The Comparison of Designs for Sequential Clinical Trials with Covariate Information

Anthony C. Atkinson *
The London School of Economics, London WC2A 2AE, UK

December 14, 2001

Abstract

The paper develops methods for the comparison of randomised rules of the biased-coin type for the sequential allocation of treatments in a clinical trial. One important characteristic is the *loss*, which measures the increase in the variance of parameter estimates due to the imbalance caused by randomisation. The other important characteristic is the *selection bias* measuring the probability of correctly guessing which treatment is to be allocated next. The combination of these two measures leads to the elucidation of admissible designs. Simulations provide clear plots of the behaviour of the designs and make it possible to distinguish good designs from those which are less good.

Keywords: admissibility; balance; Bayesian coin; bias; biased-coin design; D_A-optimum design; loss; minimisation; randomization.

1 Introduction

Patients arrive sequentially and are each to be given one of t treatments in a parallel group trial. The statistical literature describes many rules for deciding which treatment should be allocated to the patient who has just arrived. Examples are the minimisation rule of Pocock and Simon (1975) and the "biased-coin" rule of Efron (1971). The purpose of this paper is to describe a method for comparing such rules.

These two rules are rather different. The minimisation rule is concerned with balance of allocations over the covariates or prognostic factors. Because

*a.c.atkinson@lse.ac.uk

it is not known when recruitment to the trial will stop, the sequential allocation of treatments should be made so that the trial is always as balanced as possible. Ideally if, for example, age is thought to be an important factor, the same number of old people should receive each treatment. Likewise the number of young people receiving each treatment should be as equal as possible. Of course, the numbers of young and old patients may be quite different. This balance should be achieved for each prognostic factor. Then, if these factors are included in the analysis, the effect of the treatments will be estimated with minimum variance. Such a rule does not include any overt randomisation. Some randomisation is however included in the biased coin designs of Efron (1971) which allocate the next treatment with a random component which depends upon the previous history of allocations. Since the design is not forced to be as balanced as possible, stopping the trial at an arbitrary point will lead to a slight increase in the variance of the parameter estimates compared with that from the minimisation allocation. It is the relationship between the variance of parameter estimates and the bias which can result from the lack of randomisation that provides the basis for our comparisons.

In general, increasing randomisation will reduce the possibility of bias and will increase the variance of estimates. The comparisons in this paper quantify this relationship and provide a method of assessing the admissibility of the various allocation schemes. Some well-known schemes are shown to be inadmissible in this sense - they give both a higher bias and a higher variance than at least one competitive scheme.

In the original biased-coin design of Efron (1971) there are just two treatments and no measurements of prognostic factors on the patients. There have been numerous extensions of this idea to t treatments and to the presence of prognostic factors. Seven different schemes are compared in this paper, for four and nine prognostic factors. The numerical comparisons are mostly for two treatments. But the simulation methods and analytical results are applicable to any number of treatments. The emphasis is at least as much on the methods developed for making the comparisons as on the performance of any specific allocation rule.

Section 2 introduces the literature on clinical trials with sequential recruitment of patients. The linear model theory on which the comparisons of designs are based is introduced in §3, as are biased-coin designs based on D_A-optimality. As is shown in §4, the increase in the variance of estimated treatment effects due to randomisation can be interpreted as a loss, expressed as the number of patients on whom information is unavailable. Asymptotic results for the loss due to Smith (1984a) and Smith (1984b) are augmented by simulations which establish the losses associated with the allocation rules

for small sample sizes and for several distributions of prognostic factors. The rate of convergence of these losses to the asymptotic values is demonstrated graphically.

Section 5 relates the Bayesian biased-coin allocation rule of Ball, Smith, and Verdinelli (1993) to those of §4, showing that one version can be considered an extension of designs based on randomised D_A-optimality. The loss for these designs is established and exhibited by simulation. This loss, like those of §4 is an average over many simulations. §6 explores the distribution of loss, which has a remarkably stable pattern, depending only on the mean of the distribution.

The seventh section develops analogous results for selection bias, a measure of the effect of randomisation. For two treatments the bias is the difference in the average number of correct and incorrect guesses of the next treatment to be allocated. The results on loss and bias are combined in §8, where admissible rules are defined. Plots clearly show those inadmissible rules which have both loss and bias higher than another rule. Although the effects on loss are not large compared with the numbers of patients, the results provide a clear, unambiguous method for comparing methods for allocating treatments. The final section mentions unresolved issues, including the analysis of data from these sequential rules.

The sequential schemes of this paper do not depend on the outcome of the treatment, that is the observations y, an important difference from the adaptive designs reviewed in Rosenberger (1996). The large mathematical statistical literature on the allocation of treatments in sequential clinical trials is reviewed in the papers by Smith. Some further references are in Atkinson and Bailey (2001, §13).

2 Background

The design and execution of a clinical trial includes numerous practical, ethical, regulatory and data analytical aspects as well as design. Piantadosi (1997) gives much detail. The place of clinical trials in drug development is considered by Senn (1997). The theoretical purposes and practical aspects of randomization for both design and inference are given book-length treatment by Matthews (2000) and by Rosenberger and Lachin (2002). Chapter 4 of the latter book is closest to this paper in attempting to quantifying the conflicting requirements of randomness and balance.

The historical methods of imposing balance in the sequential allocation of treatments across prognostic factors (Taves 1974, Pocock and Simon 1975) consider either categorical factors, or categorise those that are continuous

such as age. The aim is then to balance the number of patients in each stratum, that is each cell of the factorial array that receives each treatment. Such balance rapidly becomes impossible for even a few covariates unless the number of patients is large. The allocation rules therefore aim for marginal balance across the factors.

Randomization can be built into the rules by the incorporation of a biased coin. If the aim is to obtain balance within each factorial cell, Efron's biased coin can be used when there are two treatments, or an extension of it if there are three or more (Atkinson 1999b). It seems to be this kind of rule which is mostly used in practice.

Altman (1991, §15.2.3), following Fentiman, Rubens, and Hayward (1983), describes a trial in which there are two treatments for patients with breast cancer. There are four prognostic factors each dichotomised: age ≤ 50 years or > 50; stage of the disease (1 or 2 against 3 or 4); time between diagnosis of cancer and effusion, again dichotomised, and whether or not the patient was post-menopausal. More recent examples have a similar structure. Eichhorn, Domanski, Krause-Steinrauf, and others (2001) again have two treatments. The patients are stratified according to four factors at each clinical site and a biased coin allocation is used. Seibold, Korn, Simms, and others (2000) have two levels of treatment plus a placebo and try to balance their 68 patients over disease duration and previous use of D-penicillamine. A more complicated treatment structure is used by Roberts, Urdaneta, Vera, and others (2000) who have four levels of radiation with or without chemotherapy. At the time of reporting 160 patients had been recruited whose treatments had been balanced and randomized over two hospitals and six disease stages. Some of the combinations of prognostic factors were, of course, sparsely represented. The simulation results in the right-hand panel of Figure 1 show the effect of such sparsity on a trial with nine dichotomised prognostic factors.

Randomization is required in order to avoid biases which can weaken or destroy the usefulness and applicability of the results of a trial. Matthews (2000) list five sources of bias. That most important at the design stage is selection bias which arises when the choice of the next patient to enter the trial is affected by knowledge of the treatment that patient will receive. A famous non-clinical example is the Lanarkshire milk experiment ("Student" 1931) in which teachers swapped treatment and controls so that undernourished children received the milk. The paramount importance of avoiding selection bias in clinical trials is emphasized by Chalmers (1990). As we see in §7, selection bias can readily be calculated for some allocation schemes and found by simulation for others. However Taves (2001) comments that selection bias is not the only form of bias which can arise. He is particularly concerned with the effect of low frequency biases, such as those due to an omitted trend or low

frequency variation. Atkinson (2001) summarises the evidence (Steele 1980, Smith 1984b) for two treatment biased-coin designs without covariates. For omitted variables, that is low frequency bias, the biased-coin design behaves as well as a permuted block design of length ten. For high frequency correlations, such as those between neighbouring responses, there is a very small increase in variance from use of a biased-coin design compared with complete randomization. The evidence is thus that selection bias is a good surrogate for many of the biases which arise in clinical trials, such as the biases in allocation which arise due to an omitted prognostic factor. We therefore use the extent of selection bias as the measure of the effective randomisation of the designs considered.

In the examples cited above of recent clinical trials in which a form of biased-coin design has been used, the prognostic factors were either categorical, or had been categorised, and there were often several of them. In an editorial on randomization in clinical trials in *The British Medical Journal*, Treasure and MacRae (1998) mention a trial with four prognostic factors. In the ensuing correspondence (Ross 1999, Treasure and MacRae 1999) the point is made that methods, such as the use of a randomised block design within each stratum, need a very large number of patients if treatment balance is to be achieved over each cell of the stratification. Minimization, or the more complicated specific scheme of Efron (1980), provides either marginal balance or a combination of marginal balance with that in each stratum by use of a more or less arbitrary function (Efron describes his function as 'peculiar'). Many of the methods compared in this paper extend to continuous prognostic factors and provide a weighting of the imbalances which is derived from optimum design theory for the linear model and so avoids arbitrariness. In their editorial Treasure and MacRae (1998) refer to minimization as the "platinum standard" for clinical trials. Despite this encomium, the comparisons given here show that this particular scheme is inadmissable. But, more importantly, the paper provides a method for quantitative comparison between the variances and biases of the various methods that have been suggested.

3 Modelling and Design

3.1 A Regression Model

When a trial protocol specifies that the analysis will use adjustment for a set of prognostic factors, we can derive results about the precision of estimated treatment effects for any given allocation of treatments. The resulting

comparison of allocation rules requires a model and an estimation method. We can then calculate the variance of the parameter estimates for each rule assuming one of a number of distributions of prognostic factors in the population.

It is assumed that each of n patients receives one of t treatments which is allocated in the knowledge of a vector of prognostic factors. In many trials the response will be approximately normally distributed, perhaps after transformation: one example would be the logarithm of survival time. It is then appropriate to use the regression model

$$E(Y_n) = G_n\omega = H_n\alpha + Z_n\gamma, \tag{1}$$

for the responses of the first n patients. In (1) H_n is the $n \times t$ matrix of indicator variables for the treatments with one non-zero entry per row, and Z_n is the $n \times (q-1)$ matrix of prognostic factors, including interactions and other terms, if required. The subscript n will only be used when it is necessary to distinguish between quantities for the $(n+1)$st patient for whom a treatment allocation is required and those for the n patients to whom treatments have already been allocated.

For patient n the model (1) is that

$$E(Y_n) = \alpha_j + z_n^T\gamma,$$

when the patient receives treatment j and has a vector z_n of $q-1$ prognostic factors: the γ are nuisance parameters. If there are only two treatments, interest is assumed to be solely in the difference between α_1 and α_2. The purpose of the trial is thus to compare the two treatments, rather than to compare both against some standard. Under these conditions a balanced design is one in which the variance of the estimated treatment difference

$$\text{var } (\hat{\alpha}_1 - \hat{\alpha}_2) \tag{2}$$

is minimised. Other designs should lead to the estimation of $\alpha_1 - \alpha_2$ with small variance.

For any t, interest is in contrasts between the α. The mean level of response therefore becomes an additional nuisance parameter making q nuisance parameters in all. With more than two treatments a set of $t-1$ contrasts orthogonal to the mean is required, for example

$$L^T = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ & & \vdots & & \\ 1 & 0 & 0 & \dots & -1 \end{pmatrix}. \tag{3}$$

For $t = 2$ these contrasts reduce to the difference between treatments (2). For three treatments the two selected contrasts are the difference between the first and second and the first and third treatments. The difference of these two is the contrast for the second and third treatments, but only two of these are linearly independent.

Since the volume of the normal theory confidence ellipsoid for least squares estimates of the contrasts is unaffected by non-singular linear transformations of the contrasts, the exact form of (3) is unimportant, provided the contrasts span the $t-1$ dimensional space orthogonal to the overall mean. Because the γ in (1) are nuisance parameters, the contrasts in (3) need augmenting by a $(t-1) \times (q-1)$ matrix of zeroes

$$A^T = (L^T \quad 0) \tag{4}$$

to reflect interest solely in contrasts in the treatment parameters.

The volume of the confidence ellipsoid for α from the model including the prognostic factors is proportional to the square root of the determinant

$$|A^T(G^TG)^{-1}A| = |L^T\{H^TH - H^TZ(Z^TZ)^{-1}Z^TH\}^{-1}L|. \tag{5}$$

This generalized variance is minimized by the balanced design, in which both an equal number of patients is allocated to each treatment, and there is balance over all prognostic factors so that $H^TZ = 0$. The determinant (5) then reduces to

$$|A^T(G^TG)^{-1}A| = t^t/(n^{t-1}).$$

Of course, given a particular sequence of prognostic factors $z_1, z_2, \ldots$ it may not be possible to obtain exact balance. This value is then a theoretical optimum which is used as a standard for comparisons. But as the simulations of the next section show, sequentially constructed designs can get close to this optimum for even small values of n. The efficiency of any other design is then the ratio of determinants

$$\mathcal{E}_n = \left(\frac{t^t/(n^{t-1})}{|A^T(G^TG)^{-1}A|} \right)^{1/(t-1)}. \tag{6}$$

Raising the ratio to the power $1/(t-1)$ gives a measure of efficiency which responds like the variance of a single parameter to an increase in sample size, for example halving when n is doubled. From (6) the effective number of trials for an unbalanced design relative to the balanced design is $n\mathcal{E}_n$. As a measure of the loss after n trials due to lack of balance in the design we therefore use

$$\mathcal{L}_n = n(1 - \mathcal{E}_n). \tag{7}$$

If the design is exactly balanced, $\mathcal{L}_n$ is zero. Otherwise the loss of information is expressed in terms of number of patients. For the randomized designs studied here $\mathcal{L}_n$ is a random variable. The results of Smith (1984a) and of Smith (1984b) provide asymptotic values $\mathcal{L}_\infty$ for the expected value of the loss. Burman (1996) focused attention on the expected value $\mathcal{L}_n$, using simulation to study small sample properties. Simulations, which are an extension of those of Burman are likewise used here to study the progress of the loss towards its asymptotic value. In the initial stages of the trial, imbalance may be relatively high and the loss $\mathcal{L}_n$ may be far from $\mathcal{L}_\infty$.

3.2 Biased-Coin D_A-Optimum Designs

One of the allocation rules to be considered (Atkinson 1982) makes direct use of results from optimum design theory. Since similar results are also used in the analysis of the Bayesian scheme of Ball et al. (1993), the required material is here gathered together in one section.

In the literature on experimental design for regression models, designs maximizing the determinant $|G^TG|$ are called D-optimum. 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 then $\hat{y}(g)$ with

$$\text{var}\{\hat{y}(g)\} \propto d(g,n) = g^T(G_n^TG_n)^{-1}g. \tag{8}$$

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

$$|G_{n+1}^TG_{n+1}| = \{1 + g_{n+1}(G_n^TG_n)^{-1}g_{n+1}\}|G_n^TG_n|. \tag{9}$$

It then follows that the optimal choice of the $(n+1)$st allocation is that treatment for which the variance of the predicted response after n trials is a maximum. The use of this algorithm in the construction of D-optimum designs is described by Wynn (1970).

In the clinical trials considered here, where interest is in the matrix of contrasts A given by (4), D-optimality is replaced by D_A-optimality with designs being found to minimize $|A^T(G^TG)^{-1}A|$. These designs can again be constructed iteratively, the variance (8) being replaced by

$$d_A(j,n,z_{n+1}) = g_{n+1}^T(G_n^TG_n)^{-1}A\{A^T(G_n^TG_n)^{-1}A\}^{-1}A^T(G_n^TG_n)^{-1}g_{n+1}. \tag{10}$$

In the iterative construction of D_A-optimum designs the next trial would be added where the variance (10) was a maximum over the design region, which consists of allocation of one of the treatments. The design will depend not

only on the previous allocations and the matrix of prognostic factors Z_n, contained in G_n, but also, as in the discussion of (9), on g_{n+1}. This vector includes both the indicator for the allocated treatment and also z_{n+1}, the vector of prognostic factors for the new patient, which are given, not chosen. To emphasize these dependencies the variances have been written $d_A(j, n, z_{n+1})$, $j = 1, \ldots, t$. Asymptotically all treatments will be allocated equally often and the variances will tend to equality. To provide a randomised form of this iterative construction, Atkinson (1982) suggests allocating treatment j with probability

$$\pi_A(j|z_{n+1}) = \frac{d_A(j, n, z_{n+1})}{\sum_{j=1}^{t} d_A(j, n, z_{n+1})}. \tag{11}$$

In (11) the variances $d_A(.)$ could be replaced by any monotone function $\psi\{d_A(.)\}$. In §5 it is shown that one version of the Bayesian biased-coin procedure of Ball, Smith, and Verdinelli (1993) uses

$$\psi(u) = (1 + u)^{1/\gamma},$$

with γ a parameter to be elucidated from the experimenter.

4 Two Treatments

4.1 Loss

With two treatments the parameter of interest is $\Delta = \alpha_1 - \alpha_2$. Following Smith (1984a) the model (1) can be written

$$E(Y) = a\Delta + 1\beta_o + Z\gamma = a\Delta + F\beta, \tag{12}$$

where a is the $n \times 1$ vector of allocations with elements +1 and -1, and the constant term and covariates are included in the $n \times q$ matrix F. Then

$$\mathrm{var}(\hat{\Delta}) = \sigma^2\{a^T a - a^T F(F^T F)^{-1} F^T a\}^{-1}. \tag{13}$$

In (13) it is meaningful to let $b = F^T a$, a "balance" vector which is identically zero when all covariates are balanced across all treatments. Also $a^T a = n$, so that (13) can be written in the revealing form

$$\mathrm{var}(\hat{\Delta}) = \frac{\sigma^2}{n - b^T(F^T F)^{-1} b} = \frac{\sigma^2}{n - \mathcal{L}_n}, \tag{14}$$

where $\mathcal{L}_n$ is the loss after n trials.

It is informative to see what happens when there are no prognostic factors. Then F is a vector of ones and $F^TF = n$. If n_1 patients receive treatment 1, for which the elements of the allocation vector a equal one and $n_2 = n - n_1$ receive treatment 2 for which $a = -1$,

$$b = F^Ta = n_1 - n_2.$$

Now, from (14)

$$var(\hat{\Delta}) = \frac{\sigma^2}{n - (n_1 - n_2)^2/n},$$

so that

$$\mathcal{L}_n = (n_1 - n_2)^2/n.$$

Thus the loss in the absence of prognostic factors is zero when the design is balanced, that is when each treatment is allocated to an equal number of patients. If prognostic factors are present the loss depends not only on n_1 and n_2 but on the other elements of b which measure balance over the factors.

In the absence of prognostic factors a simple expression can also be found for the variance $d_A(j, n, z_{n+1})$ given by (10). In this case $G_n^TG_n$ is a diagonal matrix with elements n_i and

$$A^T(G_n^TG_n)^{-1}A = 1/n_1 + 1/n_2,$$

which is indeed proportional to the variance of $\hat{\alpha}_1 - \hat{\alpha}_2$. The variances for the allocation of the two treatments then reduce to

$$d_A(1, n) = n_2/(nn_1) \quad \text{and} \quad d_A(2, n) = n_1/(nn_2), \tag{15}$$

where the notation stresses that the allocation does not depend on z_n.

4.2 Allocation Rules

With three exceptions, the allocation rules considered in this paper depend on the variances $d_A(j, n, z_{n+1})$. In order to achieve balance that treatment should be allocated for which $d_A(j, n, z_{n+1})$, $j = (1, 2)$ is larger. In the absence of prognostic factors the expressions for $d_A(j, n,)$ in (15) show that treatment 1 would be allocated if $n_2 > n_1$. By extension of this result, treatment one can be described as being "under represented" if $d_A(1, n, z_{n+1}) > d_A(2, n, z_{n+1})$.

The allocation rules are expressed in terms of probabilities

$$\pi(1) = \text{ prob } [a_{n+1} = 1|\{d_A(1, n, z_{n+1}) > d_A(2, n, z_{n+1})\}],$$

that is the probability of allocating the "under-represented" treatment one.

D: Deterministic (Sequential Design Construction)

$$\pi_D(1) = 1.$$

The treatment with larger variance $d_A(j, n, z_{n+1})$ is always selected. Asymptotically, for any reasonable distribution over time of prognostic factors, the design will be balanced over the factors and there will be no loss: $\mathcal{L}_\infty = 0$.

R: Completely Randomized

$$\pi_R(1) = 0.5,$$

with $\mathcal{L}_\infty = q$, the number of nuisance parameters, including the constant. The result that randomisation over q variates causes an expected increase in variance of q goes back at least to Cox (1951).

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

E: Efron's Biased-Coin

The "under-represented" treatment is allocated with a probability greater than one half. In particular Efron (1971) elucidated the properties of the the rule

$$\pi_E(1) = 2/3,$$

although without covariates. As for the deterministic rule with covariates $\mathcal{L}_\infty = 0$. Values other than 2/3 will give a different rate of convergence to $\mathcal{L}_\infty$ and a different probability that the clinician can guess correctly which treatment will be allocated next. Some investigations of the properties of schemes with values other than 2/3 are given by Burman (1996).

A: $\mathbf{D_A}$-Optimality

With two treatments the biased-coin allocation of Atkinson (1982) according to (11) becomes

$$\pi_A(1) = \frac{d_A(1, n, z_{n+1})}{\sum_{j=1}^{2} d_A(j, n, z_{n+1})}.$$

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

C: Balanced Covariates

This rule does not depend on the variances $d_A(j, n, z_{n+1})$. The values of the $q-1$ covariates are dichotomised about their individual medians, giving 2^{q-1} possible cells in which the value of z_{n+1} could lie. The under-represented treatment in the cell indicated by z_{n+1} is then allocated, the probability being 0.5 if the numbers of the two treatments are equal. Even if the numbers receiving treatment 1 and treatment 2 are equal in all cells, the balance vector b is most unlikely to be zero, since the criterion takes no account of the observed values of the z, merely categorising them as above or below the median. A potential practical problem is that the value of the median of each covariate is assumed known. The value of b will be inflated if the median is incorrect or if the distribution of the covariates is skewed. Some numerical illustrations of the effect of the resulting inflation of b are in §4.4.

A randomised version of the rule could have a biased coin within each cell. If q is not small, the large number of cells may be sparsely filled, particularly for the non-randomized version considered here. There will then be a lack of balance over the margins of the table.

M: Minimisation - Pocock and Simon

The family of rules introduced by Pocock and Simon (1975) are concerned with marginal balance. For the ith element of the covariate vector z_{n+1} let $k(i, n+1)$ be 1 if the element is below the median and 2 if it is above. The marginal totals $m(i, k, j)$ record the number of times a patient with level k of the ith covariate is allocated treatment j. A simple, non-randomised rule exemplified by Pocock and Simon calculates the effect on the balance of allocating treatment 1 as

$$C_1 = \sum_{i=2}^{q} |m\{i, k(i, n+1), 2\} - m\{i, k(i, n+1), 1\} - 1|,$$

with the effect of allocating treatment 2 being

$$C_2 = \sum_{i=2}^{q} |m\{i, k(i, n+1), 2\} - m\{i, k(i, n+1), 1\} + 1|.$$

Treatment 1 is allocated if $C_1 < C_2$ and vice versa, with random allocation when $C_1 = C_2$.

4.3 Simulation Results

The left-hand panel of Figure 1 shows the average results of one thousand simulations of sequential designs for n up to 200. Some details of the simu-

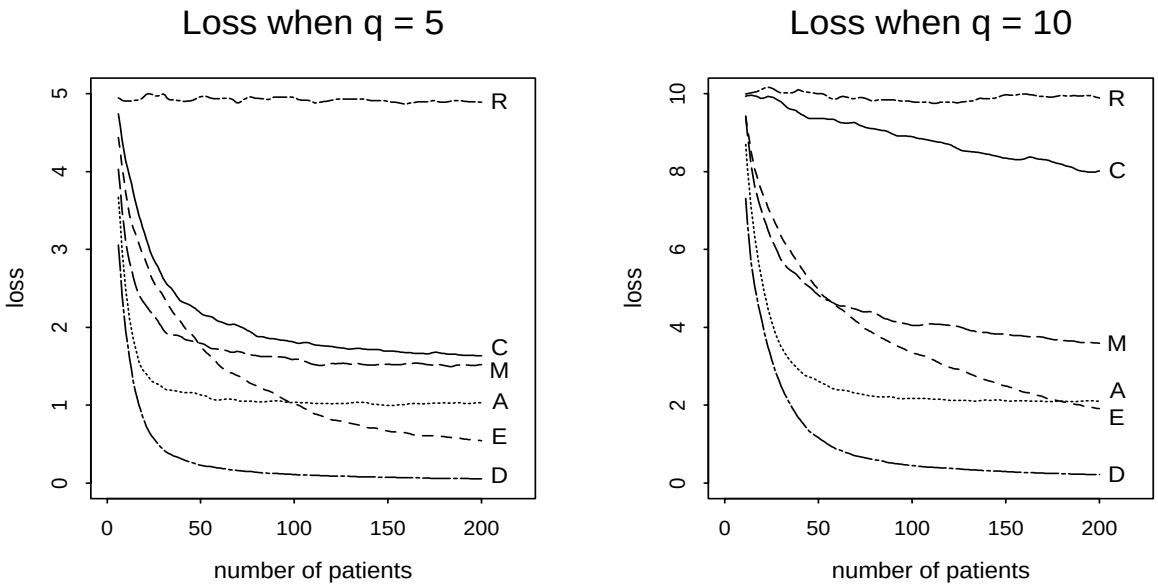

Figure 1: Loss $\mathcal{L}_n$ for six strategies for sequential allocation of treatments with $t = 2$: A, D_A-optimality; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin; M Minimisation and R, Random. Means of 1,000 simulations: (a) $q = 5$, (b) $q = 10$

lation method are given in the Appendix. There are four prognostic factors ($q = 5$) simulated independently from the standard normal distribution. The results support the theoretical conclusions above. For the deterministic allocation, D, the loss rapidly declines to almost zero, whereas for the completely random allocation, R, the value is near 5. The randomized D_A-optimum design, A, quickly achieves a loss close to $q/5 = 1$. Efron's rule, E, has a loss less than that of A for n greater than about 100, a loss which continues to decline as n increases. At 200 trials $\mathcal{L}$ for the covariate balanced strategy C is about 1.63, the second largest value. The behaviour of the minimisation rule M is similar to that of C.

The designs for $q = 10$ in the right-hand panel of Figure 1 show similar features for rules R, A and D. Efron's biased-coin strategy, E, now does not become comparable in loss with A until n is near 200 - for small n and large q the probabilities given by (11) can be much greater than 2/3, reflecting large values of the variance $d_A(j, n, z_{n+1})$ for one of the treatments. However, the most striking difference between $q = 5$ and $q = 10$ is in the behaviour of the covariate balancing designs C. With $2^9 = 512$ cells to be filled, a large amount of imbalance is possible, reflected in the slow decline of the loss. Minimisation M now shows a larger loss than A, although much less than that of C.

4.4 Non-normal Covariates

The simulations of the preceding section for normal covariates were repeated with discrete covariates and with a skewed continuous covariate distribution. The two schemes which depend on categorisation of the variables, C and M, were the most sensitive to these changes.

The covariates were again sampled independently. The discrete distribution was a Bernoulli distribution giving values of 1 and -1 with equal probabilities. The skewed distribution was a lognormal formed by exponentiating the standard normal variables of earlier simulations. For this distribution the median is e^0, that is one.

Figure 2 gives the average values of $\mathcal{L}_n$ from 1,000 simulations for $q = 5$. The results for Bernoulli covariates are in the left-hand panel, with those for the lognormal distribution on the right. This figure is to be compared with the left-hand panel of Figure 1 for normal covariates. The results for $\mathcal{L}_{200}$ for all three covariate distributions are in Table 1. The figure shows that all rules, except, E, are close to their asymptotic values. For rules R, A, E and D there is little difference between the loss at $n = 200$ for the normal distribution of covariates and that for the Bernoulli. In all four cases the skewed lognormal distribution leads to a slight increase in the loss. These

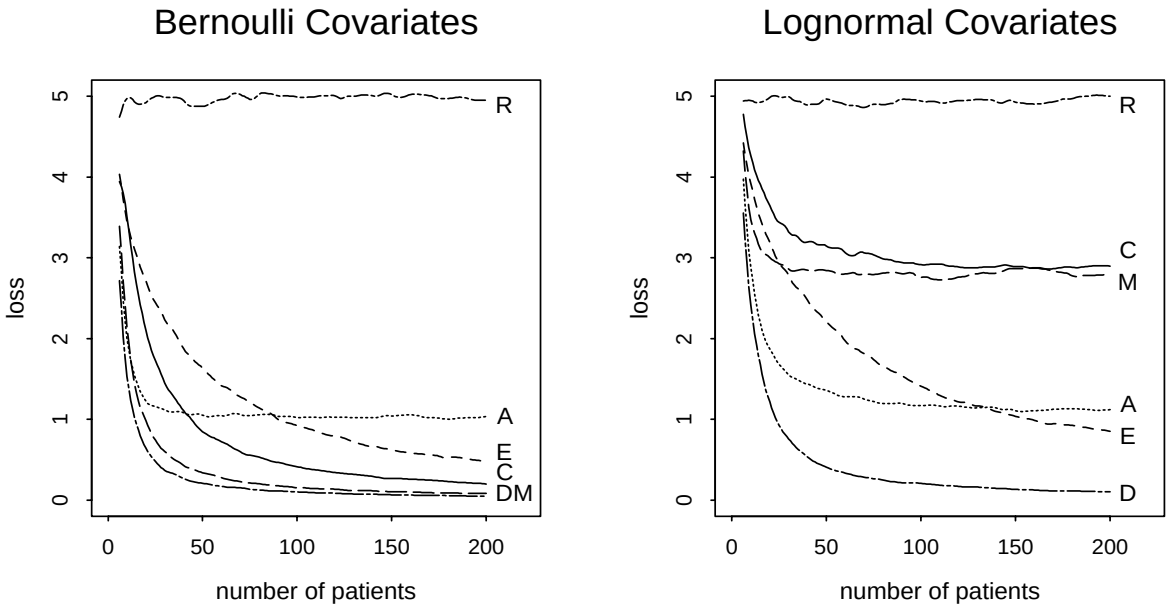

Figure 2: Effect of Covariate Distribution. Loss $\mathcal{L}_n$ for six strategies for sequential allocation of treatments with $t = 2$ and $q = 10$: (a) Bernoulli covariates and (b) lognormal covariates. A, D_A-optimality; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin; M Minimisation and R, Random. Means of 1,000 simulations

results hold for both $q = 5$ and $q = 10$. The rules are little affected by the distribution of covariates because the allocation rules, except R, respond to the imbalance vector b which is part of the analysis used in calculating loss. C and M, the two rules depending only on counting numbers of patients in categories behave very differently.

For $q = 5$ both C and M have very small losses $\mathcal{L}_{200}$ for Bernoulli variables. The schemes balance the numbers above and below the median. Since there are only two values of the covariate, this must also lead to a small value for b. For $q = 10$, both rules show reduced losses, when compared with the normal, although the loss for C is still large, because of the number of cells over which balance is required.

If the Bernoulli distribution of covariates improves the performance of

Table 1: Covariate Distribution and Loss $\mathcal{L}_{200}$

Distribution	Allocation Rule					
	R	C	M	A	E	D
	$q = 5$					
Bernoulli	4.95	0.20	0.08	1.04	0.48	0.05
Normal	4.90	1.63	1.52	1.03	0.54	0.05
Lognormal	5.01	2.90	2.80	1.12	0.85	0.10
	$q = 10$					
Bernoulli	10.05	6.95	0.46	2.08	1.87	0.20
Normal	9.89	8.01	3.60	2.09	1.91	0.21
Lognormal	10.06	8.80	6.45	2.34	2.65	0.33

rules C and M, it is severely degraded by the skewed lognormal distribution. For $q = 5$ the loss is almost doubled compared with that for the normal distribution. For $q = 10$ the losses also increase, that of M again almost doubling. The loss of C is already close to 8 so that doubling is not possible.

These results show the sensitivity of rules C and M to the distribution of the covariates. They respond well to variables which really are dichotomised and badly to skew ones. In addition, these rules depend on knowing the population median. If it is incorrect, their performance can be degraded. For example, for standard normal covariates with $q = 5$, suppose that the median was thought to be one rather than zero. Then $\mathcal{L}_{200}$ for both C and M is 3.93. This large value is moving towards that for random allocation since, with a very incorrect median, the allocation virtually ignores the value of the covariates.

4.5 Extensions

4.5.1 Correlated Covariates

In the preceding simulations it was assumed that the prognostic factors were independent. However measurements of these variables on patients are likely to be correlated. The simulations for normal variables were therefore repeated for variables with a correlation of 0.8, but still with a variance of one. The results for $n = 200$ are in Table 2, together with the uncorrelated results

Table 2: Covariate Correlation and Loss $\mathcal{L}_{200}$

	$q = 5$					
	Allocation Rule					
	R	C	M	A	E	D
Independent	4.90	1.63	1.52	1.03	0.54	0.05
Correlated	5.04	2.43	2.35	1.00	0.52	0.05
	$q = 10$					
Independent	9.89	8.01	3.60	2.09	1.91	0.21
Correlated	10.06	7.78	6.44	2.14	1.98	0.21

for comparison.

The effect of correlation is negligible on rules D, A and E. There is a slight increase in the loss for random allocation, both for $q = 5$ and $q = 10$. The main effects are again on the rules relying on dichotomisation, C and M, for both of which the loss increases, except for C when $q = 10$. However Atkinson (1999b) shows that if the simulations are extended to larger values of n rule C is also adversely affected by correlation when $q = 10$. The plot of $\mathcal{L}_n$ also shows that, even by $n = 2,000$ the loss for rule C when $q = 10$ has not reached its asymptotic value, but is still slowly decreasing.

4.5.2 Three Treatments

It is straightforward to extend four of the two-treatment allocation rules to three or more treatments. In random allocation, R, each treatment has a probability of $1/t$ of being selected. Procedures D, E and A require the variances (10). For D the treatment with largest variance is allocated; A requires the probabilities (11). For two treatments Efron's biased coin allocated treatments with odds in the ratio 1:2. One way of extending this to t treatments is to take the odds in the ratio 1:2: $\ldots$:t after ordering by the values of $d_A(j, n, z_{n+1})$. Thus, for three treatments, the probabilities are 1/6, 1/3 and 1/2. For covariate balance C there are $t2^{q-1}$ cells in which the treatment allocations are counted. In the extension of the minimization rule M the differences in the formulae for two treatments are replaced by searching over a range of values. The details are in Pocock and Simon (1975).

Simulation results of the losses (7) for $t = 3$ and $q = 5$ and 10 are given

by Atkinson (1999b). The quantitative behaviour of the procedures is close to that when $t = 2$, the major difference being an increase in the loss of A when $q = 10$. As a result A and C have similar losses for both values of q.

5 A Bayesian Biased-Coin

5.1 Theory

The rules of the previous section were derived with the idea of reducing variance, most explicitly for rule A, derived from optimum design theory. In order to reduce bias, some randomisation was introduced, but in a fundamentally *ad hoc* manner. The criteria do not explicitly include the balance between variance and bias. To include both aspects of the problem in a single criterion Ball et al. (1993) suggest that the probabilities of treatment selection $\pi_B(j|z_{n+1})$ be chosen to maximize a utility which combines both the variance of parameter estimates and randomness. It is helpful to write this utility as

$$\begin{aligned} U &= U_V - \gamma U_R \\ &= \sum_{j=1}^{t} \pi_B(j|z_{n+1})\phi(M_{j,n+1}) - \gamma \left\{ \sum_{j=1}^{t} \pi_B(j|z_{n+1}) \log \pi_B(j|z_{n+1}) \right\}, \end{aligned} \quad (16)$$

where the contribution of U_V is to provide estimates with low variance, whereas U_R contributes randomness. The parameter γ provides a balance between these two desiderata.

In U_V in (16)

$$M_{j,n+1} = G_{j,n+1}^T G_{j,n+1},$$

the information matrix if treatment j were allocated to the $(n+1)$st patient. Both parts of the utility are functions of the probabilities $\pi_B(j|z_{n+1})$. The utility U_V is maximised by putting the probability equal to one for that treatment for which $\phi(M_{j,n+1})$ is a maximum. Thus when $\gamma = 0$ in (16) the sequential allocation is non-random, the next treatment being chosen to maximize $\phi(M_{j,n+1})$. The function $\phi(M_{j,n+1})$ can be chosen to reflect the purpose of the experiment. For D-optimality

$$\phi(M_{j,n+1}) = \log |G_{j,n+1}^T G_{j,n+1}| \quad (17)$$

and for the D_A-optimality of earlier sections

$$\phi(M_{j,n+1}) = -\log |A^T (G_{j,n+1}^T G_{j,n+1})^{-1} A| \quad (18)$$

so that the sequential construction of the optimum designs is recovered.

The second part of the utility function, $-U_R$, provides randomness and is maximised by equalising the probabilities of allocating the individual treatments. Thus as $\gamma \to \infty$, the procedure tends towards the random allocation rule of §4. These two values of γ thus provide procedures which respectively minimize variance by maximising balance and minimize potential bias by maximising randomness. It is however not clear what are the properties of the procedure for any intermediate value of γ. In order to calibrate this Bayesian family against the non-Bayesian procedures of earlier sections, simulation is used to find the loss $\mathcal{L}_n$ for a range of values of γ. But first the criterion can be informatively re-written in terms of variances.

Differentiation of (16) using a Lagrange multiplier to ensure that the probabilities sum to unity, shows that the optimum treatment allocation probabilities are given by

$$\pi_B(j|z_{n+1}) = \exp\{\phi(M_{j,n+1})/\gamma\} / \sum_{k=1}^{t} \exp\{\phi(M_{k,n+1})/\gamma\}. \tag{19}$$

For D-optimality (17) the probabilities are therefore given by

$$\pi_B(j|z_{n+1}) = |G_{j,n+1}^T G_{j,n+1}|^{1/\gamma} / \sum_{k=1}^{t} |G_{k,n+1}^T G_{k,n+1}|^{1/\gamma}, \tag{20}$$

with a similar form for D_A-optimality. A simpler and more informative form of these allocation probabilities is found by using (9), the relationship between variance and determinant that leads to the iterative construction of optimum designs. The allocation probabilities (20) then become

$$\pi_B(j|z_{n+1}) = \{1 + d(j, n, z_{n+1})\}^{1/\gamma} / \sum_{k=1}^{t} \{1 + d(k, n, z_{n+1})\}^{1/\gamma}. \tag{21}$$

More importantly, for the comparisons of the present paper, similar results hold for D_A-optimality so that

$$\pi_B(j|z_{n+1}) = \frac{\{1 + d_A(j, n, z_{n+1})\}^{1/\gamma}}{\sum_{k=1}^{t} \{1 + d_A(k, n, z_{n+1})\}^{1/\gamma}}. \tag{22}$$

The Bayesian criterion is thus one generalization of the biased-coin D_A-optimum design algorithm (11) with $\psi(u) = (1 + u)^{1/\gamma}$.

Comparison of (22) with (11) is informative about the asymptotic behaviour of the Bayesian procedure. In both formulae the variance $d_A(.)$ has not been normalized: for a balanced design it will have the value $(t - 1)/n$.

In (11) multiplication of the variances by n will leave the criterion unchanged and will recover the original formulation of Atkinson (1982). But the Bayesian criterion (22) would be affected by this multiplication. The probabilities in (11) converge to values which, as the simulations show, give an asymptotic loss of $q/5$. But those in (22) converge to the value $1/t$, leading to random allocation and a loss of q as $n \to \infty$.

5.2 Simulation Results

Simulations support the interpretation that the Bayesian rules converge to random allocation for large n. The left-hand panel of Figure 3 is the average loss for 10,000 simulations when $q = 5$ and $t = 2$. The values of γ were chosen to give a wide distribution of values of $\mathcal{L}_{200}$. As the theoretical considerations above suggested, the Bayesian procedure gives a loss which behaves in a different way from those of the other five procedures. Initially the design responds strongly to the values of $d_A(.)$, particularly for small γ. The criterion is then almost that for sequential design construction and the loss decreases rapidly. But, as n increases, the allocation becomes increasingly random and the loss begins to rise towards q. For given values of the variances $d_A(.)$ decreasing the value of γ makes the allocation probabilities less uniform and so decreases the loss. For $\gamma = 0.03$ $\mathcal{L}_{200} = 2.46$.

Simulations for $q = 10$ and, again $t = 2$, are in the right-hand panel of Figure 3. Not only are the values of γ the same in the two panels, but the values of the losses for $q = 10$ are very close to twice those for $q = 5$, suggesting an invariance. Now for $\gamma = 0.03$ $\mathcal{L}_{200} = 4.96$. In fact it is hard to tell apart the right hand halves of the two figures when allowance is made for the doubling of the loss with $q = 10$. However the lower initial losses for $q = 5$ compared with those for $q = 10$ show that the designs with fewer prognostic factors initially achieve the greater degree of balance.

The results of the simulations for $t = 3$ are summarized by values of $\mathcal{L}_{200}$ in Table 3. Plots of the losses are given by Atkinson (1999a). What is most remarkable about these results is their stability to the change in the number of treatments. The effect of increasing the number of treatments from 2 to 3 is less for the Bayesian coin than for several of the procedures of §4.2 and their three treatment generalisations in Atkinson (1999b). The effect is largest for $\gamma = 1$. Again the values for $q = 10$ are roughly twice those for $q = 5$. Apart from the doubling of the loss with increased q, the four plots are practically identical.

The simulations here indicate the loss as a function of γ over the range of t and q likely to be of practical importance. The adjustable parameter γ makes it possible to provide a procedure with specified loss for a particular

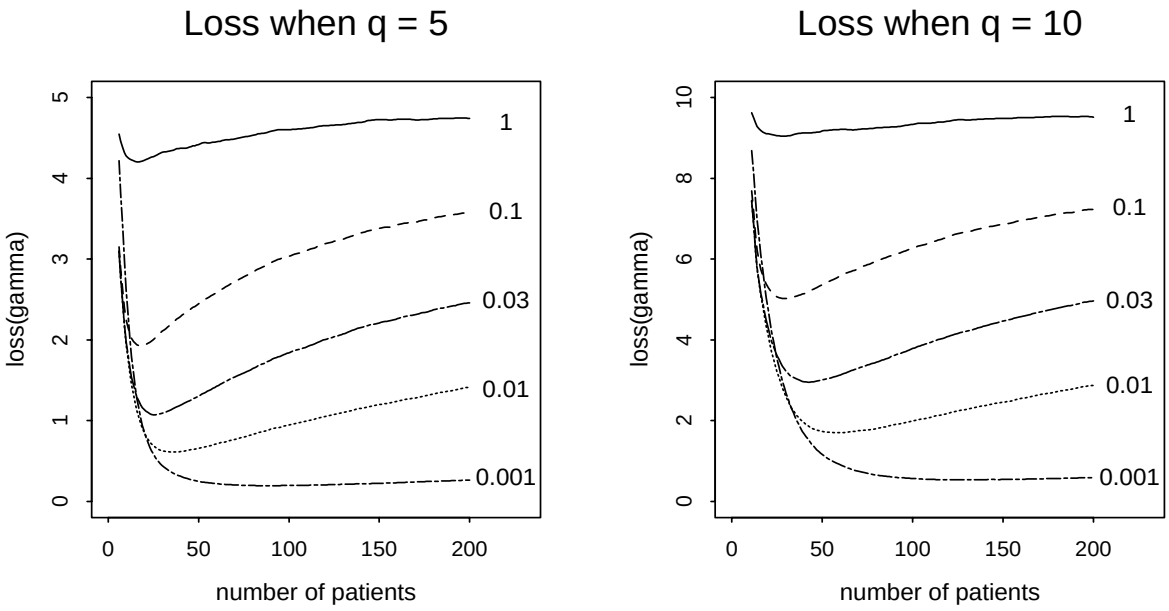

Figure 3: Loss $\mathcal{L}_n$ for five Bayesian strategies for sequential allocation of treatments with $t = 2$ and $\gamma = 1$, 0.1, 0.03, 0.01 and 0.001. Means of 10,000 simulations: (a) $q = 5$, (b) $q = 10$

Table 3: The Loss $\mathcal{L}_{200}$ for Bayesian Biased Coins

			q = 5		
γ	1.0	0.1	0.03	0.01	0.001
$t = 2$	4.739	3.573	2.456	1.414	0.264
$t = 3$	5.079	3.744	2.539	1.444	
			q = 10		
$t = 2$	9.517	7.229	4.960	2.872	0.589
$t = 3$	9.799	7.280	4.978	2.878	

value of n. However the general shape of the figures differs from those for the non-Bayesian procedures, indicating that, after an initial decline, the loss increases with n, rather than converging to an asymptote less than q.

5.3 Proportional Efficiency

Although the loss for the Bayesian procedures increases with n to a value of q, the variance of the estimated treatment difference decreases smoothly with n. To emphasize this point it is informative to look at the proportional efficiency, which, from (14) can be written

$$\mathcal{P}_n = \frac{n - \mathcal{L}_n}{n} = \frac{\sigma^2}{n \text{ var } \hat{\Delta}}. \tag{23}$$

For random allocation $\mathcal{P}_n = 1 - q/n$, which is a lower bound for all rules considered in this paper. Figure 4 shows the values of $\mathcal{L}_n$ from Figure 3 re-expressed as proportional efficiencies. In Figure 3 the rule for $\gamma = 1$ was shown to give losses close to those for random allocation. It is therefore this value of γ which provides the lowest simulated proportional efficiencies, plotted as a continuous line in the figure. The dotted line just below gives the theoretical values $1 - 5/n$ for random allocation. All rules considered, not just the Bayesian rules, give proportional efficiencies between this curve and one. So, for example, all rules are at least 95% efficient when $n = 100$ and $q = 5$. The indication is that the effect of the allocation rule on loss will only be important for small trials.

6 The Distribution of Loss

The results presented so far for the loss $\mathcal{L}_n$ are all for average values. The plots show smoothed averages over 1,000 trials. But, even so, the curves, especially those for R and C, show evidence of random fluctuation. However, in conducting a trial, it is the behaviour of the one design that is used that is of interest, rather than the average behaviour of the designs produced by a particular randomisation rule, a point stressed by Cox (1982a).

Figure 5 gives boxplots for two allocation rules, E and R, of the results for individual trials from the 1,000 used in calculating the plots of average loss. The distributions are summarized for 8 values of n from 25 to 200. The vertical scales in the two plots are not the same.

The most extreme losses occur for some trials using rule R with a maximum over all trials close to 25. The distribution of values for this random allocation in the left-hand panel of Figure 5 shows no discernible trend with

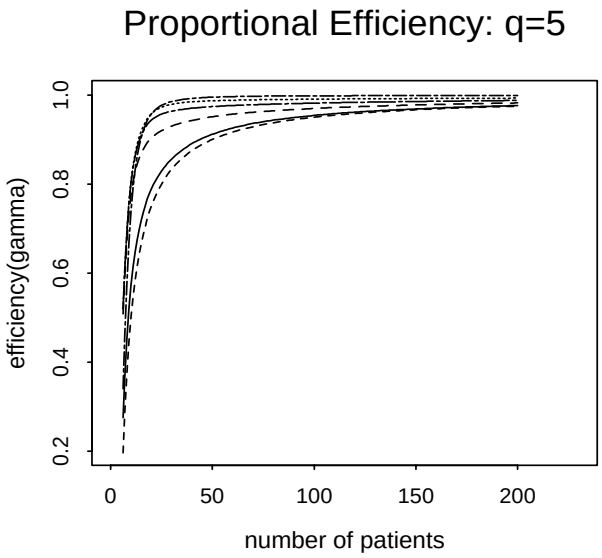

Figure 4: Proportional Efficiency $\mathcal{P}_n$ for five Bayesian strategies for sequential allocation of treatments with $t = 2$, $q = 5$ and $\gamma = 1$, 0.1, 0.03, 0.01 and 0.001. Means of 10,000 simulations. The bottom dashed line is $1 - 5/n$, the efficiency for random allocation, which is immediately below the curve for $\gamma = 1$.

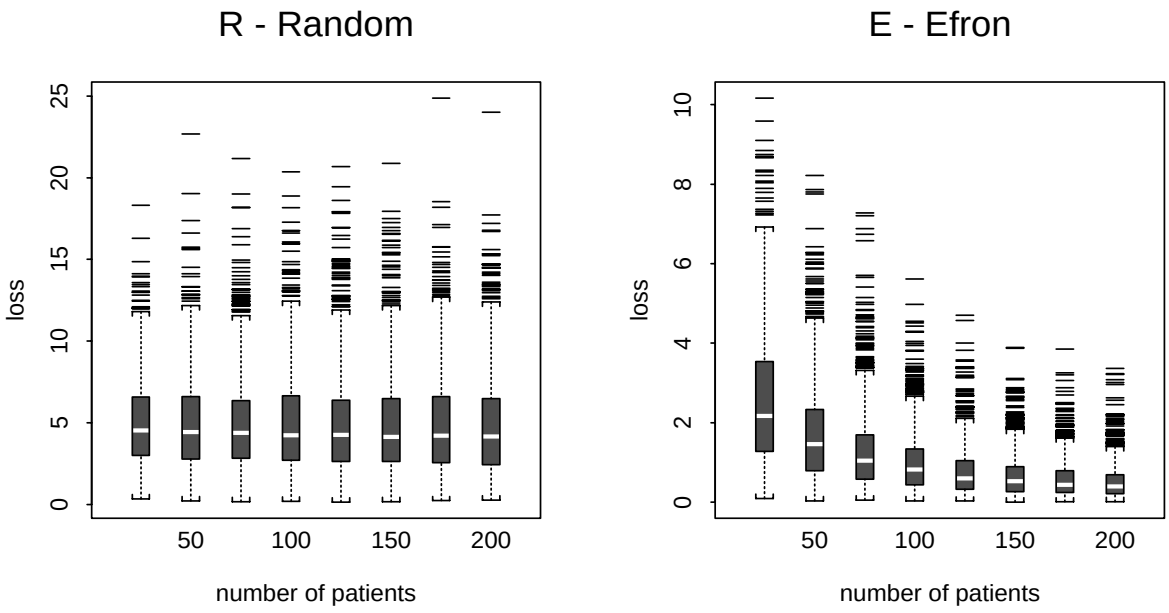

Figure 5: Boxplots of the distribution of loss $\mathcal{L}_n$ for (a) R, Random Allocation and (b) E, Efron's Biased Coin. 1,000 simulations: $q = 5$

n, whereas those for the Bayesian coin with $\gamma = 0.1$ (not shown here) increase as the allocation becomes more random for larger n. Intermediate values arise from rule C. As would be expected from the plot of $\mathcal{L}_n$ in the left-hand panel of Figure 1, the distribution initially decreases in mean and variance, but then becomes nearly constant. The results for M are similar and again are not shown. Efron's rule, in the right-hand panel of Figure 5, begins with a spread of values similar to those from M, but the variability of loss decreases with increasing n. For the D_A-optimum rule the distribution changes little after $n = 25$. The maximum loss is less than five, whereas for the deterministic allocation the distribution of loss rapidly becomes close to zero with small variability.

There is a surprisingly constant structure over all these simulations. The distributions of loss are, of course, skew. In all examples, the upper boundary of the central part of the distribution lies below $3\mathcal{L}_n$ and no trial is so unbalanced as to produce a loss greater than $5\mathcal{L}_n$. Selection of a rule can thus be made in the knowledge of the variability in loss that is to be expected.

7 Selection Bias

A general and unsurprising conclusion from the results so far is that the most random allocation rule, R, results in an appreciably greater loss than less random rules such as D. It is generally agreed that there should be some random component in the allocation of treatments and the requirement of randomisation is customarily included in protocols for clinical trials. But, in order to compare the effects of the differing amounts of randomness in the allocation rules considered here, it is essential to be able to model the effect of randomisation as it has been possible to do for balance. The two contrasting aspects can then be traded off, one against the other.

The discussion of the background to clinical trials in §2 mentioned that one use of randomisation in the design of agricultural or industrial experiments is to guard against factors that have been omitted. If recruitment to the trial takes place over an appreciable time period, randomisation of treatments will help to guard against secular trends in the population's health or in the virulence of a disease. Smith (1984b) considers randomisation against smooth trends and correlated errors as well as "selection bias", which arises from the ability of the experimenter to guess the next treatment to be allocated.

Selection bias in this context was introduced by Blackwell and Hodges (1957) who considered an artificial example in which the number of patients was known in advance, exact balance was required and there were no prog-

nostic factors. More recent authors such as Efron and Smith also calculated this bias for a number of schemes. Of course, in a double-blind trial the clinician is not able to guess the next treatment to be allocated and so is unable to influence which patient receives which treatment. This is even more so in the case of a multi-centre trial in which treatments are allocated centrally. Selection bias should therefore be considered as a calculable surrogate for all the reasons for which randomness is required. A design with the kind of balance provided by the rules considered here with, in addition, a low selection bias should behave well in the presence both of smooth trends and of short-range cyclical patterns.

To begin the calculation of bias let

$$\mathcal{W}_n = \{E(\text{number of correct guesses}) - E(\text{number of incorrect guesses})\}/n. \tag{24}$$

With two treatments and the deterministic allocation D, the treatment is always guessed correctly and $\mathcal{W}_n = 1$. The other extreme is random allocation, R, when , again for $t = 2$, the number of correct and incorrect guesses is equal, so that $\mathcal{W}_n = 0$ and the procedure is unbiased.

To extend the definition to more than two treatments needs some scaling. For rule D, $\mathcal{W}_n = 1$, as we require. But, with random allocation, the probability of a correct guess is $1/t$, so that $\mathcal{W}_n = (2-t)/t$. To obtain a value of zero for the bias of this unbiased design, introduce the bias $\mathcal{B}_n$, defined as

$$\mathcal{B}_n = \{t(\mathcal{W}_n + 1) - 2\}/\{2(t-1)\}. \tag{25}$$

This quantity in addition has the correct value of one for the deterministic allocation and reduces to $\mathcal{W}_n$ when $t = 2$. In this section values of $\mathcal{B}_n$ are found for the rules of §§4 and 5. In the next section information on $\mathcal{B}_n$ is combined with that on $\mathcal{L}_n$ to help in the choice of allocation rules.

The bias for some further allocation rules for $t = 2$ can also be found explicitly. For Efron's rule the under-allocated treatment is selected with probability 2/3, so $\mathcal{B}_n = 1/3$. For the covariate balance rule C, once initial conditions have worn off, the number of patients receiving the two treatments, for a particular combination of prognostic factors, is either equal or differs by one. If it is equal, the next allocation is made at random, and the numbers then differ by one. If they are not equal, the treatment is made deterministically (and so can be guessed) and balance restored. The probability of being either balanced or unbalanced is thus one half and $\mathcal{B}_n = 0.5$.

The other rules, including the Bayesian family, require simulation to establish their properties. We find the average values, over many simulations, of indicator variables which are one if the allocation is correctly guessed and

minus one otherwise. The guess is always the sensible one that the allocation is of the treatment with the highest probability of being allocated, with random choice if the two treatments have equal probability.

The left-hand panel of Figure 6 shows the smoothed averages of the biases from 100,000 simulations for the six non-Bayesian rules when $q = 5$. Such a large number of simulations is necessary because of the binary nature of the variable being averaged. Rules which have a low bias are highly randomised and so may be expected to have a high loss. Figure 6 therefore needs to be interpreted alongside Figure 1. The biases for the rules D and R are respectively 0 and 1, the reverse of the ordering by loss. The pair of rules A and E also reverse between the two figures - for n greater than 100, A has a higher loss than E, but a lower bias. The balancing and minimisation rules C and M are both above A and E in both figures, except at the very beginning of the plots. This indicates that bias and loss can both be reduced by using A or E rather than C or M. A feature of the bias from the minimisation rule that does not show on the smoothed plot is that the bias alternates between 0.78 when n is odd and 0.85 when n is even. The lower bias occurs when a greater proportion of the allocations are made at random, which is when there is equality between the criteria for allocation of the two treatments. Numerical experience shows that this is slightly more likely to occur when n is odd.

The right-hand panel of Figure 6 shows the same plot for $q = 10$. The extreme rules R and D behave as before with biases of zero and one. Efron's rule again has a bias of 1/3, whereas Atkinson's rule has a slightly higher bias for $q = 10$ than it did for $q = 5$. The same is true for M, which continues with a high bias, now around 0.9 rather than 0.8. Simulations over a range of values of q indicate that $(q-1)/q$ as a rough approximation to the bias for this rule. The major change in going from $q = 5$ to $q = 10$ is for rule C. With a large number of empty cells this starts off as virtually a random rule and so has low bias, combined with high loss. But, as n grows, some of the cells contain one patient, and so the allocation is known. The figure shows the bias increasing with n, although at $n = 200$ it is still far from the asymptotic value of 0.5. Figure 1 shows how the loss for this rule decreases with n.

As Figure 7 shows, the plots for the Bayesian rules have a rather different shape. The left-hand panel is for $q = 5$. Initially, for any γ, the rules are virtually deterministic, forcing balance and a low loss at the cost of bias. As n increases, the rules move toward random allocation with a decrease in bias. The bias is seen to be highest for small γ. The plot for $q = 10$ in the right-hand panel of Figure 7 is similar, although the biases are slightly higher than those for $q = 5$. Increasing q has more effect on the structure of the

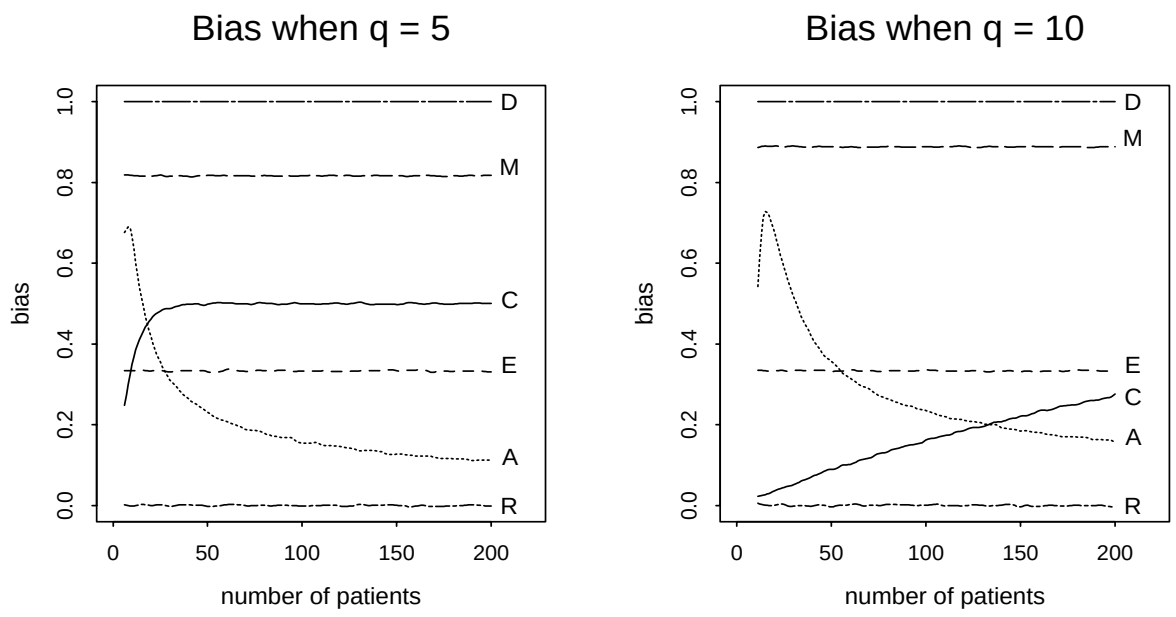

Figure 6: Bias $\mathcal{B}_n$ for six strategies for sequential allocation of treatments with $t = 2$: A, D_A-optimality; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin; M Minimisation and R, Random. Means of 100,000 simulations: (a) $q = 5$, (b) $q = 10$

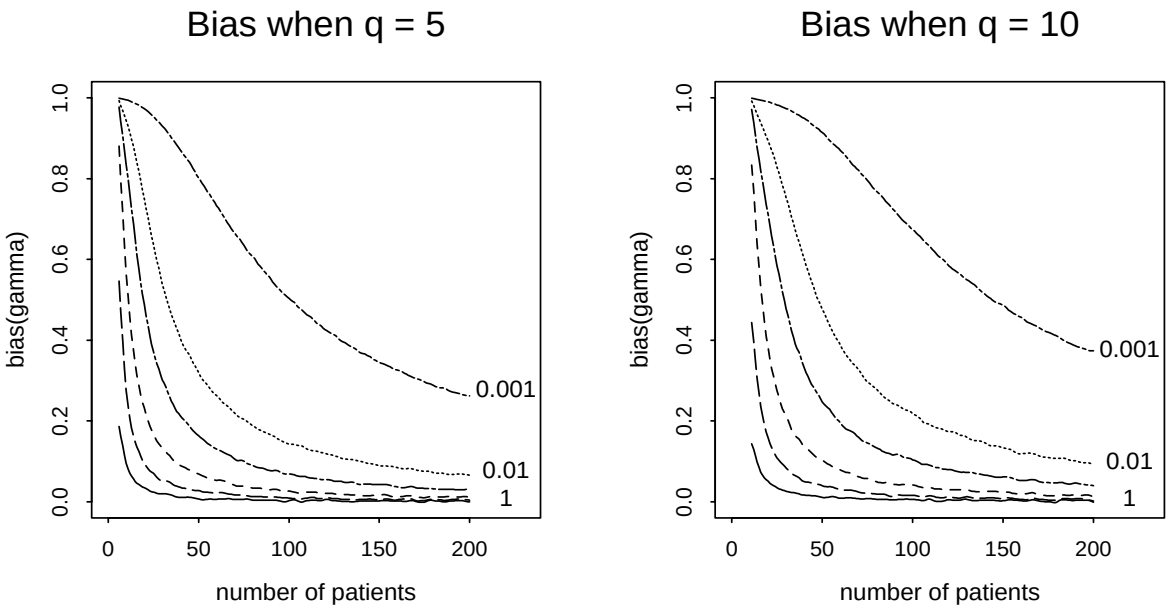

Figure 7: Bias $\mathcal{B}_n$ for six Bayesian strategies for sequential allocation of treatments with $t = 2$ and $\gamma = 0.001$, 0.01, 0.03, 0.1, 0.3 and 1. Means of 100,000 simulations: (a) $q = 5$, (b) $q = 10$

plots of $\mathcal{B}_n$ than it does on those of $\mathcal{L}_n$, once allowance has been made for the approximate doubling of the values of loss.

8 Admissibility

The preceding plots illustrate the separate behaviour of bias and loss for the various allocation rules. However, to compare rules, both loss and bias need to be considered. The plot of loss against bias for n up to 200 for each rule yields a curve in the bias-loss space. A good rule will have low loss and low bias and so will be in the lower left-hand corner of such a plot. The plots for all rules will move downwards, or to the left, as long as bias or loss continues to decrease with n. If, for a particular n, one rule plots below and to the left of another rule, then the second rule is not admissible by this criterion, since the first rule has both lower loss and lower bias than the first for the same sample size.

The plots for the Bayesian rules for six values of γ are in Figure 8. The left-hand panel shows the curves for $q = 5$. To aid interpretation, symbols have been plotted on the curves for $n = 20$, 50 and 200. The rules start deterministically, and so have high biases. These decrease as n increases. Finally, when the biases are low, the losses start to increase as the allocations become more nearly random. The results for $q = 10$ in the right-hand panel are similar, but both losses and biases are higher for all values of n. The choice of γ, for a particular value of q, yields a rule which lies at a specific point between random and deterministic allocation. What are the exact values of loss and bias depends both on q and on n.

Finally Figures 9 and 10 combine information on the Bayesian rules with that on the six non-Bayesian rules. For each of the three values of n (20, 50 and 200) marked in Figure 8, the bias and loss of the Bayesian rules have been connected, to give an approximation to the envelope that would be achieved for the particular n by considering a range of values of γ. As we shall see, these envelopes are curves of best performance: for a given n no other rules plot below and to the left of them, although some seem to be as good and therefore are admissible.

The plots for all twelve rules for $q = 5$ are in Figure 9. As is clear from Figure 6, four of the non-Bayesian rules have constant bias. They therefore plot as straight lines in Figure 9, reduced to a point for random allocation R. Since R has the lowest possible bias, zero, and D the lowest possible loss, again zero, both R and D are admissible. Comparison of E and C at the three chosen values of n shows that E dominates C, as A does M. The relationship between A and E is more complicated. The plots of loss in Figure 1 and of

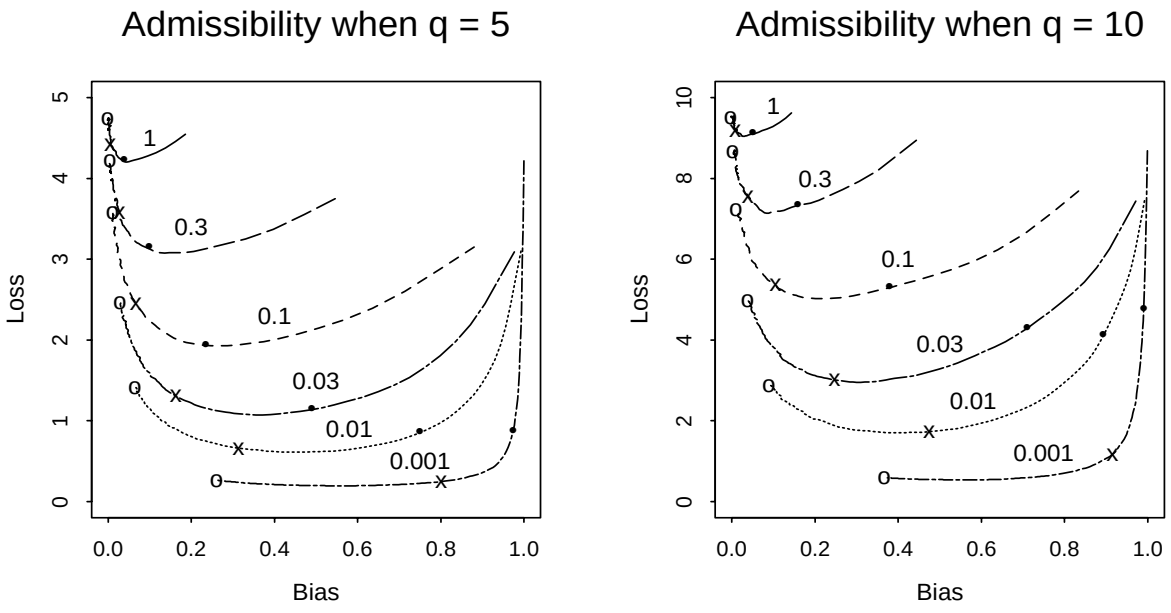

Figure 8: Bias $\mathcal{B}_n$ and loss $\mathcal{L}_n$ for six Bayesian strategies for sequential allocation of treatments with $t = 2$ and $\gamma = 0.001$, 0.01, 0.03, 0.1, 0.3 and 1. Means of 100,000 simulations: •, $n = 20$; X, $n = 50$; O, $n = 200$: (a) $q = 5$, (b) $q = 10$

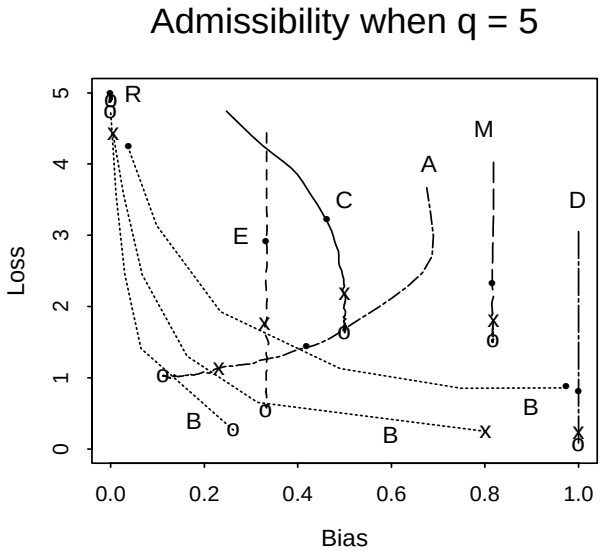

Figure 9: Admissibility: bias $\mathcal{B}_n$ and loss $\mathcal{L}_n$ for twelve strategies for sequential allocation of treatments with $t = 2$ and $q = 5$: A, D_A-optimality; B, Bayesian rules for specific n; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin; M Minimisation and R, Random. Means of 100,000 simulations: •, $n = 20$; X, $n = 50$; O, $n = 200$

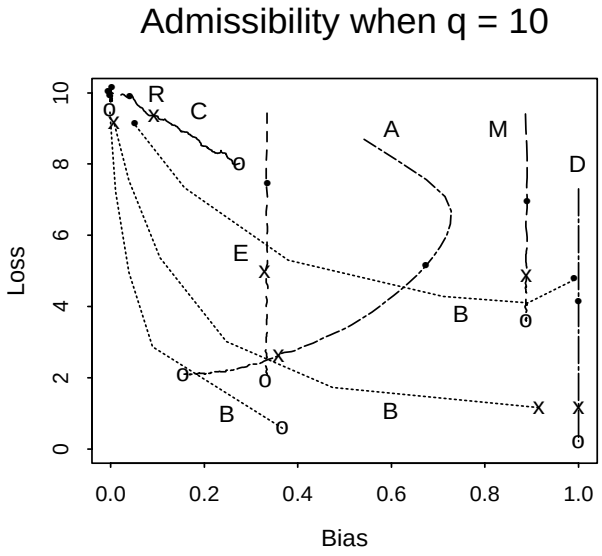

Figure 10: Admissibility: bias $\mathcal{B}_n$ and loss $\mathcal{L}_n$ for twelve strategies for sequential allocation of treatments with $t = 2$ and $q = 10$: A, D_A-optimality; B, Bayesian rules for specific n; C, Covariate Balance; D, Deterministic; E, Efron's Biased Coin; M Minimisation and R, Random. Means of 100,000 simulations: •, $n = 20$; X, $n = 50$; O, $n = 200$

bias in Figure 6 show that these curves cross each other for different values of n. Figure 9 shows that A dominates E for $n = 50$, but not for the other two values considered. For large n the bias of A decreases, with the loss remaining constant, whereas for E the loss decreases, but the bias remains fixed at 1/3. Neither dominates the other. The most striking feature is that, even for 200 patients, C and M lie well inside the envelope for $n = 20$ [sic]. Efron's biased-coin rule E lies inside the envelope for each n and so is dominated by rules with a range of values of γ. Only A lies on these envelopes and so is as good as some Bayesian rule.

The results in Figure 10 for $q = 10$ are similar to those in Figure 9, apart from the curve for C, which initially behaves almost like random allocation. However, by the time $n = 200$, it is dominated by A. The minimisation rule M is again dominated by A. Once more, A and E are different, but neither dominates the other. A general conclusion from this and the previous figure is that the two rules, C and M, which depend on balance achieved by dichotomising the prognostic factors, waste resources compared with rules which use the actual values of the factors.

The structure of Figure 10 is similar to that of Figure 9, although the bias is generally higher for all rules and the loss more than twice as much. Only rules A and E follow identical trajectories in the two plots, in the case of A after the first few trials and after the loss for $q = 5$ has been doubled. Again C and M are well inside the Bayesian envelopes, E is slightly inside and A, apart from $n = 20$, is admissible.

9 Discussion

Figures 9 and 10 provide a powerful summary of the properties of rules for sequential treatment allocation. However the figures should not be over-interpreted. Because, in Figure 9, the line for rule M for all n up to 200 M lies on the inadmissible side of the Bayesian curve for $n = 20$, this does not mean that less information is obtained by using rule M for $n = 200$ than by using a Bayesian rule with $n = 20$. The comparisons are of loss of information, that is increase in variance (14). The loss when $n = 200$ for M is 1.52, which would seem to be a small loss compared with the best that can be achieved, which is near to zero for this bias.

The discussion of the plots of proportional efficiency in Figure 4 indicated that differences in loss between the various rules may be relatively unimportant for large trials. However the bias is not so affected by changes in sample size and it may be that differences in selection bias should be at least as important in selecting a rule as are differences in loss. Even if with a mul-

ticentre trial and central treatment allocation the opportunities to bias the allocation are negligible, randomisation does serve to make the results of the trial more objective and scientifically acceptable.

There are some other matters. One is that, with three or more treatments, there may be a set of specific contrasts which are of interest, rather than the general set given by (3). The analysis given here could be repeated for such a set. Secondly, for the Bayesian rules, selection of a particular γ causes the performance of the rule to lie on curves like those of Figure 8, the exact point on the curve depending on n. If a trial persists longer than expected, the loss might be higher than desired. A lower loss, offset by higher bias, could be achieved by decreasing the value of γ during the trial.

The emphasis in this paper has been on the comparison of designs. Little has been said about the analysis of the data, apart from the assumption that the appropriate variances to use are those from least squares. But sometimes regulatory agencies specify the form of analysis to be carried out, which may be a permutation analysis. The effect of the sequential allocation on such tests has been investigated by several authors including Cox (1982b) and Smith (1984b). The general conclusion is that difficulties with the randomisation analysis only arise with very unbalanced designs. That the subject remains controversial is indicated by the report of Ebbutt et al. (1997) on the analysis of the three treatment Caesar trial with HIV positive patients. Procedures for design and methods of inference were both questioned by the FDA, an American regulatory agency. The methods of the present paper would at least have provided a basis for the objective discussion of the properties of the design used, and of that to which it was changed.

Acknowledgements

I remain grateful to Professor Stephen Senn for his interest in this work and for stimulating suggestions and references. I am also grateful to Gillian Raab and to those referees whose positive comments led to a clarification of the argument of the paper.

References

Altman, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman and Hall.

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

Atkinson, A. C. (1999a). Bayesian and other biased-coin designs for sequential clinical trials. *Tatra Mountains Mathematical Publications 17*, 133–139.

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

Atkinson, A. C. (2001). Reply to letter to the editor (Taves 2001). *Statistics in Medicine 20*, 816–818.

Atkinson, A. C. and R. A. Bailey (2001). One hundred years of the design of experiments on and off the pages of. *Biometrika. Biometrika 88*, 53–97.

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.

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

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

Chalmers, T. C. (1990). Discussion of "Biostatistical collaboration in medical research" by J.H. Ellenberg. *Biometrcs 46*, 20–22.

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

Cox, D. R. (1982a). Randomization and concomitant variables in the design of experiments. In G. Kallianpur, J. K. Krishnaiah, and J. K. Ghosh (Eds.), *Statistics and Probability: Essays in Honor of C. R. Rao*, pp. 197–202. Amsterdam: North-Holland.

Cox, D. R. (1982b). A remark on randomization in clinical trials. *Utilitas Mathematica 21A*, 245–252.

Ebbutt, A., R. Kay, J. McNamara, and J. Engler (1997). The analysis of trials using a minimisation algorithm. In *Statisticians in the Pharmaceutical Industry Annual Conference Report, 1997*, pp. 12–15. London: PSI.

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

Efron, B. (1980). Randomizing and balancing a complicated sequential experiment. In R. J. Miller, B. Efron, B. W. Brown, and L. E. Moses (Eds.), *Biostatistics Casebook*, pp. 19–30. New York: Wiley.

Eichhorn, E., M. Domanski, H. Krause-Steinrauf, and others (2001). A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *New England J. Medicine 344*, 1659–1667.

Fentiman, L. S., R. D. Rubens, and J. L. Hayward (1983). Control of pleural effusions in patients with breast cancer. A randomized trial. *Cancer 52*, 737–739.

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

Piantadosi, S. (1997). *Clinical Tials*. New York: Wiley.

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

Roberts, K. B., N. Urdaneta, R. Vera, and others (2000). Interim results of a randomized trial of mitomycin C as an adjunct to radical radiotherapy in the treatment of locally advanced squamous-cell carcinoma of the cervix. *International Journal of Cancer 90*, 206–223.

Rosenberger, W. F. (1996). New directions in adaptive designs. *Statistical Science 11*, 137–149.

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

Ross, N. (1999). Randomised block design is more powerful than minimisation. *British Medical Journal 318*, 263.

Seibold, J. R., J. H. Korn, R. Simms, and others (2000). Recombinant human relaxin in the treatment of scleroderma - a randomized, double-blind, placebo-controlled trial. *Annals of Internal Medicine 132*, 871–879.

Senn, S. J. (1997). *Statistical Issues in Drug Development*. Chichester: Wiley.

Smith, R. L. (1984a). Properties of biased coin designs in sequential clinical trials. *The Annals of Statistics 12*, 1018—1034.

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

Steele, J. M. (1980). Efron's conjecture on vulnerability to bias in a method for balancing sequential trials. *Biometrika 67*, 503–504.

"Student" (1931). The Lanarkshire milk experiment. *Biometrika 23*, 398–406.

Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics 15*, 443–453.

Taves, D. R. (2001). Letter to the editor about Atkinson 1999b. *Statistics in Medicine 20*, 813–816.

Treasure, T. and K. D. MacRae (1998). Minimisation: the platinum standard for trials? *British Medical Journal 317*, 362–363.

Treasure, T. and K. D. MacRae (1999). Minimisation is much better than the randomised block design in certain cases. *British Medical Journal 318*, 1420.

Wynn, H. P. (1970). The sequential generation of D-optimal experimental designs. *Annals of Mathematical Statistics 41*, 1055–1064.

10 Appendix: Simulation

The properties of the allocation rules were found by simulating a series of clinical trials using one rule at a time. Suppose n patients have already been allocated their treatments. Allocation for the $n+1$st patient starts with sampling the vector of $q-1$ prognostic factors z_{n+1}. The previous history of the trial is contained in the matrix G_n, which includes the first n allocations and the n vectors of prognostic factors. The results of §3 can then be used to calculate the variances (10) if these are needed for the particular allocation rule of §4.2 being studied. The variances are used to calculate the probabilities of allocating each treatment. For the Bayesian designs, for example, these are given by (22). For rules C and M the probabilities are calculated by counting over the numbers of allocations in cells determined by the z_n. For the new patient this cell is determined by z_{n+1}. A random number using these calculated probabilities then determines which treatment is allocated.

Once the $n+1$st treatment has been allocated, the history G_n is updated and the new loss calculated, together with a one or zero depending on whether or not the treatment allocation was correctly guessed. Each simulation of a 200 trial design gives $200-q$ values of the loss and the bias. The trial is repeated 1,000 or 10,000 times for calculations of loss, depending on the rule being used, and the averages give the values of $\mathcal{L}_n$ plotted here. For calculations of bias the results of 100,000 simulations were averaged.

A small difficulty is that the designs are singular until $n = q$. Thus, for $n < q$, $G_n^T G_n$ cannot be inverted and the allocation rules cannot operate. To overcome this difficulty the simulated designs are initially regularised by the addition of a small quantity to the diagonal of $G_n^T G_n$, which can then be inverted. With two treatments interest is in estimation of the difference $\alpha_1 - \alpha_2$. The variance of this difference is unaffected by the regularisation. However, for three or more treatments, the regularisation does affect the variances and so the calculated loss. The regularisation should therefore be removed or corrected for once the design is non-singular. Numerical results on this point are summarised in Atkinson (1999b).

The simulations used a Fortran program which is available from the author, although the figures were drawn in S-Plus. However S-Plus was found to be too slow for simulation of the trials.

The application of these methods to a sequential trial requires appreciably less calculation than their comparison. Allocation of the next treatment depends on the vector of prognostic factors of the new patient. Formulae for updating the matrices in the calculation of the general variance for D_A-optimality (10) are discussed by Atkinson (1982, §6). The calculations for the Bayesian family of rules (22) follow directly from the update of (10).