the upper tail of the distribution also increases dramatically - the highest loss is around 500 at $n = 700$, rather than 50. The plot of the proportion r_1 in the right-hand panel now shows that the average value

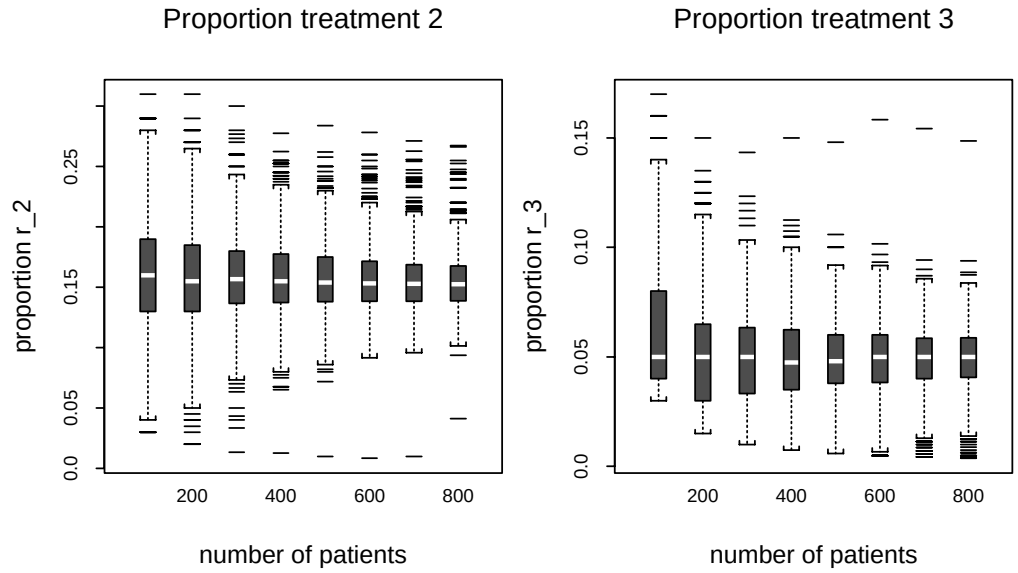

Figure 9: 1,000 individual unregularised adaptive designs for rule A when $q = 5$ and $\sigma = 1$. Left-hand panel: proportion r_2 of patients receiving the second treatment. Right-hand panel: proportion r_3 of patients receiving the third treatment

approaches 0.8 more rapidly than for the regularised design, but that there are now trials with initially high values of r_2 or r_3.

The plot of r_2 in the left-hand panel of Figure 9 is centred around the value of 0.15, but with a higher scatter than for the regularised design. In particular there seem to be several trials with values of 0.25 or above for r_2, where the maximum for the regularised design is around 0.22. The greatest effect of the regularisation is on the value of r_3 which is centred close to 0.05 throughout for the unregularised design. But there are now many trials giving rise to values of r_3 close to zero, even at $n = 800$. A final feature of these plots is that it is possible to follow the progress of one trial that has a very small value for r_2, a large one for r_3 and gives a very large value of loss throughout. However, for $n = 800$, it is starting to move in the direction of the optimum allocation proportions.

The distribution of loss for several rules is explored by Atkinson (2003) for unskewed designs. But, due to the adaptive nature of the designs considered here, the distributions are much longer tailed; occasional trials have a proportion of patients receiving, in particular, treatment three that is far from the target. Such long tails are a general feature of adaptive methods: frequent good behaviour is offset by occasional trials in which a large number of patients is needed.

7 Discussion

The main contribution of this paper is to extend the two-treatment designs of Atkinson and Biswas (2004b) to any number of treatments. Examples have been given of the construction of skewed designs for three treatments when some randomization is required across prognostic factors. This procedure has been combined with adaptive estimation of the skewing proportion to obtain adaptive designs. A subsidiary purpose of the present paper, as an extension of Atkinson (2002), is to provide a methodology for the evaluation of designs strategies for adaptive clinical trials. We calculate loss, which is the number of patients on whom information is effectively lost due to the imbalance of the design. We can also calculate the number of patients receiving the inferior treatments, so both measures of design performance are in the same units.

Although the loss for the adaptive designs can be appreciably greater than that for the designs with known skewed allocation, the curves of average loss in Figure 3 must converge, for large n or small σ, to those of Figure 1. The results of Atkinson (2003) show that rule R and the Bayesian rules ultimately have losses with a χ^2_q distribution. That for rule D degenerates to a point distribution as the loss becomes identically zero. However, the simulations shown here indicate that the value of n would have to be so large, or the value of σ so small, that these conditions are unlikely to be met in practice. The plots of the simulations of §6 provide a cogent way of representing the distribution of possible outcomes for these adaptive design schemes.

Finally, we comment briefly on a theoretical aspect of our adaptive designs. In deriving these designs we have used optimum design theory for regression models with independent errors and then estimated the unknown value of p from results on earlier patients. A similar strategy of substitution of parameter estimates is used, for a different criterion and example, by Rosenberger et al. (2002). Since each allocation depends on the earlier responses, the observations are not independent and the likelihood is complicated. However, Antognini and Giovagnoli (2004) prove that, for responses modelled by the exponential family, the optimum adaptive designs obtained by sequential use of parameter estimates are indeed optimum. For our adaptive procedure to be covered by similar results we would need a result on the temporal distribution of covariates to assure the convergence of the normalised information matrix $G_n^T G_n/n$ to a positive definite matrix. We would also require that the proportion r_j for any j does not tend to zero. Our regularisation assures that this holds.

References

Antognini, A. B. and A. Giovagnoli (2004). On the large sample optimality of sequential designs for comparing two treatments. *Journal of the Royal Statical Society, Series A 167*. (in press).

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

Atkinson, A. C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine 18*, 1741–1752.

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

Atkinson, A. C. (2003). The distribution of loss in two-treatment biased-coin designs. *Biostatistics 4*, 179–193.

Atkinson, A. C. and A. Biswas (2004a). Bayesian adaptive biased-coin designs for clinical trials with normal responses. (Submitted).

Atkinson, A. C. and A. Biswas (2004b). Optimum design theory and adaptive-biased coin designs for skewing the allocation proportion in clinical trials. (Submitted).

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

Bandyopadhyay, U. and A. Biswas (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika 88*, 409–419.

Burman, C.-F. (1996). *On Sequential Treatment Allocations in Clinical Trials*. Göteborg: Department of Mathematics.

Cox, D. R. (1988). A note on design when response has an exponential family distribution. *Biometrika 75*, 161–164.

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

Matthews, J. N. S. (2000). *An Introduction to Randomized Controlled Clinical Tials*. London: Edward Arnold.

Rosenberger, W. F. and J. L. Lachin (2002). *Randomization in Clinical Trials: Theory and Practice*. New York: Wiley.

Rosenberger, W. F., N. Stallard, A. Ivanova, C. N. Harper, and M. L. Ricks (2002). Optimal adaptive designs for binary response trials. *Biometrics 57*, 909–913.

A Appendix

This appendix summarises the algebra for linear regression models when there are t treatments. The first two sections establish notation for the model and give the variance of linear combinations of parameters. The third section describes the sequential construction of optimum experimental designs; biased-coin designs are in Section A.4.

A.1 Models

The matrix form of the model (1) for n patients is

$$E(Y_n) = G_n\omega = H_n\alpha + Z_n\theta \tag{A1}$$

where Y_n is the $n \times 1$ vector of responses for the n patients and α is the vector of treatment effects. Here H_n is the $n \times t$ matrix of indicator variables for treatment allocation, with one non-zero entry per row, and Z_n is the $n \times (q-t+1)$ extended matrix of prognostic factors, with ith row x_i. Because of the way we have parameterised the treatment effects, Z_n does not include a constant column.

The variance of the least squares estimator of the parameter vector α is

$$\operatorname{var}\hat{\alpha} = \sigma^2\{H^TH - H^TZ(Z^TZ)^{-1}Z^TH\}^{-1}, \tag{A2}$$

where σ^2 is the variance of the errors, assumed additive in (A1). Here, as elsewhere, we have suppressed the subscript n when it is not of importance.

A.2 Variances of Linear Combinations of Parameters

A skewed allocation is found by designing to estimate a particular linear combination of the elements of α. For estimation of the general linear combination $l_1^T\alpha$,

$$\operatorname{var}(l_1^T\hat{\alpha}) = \sigma^2 l_1^T\{H^TH - H^TZ(Z^TZ)^{-1}Z^TH\}^{-1}l_1, \tag{A3}$$

where l_1 is $t \times 1$. A more compact expression is obtained by instead writing the general linear combination as $l^T\omega = l_1^T\alpha + l_2^T\theta$, which is (2). Since the θ

are nuisance parameters, l_2 is a vector of zeroes and an equivalent expression for (A3) is

$$\text{var}\,\{l^T\widehat{\omega}\} = \sigma^2 l^T (G_n^T G_n)^{-1} l, \tag{A4}$$

a scalar. When there are three treatments the linear combination (4) is taken as

$$l_1^T = (p_1 \quad -p_2 \quad 1-p_1-p_2). \tag{A5}$$

The vector l is used to calculate the predicted variance $d_c(j, n, x_{n+1})$ that is needed in the next section as a component of our design algorithms.

A.3 Optimum Experimental Designs

D-optimum experimental designs for the linear regression model $E(Y) = G\omega$ maximize the determinant $|G^T G|$; they minimize the generalized variance of the parameter estimates and provide a normal theory confidence region of minimum volume. Such optimum designs can be constructed sequentially. After n trials the matrix of allocations and prognostic factors is G_n. The predicted response at the point g is $\hat{y}(g)$ with

$$\text{var}\,\{\hat{y}(g)\} \propto d(g, n) = g^T (G_n^T G_n)^{-1} g. \tag{A6}$$

If the vector of allocation and prognostic factors for the $(n+1)$st patient is g_{n+1}, G_{n+1} is formed by adding the row g_{n+1}^T to G_n. A useful matrix result is that

$$\begin{aligned} |G_{n+1}^T G_{n+1}| &= \{1 + g_{n+1}^T (G_n^T G_n)^{-1} g_{n+1}\} |G_n^T G_n| \\ &= \{1 + d(g_{n+1}, n)\} |G_n^T G_n|. \end{aligned} \tag{A7}$$

The optimum treatment for the $(n+1)$st patient is therefore that for which $d(g_{n+1}, n)$, the variance of the predicted response after n trials, is a maximum.

In the clinical trials considered in this paper, where interest is in the vector of coefficients l, D-optimality is replaced by c-optimality with designs being found to minimise $l^T (G_{n+1}^T G_{n+1})^{-1} l$. These designs can again be constructed iteratively, the variance $d(g_{n+1}, n)$ being replaced by

$$\begin{aligned} d_c(j, n, x_{n+1}) &= \{g_{n+1}^T (G_n^T G_n)^{-1} l\}^2 / l^T (G_n^T G_n)^{-1} l, \\ &\quad (j = 1, \ldots, t), \end{aligned} \tag{A8}$$

where g_{n+1} combines the allocation indicator h_{n+1} for the $(n+1)$st patient and z_{n+1}, the extended vector of prognostic factors, known for the new patient. In the iterative construction of c-optimum designs, patient $n+1$ would, in the absence of randomization, receive the treatment for which $d_c(j, n, x_{n+1})$ is a maximum, where j runs over all t treatments. In the numerical examples of this paper, $t = 3$.

A.4 Sequential Biased-Coin c-Optimum Design

We use optimum design theory to generate designs which are unlikely to be far from balance over the covariates if the trial ceases at an arbitrary time point. Inclusion of a biased coin provides some randomness.

The vector l forces unequal allocation. At the optimum design, which allocates a fraction p_j of the patients to treatment j, all $d_c(j, n, x_{n+1})$ are equal. In other designs, a larger value of $d_c(j, n, x_{n+1})$ indicates a treatment which is under-represented. The original suggestion of Atkinson (1982) for D_A-optimality reduces to allocation of treatment j with probability

$$\pi_A(j|x_{n+1}) = \frac{p_j d_c(j, n, x_{n+1})}{\sum_{s=1}^{t} p_s d_c(s, n, x_{n+1})}. \tag{A9}$$