Covariate-Balanced Response-Adaptive Designs for Clinical Trials with Continuous Responses that Target Allocation Proportions

Anthony C. Atkinson, Department of Statistics,
London School of Economics, London WC2A 2AE, UK,
Atanu Biswas, Applied Statistics Unit,
Indian Statistical Institute, 203 Barrackpore Trunk Road,
Kolkata 700 108, India
and
Luc Pronzato, Laboratoire I3S
CNRS/Université de Nice-Sophia Antipolis
Bât Euclide, Les Algorithmes, 2000 route des lucioles, BP 121
06903 Sophia Antipolis cedex, France

September 14, 2011

Abstract

We provide adaptive designs for the sequential allocation of treatments to patients in a clinical trial when the responses are approximately normally distributed. The skewing of the allocations towards the better treatments depends on target proportions for the ranked treatment effects. In addition we combine balance across prognostic factors with a controllable amount of randomization. The asymptotic properties of our rule are established. Examples are given of the design of two- and three-treatment trials and the importance of regularization in avoiding extreme allocations is stressed. We use simulation to illustrate the properties of our procedure which compares favorably with other rules. Redesign of an adaptive trial shows an appreciable increase in allocation of patients to the better treatment combined with a negligible loss in power.

Keywords: asymptotic normality; biased-coin design, CARAB design,

ethical allocation, limiting allocation proportion, pseudo-null distribution, selection bias, randomization, regularisation

1 Introduction

We provide adaptive designs for the sequential allocation of treatments to patients in a clinical trial with continuous responses, typically normal. Such adaptive designs are intended to provide information about treatment comparisons whilst keeping low the number of patients receiving inferior treatments. We develop designs that target specified allocation proportions for the ranked treatments and derive asymptotic properties. For efficient estimation of the treatment effects the allocation is approximately balanced over the prognostic factors and covariates of the individual patients. In order to avoid bias there is also a controllable amount of randomization in the allocation.

Early papers on adaptive design include Robbins (1952) and Zelen (1969). The emphasis thereafter was on binary responses, with continuous responses virtually excluded. The historical summary at the beginning of Hu and Zhang (2004) further indicates the extent to which prognostic factors have been ignored. The more recent references in Zhang and Rosenberger (2007) support this impression.

Books on randomization and adaptive design in clinical trials include Rosenberger and Lachin (2002) and Matthews (2006), where the emphasis again is on binary responses in the absence of prognostic factors, with designs generated from urn models. The coverage of Hu and Rosenberger (2006) is more general, as is that of Chow and Chang (2007). Gallo *et al.* (2006) stress the growing importance of adaptive designs. Rosenberger and Sverdlov (2008) discuss the handling of prognostic factors in clinical trials and Dumville *et al.* (2006) survey the use of unequal randomization ratios.

We base our designs on randomized versions of the sequential construction of optimum designs for linear models, so that we include prognostic factors in a natural way. Our model is introduced in §2.1 and applied in §2.2 to non-adaptive designs targeting specified unequal allocations of treatments whilst providing some covariate balance and randomization. Our adaptive design criterion applies these target allocations to the ranked treatments. The criterion is presented in §2.3. In §3 we use results of Lai and Wei (1982) to prove the convergence of our designs to the targets. We show that the parameter estimates have an asymptotically normal distribution with covariance matrix that for least squares. One criterion for the comparison of designs with randomization, loss, is described in §4. Section 5 stresses the importance of regularization in the avoidance of very unbalanced designs.

Since our designs are adaptive, they will vary from trial to trial due to the random nature of the responses. In §6 we find the average properties of our method by simulations of a three-treatment design. However, in any trial, only one design will be used. The clinically important properties of that individual design are explored in §7. The inferential properties of our adaptive design procedure are presented in §8 through the use of "pseudo-null" distributions.

The cost of an adaptive design in which fewer patients receive inferior treatments is that power will be lost for treatment comparisons. Accordingly, in §9 we compare our allocation rule with four others in terms of loss, selection bias, allocation proportion and power for a test of treatment differences. To illustrate the ethical imperative for the use of adaptive designs, we devote §10 to redesigning an 88 patient adaptive trial. Use of our design leads, on average, to a further 18.5 patients receiving the better treatment at a negligible cost in power. We close in §11 with comments on extension to other generalized linear models, including binomial responses.

2 An Adaptive Allocation Rule

2.1 Models

There are 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. We assume that the results of the trial, perhaps after data transformation, will be analysed using the regression model

$$E(y_i) = f_i^T \beta = h_i^T \alpha + z_i^T \theta, \tag{1}$$

with additive independent errors of constant variance. Here h_i is a vector of t indicator variables, the one non-zero element indicating which treatment the patient received. The $v \times 1$ 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.

The model (1) for n patients in matrix form is

$$E(Y^n) = F^n \beta = H^n \alpha + Z^n \theta \tag{2}$$

where Y^n is the $n \times 1$ vector of responses for the n patients. Because of the way we have parameterized the treatment effects, Z^n does not include a constant column. The effects of the variables z_i are usually not of interest in themselves, although we do want the potential to adjust for them in the analysis of the trial.

In order to reduce selection bias Atkinson (1982) introduced a randomised version of the sequential construction of optimum designs for estimation of linear combinations of the parameters in (2). It is sufficient to consider a single linear combination

$$a^T\beta = l^T\alpha + w^T\theta. \tag{3}$$

If the values of θ are not of interest, the v elements of w are zero. The elements of l are chosen to give the required frequency of allocation of the various treatments.

The variance of the estimated combination of coefficients is

$$\text{var}\,\{a^T\widehat{\beta}^n\} = \sigma^2 a^T(F^{nT}F^n)^{-1}a, \tag{4}$$

where $\widehat{\beta}^n$ is the least squares estimate of β after the results of all n patients have been analysed and σ^2 is the variance of the errors, assumed additive in (2). Minimization of this variance is central to our adaptive design of clinical trials.

We design to minimise the variance of a single linear combination of the $t+v$ parameters. There is then a space of nuisance parameters of dimension $q = t+v-1$; the v parameters θ and the space of dimension $t-1$ of linear combinations of the α that are not of interest. We see in §4 that the properties of the design depend on the value of q.

For example, when there are two treatments and $l^T = (0.5 \quad -0.5)$, the design minimizing (4) provides balance over the covariates and equal allocation to the two treatments, so that the variance of the treatment difference is minimized. In a comparative study the estimate of the mean treatment effect $(\alpha_1+\alpha_2)/2$ is not of direct interest. To generate response adaptive designs we require vectors l that lead to unequal allocation.

2.2 Skewed Allocations

To obtain skewed allocation combined with efficient parameter estimation Atkinson and Biswas (2005) introduced designs for estimation of the linear combination

$$l^T\alpha = \pm p_1\alpha_1 \mp \ldots \pm p_t\alpha_t, \tag{5}$$

where the coefficients $p_j, j = 1,\ldots,t$ are such that $0 < p_j < 1$ and $\sum p_j = 1$. It is straightforward to show that the variance of $l^T\hat{\alpha}$, in the absence of covariates, is minimized when r_j, the proportion of patients receiving treatment j, is p_j, as it is when the design is balanced across treatments. We discuss balance further in §3. The signs in (5) are a generalization to t treatments of the weights 0.5 and -0.5 that give efficient designs for the treatment difference.

In our designs the allocation weights p_j depend on the ranking of the treatments. Let the target proportion of patients receiving treatment ranked j be p_j^*. Then we require that

$$p_1^* \geq p_2^* \geq \ldots \geq p_j^* \geq \ldots \geq p_t^*, \tag{6}$$

with, to avoid uniform allocation, at least one inequality. With $p_1^* = 1$ (and all other $p_j^* = 0$) we obtain a form of play the winner rule that includes all past history. Less extreme rules have the p_j^* a decreasing function of j. For two treatments a generalization of the rule of Efron (1971) would be to have a target allocation of 2/3 for the better treatment.

Let the ranking of treatment j be $R(j)$. Then in (5) we take

$$p_j = p_{R(j)}^*. \tag{7}$$

Of course, in practice the ranking of the treatments is not known but is that of the estimated treatment effects $\hat{\alpha}_j$, so that in (7) $p_j = p_{\widehat{R}_j^n}^*$, where $\widehat{R}_j^n$ is the estimated rank of treatment j after n patients. The difference from Efron's rule is that we rank the treatments by the values of the $\hat{\alpha}$ whereas in his rule, which does not include covariates, ranking is by inverse frequency of allocation to the treatments.

2.3 Adaptive Design: Rule G

In the theory of optimum experimental design, designs that minimize the variance of linear combinations of parameter estimates are called c-optimum, a special case of D_A-optimality (see, for example, Atkinson, Donev, and Tobias, 2007, §10.2). In sequential trials the extended design matrix F^n in (2) is known. Patient $n+1$ arrives with a vector of covariates x_{n+1}, a function of which forms the last row of Z^{n+1}. In the sequential construction of c-optimum designs minimizing (4) we allocate the treatment for which

$$d_c(j, n, z_{n+1}) = \{f_{n+1,j}^T (F^{nT} F^n)^{-1} a\}^2 \quad (j = 1, \ldots, t) \tag{8}$$

is a maximum. The quantity $d_c(.)$ is related to the decrease in the variance of the estimated combination $a^T\alpha$ when treatment j is allocated to patient $n+1$ with covariate vector x_{n+1}.

There is no randomness in such an allocation rule. Atkinson and Biswas (2005) use the utility given by Ball, Smith, and Verdinelli (1993) to extend the allocation rules of Atkinson (2002) to the skewed allocation of (5). These designs introduce a balance between the variance of parameter estimation and randomness that is controlled by a single parameter γ. This rule can be generalized, using methods similar to those in §3 of Atkinson and Biswas (2005),

to maximization of the gain G_j resulting from the allocation of treatment j. We take $G_j = p^*_{\hat{R}^n j}$. The resulting generalised biased-coin rule G combines this randomness with c-optimality; the probability of allocation of treatment j for rule G is

$$\pi(j|x_{n+1}) = \frac{\{1 + d_c(j, n, z_{n+1})\}^{1/\gamma} p^*_{\hat{R}^n(j)}}{\sum_{s=1}^{t} \{1 + d_c(s, n, z_{n+1})\}^{1/\gamma} p^*_{\hat{R}^n(s)}}, \tag{9}$$

for $\gamma > 0$. In §3 we show, under mild conditions, that for all j the allocation proportion for treatment j converge to the target p^*_j. In an extension of the usage of Chapter 9 of Hu and Rosenberger (2006), this rule provides covariate-adjusted and balanced response-adaptive (CARAB) randomization procedures.

It is important that the variances $d_c(j, n, z_{n+1})$ are not normalized by n. Consequently they tend to zero as $n \to \infty$ and the effect of the parameter γ vanishes, the rule becoming random allocation with probabilities p^*_j. For small n the rule forces balance over the covariates, the effect being larger for small γ. Figure 1 shows an example for four values of γ from 0.01 to 1. Atkinson (2002) gives numerous simulation results for the effect of different values of γ for non-adaptive designs.

It is a characteristic of this scheme that the probability of allocating the treatments depends on p^* and on the ordering of the α_j, but not on the differences between them. Suppose there are two treatments. Then, if $\alpha_1 > \alpha_2$, treatment 1 will eventually be allocated in a proportion p^*_1 of the trials regardless of the value of $\Delta = \alpha_1 - \alpha_2$. Of course, if Δ is small relative to the measurement error, in many of the initial trials, $\hat{\alpha}_1 < \hat{\alpha}_2$ and it will seem that treatment 2 is better. Then some individual allocations will be skewed in favour of treatment 2 with target p^*_1, that is $p_{\hat{R}^n(2)}$. When $\hat{a}_1 > \hat{\alpha}_2$, treatment 1 will be preferred. If the trial is terminated before a clear difference between the treatments has been established, each treatment may have been be allocated to around half the patients. Under such conditions a predetermined allocation of p^*_1 far from 0.5 does little ethical harm.

The purpose of Rule G is to ensure a specified ethical gain without going through possible extreme designs even if, for a two treatment design, $\alpha_1 \gg \alpha_2$. By using ranks, we ensure both a prefixed allocation which is ethically skewed and sufficient allocation to each treatment to ensure that the design is not too inefficient for estimation of the treatment difference.

3 Asymptotic Properties

Inference for effects in adaptive designs requires care. See, for example, the discussion of the ECMO trial in Begg (1990). In our model (1) the errors are independent and we use least squares to estimate the parameters. However, the allocation depends on the earlier responses and so the observations are not independent. We use results of Lai and Wei (1982) on stochastic regression models to give an asymptotic justification for least squares and to prove the convergence of the allocation probabilities to the targets π_j^*.

The choice of $\pi(j|x_{n+1})$ in (9) guarantees that

$$\pi(j|x_{n+1}) > c \text{ for all } n, j$$

with c some positive constant. Therefore, from Borel-Cantelli arguments, the number $n_{n,j}$ of individuals having received treatment j after n allocations satisfies $n_{n,j} \to \infty$ for all j. To obtain limiting results for the constructed designs and estimators, we need to make some assumptions on the covariates z_i in (1). We suppose that they are distributed among trials independently of the treatments and that $(1/n)\sum_{i=1}^n z_i \stackrel{\text{a.s.}}{\to} \mu$, $(1/n)\sum_{i=1}^n z_i z_i^T \stackrel{\text{a.s.}}{\to} M_Z$, with the asymptotic variance-covariance matrix $M_Z - \mu\mu^T$ having full rank. Now, arranging the covariates according to treatments and denoting by z_{ij} the vector of covariates for the i-th trial with treatment j and $\mu_j = (1/n_{j,n})\sum_{i=1}^{n_{j,n}} z_{ij}$, we can write

$$\frac{1}{n}F^{nT}F^n = \begin{pmatrix} D^n & D^nH^n \\ H^{nT}D^n & M_Z^n \end{pmatrix},$$

where $D^n = \text{diag}\{n_{j,n}/n,\ j = 1, \ldots, t\}$, $H^n = (\mu_1, \ldots, \mu_t)^T$, $M_Z^n = (1/n)\sum_{i=1}^n z_i z_i^T$ and $M_Z^n \stackrel{\text{a.s.}}{\to} M_Z$, $H^n \stackrel{\text{a.s.}}{\to} u\mu^T$ when $n \to \infty$, with u the t-dimensional vector of ones. Therefore, $\det[(1/n)F^{nT}F^n] = \det(D^n)\ \det(M_Z^n - H^{nT}D^nH^n)$ with

$$M_Z^n - H^{nT}D^nH^n = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n_{j,n}}(z_{ij} - \mu_j)(z_{ij} - \mu_j)^T \stackrel{\text{a.s.}}{\to} M_Z - \mu\mu^T\,,\ n \to \infty\,.$$

Since $M_Z - \mu\mu^T$ has full rank, the sufficient conditions of Lai and Wei (1982, Corollary 3) for the strong consistency of $\hat{\beta}^n$ are satisfied: with $\lambda_{\min}$ and $\lambda_{\max}$ denoting minimum and maximum eigenvalues,

$$\begin{cases} \lambda_{\min}(F^{nT}F^n) \stackrel{\text{a.s.}}{\to} \infty \\ [\log \lambda_{\max}(F^{nT}F^n)]^{1+\rho} = o[\lambda_{\min}(F^{nT}F^n)]\,,\ \text{a.s.} \end{cases}$$

for some $\rho > 0$. In particular, $\hat{\alpha}^n \stackrel{\text{a.s.}}{\to} \alpha$, $n \to \infty$ and, if $\alpha_j < \alpha_{j+1}$ for all $j = 1, \ldots, t-1$, then there exists n_0 such that, for all $n > n_0$, we have

$\hat{R}^n_{(j)} = R_{(j)}$. The asymptotic allocation rule thus coincides with that using the true ordering or treatments and $n_{j,n}/n \to p^*_j$ for all j. Notice, moreover, that the conditions Lai and Wei (1982, Th. 3) for asymptotic normality of $\hat{\beta}^n$ are satisfied, so that

$$(F^{nT}F^n)^{1/2}(\hat{\beta}^n - \beta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2 I)\,,$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution and I is the $(t+v)$-dimensional identity matrix.

The asymptotics of our model are simpler to analyse than those for the large class of allocation rules analysed by Zhang *et al.* (2007) which includes generalized linear models. Because our rule is asymptomatically independent of the values of the z_i the information matrix is that for ordinary least squares; equation (3.4) of Zhang *et al.* (2007) rather than equation(3.3). In §8 we see how good this information matrix is for n as small as 50.

4 Loss and the Assessment of Designs

Our adaptive designs favour the best treatment. At the same time they provide estimates of the treatment parameters and some randomness in allocation. To compare designs we need small-sample measures of performance.

Let the treatments be correctly ordered. Then the linear combination of the parameters corresponding to the proportions p^*_j is $a^T_* \beta$. From (4) the variance of the estimated linear combination has a minimum value of

$$\mathrm{var}\,\{a^T_* \hat{\beta}_*\} = \sigma^2/n, \tag{10}$$

where $\hat{\beta}_*$ is the estimate from the optimum design with treatment proportions p^*_j.

For other designs we find the variance of the same linear combination from (4). Comparisons can use either the ratio of variances, that is the efficiency E_n, or the loss (Burman, 1996), calculated by Atkinson (2002) for eleven rules for unskewed treatment allocation. The efficiency of any design is then

$$E_n = 1/\left\{n a^T_* (F^{nT}F^n)^{-1} a_*\right\}. \tag{11}$$

The loss L_n is defined by writing the variance (4) as

$$\mathrm{var}\,\{a^T_* \widehat{\beta}\} = \frac{\sigma^2}{n - L_n}, \tag{12}$$

so that

$$L_n = n(1 - E_n). \tag{13}$$

With a random element in treatment allocation, the loss L_n is a random variable, the value of which depends upon the particular trial and pattern of covariates. Let $E(L_n) = \mathcal{L}_n$. For random allocation of two treatments in the unskewed case ignoring balance, $\mathcal{L}_n \to q$, the number of nuisance parameters, as n increases, a result that goes back at least to Cox (1957). Designs that force more balance have lower values of $\mathcal{L}_n$. The loss can be interpreted as the number of patients on whom information is lost due to the lack of optimality of the design.

To monitor the effectiveness of the adaptive nature of the designs we supplement study of L_n by also looking at the evolution with n of $r_{j,n}$, the proportion of patients receiving treatment j.

5 Regularized Designs

We begin the study of the properties of our allocation rule with a simulation study of a three-treatment design. We only report the results of employing regularized designs, in which we set a lower bound, depending on n, on the minimum number of allocations of each treatment. The purpose is to avoid extreme designs in which an appreciable number of patients receive poor treatments. In our analysis, such designs would have high losses. An example is the two-treatment Michigan ECMO trial in which only one out of 12 children was given the standard therapy (Bartlett *et al.*, 1985).

Regularization ensures that each treatment is allocated throughout the trial, although with a decreasing frequency if the treatments differ. For three-treatment trials we allocate three of the first nine patients to each treatment. Thereafter, if the number allocated to either treatment is below $\sqrt{n}$, the under-represented treatment is allocated when n is an integer squared. For an 800 trial design the first regularization could occur when $n = 16$, when one treatment had only been allocated three times. The last regularization would be when $n = 784 (= 28^2)$ and one treatment had been allocated 27 times. The regularization ensures that the treatment estimates and their linear combinations such as (5) at least have root n consistency (Neyman, 1959). There is nothing special about $\sqrt{n}$: we only require a bounding sequence that avoids very extreme allocations.

6 Three Treatments

This section gives the results of a small part of our numerical assessment of the properties of Rule G (9) for three treatments. To use this rule we need

to specify three values of p_j^*. If interest is solely in the best treatment, both p_2^* and p_3^* could be put equal to zero. We however assume decreasing interest in worse treatments.

To start the numerical investigation of designs we take the treatment effects as $\alpha = (6.0,\ 2.65,\ 2.0)^T$. It is therefore relatively easy to identify the best treatment, but the performance of treatments 2 and 3 is not so easily ranked. The responses are normally distributed with error standard deviation $\sigma = 1.0$; there are three covariates, independently normally distributed with zero mean and unit variance; the dimension of the space of nuisance parameters q is therefore 5. We take $p^* = (0.8,\ 0.15,\ 0.05)^T$. For numerical stability in the sequential calculation of designs we take $p_{R(1)} = p_1^*, p_{R(2)} = -p_2^*$ and $p_{R(3)} = p_3^*$; the negative sign does not affect the limiting proportions of treatments in the design.

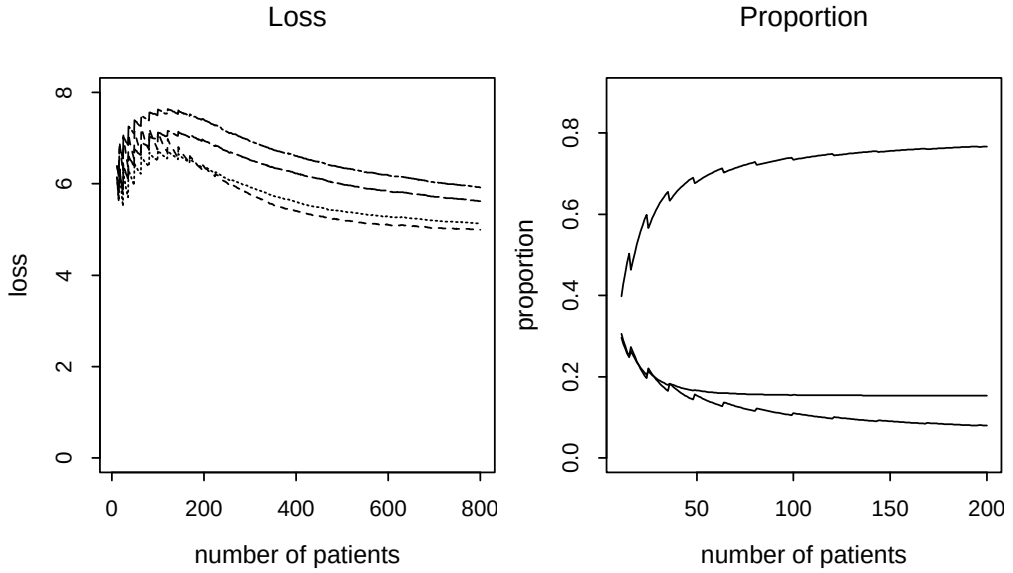

Figure 1: Rule G: regularized designs for three treatments. Left-hand panel: average losses $\bar{L}_n$ for four values of γ: reading downwards 1, 0.1, 0.03 and 0.01. Right-hand panel: average proportion $\bar{r}_{j,n}$ receiving each treatment when $\gamma = 0.01$. Averages of 10,000 simulations, $\alpha = (6.0,\ 2.65,\ 2.0)^T$, $p^* = (0.8,\ 0.15,\ 0.05)^T$, $\sigma = 1.0$, $q = 5$. The zig-zag effects are caused by regularization.

We only consider regularized designs. The left-hand panel of Figure 1 gives the average losses $\bar{L}_n$ from 10,000 simulations for sequential allocation using rule G for up to 800 patients for four values of γ: 1, 0.1, 0.03 and 0.01.

For known ordering of the treatments and large n the rule becomes random allocation with probabilities p_j^*. The asymptotic loss for this non-

adaptive design is $\mathcal{L}_\infty = q$, here 5. Small values of γ initially give rules with greater emphasis on balance and so lower loss, although the effect of γ decreases as n grows. The left-hand panel of Figure 1 shows some of this structure, although the adaptivity of the design increases the value of loss since, initially, identification of the order of the treatments is uncertain. For all values of γ the loss for rule G decreases after its maximum, around $n = 150$ because, once the best treatment has been identified this is allocated with probability $p_1^* = 0.8$; this allocation dominates in the calculation of loss.

The right-hand panel of the figure, for n up to 200 when $\gamma = 0.01$, shows the design approaching the target value of 0.8 ($\bar{r}_{1,100} = 0.735$). The proportions for treatment 2, and particularly treatment 3, approach the target values more slowly; $\bar{r}_{3,100} = 0.110$, although the target os 0.05. Particularly for treatment 3, the effect of any early over allocation due to misranking of the treatments is slow to die out when only 5% of the patients are expected to have this treatment.

The fine structure of zig-zags in the figures is caused by the operation of our regularization rule. The jumps in the lowest curve in the right-hand panel, that for $\bar{r}_{3,n}$, occur because of a forced allocation to treatment three in some designs to ensure that the allocation is at least $\sqrt{n}$ for $n = 16, 25, \ldots$. The initial values for $\bar{r}_{2,n}$ also show evidence of regularization. When the allocations $n_{3,n}$ or $n_{2,n}$ are increased by the regularization rule, the value of $n_{1,n}$ must decrease. This effect is also evident in the figure.

The left-hand panel of the figure also shows the effect of regularization, but on loss. This increases slightly at each regularization as the allocation proportions are, on average, forced slightly away from p^*.

A final point about the overall effect of regularization is that, by increasing the allocations to treatments 2 and 3, the rate of convergence of the $\bar{r}_{j,n}$ to p^* is reduced. Above, for the regularized design we had $\bar{r}_{100} = 0.735$, 0.155 and 0.110. For an unregularized design the numbers were 0.788, 0.146 and 0.066. The benefit from this slower convergence is the avoidance of extremely unbalanced designs.

7 Properties of Individual Designs

So far we have only considered the average properties of designs. But we require to use design methods which are not only good on average, but also have good properties for the one design that will actually be used in a clinical trial. We therefore look briefly at the results of 1,000 simulations of individual regularized trials that form part of the averages in Figure 1.

Figure 2 gives boxplots showing the distribution of loss and allocation

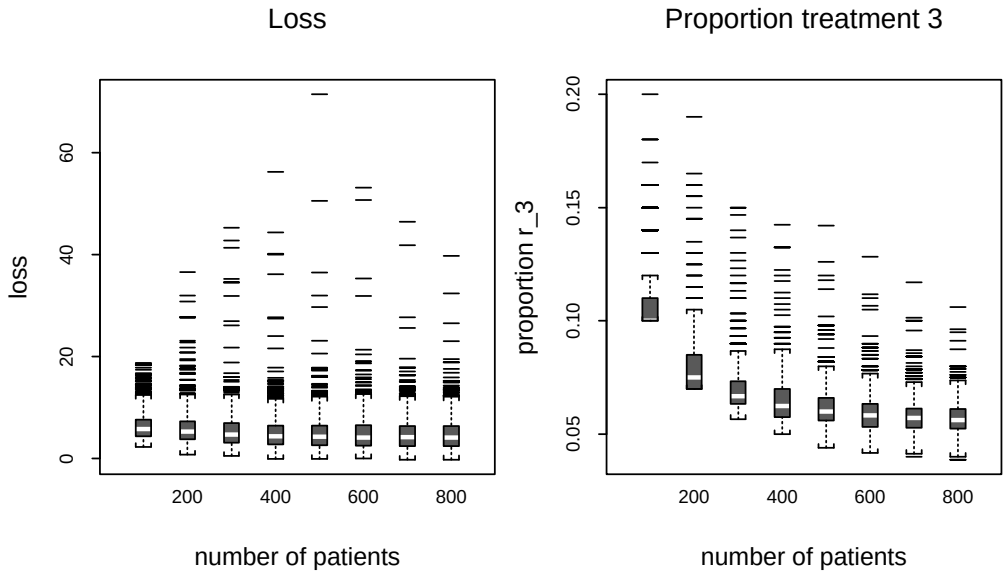

Figure 2: Three treatments: 1,000 individual regularized adaptive designs for rule G when $q = 5$, $\sigma = 1$ and $\gamma = 0.03$. Left-hand panel: boxplots of loss L_n. Right-hand panel: proportion r_3 of patients receiving the third treatment

proportion for eight values of n from 100 to 800. The losses in the left-hand panel of the figure show a steady central pattern that gradually decreases, as would be expected from the average losses in the left-hand panel of Figure 1, but with a few trials giving high losses. The right-hand panel gives the proportional allocations to treatment three. The effect of regularization is clear by the absence of low allocations for small values of n. For these regularized designs the minimum proportion when $n = 100$ is 0.1, which does not fall to the target value of 0.05 until $n = 400$. The right-hand panel also shows that, for several trials, treatment 3 around $n = 400$ has a value of r_3 of at least 0.10; treatments 2 and 3 have been misordered in some trials over an appreciable range of n, with a consequent inflation of loss.

The maximum value of loss in the left-hand panel of Figure 2 is 71.4 at $n = 500$. A similar plot for the unregularized design has a value of 140.5 at $n = 400$. For both rules there are a few trials that give high values, but, for the regularized design, the distribution is much less dispersed. For example, for $n = 100$, there are no losses above 20 for the regularized design. For the unregularized design there are 22 such values out of 1,000 in our simulations, with a maximum of 39.3.

The effect of regularization is to increase the proportion of patients receiving treatment 3 when n is small. This is achieved at a slight increase in

average loss. However, this increase depends on the way in which the target proportions p_j^* have been chosen. If we take $p_1^* = 0.8$, but with $p_2^* = p_3^* = 0.1$ we will have a sophisticated play the winner rule in which the loss will be unaffected if treatment 3 is preferred to treatment 2.

8 Inference

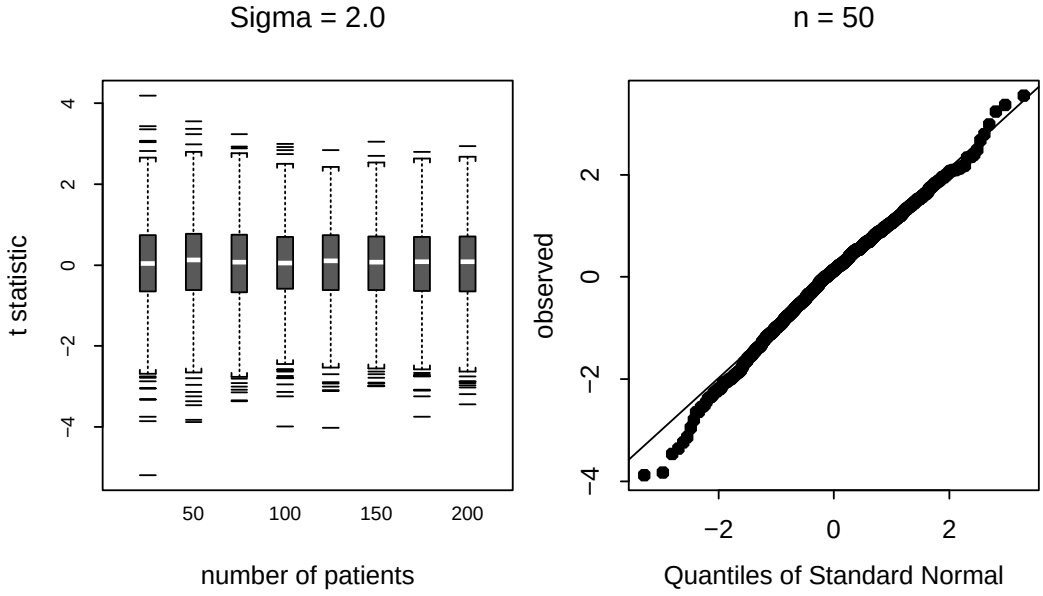

Figure 3: Rule G, pseudo-null distribution of the t-test for treatment equality: 1,000 individual regularized adaptive designs with two treatments when $q = 5$, $\gamma = 0.03$, $\mu = 0.65$ and $\sigma = 2$ ($p_1^* = 0,75$). Left-hand panel: boxplots of empirical distribution. Right-hand panel: normal QQ-plot when $n = 50$

The asymptotic results of §3 show that, for large n, the parmammteres $\hat{\beta}^n$ have a normal distribution with covariance matrix $\sigma^2(F^{nT}F^n)^{-1}$. We now empirically investigate how rapidly this convergence occurs by simulating the distribution of the t statistic for equality of means in a two-treatment trial for a series of values of n.

Since we are investigating adaptive designs, we require the design to be skewing the allocation and so the null distribution of the statistic is not of interest. Instead we investigate the "pseudo-null" distribution by subtracting off the known value of $\alpha_1 - \alpha_2$. We take $p_1^* = 0.75$, so that at balance $3n/4$

patients will receive treatment 1. With $\mu = 0.65$ and $\sigma = 2$,

$$E(t_\mu) \simeq \frac{\alpha_1 - \alpha_2}{\sigma\sqrt{4/n + 4/(3n)}} = 0.141\sqrt{n}. \tag{14}$$

When $n = 200$, $E(t_\mu) = 1.99$.

Figure 3 shows boxplots of 1,000 simulated values of the pseudo-null distribution of t, that is the sampling version of (14) with numerator $(\hat{\alpha}_1 - \hat{\alpha}_2) - (\alpha_1 - \alpha_2)$. We used rule G with $\gamma = 0.3$ as n goes from 25 to 200. Even with 25 patients, the statistics in the left-hand panel are centred close to zero with a symmetrical distribution. The normal QQ plot in the right-hand panel confirms the impression that the distribution is close to normal when $n = 50$. In particular, observed values between 2 and -2 occur with near to correct frequency, although the lower tail of the distribution beyond this range is slightly long. As n increases the boxplots show that the lower tail shrinks and that the distribution approaches normality with unit variance. The adaptive nature of the design seems to have a negligible effect on inference, even for small n, and standard normal, or t, tests can be used without modification in line with the asymptotic results of §3.

9 Comparisons

This section compares our new allocation rule with four others in the literature. In two rules (F and B) the specified allocation targets of our rule are replaced by targets calculated from a link function operating on treatment differences. The five rules, for two treatments, are:

1. Rule G given by (9) with $\gamma = 0.03$

2. "Doubly-adaptive" Rule H. Hu and Zhang (2004) investigate the properties of a doubly-adaptive allocation rule that ignores covariates. With $r_{1,n}$ the proportion of allocations to treatment 1 let

$$b = r_{1,n} \quad \text{and} \quad c = p^*_{\hat{R}(1)}.$$

Then the probability of allocating treatment 1 is

$$\pi_H(1, n+1) = \frac{c(c/b)^\nu}{c(c/b)^\nu + (1-c)\{(1-c)/(1-b)\}^\nu}. \tag{15}$$

When the treatments are correctly ordered and $r_{1,n}$ equals the target value $p^*_{R(1)}$ the allocation probability $\pi_H(1, n+1) = p^*_{R(1)}$. In (15) ν is a non-negative constant which determines the strength of forcing balance. Our comparisons are not sensitive to its value and we take $\nu = 1$.

In the applications presented by Hu and Zhang (2004) c is also estimated sequentially from the data, for example a function of the standard deviations of samples from two normal populations, hence the "double" adaptivity.

3. Random Allocation Rule R. Treatment 1 is allocated with probability $p^*_{\widehat{R}(1)}$.

4. Link-function Rule F. In this rule the target probabilities depend on the estimated difference in treatment means $\widehat{\Delta} = \hat{\alpha}_1 - \hat{\alpha}_2$. Atkinson and Biswas (2005) use a link function to relate $\widehat{\Delta}$ to the p_j. As do Bandyopadhyay and Biswas (2001) they take $p_1 = \Phi(\widehat{\Delta}/T)$, where $\Phi(.)$ is the standard normal c.d.f. The value of p_1 may be greater or less than 0.5, with the parameter T controlling the degree of skewing of the allocation. Randomness and covariate adaptation are achieved by taking the probabilities as in (9) but with the $p^*_{\widehat{R}(j)}$ replaced by p_1 and $1 - p_1$.

5. Link-function Rule B. The probability p_1 is found, as in Rule F, using the link function of Bandyopadhyay and Biswas (2001). Treatment 1 is then allocated with probability p_1. There is no attempt to balance over covariates.

Some general properties of these rules are clear. Rules B, H and R do not respond to the covariate pattern and so do not correct any over-allocation of one treatment to a particular set of covariates. Rules B and F increasingly skew the allocation as the treatment difference Δ increases; the remaining rules all target an allocation determined by the p^*_j. Only Rule G both responds to the covariate pattern and targets a given probability.

We compare the five rules for two values of $\Delta = \alpha_1 - \alpha_2$, 0.5 and 1. For those rules that target a fixed proportion - G, H and R - we take $p^*_1 = 0.8$. For the link-based rules F and B we take T such that $\Phi(\Delta/T) = 0.8$ when $\Delta = 0.5$. With this value of T these rules target $p_1 = 0.954$ when $\Delta = 1$. In all comparisons we adjust the treatment estimates for the covariate pattern. In general the best parameter estimates and so the highest power come from equal allocation to the two treatments. As Table 1 in §10 shows for a redesigned trial, there is a trade-off between power and the proportion of allocations to the better treatment.

The left-hand panel of Figure 4 shows the average losses for 10,000 simulations of a trial with up to 200 patients. Rules H and G have the lowest values of $\bar{L}_{200}$, 5.87 and 4.77. The value for Rule G is appreciably less than that in Figure 1 as there the average allocation probabilities have greater variability due to treatments 2 and 3 sometimes being incorrectly ranked

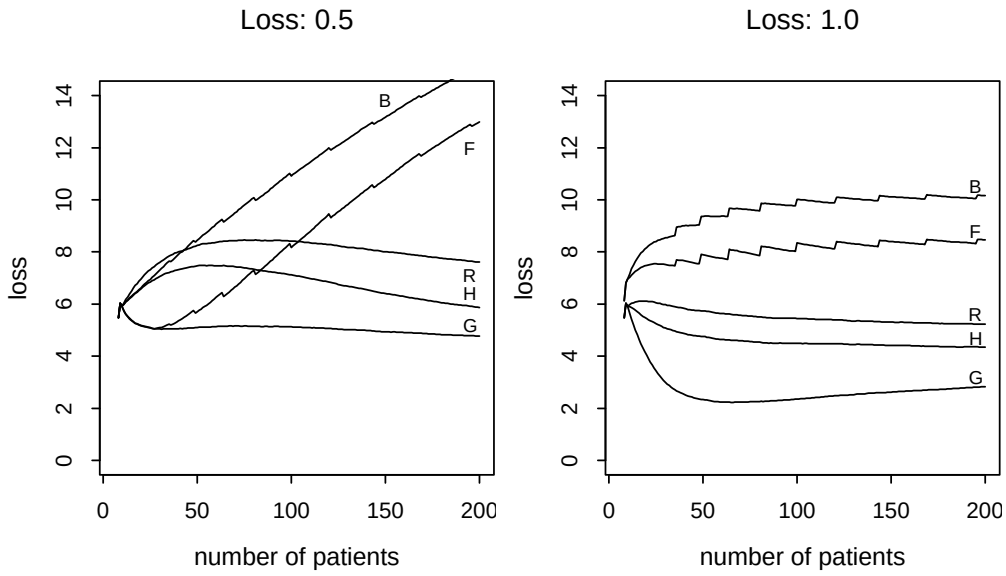

Figure 4: Average loss $\bar{L}_n$, five rules: $q = 5$, 10,000 simulations. Left-hand panel: $\Delta = 0.5$. Right-hand panel: $\Delta = 1.0$.

and allocated with the wrong probabilities. As n increases the losses for Rules G and R in Figure 4 both tend to five. The losses for the link-based rules B and F are large and increasing since the link estimate of p_1 can be greater than 0.8, although the treatments are correctly ranked. The effect of regularization is also evident for these two rules as the value of n_1 may be too close to n. All simulations start with ten observations equally allocated between the two treatments.

The right-hand panel of Figure 4 shows the losses when $\Delta = 1.0$. With the clearer difference between the treatments Rule G has a lower loss than before, which is gradually increasing as the allocation becomes more random as n increases. The losses for Rules H and R are similar, with that for Rule R close to five: $\bar{L}_{200} = 5.23$. With $\Delta = 1.0$, the target allocation for the link-based rules is 0.9538 and the values of loss for these rules were calculated using this value. With such strong evidence in favour of the first treatment, the rules allocate close to the target proportion and the losses are around 8 and 10. The effect of regularization is plainly visible.

The purpose of including randomization in these rules is to prevent various kinds of bias. Selection bias occurs when the clinician is able correctly to guess the next treatment to be allocated. For two treatments it can be

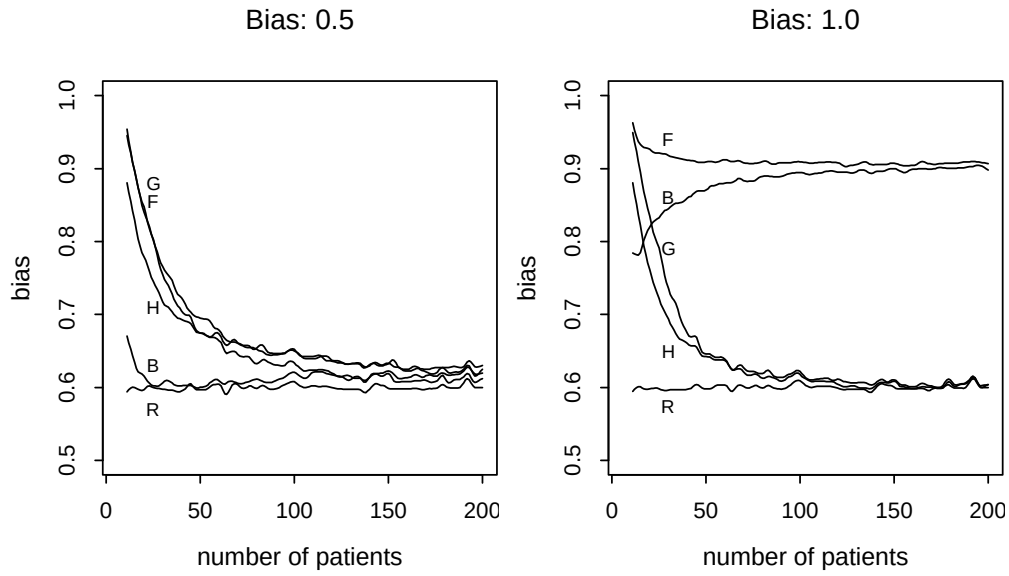

Figure 5: Smoothed bias $\bar{B}_n$, five rules: $q = 5$, 10,000 simulations. Left-hand panel: $\Delta = 0.5$. Right-hand panel: $\Delta = 1.0$.

estimated from n_{sim} simulations by

$$\begin{aligned} B_n &= (\text{number of correct guesses of allocation to patient } n \\ &\quad -\text{number of incorrect guesses})/n_{\text{sim}}. \end{aligned} \tag{16}$$

In calculating the bias we assume that the clinician has access to the allocation probabilities $\pi_V(j|x_{n+1})$, for any rule V, and chooses the treatment with the highest allocation probability. For the non-randomized sequential construction of optimum designs treatment allocation is deterministic once x_{n+1} is known so the value of B_n is one. For random allocation with two treatments $p_1^* = 0.5$ and the value is zero. For skewed random allocation the value for Rule R when guessing the treatment more likely to be allocated is $2p_1^* - 1$. With $p_1^* = 0.8$, this value is 0.6.

Figure 5 shows the values of $\bar{B}_n$ for the five rules. For $\Delta = 0.5$ all rules tend to a value of 0.6 as n increases. Rule R has this value for all n, as does B almost from the beginning. The rule with the highest initial bias is G. Rule F has almost the same value. For both rules, which have the same value of γ, allocation is virtually deterministic in the early stages of the trial. When $\Delta = 1.0$ (the right-hand panel) rules F and B have an asymptote of 0.9076 - with extreme skewing it is possible to guess with virtual certainty which treatment will be allocated. The comparison of the other three rules is similar to that in the left-hand panel. In both panels the ordering in terms

of bias for Rules G, H and R is the reverse of that for loss. The same is true when comparing the two. link-function rules B and F. The same reverse ordering holds for the non-adaptive rules compared in Atkinson (2002).

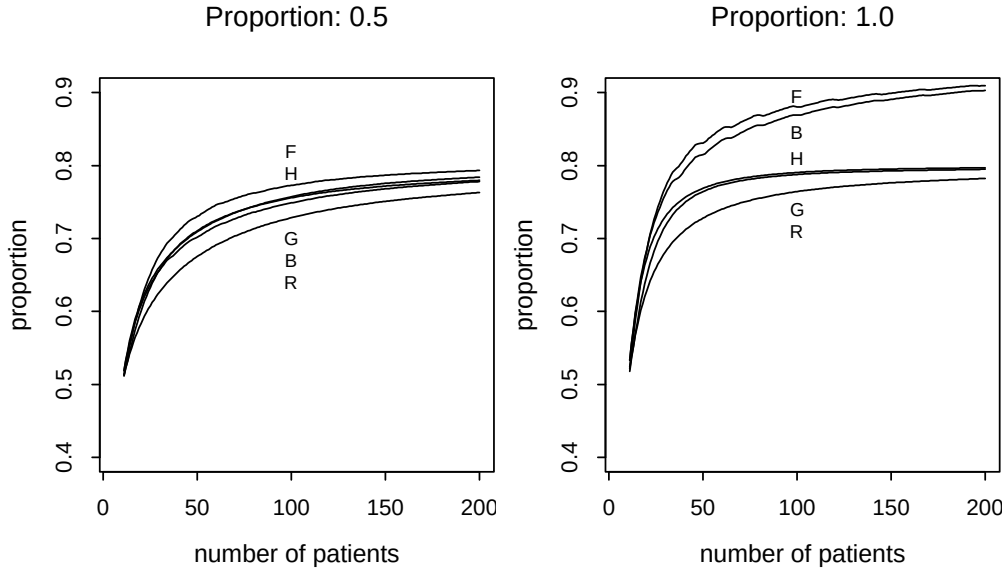

Figure 6: Smoothed allocation proportion $\bar{r}_{1,n}$, five rules: $q = 5$, 10,000 simulations. Left-hand panel: $\Delta = 0.5$. Right-hand panel: $\Delta = 1.0$.

The average proportion of patients allocated treatment 1, $\bar{r}_{1,n}$, is plotted in Figure 6. All proportions for $\Delta = 0.5$ in the left-hand panel converge to 0.8 from below, that for Rule H converging most rapidly; as (15) shows, this rule increases the probability of allocating treatment 1 when $\bar{r}_{1,n} < p_1^*$. Rule R converges most slowly with Rule G in between. The same conclusion is true for Rules H, G and R for $\Delta = 1.0$ shown in the right-hand panel of the figure. Now, however, the proportions for Rules F and B are targeting 0.9538. The effect of regularization on the proportions for these two rules is visible in the figure.

To provide an indication of the variability of the allocation proportions from trial to trial, we give in Figure 7 the standard deviations of the values of $r_{1,n}$, the averages of which are in Figure 6. The results for $\Delta = 0.5$ are in the left-hand panel. Rules R, G and H have similar standard deviations at $n = 200$, although that for H is highest around $n = 40$. Rules B and F are appreciably more variable, as would be expected from Figure 4. The situation in the right-hand panel, for the larger value of Δ, is similar. Rules R, G and H are again in decreasing order of standard deviation when $n = 200$. Rule B again shows more variability than Rule F, but the values are much reduced with the increase in Δ.

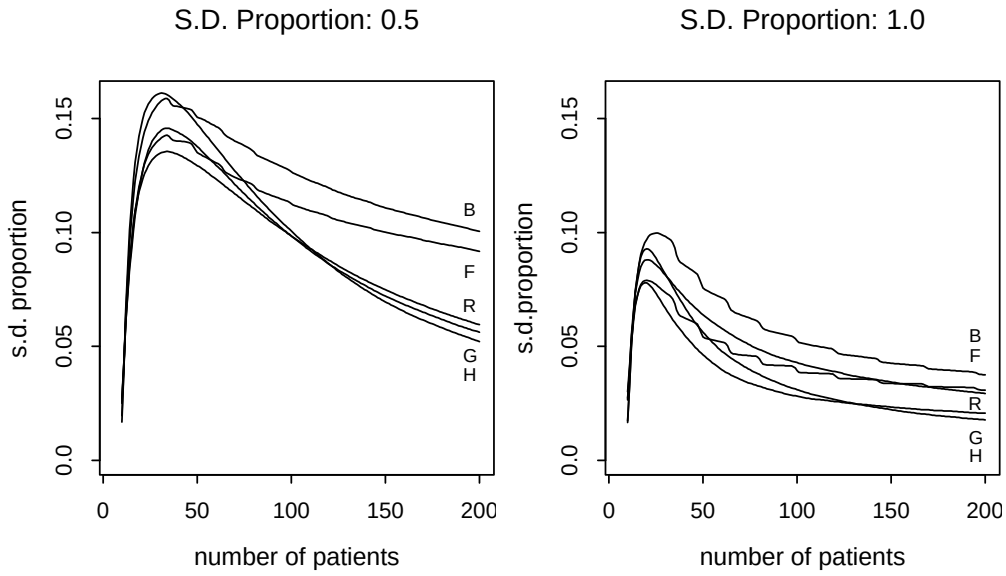

Figure 7: Smoothed standard deviation of allocation proportion, five rules: $q = 5$, 10,000 simulations. Left-hand panel: $\Delta = 0.5$. Right-hand panel: $\Delta = 1.0$.

Many comparisons of allocation rules focus on power. The comparisons in this section serve to stress the importance of considering in addition several other properties of any rule. In Figure 8 we show the results of our power calculations as the proportion of t-statistics significant at the 1% level. In the left-hand panel the ordering of the non-link rules is R, G and H as it is for $\Delta = 1$, with Rule R having higher power and, from Figure 6 the average allocation closest to 0.5. Rules F and B have good power when $\Delta = 0.5$, but perform less well when $\Delta = 1.0$. The power of Rule G could be improved up to that of Rule R by using larger values of γ, with a consequent increase in loss and decrease in the proportion of patients allocated the better treatment.

10 Redesigning a Trial: Fluoxetine Hydrochloride

We close with an example in which we redesign an existing trial, using part of the data from Tamura *et al.* (1994) on the treatment of depressive patients, for which there is a speedily available surrogate response. There are two treatments, fluoxetine and control, and two covariates. One covariate, sleep disfunction before the trial, is binary. The second gives the initial values of HAMD_{17}, a measure of depression, for each patient. The response is the

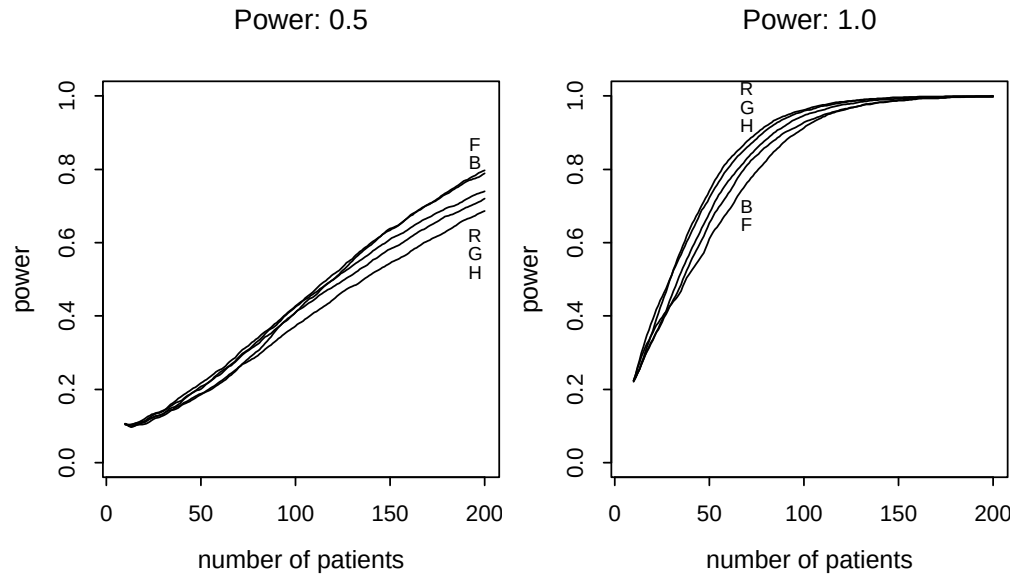

Figure 8: Power: smoothed proportion of t-statistics significant at the 1% level, five rules: $q = 5$, 10,000 simulations. Left-hand panel: $\Delta = 0.5$. Right-hand panel: $\Delta = 1.0$.

negative of the change in HAMD_{17}. Since HAMD_{17} is measured on a 53-point scale, we treat it as a continuous variable. Large values are desired. Because of the surrogate response we can assume, as did Tamura *et al.* (1994), that all responses up to that of patient n are available when the allocation is made for patient $n + 1$.

We code the binary covariate with values -1 and 1, although Tamura *et al.* (1994) used 0 and 1. Analysis of the data shows that the probability of each value is 0.5. We subtract the mean value of 21.7045 from the initial value of HAMD_{17}, which we take as normally distributed with a standard deviation of 3.514. Surprisingly, the two covariates are uncorrelated, so we model them as independent random variables. In addition, neither covariate has a significant effect on the response.

There are 88 observations since one observation in the original data set does not have a response. After adjustment for the covariates the residual mean square estimate of the standard deviation is $s = 6.97$ and the estimated treatment difference $\hat{\alpha}_1 - \hat{\alpha}_2 = 3.795$; the treatment does seem to have decreased depression, since large values of the surrogate are good. The t value for this effect is 2.55, with a nominal significance level of 1.6% when any effect of the sequential design is ignored.

Tamura *et al.* (1994) used a form of randomized play the winner rule which resulted in 43 allocations of treatment 2, the control. The adaptive

Table 1: Data on fluoxetine hydrochloride from Tamura et al. (1994). Average proportion of allocations to treatment 1 (fluoxetine) and average t-statistic from 1,000 simulations of 88 patient clinical trial.

Target p_2^*	Average proportion $\bar{r}_2$	Average statistic $\bar{t}$
0.5	0.500	2.563
0.55	0.546	2.549
0.6	0.592	2.512
0.65	0.637	2.450
0.7	0.681	2.371
0.75	0.722	2.266
0.8	0.760	2.140
0.85	0.796	1.970
0.9	0.820	1.810
0.95	0.833	1.712

scheme should preferentially allocate treatment 1. We take p_1^* over the range 0.5 (unskewed allocation) to 0.95. Interest then is in the linear combination of parameters given by $a = (p_1^* \quad 1 - p_1^* \quad 0 \quad 0)^T$.

We simulated 1,000 trials with 88 patients for p_1^* in the range 0.5 to 0.95. The results are in Table 1. They show that as the target increases from 0.5, so does the average proportion allocated to treatment 1, although more slowly than the target. We also give the average values of the simulated t-statistic. These enable us to quantify the relationship between increasingly ethical allocation from skewing and the decrease in power. The average statistic is still just greater than 1.96 when $p_1^* = 0.85$.

To check the distribution of the statistic we took $p_1^* = 0.75$. The QQ-plot of the 1,000 values of the statistic, which we do not display, is similar to that in the right-hand panel of Figure 3, but with less curvature since $n = 88$ rather than 50. Standard inferences can be used to analyse this trial.

For $p_1^* = 0.75$ the average values of the statistic is 2.27, compared with the value of 2.55 of Tamura *et al.* (1994). This slight decrease in power is in line with the results of Pocock (1983) on the effects of imbalance. Our rule however allocates an average of 63.5 patients to treatment 1, that is 18.5 more than received the better treatment in the original trial.

11 Discussion

Our rule G requires only that the treatments be ordered by desirability. Many factors, such as efficacy, cost, toxicity and long-term side-effects could be included in such a ranking. In exploring the numerical properties of our designs we have looked at only two or three treatments and five nuisance parameters. However, the results of Atkinson (1999) for non-adaptive designs show that results for $q = 10$ are similar in structure to those for $q = 5$. Related calculations for rule G, not reported here, suggest the same is true for this adaptive rule. Atkinson (2002) also shows that the effect of non-normal covariates on the aggregate properties of biased-coin designs is slight.

Although the ordering of the treatments does not depend on the assumption of normality, the variance $d_(\cdot)$ used in calculating the probabilities in (8) does assume that the linear model (2) is appropriate for analysis with errors of constant variance. Our analysis therefore also extends to data which can be made approximately normal by transformation (Box and Cox, 1964; Yeo and Johnson, 2000). For generalized linear models the main application has been binomial responses (Tymofyeyev, Rosenberger, and Hu, 2007). To use the methods of optimum design the expressions for variances need extending to include the iterative weights used in parameter estimation. Exceptions are where the treatment effects are sufficiently small that the effect on the design of the iterative weights can be ignored (Cox, 1988) and gamma models with the log link. In neither case is a modification required. If responses are delayed, information on those responses that are available is used to order the treatments, with the calculation of $d_c(\cdot)$ using information on all $n + 1$ vectors of prognostic factors.

Related forms of rule are possible. For example, we could extend Rule G by introducing a tolerance region into the calculation of the p_j^*. If the $\hat{\alpha}_j$ suggest that the differences in some α_j are not technically significant, we could take those values of p_j^* equal, so ensuring equal target allocation probabilities. Decreasing the size of this tolerance region as $n^{-1/2}$ would relate the targets p_j^* to statistical significance.

Acknowledgment This paper was completed at the Isaac Newton Institute for Mathematical Sciences in Cambridge, England, during the 2011 programme on the Design and Analysis of Experiments.

References

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. and Biswas, A. (2005). Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics*, **61**, 118–125.

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

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

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

Bartlett, R., Roloff, D. W., Cornell, R. G., Andrews, A. F., Dillon, P. W., and Zwischenberger, J. B. (1985). Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. *Pediatrics*, **76**, 479–487.

Begg, C. B. (1990). On inference from Wei's biased coin design for clinical trials (with discussion). *Biometrika*, **77**, 467–484.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B*, **26**, 211–246.

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

Chow, S.-C. and Chang, M. (2007). *Adaptive Design Methods in Clinical Trials*. Chapman and Hall/ CRC Press, Boca Raton.

Cox, D. R. (1957). The use of a concomitant variable in selecting an experimental design. *Biometrika*, **44**, 150–158.

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

Dumville, J. C., Hahn, S., Miles, J. N. V., and Torgerson, D. J. (2006). The use of unequal randomisation ratios in clinical trials: a review. *Contemporary Clinical Trials*, **27**, 1–12. doi:10.1016/j.cct.2005.08.003.

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

Gallo, P., Chuang-Stein, C., Dragalin, V., Gaydos, B., Krams, M., and Pinheiro, J. (2006). Adaptive designs in clinical drug development: An executive summary of the PhRMA working group. *Journal of Biopharmaceutical Statistics*, **16**, 275–283.

Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, New York.

Hu, F. and Zhang, L.-X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Annals of Statistics*, **32**, 268–301.

Lai, T. Z. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics*, **10**, 154–166.

Matthews, J. N. S. (2006). *An Introduction to Randomized Controlled Clinical Trials, 2nd edition*. Edward Arnold, London.

Neyman, J. S. (1959). Optimal asymptotic tests of composite statistical hypotheses. In U. Grenander, editor, *Probability and Statistics: the Harald Cramér Volume*. Almquist and Wiksell, Uppsala.

Pocock, S. J. (1983). *Clinical Trials*. Wiley, New York.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, **58**, 527–535.

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

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

Tamura, R. N., Faries, D. E., Andersen, J. S., and Heiligenstein, J. H. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association*, **89**, 768–776.

Tymofyeyev, Y., Rosenberger, W. F., and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association*, **102**, 224–234.

Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**, 954–959.

Zelen, M. (1969). Play the winner and the controlled clinical trial. *Journal of the American Statistical Association*, **64**, 131–146.

Zhang, L. and Rosenberger, W. F. (2007). Response-adaptive randomization for survival trials: the parametric approach. *Applied Statistics*, **56**, 153–165.

Zhang, L.-X., Hu, F., Cheung, S. H., and Chan, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *The Annals of Statistics*, **35**, 1166–1182.