Adaptive Designs
IMS Lecture Notes - Monograph Series (1995) Volume 25

BIASED COIN DESIGNS: SOME PROPERTIES AND APPLICATIONS

BY JEFFREY R. EISELE*

CIBA-GEIGY AG

Abstract

A brief review of biased coin designs is presented. Some asymptotic properties of the adaptive biased coin designs of Wei (1978) and Eisele (1994) are given. Applications of biased coin designs to estimation and testing problems are also given.

1. Introduction: biased coin designs. In designing a clinical trial, the method of treatment allocation is a primary consideration. Because patients arrive sequentially from a population that may be very heterogeneous it is only in completion of the study that the characteristics of the patient population are known. In addition, the size of the study cannot be determined in advance. Because of this, much of the traditional experimental design methodology is inapplicable. Randomization has been used in treatment allocation due to the following advantages. First, it minimizes the possibility of selection bias which may occur if the experimenter is aware of which treatment the next patient will receive. A second advantage is freedom from accidental bias which could result, for example, if time trends are present in the

Received September 1992; revised April 1993.

*Research supported by the U.S. Army Research Office under DAAL-03-88-0122 and by NIH/NEI Grant T32-EY07119.

AMS 1991 subject classification. Primary 62L05, secondary 62E20.

Key words and phrases. Clinical trial, exponential families, invariance principle, martingale central limit theorem, restricted randomization, sequential allocation, sequential probability ratio test.

data. Finally, randomization can serve as a basis for inference. For a more detailed discussion of these advantages see Efron (1971). Another possible method for treatment allocation is the deterministic design ABABAB..., where 'A' and 'B' are two treatment groups. This design results in perfect balance but maximizes the possibility of selection bias.

In small sized experiments or interim analyses the use of complete randomization might result in a severe imbalance in treatment allocations. In fact, Pocock (1979) recommends using complete randomization only in trials with over 200 patients. Because of this, restricted randomization allocation schemes have been proposed which offer a compromise between complete randomization and the perfect balance guaranteed by a deterministic design. In this paper, the restricted randomization designs of Efron (1971), Wei (1978), and Eisele (1994) will be discussed. More detailed reviews of treatment allocation methods can be found in Pocock (1979), Simon (1979), and Kalish and Begg (1985). For a discussion of randomized clinical trials see Royall (1991).

In Section 2, the biased coin designs of Efron, Wei, and Eisele are presented. Some asymptotic properties of the Wei and Eisele coins are given in Section 3. Applications to estimation and testing problems are presented in Section 4.

2. Biased coin designs.

2.1. Efron's biased coin design. Efron (1971) introduced the biased coin design, BCD(η), to force an experiment to be balanced yet retain some randomization. The BCD(η) can be described as follows: Suppose that after ($k > 1$)patient assignments there are m_k patients assigned to treatment A and n_k patients assigned to treatment B. Let $S_k = m_k - n_k$. If $S_k = 0$, assign the next patient to treatment A with probability $1/2$; if $S_k < 0$, assign the patient to treatment A with probability η; if $S_k > 0$, assign the patient to treatment A with probability $1 - \eta$, $\eta \geq 1/2$. A criticism is that this design does not discriminate between large and small differences in treatment assignments or between large and small sized experiments. Pocock (1979) discusses guidelines for the selection of η.

2.2. Wei's adaptive biased coin design. Wei (1978) introduced the adaptive biased coin design to force an extremely unbalanced or a small-sized experiment to be balanced yet tend toward complete randomization as the size of the experiment increases. The adaptive biased coin design can be described as follows: Let $q$be a nonincreasing function from $[-1, 1]$ and $[0, 1]$ for which $q(0) = 1/2$. Then the $(k+1)$th patient

is assigned to treatment A with probability $q_k = q(S_k/k)$ and to treatment B with probability $1-q_k$. Wei recommends using $q(x) = (1-x)/2$ for its simplicity. Simon (1979) recommends the designs of Efron and Wei for stratified studies in which limited numbers of patients are expected in each stratum.

2.3. Doubly adaptive biased coin design. In many problems balance is not desired. This is the case especially in phase II clinical trials where estimation of response is of primary interest; see Pocock (1979). In fact, the desired allocation proportion may be unknown. Eisele (1994) introduced the doubly adaptive biased coin design for the case where the desired allocation proportion is unknown. The design is doubly adaptive in the sense that it takes account of both the current proportion of patients assigned to each treatment and a current estimate of the desired allocation proportion.

2.4. General model. Suppose patient responses $X_1, X_2, \ldots$ to treatment A and $Y_1, Y_2, \ldots$ to treatment B are independent random variables from d-dimensional standard exponential families. More formally, assume that

$$X_1, X_2, \ldots \sim f_\theta(x) = exp(\theta \cdot x - \psi(\theta))$$

and

$$Y_1, Y_2, \ldots \sim g_\omega(y) = exp(\omega \cdot y - \varphi(\omega)),$$

where

$$\theta = (\theta_1, \ldots, \theta_d)', \; x = (x_1, \ldots, x_d)', \; \omega = (\omega_1, \ldots, \omega_d)', \; y = (y_1, \ldots, y_d)';$$

$'$ denotes transpose and $\cdot$ denotes the inner product. Let

$$\mu = \mathbf{E}_\theta \mathrm{X} = \nabla\psi(\theta) \text{ and } \nu = \mathbf{E}_\omega \mathrm{Y} = \nabla\varphi(\omega),$$

where $\nabla\psi$ and $\nabla\varphi$ denote the gradient vectors, that is

$\nabla\psi = (\partial\psi/\partial\theta_1, \ldots, \partial\psi/\partial\theta_d)'$ and $\nabla\varphi = (\partial\varphi/\partial\omega_1, ..., \partial\varphi/\partial\omega_d)'$. Let

$$\overline{X}_{m_k} = (m_k)^{-1} \sum_{i=1}^{m_k} X_i \text{ and } \overline{Y}_{n_k} = (n_k)^{-1} \sum_{i=1}^{n_k} Y_i,$$

be the sample means. If the families are steep and the allocation rule

is noninformative, then the MLE of μ is $\overline{X}_{m_k}$ and the MLE of ν is $\overline{Y}_{n_k}$. Brown (1986) is recommended for more background on exponential families.

The goal of the allocation scheme is then to have $m_k/k = \rho$, where m_k/k is the proportion of patients assigned to treatment A at time k and $\rho \equiv \rho(\mu, \nu) : \mathbf{R}^{2d} \to [0,1]$ is the desired allocation proportion. To accomplish this, the allocation scheme is designed to sample the X population with probability less than (respectively, greater than) $\hat{\rho}_k$ when $m_k/k > \hat{\rho}_k$ (respectively, $m_k/k < \hat{\rho}_k$), where $\hat{\rho}_k \equiv \rho\left(\overline{X}_{m_k}, \overline{Y}_{n_k}\right) \in [0,1]$ is the current maximum likelihood estimate of ρ. This employs the same idea as Wei's adaptive biased coin design except that Wei fixes $\rho = 1/2$.

In addition, let

$$\sigma^2 = \rho'_{10}(\mu, \nu) \nabla^2 \psi(\theta) \rho_{10}(\mu, \nu),$$

and

$$\tau^2 = \rho'_{01}(\mu, \nu) \nabla^2 \varphi(\omega) \rho_{01}(\mu, \nu),$$

where $\nabla^2 \psi(\theta) = \mathrm{Cov}_\theta X$ and $\nabla^2 \varphi(\omega) = \mathrm{Cov}_\omega Y$, respectively, denote the Hessian matrices

$$\nabla^2 \psi(\theta) = (\partial^2 \psi / \partial \theta_\imath \partial \theta_\jmath)_{1 \le \imath \le d, 1 \le \jmath \le d}$$

and

$$\nabla^2 \varphi(\omega) = (\partial^2 \varphi / \partial \omega_\imath \partial \omega_\jmath)_{1 \le \imath \le d, 1 \le \jmath \le d},$$

and ρ_{10} and ρ_{01} are first partial derivatives of ρ.

2.5. The allocation rule. Let q be a function from $[0,1]^2 \to [0,1]$ such that the following four conditions hold:

(i) q *is jointly continuous,*
(ii) $q(r,r) = r$,
(iii) $q(p,r)$ *is strictly* $\searrow$ *in* p *and strictly* $\nearrow$ *in* r *on* $[0,1]^2$, and
(iv) q *has bounded derivatives in both arguments and*

$$q_{10}(\rho, \rho) = \partial q(x,y)/\partial x \mid_{x=\rho, y=\rho} \neq 0.$$

Let $\delta_1 = \cdots = \delta_{n_0} = 1$, $\delta_{n_0+1} = \cdots = \delta_{2n_0} = 0$, and

$$\delta_k = I\left\{U_k \leq q\left(\frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1}\right)\right\} \ \forall k \geq 2n_0 + 1,$$

where the U_k are independent identically distributed uniform$[0,1]$ random variables and are independent of $X_1, X_2, \ldots$ and $Y_1, Y_2 \ldots$. Then

$$m_k = \delta_1 + \cdots + \delta_k \text{ and } n_k = k - m_k.$$

3. Properties of adaptive biased coin designs. In this section asymptotic properties of the Wei and Eisele biased coins are presented. For Wei's biased coin, asymptotic normality of the difference between the number of patients assigned to each treatment group is established. In the proof it is shown that the difference may be magnified by appropriate constants so that the magnified difference is nearly a martingale. The approach is technically simpler than that given in Wei (1978) and delivers an invariance principle with little additional effort. In addition, the asymptotic distribution of the number of correct guesses for the assignment of the next patient, using the obvious strategy, is given as an application. The results presented are a summary of those given in Eisele and Woodroofe (1990). Smith (1984) obtained an invariance principle using similar methods.

For the Eisele biased coin, a strong law of large numbers and a central limit theorem are stated. An invariance principle follows from the central limit theorem with little additional effort. The results presented are a summary of the main results in Eisele (1994) and Eisele and Woodroofe (1994).

3.1. The Wei coin. Let $X_k = +1$ denote assignment of the kth patient to treatment A and -1 for an assignment to treatment B and let $U_1, U_2, \ldots$ denote i.i.d. uniformly distributed random variables. Then

$$X_k = 2I\left\{U_k \leq q\left(\frac{S_{k-1}}{k-1}\right)\right\} - 1 \ \forall k = 1, 2, \ldots,$$

where $S_k = X_1 + \cdots + X_k = m_k - n_k, \forall k = 1, 2, \ldots$.

THEOREM 1. *(Central Limit Theorem) Let* $\alpha = -2q'(0)$. *Then*

$$\frac{S_n}{\sqrt{n}} \Rightarrow N\left(0, \frac{1}{1+2\alpha}\right) \text{ as } n \to \infty.$$

REMARK 1. α is a measure of the steepness of the q-function near zero. A steeper q-function forces balance faster and thus the asymptotic variance of $S_n/\sqrt{n}$ decreases as α increases.

OUTLINE OF THE PROOF. The idea is to magnify S_n by the appropriate sequence of constants so that the magnified value is nearly a martingale. Let $m = [\alpha] + 1$, where $[x]$ denotes the greatest integer which is $\leq x$ for $-\infty < x < \infty$; let $c_k = 1$ for $k = 1, \ldots m$; and let

$$c_n^{-1} = \prod_{k=m+1}^{n} \left(1 - \frac{\alpha}{k-1}\right) \; \forall \, n \geq m+1.$$

For $n = 1, 2, \ldots$, let

$$Y_n = \sum_{k=1}^{n} c_k \left(X_k - \mu_k\right),$$

where $\mu_k = \mathbb{E}\left(\mathrm{X_k/F_{k-1}}\right)$ and $\mathcal{F}_k = \sigma\{U_1, \ldots, U_k\}$ is the σ-algebra representing the natural history. Then $\{Y_n, \mathcal{F}_n, n = 1, 2, \ldots\}$ is a martingale. Now use c_n as the sequence of magnifying constants and write

$$c_n S_n = c_m S_m + Y_n - Y_m + R_n \; \forall n \geq m,$$

where R_n is a remainder term of order $o\left(n^{\alpha+1/2}\right)$. Then

$$\frac{S_n}{\sqrt{n}} = \frac{c_n S_n}{\sqrt{n} c_n} = \frac{1}{\sqrt{n} c_n} \left\{Y_n - Y_m + R_n + c_m S_m\right\}.$$

Asymptotic normality now follows from the martingale central limit theorem. □

Let $\mathbb{B}$ denote a Brownian motion with drift 0 and variance 1 per unit time and let

$$\mathbb{S}_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}$$

for $0 \leq t \leq 1$ and $n = 1, 2, \ldots$ and where $[x]$ denotes the greatest integer which is $\leq x$. Then $\mathbb{S}_n$, $n \geq 1$, may be regarded as random elements in $D[0, 1]$.

THEOREM 2. *(Invariance Principle) As* $n \to \infty, \mathbb{S}_n \Rightarrow \mathbb{S}$ in $D[0,1]$, *where*

$$\mathbb{S}(t) = \frac{\mathbf{B}(t^{2\alpha+1})}{\sqrt{(2\alpha+1)t^{\alpha}}} \quad 0 < t \leq 1.$$

Using Theorem 2.5 of Durrett and Resnick (1978), the invariance principle follows from Theorem 1.

If an experimenter attempts to guess the assignments of patients to treatment and control, then an obvious strategy is to guess $X_k = +1$ or -1 accordingly as $S_{k-1} \leq 0$ or $S_{k-1} > 0$. Denote by N_n the number of correct guesses among the first n assignments. The following corollary follows from Theorem 2.

COROLLARY 1. $\lim_{n\to\infty} \mathbf{E}\left\{\frac{N_n - \frac{1}{2}}{\sqrt{n}}\right\} = \left(\frac{2}{\pi}(1+2\alpha)\right)^{1/2}$.

As an application of Corollary 2, if $\alpha = 1$ (e.g. $q(x) = (1-x)/2$) and $n = 100$, then an experimenter can expect to make about 63.8 correct guesses with a biased coin design and 50 with complete randomization.

3.2. The doubly adaptive biased coin. In addition to conditions (i)-(iv) imposed on q, the following two conditions on ρ are needed.

(v) *There are positive constants* C *and* r *for which*

$$\frac{1}{\rho} + \frac{1}{1-\rho} \leq C\{\| \mu \|^r + \| \nu \|^r\}.$$

(vi) *For sufficiently small* $\epsilon > 0$, ρ *is twice continuously differentiable on the set*

$$R = \{(x,y) : \| x - \mu \| \leq 2\epsilon, \| y - \nu \| \leq 2\epsilon\}.$$

THEOREM 3. *(Strong Law of Large Numbers) Let* $S_k = m_k - k\rho$, $\forall\ k = 2n_0 + 1, 2n_0 + 2, \ldots$. *Under conditions (i)–(vi),*

$$\lim_{n\to\infty} \frac{S_n}{n} = 0 \quad w.p.1.$$

Complete details of the proof of Theorem 3 are given in Eisele (1994). A different proof is also given in Eisele (1990) for the case of normally distributed patient responses.

THEOREM 4. *(Central Limit Theorem) Let* $S_k = m_k - k\rho, \forall\ k = 2n_0 + 1,\ 2n_0 + 2....$ *Under conditions (i)-(vi),*

$$\begin{aligned} \frac{S_n}{\sqrt{n}} &= \frac{1}{\sqrt{n}}(m_n - n\rho) \\ &\Longrightarrow N\left(0, \frac{\rho(1-\rho)}{2\alpha+1} + \frac{2\gamma^2}{(\alpha+1)(2\alpha+1)}\left(\frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho}\right)\right), \end{aligned}$$

where $\alpha = -q_{10}(\rho,\rho)$, $\gamma = q_{01}(\rho,\rho)$, *and* q_{01} *and* q_{10} *are the partial derivatives of* q.

SKETCH OF THE PROOF. The basic idea is the same as for the Wei coin although difficulties arise because there are two terms in the conditional expectation of S_k. The additional term is due to estimating ρ. In order to get a martingale, each of these terms must be magnified by different sequences of constants. This requires looking at a vector of these two terms and then magnifying this vector by a matrix of constants. □

THEOREM 5. *(Invariance Principle) Let* $\mathbb{S}_n(t) = \frac{1}{\sqrt{n}} S_{[nt]}$. *Then, as* $n \to \infty$

$$\mathbb{S}_n(t) \Rightarrow t^{-\alpha}\mathbb{S}\left(t^{\alpha+\frac{1}{2}}\right) \text{ in } D[0,1],$$

where $\mathbb{S}(t)$ *is a one dimensional Brownian motion with variance*

$$\frac{\rho(1-\rho)}{2\alpha+1} + \frac{2\gamma^2}{(\alpha+1)(2\alpha+1)}\left(\frac{\sigma^2}{\rho} + \frac{\tau^2}{1-\rho}\right)$$

per unit time.

Complete details of the proofs of Theorems 4 and 5 are given in Eisele and Woodroofe (1994).

4. Applications.

4.1. Estimation: Behrens-Fisher problem (Robbins, Simons, and Starr, 1967). Suppose it is desired to design a sequential procedure,

with a randomized allocation scheme, for the fixed width interval estimation of the difference of the means of two populations. Minimizing the total size of the experiment can be accomplished by designing the sequential procedure so that patients are allocated to the two treatments in the correct proportions.

More formally, assume that $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are independent random variables for which

$$X_1, X_2, \ldots \sim N\left(\mu, \sigma^2\right) \text{ and } Y_1, Y_2, \ldots \sim N\left(\nu, \tau^2\right),$$

where the four parameters μ, ν, σ and τ are unknown. Here, $X_1, X_2, \ldots$ denote responses to treatment A and $Y_1, Y_2, \ldots$ denote responses to treatment B. These could be, for example, blood pressure readings. Then, the correct allocation proportions for minimizing the total sample size and retaining preassigned coverage probability and interval width are [see Robbins, Simons, and Starr (1967) or Eisele (1990)] $\sigma/(\sigma+\tau)\times k$ to treatment A and $\tau/(\sigma+\tau) \times k$ to treatment B. Thus,

$$\rho\left(\sigma^2, \tau^2\right) = \frac{\sigma}{\sigma+\tau}.$$

Taking

$$\hat{\sigma}^2_{m_k} = (m_k - 1)^{-1} \sum_{i=1}^{m_k} \left(X_i - \overline{X}_{m_k}\right)^2$$

and

$$\hat{\tau}^2_{n_k} = (n_k - 1)^{-1} \sum_{i=1}^{n_k} \left(Y_i - \overline{Y}_{n_k}\right)^2$$

to be the usual estimates of σ^2 and τ^2 gives

$$\hat{\rho}_k = \frac{\hat{\sigma}_{m_k}}{\hat{\sigma}_{m_k} + \hat{\tau}_{n_k}}.$$

The sequential procedure can now be described as follows: to start, take $n_0 \geq 2$ observations on X and on Y. Then, if at any stage there are m_k observations on X and n_k on Y, with $k = m_k + n_k \geq 2n_0$, take observation $k+1$ on X if

$$U_{k+1} \leq q_{k+1} \equiv q\left(\frac{m_k}{k}, \hat{\rho}_k\right).$$

Otherwise, take observation $k+1$ on Y.

If the desired width and coverage probability of the confidence interval I, for $\theta = \mu - \nu$, are $2h$ and α, respectively, and if the constant a is defined by $2\Phi(a) - 1 = \alpha$, where Φ denotes the $N(0,1)$ distribution function, then a possible stopping rule for the sequential procedure is: stop after N observations, where

$$N = \inf\left\{k \geq 2n_0 : \frac{\hat{\sigma}^2_{m_k}}{m_k} + \frac{\hat{\tau}^2_{n_k}}{n_k} \leq \left(\frac{h}{a_k}\right)^2\right\}$$

and $\{a_k\}$ is a given sequence of positive constants such that $a_k \to a$ as $k \to \infty$. Three other variations of this stopping rule are also given in Eisele (1990).

THEOREM 6. *Under conditions (i)–(iii) on q, as $h \to 0$, for all μ, ν, σ and τ,*

(i) $$\frac{N}{\left(\frac{a}{h}\right)^2 (\sigma+\tau)^2} \to 1 \text{ a.s.},$$

(ii) $$\mathbf{E}\left(\frac{\mathrm{N}}{\left(\frac{\mathrm{a}}{\mathrm{h}}\right)^2 (\sigma+\tau)^2}\right) \to 1 \text{ (asymptotic efficiency)},$$

(iii) $\mathbf{P}(\theta \in \mathrm{I}) \to \alpha$ *(asymptotic consistency).*

The simulation results given in Table 1 illustrate the asymptotic properties of the doubly adaptive biased coin design for the case of normally distributed patient responses. The values

$$\alpha = .95, \quad a = 1.96, \quad n_0 = 5, \quad a_k^2 = \left(\frac{k+4}{k-4}\right) a^2,$$

were selected. For these values, the optimal sample size, for normally distributed data, becomes

$$n^* = \left(\frac{1.96}{h}\right)^2 (\sigma+\tau)^2.$$

The *q-function*, $q(x,y) = [1-(1/y-1)x]_+$, was selected for its simplicity.

Table 1.
Simulation Results for 2,000 Trials for Normally Distributed Subject Responses: Expected Sample Sizes, Allocation Ratios, and Coverage Probabilities for $\sigma/\tau = 1, 1/2$, and $1/4$.

	$\sigma/\tau = 1$				$\sigma/\tau = 1/2$				$\sigma/\tau = 1/4$			
n^*	m_N	n_N	$\frac{m_N}{n_N}$	CP	m_N	n_N	$\frac{m_N}{n_N}$	CP	m_N	n_N	$\frac{m_N}{n_N}$	CP
10	7.6	7.8	.974	.979	6.1	9.7	.629	.979	5.1	11.2	.455	.975
20	12.1	12.6	.960	.957	8.4	16.5	.509	.958	5.8	19.4	.299	.960
30	17.0	17.4	.977	.950	11.3	22.9	.493	.950	7.1	27.8	.255	.954
40	22.0	22.4	.982	.948	14.3	29.6	.483	.948	8.6	35.9	.240	.951
50	27.0	27.4	.985	.947	17.6	36.5	.482	.947	10.2	43.9	.232	.950
60	32.1	32.5	.988	.949	20.7	43.7	.474	.948	12.0	51.6	.233	.947
70	36.9	37.5	.984	.948	24.2	49.7	.487	.947	13.7	60.2	.228	.949
80	42.4	42.7	.993	.950	27.9	56.6	.493	.950	16.1	68.8	.234	.951
90	47.2	47.6	.992	.950	30.9	63.9	.484	.950	17.8	76.3	.233	.949
100	52.7	52.8	.998	.951	34.2	70.5	.490	.950	19.9	83.8	.237	.949
125	65.0	65.2	.997	.951	42.7	87.2	.490	.951	25.0	104.4	.239	.950
150	77.8	77.8	1.000	.951	51.1	103.8	.492	.951	30.1	124.5	.242	.950
175	90.1	90.2	.999	.952	59.3	120.5	.492	.951	35.3	144.7	.244	.951
200	102.8	103.9	.989	.952	67.9	137.1	.495	.951	40.2	164.6	.244	.951

REMARK 2. Although the *q-function* selected is not strictly increasing in the second argument on $(0,1)^2$, the proof of the strong law of large numbers given in Eisele (1990) for the case of normally distributed patient responses only requires q to be strictly increasing near the diagonal.

REMARK 3. For normally distributed observations,

$$P\left(\left|\overline{X}_{m_N} - \overline{Y}_{n_N} - \theta\right| \leq h\right) = 2\mathbf{E}\Phi\left(\frac{\mathrm{h}}{\left(\frac{\sigma^2}{\mathrm{m_N}} + \frac{\tau^2}{\mathrm{n_N}}\right)^{1/2}}\right) - 1.$$

Estimates of the coverage probabilities are found by estimating the above expectation using the simulated values of N.

REMARK 4. In terms of allocation proportions, the doubly adaptive biased coin appears to be performing as desired for the three examples given. The total sample size is roughly between 4 and 6 observations larger than the optimal total sample size. This results in high coverage probabilities for small values of n^* where there is oversampling.

REMARK 5. For more details on this sequential procedure, including derivations, other stopping rules, and simulation results, see Eisele (1990).

4.2. Testing: Sequential probability ratio test (Robbins and Siegmund, 1974). Suppose in a clinical trial where patients can be assigned

to one of two treatments A and B, it is desired to design a sequential procedure to minimize the number of patients assigned to the inferior treatment. Robbins and Siegmund (1974) derived a sequential procedure for which the error probabilities are essentially independent of the allocation rule used. In the application presented here, two allocation rules are proposed for the Robbins-Siegmund procedure using adaptive biased coin designs.

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ denote independent random variables for which

$$X_1, X_2, \ldots \sim N(\mu, 1) \text{ and } Y_1, Y_2, \ldots \sim N(\nu, 1),$$

where μ and ν are unknown parameters. Here, X_k and Y_k denote responses to treatments A and B, respectively. Let $\theta = \nu - \mu$ denote the mean difference in response. It is desired to test the hypothesis $\mathrm{H}_0 : \theta > 0$ vs. $\mathrm{H}_1 : \theta < 0$. Let $i_k = m_k n_k / k$, $\hat{\theta}_k = \overline{Y}_{n_k} - \overline{X}_{m_k}$, and $Z_k = i_k \hat{\theta}_k$.

With these definitions, a sequential probability ratio test for H_0 vs. H_1 is as follows: for a given $a > 0$, let

$$t_a = \inf \{k \geq 2 : |Z_k| > a\}$$

and accept H_0 or H_1 according as $Z_t \geq a$ or $\leq -a$, respectively.

For a given sequential design and $a > 0$, one expects to find $Z_k \cong i_k \theta$ for $k = 1, 2, \ldots$ and $Z_t^2 \cong a^2$ on $|Z_t| > a$. Therefore, one expects to find $i_t^2 \cong b^2 \equiv a^2/\theta^2$ on the event $\{|Z_t| > a\}$.

Suppose the cost of allocating to the X population is $g(\theta)$ and to the Y population is $h(\theta)$. Then the total cost is $C = g(\theta) m_k + h(\theta) n_k$ and the following minimization problem arises: given $b, g, h > 0$ minimize

$$C = g(\theta) m + h(\theta) n$$

subject to

$$\frac{mn}{m+n} = b.$$

The solution to the minimization problem is to let

$$m = b\left(1 + \sqrt{h(\theta)/g(\theta)}\right) \text{ and } n = b\left(1 + \sqrt{g(\theta)/h(\theta)}\right).$$

Thus, the desired proportion of patients allocated to treatment A is

$$\rho(\theta) = \frac{m}{m+n} = \frac{1}{1 + \sqrt{g(\theta)/h(\theta)}}.$$

Two allocation rules are presented; they differ in the nature of the q-function. The first allocation rule (Rule 1) uses a discontinuous allocation function and the second (Rule 2) uses a doubly adaptive biased coin. Let

$$q_{1,k}\left(\hat{\rho}_{k-1}, \frac{m_{k-1}}{k-1} - \hat{\rho}_{k-1}\right) = \begin{cases} (1+\hat{\rho}_{k-1})/2 & \text{if } \frac{m_{k-1}}{k-1} < \hat{\rho}_{k-1} \\ 1/2 & \text{if } \frac{m_{k-1}}{k-1} = \hat{\rho}_{k-1} \\ \hat{\rho}_{k-1}/2 & \text{if } \frac{m_{k-1}}{k-1} > \hat{\rho}_{k-1} \end{cases}$$

and

$$q_{2,k}\left(\frac{m_{k-1}}{k-1}, \hat{\rho}_{k-1}\right) = \left[1 - \left(\frac{k-1}{m_{k-1}} - 1\right)\hat{\rho}_{k-1}\right]_+,$$

where

$$\hat{\rho}_k = \rho\left(\hat{\theta}_k\right), \ k \geq 2.$$

Allocate the k^{th} patient to the treatment A if

$$U_k \leq q_{i,k-1} \text{ for } i = 1, 2 \text{ and } k \geq 3,$$

where the U_k are independent identically distributed uniform random variables that are independent of (X_k, Y_k).

The simulation results given in Table 2 were obtained for the error probability, the expected sample size on the inferior treatment, and the expected total sample size for allocation Rules 1 and 2 stated above, the allocation rule of Robbins and Siegmund (RS) and complete randomization (CR). The value $a = 6$ was selected giving an error probability of approximately 0.05 for $\theta = 0.25$. The cost functions

$$g(\theta) = \begin{cases} 1 & \text{if } \theta \leq 0 \\ 1 + i_k\theta & \text{if } \theta > 0 \end{cases} \quad \text{and} \quad h(\theta) = \begin{cases} 1 & \text{if } \theta \geq 0 \\ 1 - i_k\theta & \text{if } \theta < 0 \end{cases}$$

were selected. In the simulations, $\hat{\theta}_k$ is substituted for θ.

REMARK 6. See Remark 2.

REMARK 7. The functions $g(\theta)$ and $h(\theta)$ were selected so that the allocation probabilities would stay within reasonable bounds.

REMARK 8. The error probabilities appear to be approximately independent of the allocation rule used.

Table 2.
Simulation Results for 2,000 Trials for Normally Distributed Subject Responses: Expected Sample Sizes and Error Probabilities for Allocation Rules 1 and 2, the Allocation Rule of Robbins and Siegmund (RS), and Complete Randomization (CR).

	Rule 1			Rule 2			RS			CR			
θ	m	m+n	error	m	m+n	error	m	m+n	error	m	m+n	error	error[1]
0.05	77.4	162.7	0.366	73.0	163.2	0.344	94.2	211.2	0.345	76.0	152.2	0.350	0.354
0.10	67.2	153.2	0.219	63.4	151.8	0.207	75.7	197.8	0.231	71.7	143.1	0.242	0.231
0.15	56.0	134.7	0.124	54.9	137.8	0.121	60.0	181.7	0.142	61.8	123.6	0.127	0.142
0.20	46.9	118.1	0.086	45.0	116.8	0.080	45.9	167.2	0.078	53.6	107.0	0.070	0.083
0.25	38.1	100.9	0.047	37.5	102.2	0.036	37.1	154.1	0.041	45.8	91.6	0.049	0.047
0.30	32.4	89.6	0.020	31.8	89.1	0.024	30.0	143.1	0.025	40.8	81.5	0.021	0.027
0.40	25.3	72.5	0.009	24.6	71.8	0.007	20.3	121.5	0.004	31.8	63.8	0.007	0.008
0.50	19.1	57.2	0.002	18.8	56.8	0.002	15.6	107.9	0.002	25.3	50.6	0.001	0.002
0.75	12.2	38.0	0.000	12.5	39.2	0.000	9.7	84.6	0.000	17.3	34.5	0.000	0.000
1.00	9.4	29.3	0.000	9.3	29.6	0.000	7.2	70.6	0.000	13.1	26.1	0.000	0.000

[1] Wald approximation.

REMARK 9. For small θ, Rule 2 appears to have smaller allocations to the inferior treatment than Rule 1. Otherwise, their performances are similar. Both Rules 1 and 2 have smaller total sample sizes than the Robbins-Siegmund rule and for small θ allocate fewer observations to the inferior treatment.

REMARK 10. For all but some small values of θ, the decrease in the number of patients on the inferior treatment using allocation Rules 1 and 2 in comparison to complete randomization is roughly the same as the increase in the total sample size over complete randomization.

REMARK 11. Recent papers by Melfi (1992, 1994) have developed a renewal theory to approximate error probabilities and expected sample sizes using Efron type biased coins which might be useful in analyzing designs using Rule 1.

REMARK 12. For discussions of adaptive treatment allocation methods which incorporate accumulating information for allocating the best treatment to the most patients see Simon (1977) and Ware (1989).

Although the presentation of the doubly adaptive biased coin design given here suggests applications in clinical trials, more realistic applications may be in manufacturing.

Acknowledgments. I would like to thank Michael Woodroofe for introducing me to the topics discussed in this paper and the National Science Foundation for providing support to attend the AMS-IMS-SIAM Adaptive Designs Conference in South Hadley, MA, July 19-24, 1992.

References

BROWN, L. (1986). *Fundamentals of Statistical Exponential Families.* Hayward, CA: Institute of Mathematical Statistics.

DURRETT, R. and RESNICK, S. (1978). Functional central limit theorems for dependent random variables. *The Annals of Probability* **6** 829–847.

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

EISELE, J. (1990). An adaptive biased coin design for the Behrens-Fisher problem. *Sequential Analysis* **9** 343–359.

EISELE, J. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference* **38** 249–262.

EISELE, J. and WOODROOFE, M. (1990). A martingale proof of normality for Wei's biased coin, Department of Statistics, University of Michigan, Ann Arbor, MI (technical report).

EISELE, J. and WOODROOFE, M. (1994). Central limit theorems for the doubly adaptive biased coin design. *The Annals of Statistics,* in press.

HALL, P. and HEYDE, C.C. (1980). *Martingale Limit Theory and Its Application.* New York: Academic Press.

KALISH, L. and BEGG, C. (1985). Treatment allocation methods in clinical trials: a review. *Statistics in Medicine* **4** 129–144.

MELFI, V. (1992). Nonlinear Markov renewal theory with statistical applications. *The Annals of Probability* **20** 753–771.

MELFI, V. (1994). Nonlinear renewal theory for Markov random walks. *Stochastic Processes and Their Applications,* in press.

POCOCK, S.J. (1979). Allocation of patients to treatment in clinical trials. *Biometrics* **35** 183–197.

ROBBINS, H. and SIEGMUND, D. (1974). Sequential tests involving two populations. *Journal of the American Statistical Association* **69** 132–139.

ROBBINS, H., SIMONS, G. and STARR, N. (1967). A sequential analogue of the Behrens-Fisher problem. *The Annals of Mathematical Statistics* **38** 1384–1391.

ROYALL, R.M. (1991). Ethics and statistics in randomized clinical trials (with discussion). *Statistical Science* **6** 52–88.

SIMON, R. (1977). Adaptive treatment assignment methods and clinical trials. *Biometrics* **33** 743–749.

SIMON, R. (1979). Restricted randomization designs in clinical trials. *Biometrics* **35** 503–512.

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

WARE, J.H. (1989). Investigating therapies of potentially great benefit: ECMO (with discussion). *Statistical Science* **4** 298–340.

WEI, L.J. (1978). The adaptive biased coin design for sequential experiments. *The Annals of Statistics* **6** 92–100.

BIOMETRICS
CIBA-GEIGY AG
CH-4002
BASEL, SWITZERLAND