Biometrics 61, 118–125
March 2005

Bayesian Adaptive Biased-Coin Designs for Clinical Trials with Normal Responses

Anthony C. Atkinson

Department of Statistics, London School of Economics, London WC2A 2AE, U.K.
email: A.C.Atkinson@lse.ac.uk

and

Atanu Biswas

Applied Statistics Unit, Indian Statistical Institute, Barrackpore Trunk Road, Kolkata 700 108, India

Summary. Adaptive designs are used in phase III clinical trials for skewing the allocation pattern toward the better treatments. We use optimum design theory to derive a skewed Bayesian biased-coin procedure for sequential designs with continuous responses. The skewed designs are used to provide adaptive designs, the performance of which is studied numerically and theoretically. Important properties are loss and the proportion of allocation to the better treatment.

Key words: Bayesian biased-coin; D_A-optimal design; Limiting allocation proportion; Minimization; Randomization; Regularization.

1. Introduction

Patients arrive sequentially for a phase III clinical trial, and are to be allocated one of t treatments. Adaptive designs skew the allocation proportion to the eventually best treatment by using earlier responses to determine the next allocation. For efficient estimation of the treatment effect, the allocation needs to be approximately balanced over the prognostic factors and covariates of the individual patients. There also needs to be some randomization in the allocation.

Book length treatments of randomization in clinical trials include Matthews (2000) and Rosenberger and Lachin (2002). Typical adaptive designs (Rosenberger and Lachin, 2002; Rosenberger et al., 2002) are for binary responses in the absence of prognostic factors. We use optimum design theory to provide balance, augmented by a biased coin (Atkinson, 1982) for randomization. We introduce optimum designs that give a skewed allocation and show how they can be used to provide adaptive designs that favor the best treatment. We then extend the Bayesian biased-coin design theory of Ball, Smith, and Verdinelli (1993) to skewed allocations and construct adaptive designs for normally distributed responses.

The gains from using adaptive design can be appreciable. Zelen and Wei (1995) describe a clinical trial, reported by Connor et al. (1994), to evaluate the antiviral therapy AZT for reduction of the risk of maternal-to-infant HIV transmission. Standard randomization led to 238 pregnant women receiving AZT and 238 receiving a placebo. Sixty newborns in the placebo group were HIV-positive against 20 in the AZT group; the estimated success probabilities are therefore 0.9160 using AZT and 0.7479 from placebo. Using these as the true values, Yao and Wei (1996) show that an adaptive randomized play-the-winner rule could result in a 300:176 allocation with the greater allocation to AZT. In this process, the lives of 11 newborns would have been saved.

The model and parameter estimates for our procedure are presented in Section 2, together with a method for skewing treatment allocations. The main theoretical development is in Section 3 where we provide a Bayesian method for skewing allocations in biased-coin designs. In Section 4, we compare the average losses from Bayesian and classical rules for skewed designs. The average properties of adaptive designs are explored in Section 5. Section 6 presents the distribution of loss for individual clinical trials. These two sections show that the Bayesian designs force balance over covariates in the early stages of the trial, but allow increasing randomness in allocation as the trial progresses. Section 7 uses published results on a clinical trial to illustrate design construction. This article concludes with comments on inference for adaptive designs. Mathematical results on optimum design are given in the Appendix.

2. Models, Variances, and Loss

2.1 *Skewed Allocations*

In general, let there be t treatments, one of which may be the control. The vector of unknown treatment effects is α and the

patient presents with a vector x_i of covariates. The results of the trial will be analyzed using the regression model

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

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

We consider designs to estimate particular linear combinations of the α_j, which we can write as $l^T\alpha$. In Atkinson (2002), with just two treatments, interest was in estimation of the treatment difference $\alpha_1 - \alpha_2$, so that $l = (1\ -1)^T$. Here we require a procedure that will yield a skewed allocation in which a proportion p of the patients receive treatment 1. This is obtained by designing to estimate the linear combination

$$l^T\alpha = p\alpha_1 - (1-p)\alpha_2 \quad (0 \le p \le 1), \tag{2}$$

with minimum variance. So now $l = \{p\ -(1-p)\}^T$. If n_1 patients receive treatment 1 and $n_2 = n - n_1$ receive treatment 2, the variance of the estimated linear combination $l^T\hat{\alpha}$, in the absence of covariates, is minimized when

$$n_1 = pn \quad \text{and} \quad n_2 = (1-p)n. \tag{3}$$

With two treatments we are interested in one linear combination of α_1 and α_2. There is therefore a second linear combination the coefficient of which is not of interest. Together with the $q\ \ 1$ elements of θ there are therefore in all q nuisance parameters. As we will see, the properties of the designs depend on q.

2.2 *Loss*

Let the estimate of $l^T\alpha$ from the optimum skewed design (3) be $l^T\hat{\alpha}_*$. Then, in the absence of covariates, we obtain the simple result

$$\mathrm{var}\left\{l^T\hat{\alpha}_*\right\} = \sigma^2/n. \tag{4}$$

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

For other designs we find the variance from (A.2) with the linear combination given by (A.3). We can compare these designs using either the ratio of variances, that is the efficiency E_n, or we can use the loss (Burman, 1996). From (A.2), with the matrix A replaced by the vector l, the efficiency of any design is

$$E_n = 1/\left\{nl^T\left(G_n^TG_n\right)^{-1}l\right\}. \tag{5}$$

The loss L_n is defined by writing the variance (A.2) as

$$\mathrm{var}\left\{l^T\hat{\alpha}\right\} = \sigma^2/(n - L_n), \tag{6}$$

so that

$$L_n = n(1 - E_n). \tag{7}$$

With a random element in treatment allocation, the loss L_n is a random variable, depending upon the particular trial and pattern of covariates. Let $E(L_n) = \mathcal{L}_n$. The results of Table 1 give the asymptotic values $\mathcal{L}_\infty$. The loss can be interpreted as the number of patients on whom information is lost due to the lack of optimality of the design.

Table 1
Classical allocation rules

Rule	Probability	$\mathcal{L}_\infty$
D: Deterministic	$\pi_D(1 \mid x_{n+1}) = 1$	0
R: Completely randomized	$\pi_R(1 \mid x_{n+1}) = p$	q
A: D_A optimality	$\pi_A(1 \mid x_{n+1})$	$q/5$
	$= \frac{pd_A(1,.)}{pd_A(1,.)+(1-p)d_A(2,.)}$	
E: Efron's biased-coin		0
If $d_A(1,.) > d_A(2,.)$	$\pi_E(1 \mid x_{n+1}) = \frac{2p}{1+p}$	
otherwise	$\pi_E(1 \mid x_{n+1}) = \frac{p}{2-p}$	

3. Bayesian Biased-Coin Designs

3.1 *Skewed Designs*

The sequential optimum designs given by (A.5) provide a method of balancing the trial as rapidly as possible, producing designs with minimum loss. However, such designs are susceptible to selection bias. The biased-coin designs of Atkinson (2002) use (A.6) to provide some randomization and reduce bias at the cost of a small increase in loss. There is however no theoretical justification for this randomizing mechanism. Because the comparisons of Atkinson (2002) for unskewed designs show that the family of Bayesian biased-coin designs derived from the results of Ball et al. (1993) have better performance in terms of bias and loss than (A.6), we now derive skewed Bayesian designs when it is required that, in the long run, a proportion p_j of the patients receive treatment j.

The designs are found to maximize the utility

$$U = U_V - \gamma U_R, \tag{8}$$

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

With $\pi_B(j)$ the probability of allocating treatment j, let

$$U_R = \sum_{j=1}^{t} \pi_B(j)\log\{\pi_B(j)/p_j\}, \tag{9}$$

where the p_j are the desired allocation proportions. We want to minimize U_R subject to the constraint that $\sum_{j=1}^{t}\pi_B(j) = 1$. We introduce the Lagrange multiplier λ and minimize

$$\sum_{j=1}^{t} \pi_B(j)\log\{\pi_B(j)/p_j\} + \lambda\left(\sum_{j=1}^{t}\pi_B(j) - 1\right). \tag{10}$$

The derivative of (10) with respect to $\pi_B(j)$ is

$$1 + \log(\pi_B(j)/p_j) + \lambda, \tag{11}$$

so all ratios $\pi_B(j)/p_j$ must be equal. We then have the desired result for skewed random allocation that $\pi_B(j) = p_j$, $j = 1, \ldots, t$.

We now turn to consideration of the utility U. As in the work on balanced allocation, we write

$$U_V = \sum_{j=1}^{t} \pi_B(j)\phi_j,$$

where ϕ_j is a measure of the information from applying treatment j. Shortly, we define this in terms of D_A optimality. Then we require designs to maximise the utility

$$U = \sum_{j=1}^{t} \pi_B(j)\phi_j - \gamma \sum_{j=1}^{t} \pi_B(j) \log\{\pi_B(j)/p_j\}. \quad (12)$$

As before, the introduction of the Lagrange multiplier λ followed by differentiation with respect to $\pi_B(j)$ leads to t relationships which are now

$$\phi_j - \gamma\{1 + \log(\pi_B(j)/p_j)\} + \lambda = 0, \quad (13)$$

so that all quantities

$$\phi_j/\gamma - \log(\pi_B(j)/p_j)$$

must be constant. That is,

$$\pi_B(j)/p_j = \kappa \exp(\phi_j/\gamma),$$

for a constant κ. Because $\sum_{j=1}^{t} \pi_B(j) = 1$, we obtain

$$\pi_B(j) = \frac{p_j \exp(\phi_j/\gamma)}{\sum_{j=1}^{t} p_j \exp(\phi_j/\gamma)}. \quad (14)$$

The results of Ball et al. (1993) are obtained when all p_j's are equal, that is, in the unskewed case.

3.2 *Sequential Designs*

Probabilities for the allocation of each treatment are given by (14) for some general design criterion ϕ. We now derive specific expressions for the probabilities for sequential D_A-optimum designs, assuming that the trial continues long enough to overwhelm any prior information about the parameters ω.

For D_A optimality, it is convenient to work with maximization of the log of the determinant of the information matrix, a convex function, rather than with maximization of the determinant itself. For the current n-trial design, which is to be updated, we take

$$\phi = \log\det M = -\log\left|A^T\left(G_n^T G_n\right)^{-1} A\right|. \quad (15)$$

If the $(n+1)$st patient receives treatment j, (15) leads to

$$\phi_j = \log\det M_j. \quad (16)$$

But, from (A.4)

$$\log\det M_j = \log\{1 + d_A(j, n, x_{n+1})\} + \log\det M.$$

The combination of these results with (A.4) and (A.5) leads to the expression for the probabilities,

$$\pi_B(j\,|\,x_{n+1}) = \frac{p_j\{1 + d_A(j, n, x_{n+1})\}^{1/\gamma}}{\sum_{j=1}^{t} p_j\{1 + d_A(j, n, x_{n+1})\}^{1/\gamma}}, \quad (17)$$

At the optimum design all $d_A(j, n, x_{n+1})$ are equal, so that $\pi_B(j\,|\,x_{n+1}) = p_j$.

4. A Comparison of Sequential Designs for Skewed Allocation

4.1 *Classical Allocation Rules*

We continue to denote by $\pi(j\,|\,x_{n+1})$ the conditional probability that the $(n+1)$st patient, with prognostic factors x_{n+1}, receives treatment j. When these probabilities depend upon the ordering of the treatments by the variances $d_A(j, n, x_{n+1})$, we use $\pi([j]\,|\,x_{n+1})$ to represent the probability of allocating the treatment with the jth largest value of the variance. The rules for skewed allocation when $t = 2$ are given in Table 1.

In rule D, the Deterministic construction of sequential designs, that treatment should be allocated for which $d_A(j, n, x_{n+1})$, $j = (1, 2)$ is larger. For the completely Randomized rule R, the first treatment is allocated with probability p.

These two rules represent the extremes of rules which aim for skewing and balance over both the short and the long term. The losses of the other rules are bounded by these values. Rule A, D_A Optimality, is defined in (A.6). The extension of rule E, Efron's biased coin, depends on which treatment would be allocated by an unrandomized rule. If $d_A(1, n, x_{n+1}) > d_A(2, n, x_{n+1})$, we find the probabilities of allocating treatment 1 by replacing $d_A(1, n, x_{n+1})$ in (A.6) by 2/3: if treatment 2 is underrepresented we replace $d_A(1, n, x_{n+1})$ with 1/3.

4.2 *Numerical Results*

We compared these four allocation rules with the Bayesian rule for four values of γ. We took one known value of p, 0.75, and calculated the loss for $q = 5$ and 10. The results shown are the averages of 10,000 simulations of the 800 patient trials.

The plots of Figure 1 show the losses, as functions of patient number, when $q = 5$. The left-hand panel is for the four classical rules. Although the designs found here are for $p = 0.75$, the plot is similar to the left-hand panel of Figure 1 of Atkinson (2002) for unskewed designs with $p = 0.5$, but extended to 800 patients. Comparatively there is a slight increase in all losses for the skewed allocation. This arises because, with a 3:1 ratio of allocation, the skewed designs are on average slightly less well balanced than those for $p = 0.5$. If we read down from the top of the plot at $n = 800$, the loss for rule R is approximately five throughout and that for D is close to zero for n above 100, while that for A is close to one, which is $q/5$, from $n = 50$. The numbers for L_{800} are in Table 2.

The right-hand panel of the figure shows the average loss for the Bayes rule with $\gamma = 1$, 0.1, 0.03, and 0.01. Table 2 gives the losses for $n = 800$. These increase with increasing γ. The losses for all Bayesian rules first decrease rapidly and then increase gradually from the minimum. At the start of the trial, the Bayes rule (17) behaves much as the deterministic rule, allocating the underrepresented treatment with high probability: Balance is forced and the loss decreases rapidly. But, as the trial proceeds, the $d_A(j, .)$ decrease until all values of $\{1 + d_A(j, .)\}$ are close to one. Then allocation is virtually random and the loss increases gradually toward q, the value for rule R. The appearance of $1/\gamma$ as a power in (17) indicates that the decrease and gradual increase in the losses for the Bayes rules is greater for smaller γ, as can be seen in the right-hand panel of Figure 1. The losses for $q = 10$ are

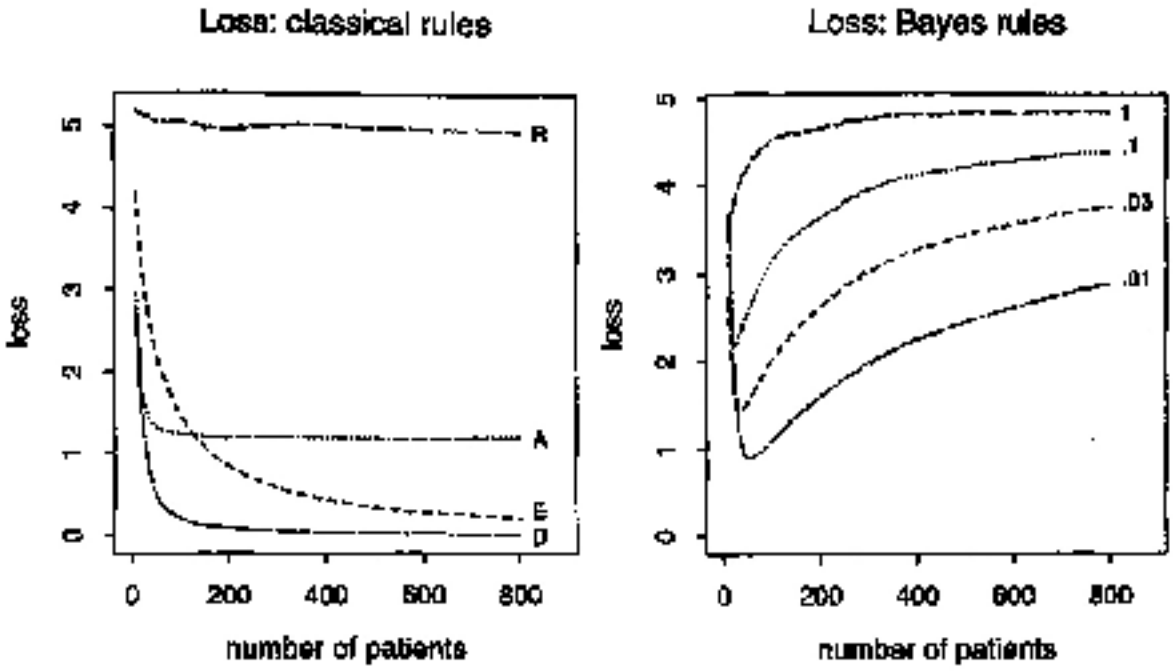

Figure 1. Designs for skewing proportion $p = 0.75$. Average losses L_n when $q = 5$ for eight allocation rules. Left-hand panel: A, D_A-optimality; D, deterministic; E, Efron's biased coin, and R, random. Right-hand panel, Bayes rules: reading downward, $\gamma = 1$, 0.1, 0.03, and 0.01. Averages of 10,000 simulations.

approximately twice those for $q = 5$, as is shown by the numbers for all rules in Table 2.

Comparison of these numerical results for skewed Bayesian allocations with those in Atkinson (2002) shows that the extension to skewed allocations does not greatly increase the loss due to imbalance. The adaptive use of these designs, when p is estimated from the data, does however lead to appreciable increases in loss.

5. An Adaptive Design

5.1 *Link-Function-Based Adaptive Design*

We use the link-function-based adaptive design of Bandyopadhyay and Biswas (2001) to provide an estimate of p. Assume that large values of the response y are desired and let $\hat{\Delta} = \hat{\alpha}_1 - \hat{\alpha}_2$. They suggest that the adaptive probability of allocating treatment 1 should be

$$\pi_{BB}(1 \mid \boldsymbol{x}_{n-1}) = J(\hat{\Delta}),$$

where $J(\cdot)$ is a suitably chosen distribution function of a random variable, symmetric about zero. We use $\Phi(x/T)$, the distribution function of a $N(0, T^2)$ random variable. If $\hat{\Delta}$ is positive, that is, $\hat{\alpha}_1 > \hat{\alpha}_2$, the probability of allocating the first treatment is >0.5.

To obtain dependence of our designs on $\boldsymbol{x}_{n+1}$, we take $p = J(\hat{\Delta})$ in calculating the quantity $\boldsymbol{d}_A(j, n, \boldsymbol{z}_{n+1})$ for the $(n+1)$st patient. We can then apply the allocation rules of Sections 3 and 4.1. As it becomes clearer that treatment 1 is superior to treatment 2, the allocation proportion converges to $J(\alpha_1 \quad \alpha_2)$; the speed of convergence depends on the allocation rule.

Table 2

Average loss L_{500} for unskewed and skewed allocations from 10,000 simulations

Rule	$q = 5$	$q = 10$	Bayes: γ	$q = 5$	$q = 10$
A	1.197	2.469	1	4.837	9.807
D	0.024	0.103	0.1	4.387	8.898
E	0.221	0.859	0.03	3.760	7.668
R	4.942	10.016	0.01	2.883	5.881

5.2 *Regularization*

In simulating adaptive designs, we found it was necessary to regularize the design to ensure that each treatment continued to be allocated throughout the trial. We therefore allocated 5 of the first 10 patients to treatment 1 and the other 5 to treatment 2. Thereafter, if the number allocated to either treatment is below $n^{1/2}$, that treatment is allocated when n is an integer squared. For our 800 trial design with five patients allocated initially to each treatment, the first regularization could occur when $n = 36$ and the last when $n = 784$. In our simulations, we had a target allocation of 0.75 to the better treatment. The effect of the regularization is to speed the convergence of the skewed adaptive allocation to this value. There is nothing special about $n^{1/2}$: We only require a sequence that avoids too extreme allocations.

5.3 *Numerical Results*

The properties of the design depend on the parameters in the function J. We take $J(x) = \Phi(x/T)$ and $\alpha_1 - \alpha_2 = 0.6745$ which, with $T = 1$, gives a value of 0.75 for p, so that we can make comparisons with the results of Section 4.2. In practice, T would be chosen to give the desired skewing proportion p for the expected difference in treatment effects. The properties of the design also depend on the error of measurement σ^2, larger values leading to designs which converge more slowly to the desired value of p.

Figure 2 shows the plots of average loss from 10,000 simulations when $\sigma = 0.8$ and $q = 5$. These plots are broadly similar to those in Figure 1 except that all losses have increased by amounts between two and three. There are however some important details that are different.

Rules A and D in the left-hand panel of the figure show a similar structure: An initial decrease in loss is followed by

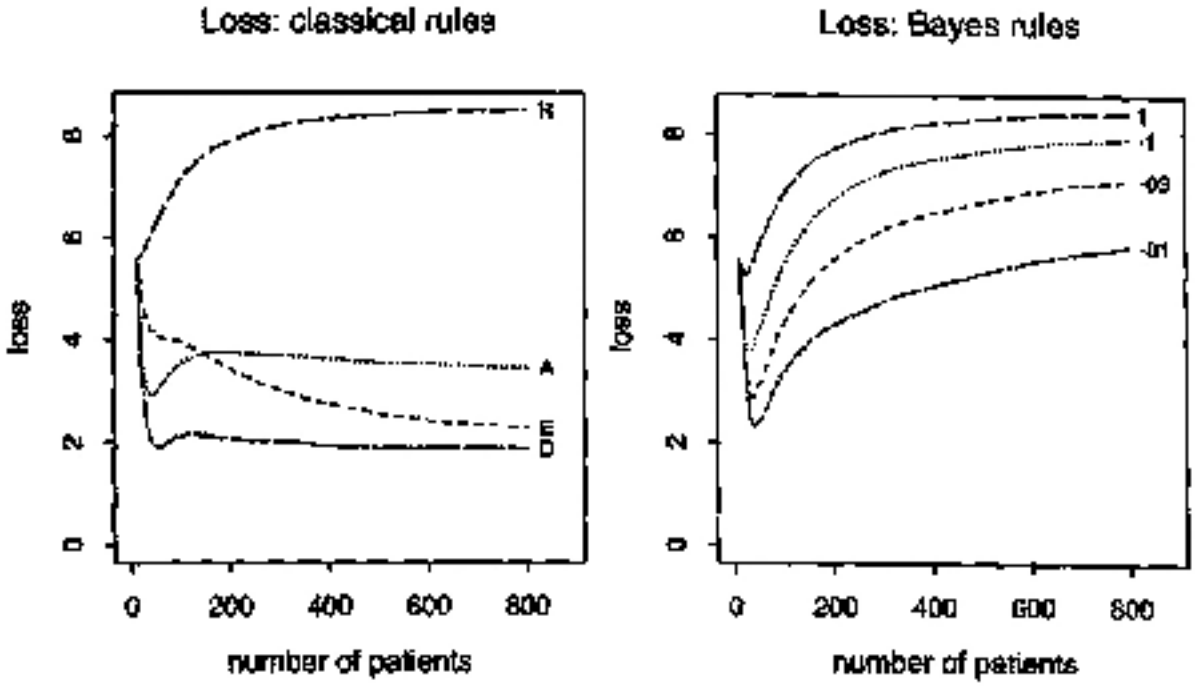

Figure 2. Adaptive designs: average losses L_n when $q = 5$ for eight allocation rules. Left-hand panel: rules D, E, A, and R. Right-hand panel, Bayes rules: reading downward, $\gamma = 1$, 0.1, 0.03, and 0.01. Averages of 10,000 simulations, $\sigma = 0.8$.

a period of increase due to designs for incorrect estimates of p. Eventually, as the trial progresses, the estimates improve and the average loss starts to decrease gradually. The other rules show slightly different behavior. Most importantly, rule R and the four Bayesian rules show average losses which are still increasing at the end of the simulations. Rule R is that of Bandyopadhyay and Biswas (2001) since allocation does not depend on x_{n+1}.

Figure 2 clearly shows the effect of not knowing p on the efficiency of the designs, which is calculated with $p = 0.75$ in Figure 1. This figure provides information for the choice of γ.

6. The Distribution of Loss

Although the average properties of the design are of interest, it is also important to look at the individual trials. We only consider in detail the results for the Bayesian rule for $\gamma =$ 0.03. Figure 2 shows that the average loss from this Bayesian rule lies between those for the two classical rules R and D.

Figure 3 shows boxplots of the distributions of proportions r and losses for 1000 simulations of an 800-trial design using the Bayes rule with regularization and with $\gamma = 0.03$. As the trial progresses, the average value of r converges to around 0.75, with a distribution of values with a gradually decreasing standard deviation. The values of the loss in the boxplots of the right-hand panel of Figure 3 have a mean which slowly increases with the number of patients. The distribution becomes markedly more dispersed from $n = 100$ to $n = 200$. Thereafter, the dispersion increases slightly with n. Also evident in the two panels is the effect of an outlying trial for which regularization is invoked at least up to $n = 500$.

The distribution of loss for several rules is explored by Atkinson (2003) for unskewed designs. But, due to the adaptive nature of the designs considered here, the distributions

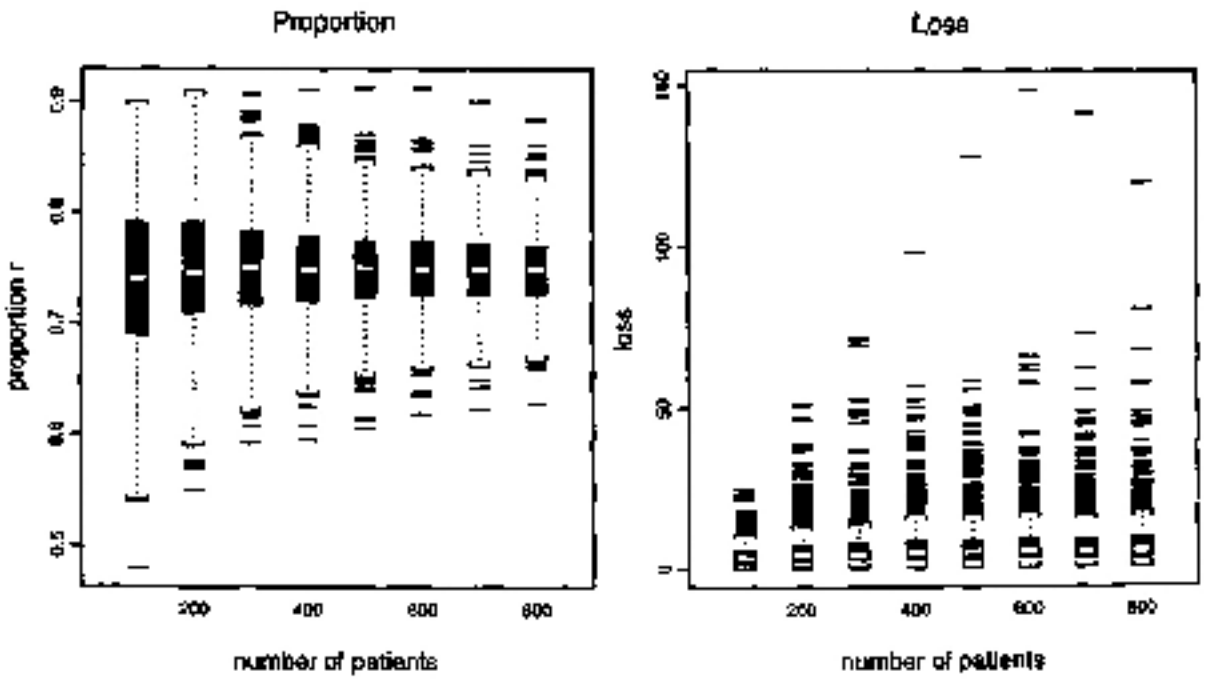

Figure 3. One thousand individual regularized adaptive designs. Left-hand panel: boxplots of proportion of patients receiving the better treatment. Right-hand panel, loss L_n. Bayesian rule with $\gamma = 0.03$, $q = 5$, $\sigma = 0.8$.

are long tailed, because of the occasional occurrence of trials for which the proportion of patients receiving the first treatment is far *from* the target. Figure 3 shows the effect of one such trial. Such long tails are a general feature of adaptive methods: Frequent good behavior is offset by occasional trials in which a large number of patients is needed to obtain good parameter estimates.

7. A Numerical Example: Fluoxetine Hydrochloride

As an example of the construction of designs, we use part of the data from Tamura et al. (1994) on the treatment of depressive patients. In order to correspond to our formalization of the model (1), the first two variables are an indicator for the treatment: column 1 control; column 2 fluoxetine. Column 3 is a categorical covariate with values -1 and 1 dividing the patients by sleep dysfunction before the trial (0 and 1 in the paper). Column 4 contains the initial values of HAMD_{17}, a measure of depression, for each patient, from which the mean value of 21.7045 has been subtracted. The fifth column, the response, is the negative of the change in HAMD_{17}. Because HAMD_{17} is measured on a 53-point scale, we treat it as a continuous variable. Large values are desired.

The data are summarized by the matrix in Table 3. The upper-left 4×4 matrix is $G_n^T G_n$, the right-hand column (and the bottom row) are the sufficient statistics $G_n^T y$, and the bottom-right element is $\sum y_i^2$. There are 88 observations since one observation in the original dataset does not have a response. This matrix contains all the information needed to fit the model and estimate and test its parameters.

The estimated treatment difference $\bar{a}_1 - \bar{a}_2 = -3.795$, so the treatment seems to have decreased depression. The t value for this effect is 2.55, with a significance of 1.6%; the treatment is effective if any effect of the design on inference is ignored. The adaptive scheme should preferentially allocate treatment 2. For these data, the mean square estimate of the standard deviation is $s = 6.95$. If we take T in Section 5.1 slightly less as 4, then the skewing proportion $p = \Phi(-3.795/4) = 0.1714$. Interest is in one contrast of the α so that A in (15) becomes the vector $a = (0.1714 \ \ -0.8286 \ \ 0 \ \ 0)^T$.

Calculation of the variance $d_A(j,.)$ in (A.5) also requires g_{n+1}, the vector of proposed treatment allocation and covariates for the next patient. Suppose the patient scores 1 on sleep dysfunction and has an initial HAMD_{17} score of 16, which, after removing the mean of 21.7045 becomes -5.7045. If treatment 1 is to be allocated, $g_{n+1} = (1 \ \ 0 \ \ 1 \ \ -5.7045)^T$, whereas, if treatment 2 is allocated $g_{n+1} = (0 \ \ 1 \ \ 1 \ \ -5.7045)^T$. Then, from (A.5)

$$d_A(1,.) = 0.002205 \quad \text{and} \quad d_A(2,.) = 0.017388. \tag{18}$$

The probability of allocating treatment 1 using the biased-coin rule (A.6) is 0.0256. This minute probability arises because analysis of the results yields $p = 0.171$ whereas Table 3 shows that 43 of the 88 patients received treatment 1.

The Bayesian rule (17) adjusts these probabilities. The right-hand panel of Figure 1 shows that for n around 90 and $\gamma = 0.1$ the rule is starting to move toward randomness and lack of bias rather than solely emphasizing balance. Using the values of $d_A(j,.)$ in (18) yields a probability of 0.151 of allocating treatment 1, slightly less than the value of 0.171 for this vector a when the design is balanced over treatment allocation and covariates.

8. Discussion

Our optimum design theory is for regression models with responses that are either normal or can be made so by transformation. Our results also cover generalized linear models where the treatment effects are sufficiently small that the effect on the design of the iterative weights used in parameter estimation can be ignored (Cox, 1988). Thus, survival times are covered by our theory.

Delayed responses are readily incorporated, since they do not affect the covariate information in G_n and g_{n+1}. The parameters α are estimated using those responses that are available when the allocation is made.

The parameter estimates $\hat{\alpha}$ are derived assuming regression models with independent errors. As each allocation depends on the earlier responses, the observations are not independent and the likelihood is complicated. A similar strategy of substitution of parameter estimates is used by Rosenberger et al. (2002). The asymptotic normality of the parameter estimates is not affected by the adaptive nature of the design (Bai, Hu, and Rosenberger, 2002) and, for responses modeled by the exponential family, the optimum adaptive designs obtained by sequential use of parameter estimates are indeed optimum (Antognini and Giovagnoli, 2004).

Inference, following adaptive allocation, is an important issue, typically discussed for binary responses. Inference for the adaptive ECMO trial is explored by Begg (1990) and discussants. An exact two-sample permutation test is proposed by Wei (1988), with the asymptotic version given by Rosenberger (1993). Exact confidence intervals for treatment difference and odds ratios are obtained by Wei et al. (1990). Estimation for an adaptive design is discussed by Rosenberger and Sriram (1997) and optimality, variability, and power were explored by Hu and Rosenberger (2003). Asymptotic normality of the relevant statistics is derived by several authors including Bandyopadhyay and Biswas (2001) for normal responses. These and other results are discussed in greater detail in the last three chapters of Rosenberger and Lachin (2002). The nonparametric results holding for binary responses also hold when the response, as here, is normally distributed.

For their adaptive trial, Tamura et al. (1994) simulate the adaptive design 500,000 times to find the distribution of the test statistic, which took 20 minutes of computer time. We recommend a similar procedure for the analysis of our

Table 3

Fluoxetine data: summary after 88 trials. The upper-left 4×4 matrix is $G_n^T G_n$, the right-hand column (and the bottom row) are the sufficient statistics $G_n^T y$, and the bottom-right element is $\sum y_i^2$.

$$\begin{pmatrix} 43 & 0 & 1 & 16.7 & 302 \\ 0 & 45 & -1 & -16.7 & 479 \\ 1 & -1 & 88 & 24 & -39 \\ 16.7 & -16.7 & 24 & 1074 & 200.75 \\ 302 & 479 & -39 & 200.75 & 11389 \end{pmatrix}.$$

designs. Our simulations of designs using rule A with $p = 0.75$ produce t-tests with a distribution that, even for $n = 50$, we are not able to distinguish from the distribution coming from nonadaptive designs. In the same spirit, Yao and Wei (1996) show that the issue of efficiency is not critical for inference procedures with an adaptive design. And, of course, we can have a significant ethical gain when a suitable adaptive design is used.

Résumé

Des schémas adaptatifs sont utilisés dans des essais cliniques de Phase III pour biaiser l'attribution des traitements en faveur des meilleurs. Nous utilisons la théorie des schémas optimaux pour établir une procédure bayésienne de tirage au sort biaisé pour des schémas séquentiels avec réponses continues. La performance de ces schémas adaptatifs est étudiée à la fois numériquement et théoriquement. Des propriétés importantes sont la perte et la proportion de sujets alloués au meilleur traitement.

References

Antognini, A. B. and Giovagnoli, A. (2004). On the large sample optimality of sequential designs for comparing two treatments. Submitted.

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

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

Atkinson, A. C. (2003). Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Applied Statistics* **52**, 261–278.

Bai, Z. D., Hu, F., and Rosenberger, W. F. (2002). Asymptotic properties of adaptive designs for clinical trials with delayed response. *Annals of Statistics* **30**, 122–139.

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.

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

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

Connor, E. M., Sperling, R. S., Gelber, R., Kiselev, P., Scott, G., O'Sullivan, M. J., VanDyke, R., Bey, M., Shearer, W., Jacobson, R. L., Jimenez, E., O'Neill, E., Bazin, B., Delfraissy, J., Culnane, M., Coombs, R., Elkins, M., Moye, J., Stratton, P., and Balsley, J. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *New England Journal of Medicine* **331**, 1173–1180.

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

Hu, F. and Rosenberger, W. F. (2003). Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* **98**, 671–678.

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

Rosenberger, W. F. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *Annals of Statistics* **21**, 2098–2107.

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

Rosenberger, W. F. and Sriram, T. N. (1997). Estimation for an adaptive allocation design. *Journal of Statistical Planning and Inference* **59**, 309–319.

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

Tamura, R. N., Faries, D. F., 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.

Wei, L. J. (1988). Exact two-sample permutation tests based on the randomized play-the-winner rule. *Biometrika* **75**, 603–606.

Wei, L. J., Smythe, R. T., Lin, D. Y., and Park, T. S. (1990). Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association* **85**, 156–162.

Yao, Q. and Wei, L. J. (1996). Play the winner for phase II/III clinical trials. *Statistics in Medicine* **15**, 2413–2423.

Zelen, M. and Wei, L. J. (1995). Foreword. In *Adaptive Designs*, N. Flournoy and W. F. Rosenberger (eds). Hayward, California: IMS.

Received October 2003. *Revised May* 2004. *Accepted May* 2004.

Appendix

Models and Variances

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

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

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

Consider the $t-1$ linear combinations $L^T\alpha$, where L^T is $(t-1) \times t$. The matrix of coefficients L^T is augmented by a $(t-1) \times q$ matrix of 0's as $A^T = (L^T \; 0)$ to reflect interest solely in the treatment parameters. The parameters θ of the prognostic factors in the linear model (1) are thus treated as nuisance parameters.

The covariance matrix of these estimated coefficients is

$$\mathrm{var}[A^T\hat{\alpha}] = \sigma^2 A^T\left(G_n^T G_n\right)^{-1} A, \tag{A.2}$$

where σ^2 is the variance of the errors, assumed additive in (A.1). Here, as elsewhere, we have suppressed the subscript n when it is not of importance. When there are two treatments

and the linear combination (2) is to be estimated, the matrix A becomes

$$A^T = (p - (1-p) \quad 0 \cdots 0), \tag{A.3}$$

a vector with $q + 1$ elements. The variance (A.2) is then a scalar.

Optimum Experimental Designs

D_A-optimum experimental designs for the linear regression model $E(Y) = G\omega$ maximize $|A^T(G^TG)^{-1}A|^{-1}$, the reciprocal of the determinant of the covariance matrix of the combinations of parameter estimates. That is, they minimize the generalized variance of these linear combinations and provide a normal theory confidence region of minimum volume. Such optimum designs can be constructed sequentially. After n trials, the matrix of allocations and prognostic factors is G_n. 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 for D-optimum designs maximizing $|G^TG|$ is that:

$$\begin{aligned}\left|G_{n+1}^T G_{n+1}\right| &= \left\{1 + g_{n+1}^T \left(G_n^T G_n\right)^{-1} g_{n+1}\right\} \left|G_n^T G_n\right| \\ &= \left\{1 + d(g_{n+1}, n)\right\} \left|G_n^T G_n\right|.\end{aligned} \tag{A.4}$$

where g_{n+1} combines the allocation indicator h_{n+1} for the $(n+1)$st patient and z_{n+1}, the extended vector of prognostic factors, is known for the new patient. In the iterative construction of D_A-optimum designs, with

$$\begin{aligned}d_A(j, n, x_{n+1}) = g_{n+1}^T \left(G_n^T G_n\right)^{-1} A \left\{A^T \left(G_n^T G_n\right)^{-1} A\right\}^{-1} \\ \times A^T \left(G_n^T G_n\right)^{-1} g_{n+1}, \quad (j = 1, \ldots, t),\end{aligned} \tag{A.5}$$

patient $n + 1$ would, in the absence of randomization, receive the treatment for which $d_A(j, n, x_{n+1})$ is a maximum, where j runs over all t treatments.

In this article, we use a matrix A which forces unequal allocation. At the optimum design, which allocates a fraction p_j of the patients to treatment j, all $d_A(j, n, x_{n+1})$ are equal. In nonoptimum designs, a larger value of $d_A(j, n, x_{n+1})$ indicates a treatment which is underrepresented. The original suggestion of Atkinson (1982) was to allocate treatment j with probability

$$\pi_n(j \mid x_{n+1}) = \frac{p_j d_A(j, n, x_{n+1})}{\sum_{s=1}^{t} p_s d_A(s, n, x_{n+1})}. \tag{A.6}$$