The Annals of Statistics
1984, Vol. 12, No. 3, 1018–1034

PROPERTIES OF BIASED COIN DESIGNS IN SEQUENTIAL CLINICAL TRIALS

BY RICHARD L. SMITH

Imperial College, London

Martingale methods and the martingale invariance principle are used to derive central limit theorems and related results for biased coin designs of the kind previously studied by Efron, Wei and many others. The results are applied to the study of selection bias. The method is developed for the simplest two-treatment case and then extended, first to the case of several treatments, and secondly to the case of two treatments with prognostic factors.

1. Introduction. In conducting a clinical trial it is necessary to have some rule for allocating patients to the different treatments under test. It is desirable that the trial should be balanced, not only with respect to the overall assignment of patients to treatments but also with respect to the various prognostic factors, such as age, sex and major indicators of clinical condition, which are always taken into account in conducting such a trial. On the other hand, randomisation is desirable for a number of reasons including the selection bias which may arise if the person in charge of selecting patients for the trial has advance knowledge of the treatment assignments.

Blackwell and Hodges (1957) established a procedure to minimise the selection bias under the constraint of perfect balance between two treatments, assuming the total number of patients is known in advance and ignoring prognostic factors. In practice, prognostic factors might be incorporated by dividing the patients into several strata, according to the prognostic factors, and applying the Blackwell-Hodges procedure to each. However the total number of patients in each stratum is rarely known in advance, nor is it clear that the constraint of perfect overall balance is necessary, as it may be sufficient for the trial to be nearly balanced in some sense.

Efron (1971) proposed a *biased coin design* in which the patients are allocated to two treatments according to the following rule. Suppose the first n patients are divided between two treatments T_1 and T_2 so that n_i are on treatment T_i ($i = 1, 2$). If $n_1 > n_2$ then patient $n + 1$ is allocated to T_2 with probability p and T_1 with probability $1 - p$, where $p \geq 1/2$. If $n_1 < n_2$ the assignment probabilities are reversed, and if $n_1 = n_2$ then patient $n + 1$ is assigned to either treatment with equal probabilities. As a specific example, Efron proposed taking $p = 2/3$. In his paper Efron discussed how to measure the imbalance and selection bias in such a design, and also considered the accidental bias arising from unknown

Received September 1982; revised February 1984.
AMS 1980 *subject classifications.* Primary 62L05; secondary 60F17.
Key words and phrases. Biased coin designs, sequential clinical trials, randomisation, martingale invariance principle, weak convergence.

disturbances to the experiment, and the possibility of randomisation inference based on permutation tests.

In recent years a number of alternative procedures have been proposed; we mention in particular the procedure analysed by Wei (1978a). An extension to $R > 2$ treatments was proposed by Wei (1978b). Procedures incorporating prognostic factors were proposed by Pocock and Simon (1975), Begg and Iglewitz (1980) and Efron (1980), and a general procedure applicable to any number of treatments and prognostic factors has been put forward by Atkinson (1982).

Our purpose here is to show how martingale methods may be used to study the procedure of Wei (1978a) and a number of generalisations of it. The results are applicable to a large class of designs; for example, they are used to obtain a number of properties of Atkinson's class of designs. The results are presented in the next three sections, Section 2 on two treatments with no prognostic factors, Section 3 on a generalisation to an arbitrary number of R treatments with no prognostic factors, and Section 4 on a procedure for two treatments which includes an arbitrary number of prognostic factors. The key to all our results is the martingale invariance principle, which is discussed in detail in the recent monograph of Hall and Heyde (1980).

2. Two treatments. Let δ_n be $+1$ or -1 according as the nth patient is assigned to T_1 or T_2. Let $D_0 = 0$, $D_n = \sum_{k\le n} \delta_k$. Wei (1978a) studied the class of treatment assignment rules of the form

$$P\{\delta_{n+1} = 1 \mid D_n\} = p(D_n/n) \tag{2.1}$$

where $p(x)$ is a nonincreasing function on $-1 \le x \le 1$ with $p(x) + p(-x) = 1$. The rule of Efron corresponds to making p constant on $[-1, 0)$ and on $(0, 1]$. Wei considered the case of a continuous function p and proved a number of properties. There are a number of arguments for using a continuous function p, including the fact that mean selection bias and mean accidental bias then tend to zero asymptotically. This is also consistent with the procedure proposed by Atkinson (1982). Under Atkinson's procedure, if one is interested in the mean responses of both treatments then T_1 is selected with probability n_2/n, while if one is only interested in mean treatment difference then T_1 is selected with probability $n_2^2/(n_1^2 + n_2^2)$. Here n_1 and n_2 are the numbers of patients on T_1 and T_2 respectively. It is natural to consider the general class of rules

$$P\{\delta_{n+1} = 1 \mid n_1, n_2\} = n_2^\rho/(n_1^\rho + n_2^\rho), \quad \rho > 0,$$

which corresponds (since $D_n = n_1 - n_2$) to (2.1) with

$$p(x) = (1 - x)^\rho/\{(1 + x)^\rho + (1 - x)^\rho\}. \tag{2.2}$$

One of Wei's results was to show that $n^{-1/2}D_n$ has asymptotically a normal distribution with mean zero and variance $\{1 - 4p'(0)\}^{-1}$. In the case (2.2), this variance becomes $(1 + 2\rho)^{-1}$. In the general case, we *define* the parameter ρ to be $-2p'(0)$; this is of course consistent with (2.2). Our first objective is to prove a theorem which generalises Wei's result to obtain weak convergence on a function space. The space is $C[0, 1]$, the space of continuous functions on $[0, 1]$, endowed with the topology generated by the metric

$$d_\rho(f, g) = \sup_{0<t<1} t^\rho |f(t) - g(t)|.$$

We shall denote this space with its topology by C_ρ. Note that the topology is different from the usual supremum norm topology, which in our notation is C_0. In later sections we shall generalise the notation to several dimensions; we denote by C_ρ^n the space of continuous functions on $[0, 1]^n$ with topology being the product topology generated by the metric d_ρ.

THEOREM 1. *Suppose the function* $p(\cdot)$ *is twice continuously differentiable, with second derivative uniformly bounded on* $[-1, 1]$. *For each* $n \geq 1$, *define a stochastic process* Z_n *by*

$$Z_n(t) = n^{-1/2}(1 + 2\rho)^{1/2}[D_k + (t - k/n)\delta_{k+1}], \quad k/n \leq t \leq (k + 1)/n. \tag{2.3}$$

Then Z_n *converges weakly on* C_ρ *to a limiting process* Z, *which is a zero-mean Gaussian process with continuous sample paths and covariance function*

$$E\{Z(s)Z(t)\} = s^{1+\rho}t^{-\rho}, \quad 0 \leq s \leq t \leq 1.$$

PROOF. First we state the following:

LEMMA 1. $E\{D_n^2\} \leq n$, $E|D_n| \leq n^{1/2}$ *and* $D_n/n \to_p 0$.

This follows from Lemmas 1–3 of Wei (1978a); Wei also proves convergence almost surely but convergence in probability suffices for the following.

Let $\mathscr{F}_n$ denote the sigma algebra generated by $\{\delta_1, \cdots, \delta_n\}$. From (2.1),

$$E\{D_{n+1} | \mathscr{F}_n\} = \alpha_n D_n + \beta_n$$

where $\alpha_n = 1 - \rho/n$, $\beta_n = 2[p(D_n/n) - p(0) - p'(0)D_n/n]$. Let n_0 be some fixed positive integer with $n_0 > \rho$, and define

$$A_n = \pi_{k=n_0}^{n-1} \alpha_k^{-1}, \quad B_n = \sum_{k=n_0}^{n-1} A_{k+1}\beta_k \quad \text{for} \quad n > n_0.$$

Then the process

$$M_n = A_n D_n - B_n, \quad n > n_0,$$

is an $\mathscr{F}_n$ martingale, i.e. $E\{M_{n+1} | \mathscr{F}_n\} = M_n$. It is readily verified that $n^{-\rho}A_n$ converges to a positive finite constant, A_0 say, as $n \to \infty$, a result we rewrite in the form

$$A_n \sim A_0 n^\rho, \quad n \to \infty \ (0 < A_0 < \infty). \tag{2.4}$$

The following form of the Martingale invariance principle is due to Brown (1971) (see also page 99 of Hall and Heyde, 1980).

Martingale invariance principle. Suppose M_n is a square integrable martingale with martingale differences $X_n = M_n - M_{n-1}(M_0 = 0)$. Let $s_n^2 = E\{M_n^2\} =$

$\sum_{k=1}^{n} E\{X_k^2\}$. Define ξ_n to be the random element of $C[0, 1]$ given by

$$\xi_n(t) = s_n^{-1}\{M_k + (s_{k+1}^2 - s_k^2)^{-1}(t\, s_n^2 - s_k^2)X_{k+1}\}, \quad s_k^2 \le ts_n^2 < s_{k+1}^2.$$

Suppose that

$$s_n^{-2} \sum_{k=1}^{n} E\{X_k^2 I(|X_k| > \varepsilon s_n)\} \to 0 \quad \forall \varepsilon > 0, \tag{2.5}$$

$$s_n^{-2} \sum_{k=1}^{n} X_k^2 \to_p 1 \tag{2.6}$$

as $n \to \infty$. Then ξ_n converges weakly (in the supremum topology) to a Wiener process ξ.

In order to apply this result, we must verify conditions (2.5) and (2.6). First, for $n > n_0$ we have

$$\begin{aligned} A_n^{-1}X_n &= A_n^{-1}(A_nD_n - A_{n-1}D_{n-1} - A_n\beta_{n-1}) \\ &= \delta_n + \rho D_{n-1}/(n-1) - \beta_{n-1}. \end{aligned}$$

But we know $D_{n-1}/n \to_p 0$ by Lemma 1. Also, since p'' is bounded, β_{n-1} is bounded by a constant times $\{D_{n-1}/(n-1)\}^2$ which also tends to 0 in probability. Also $\delta_n^2 = 1$. Thus the sequence $\{A_n^{-1}X_n\}$ is uniformly bounded by some absolute constant and satisfies

$$A_n^{-2}X_n^2 \to_p 1.$$

Now

$$n^{-1-2\rho} \sum_{k=1}^{n} X_k^2 = n^{-1-2\rho} \sum A_k^2(A_k^{-1}X_k)^2 \sim n^{-1-2\rho} \sum A_0^2 k^{2\rho}$$

in probability. (The uniform boundedness is used here.) But as $n \to \infty$ we have

$$n^{-1} \sum_{k \le n} \left(\frac{k}{n}\right)^{2\rho} \to \int_0^1 x^{2\rho}\, dx = (1 + 2\rho)^{-1}.$$

Hence

$$n^{-1-2\rho} \sum_{k=1}^{n} X_k^2 \to_p (1 + 2\rho)^{-1}A_0^2.$$

Moreover

$$n^{-1-2\rho} \sum_{k=1}^{n} X_k^2 < \text{constant}[1 + n^{-1} \sum_{k=n_0+1}^{n} (\delta_k + \rho D_{k-1}/(k-1) - \beta_{k-1})^2]$$

which is uniformly bounded, so the dominated convergence theorem implies also

$$n^{-1-2\rho}s_n^2 \to (1 + 2\rho)^{-1}A_0^2.$$

This proves (2.6). To show (2.5), since $s_n = O(n^{1/2+\rho})$ it suffices that $E\{(n^{-\rho}X_n)^2 I(|n^{-\rho}X_n| > \varepsilon n^{1/2})\} \to 0$. But this follows at once from (2.4) and the uniform boundedness of $A_n^{-1}X_n$. Hence the conclusion of the martingale invariance principle holds.

This result is the key to the theorem, but it is still some way from the result we require. First, note that

$$s_n^{-1}M_k \sim n^{-1/2-\rho}(1 + 2\rho)^{1/2}A_0^{-1}\{A_0k^\rho D_k + B_k\}$$

which suggests replacing $s_n^{-1}M_k$ with $n^{-1/2}(1 + 2\rho)^{1/2}(k/n)^\rho D_k$ in the definition of

ξ_n. With this in mind, define

$$\xi_n^*(t) = n^{-1/2}(1 + 2\rho)^{1/2}\left[\left(\frac{k}{n}\right)^\rho D_k + \frac{ts_n^2 - s_k^2}{s_{k+1}^2 - s_k^2}\left(\left(\frac{k+1}{n}\right)^\rho D_{k+1} - \left(\frac{k}{n}\right)^\rho D_k\right)\right],$$

$$s_k^2 \le ts_n^2 < s_{k+1}^2.$$

We claim that ξ_n^* also converges weakly to the Wiener process ξ as $n \to \infty$. This will follow from:

LEMMA 2. $\sup_t |\xi_n(t) - \xi_n^*(t)| \to_p 0$ *as* $n \to \infty$.

PROOF. It suffices to show

$$\sup_{k\le n} |s_n^{-1} M_k - n^{-1/2}(1 + 2\rho)^{1/2}(k/n)^\rho D_k| \to_p 0. \tag{2.7}$$

This will be broken up into several steps. Noting that $M_k = A_k D_k - B_k$, we show separately that

$$\sup_{k\le n} |s_n^{-1} A_k D_k - n^{-1/2}(1 + 2\rho)^{-1/2}(k/n)^\rho D_k| \to_p 0,$$

$$s_n^{-1} \sup_{k\le n} |B_k| \to_p 0.$$

The latter statement will follow at once if we can show that B_n/A_n is bounded in probability as $n \to \infty$. However

$$\begin{aligned} E|B_n/A_n| &\le \textstyle\sum_{k\le n} A_k A_n^{-1} E|\beta_{k-i}| \\ &\le \text{constant} \textstyle\sum_{k\le n} (k/n)^\rho E(D_k^2/k^2) \\ &\le \text{constant } n^{-\rho} \textstyle\sum_{k\le n} k^{\rho-1} \end{aligned}$$

which is bounded, so the required result follows at once from the Markov inequality.

Now let us write

$$\begin{aligned} & s_n^{-1} A_k D_k - n^{-1/2}(1 + 2\rho)^{1/2}(k/n)^\rho D_k \\ &\quad = n^{-1/2}(1 + 2\rho)^{1/2}(k/n)^\rho D_k[(s_n^{-1}(1 + 2\rho)^{-1/2} A_0 n^{1/2+\rho})(A_0^{-1} k^{-\rho} A_k) - 1]. \end{aligned} \tag{2.8}$$

The part in square brackets is uniformly bounded in k and n, and tends to zero if both $k \to \infty$ and $n \to \infty$. Let $\{k_n, n > 1\}$ be an arbitrary sequence such that $k_n \to +\infty$, $k_n = o(n^{((1+2\rho)/(2+2\rho))})$. Break up (2.8) into two parts, $k \le k_n$ and $k > k_n$. On $k \le k_n$ we have $n^{-1/2}(k/n)^\rho D_k \to 0$ uniformly, since $|D_k| \le k$. On $k > k_n$ the part of (2.8) in square brackets tends to zero as $n \to \infty$ uniformly in k, so it suffices to show that $\sup_{k\le n} n^{-1/2}(k/n)^\rho D_k$ is bounded in probability as $n \to \infty$.This follows from Doob's inequality applied to M_n, together with (2.4) and the fact (just shown) that B_n/A_n is bounded in probability as $n \to \infty$. The proof of Lemma 2 is complete.

We now know that ξ_n^* converges weakly to ξ. By the result of Wichura (1970), it is possible to construct the processes $\{\xi_n^*\}$ and ξ on a common probability space so that

$$\lim_{n\to\infty} \sup_{0\le t\le 1} |\xi_n^*(t) - \xi(t)| = 0 \text{ a.s.} \tag{2.9}$$

Our final transformation is a change of time scale. Define $Z_n(t)$ by (2.3), $Z(t) = t^{-\rho}\xi(t^{1+2\rho})$. Note that Z satisfies all the conditions in the statement of the theorem. Now $Z_n(t) = t^{-\rho}\xi_n^*(\tau_n(t))$ for some function τ_n which satisfies $\tau_n(k/n) = (s_k/s_n)^2$ and is monotonic increasing. However $(s_k/s_n)^2 \sim (k/n)^{1+2\rho}$ so that

$$\sup_t |\tau_n(t) - t^{1+2\rho}| \to \text{a.s. } 0.$$

Then we can write

$$\begin{aligned}\sup_t t^\rho |Z_n(t) - Z(t)| &= \sup_t |\xi_n^*(\tau_n(t)) - \xi(t^{1+2\rho})| \\ &\le \sup_t |\xi_n^*(\tau_n(t)) - \xi(\tau_n(t))| \\ &\quad + \sup_t |\xi(\tau_n(t)) - \xi(t^{1+2\rho})| \to \text{a.s. } 0,\end{aligned} \tag{2.10}$$

the first term by (2.9) and the second by continuity of sample paths. With this the proof is complete; we have shown that $d_\rho(Z_n, Z) \to 0$ almost surely on this particular probability space, and that suffices to prove that Z_n converges weakly to Z on the space C_ρ.

2.1 *Lack of balance.* Using Theorem 1, we are able to study many of the features of the design which are important for applications. The first of these is the imbalance in the generated design. Wei (1978a) proposed $\gamma_n = n_1^{-1} + n_2^{-1}$ as a measure of the imbalance of the generated design. He pointed out that this is equivalent to Box and Draper's (1975) measure of the robustness of a design against wild observations. It is also proportional to the variance of the estimated treatment difference. Then

$$n_1^{-1} + n_2^{-1} = 4n^{-1}\{1 + D_n^2/n^2 + O(D_n^3/n^3)\} \tag{2.11}$$

so that the increase compared with the optimum value when $D_n = 0$ is approximately $4D_n^2/n^3$, with asymptotic expectation $4/\{n^2(1 + 2\rho)\}$. The consequences of this are discussed further in Section 4 of Wei's paper.

2.2 *Selection bias.* Theorem 1 also enables us to obtain quantitative expressions for selection bias and accidental bias. We consider selection bias first. Recall that this is the bias which arises as a result of the experimenter's ability to guess the sequence of treatment allocations. A reasonable measure of selection bias is the mean number of correct guesses minus the mean number of incorrect guesses, assuming the experimenter always guesses the treatment which is more likely to come next. This measure is then

$$U_n = n^{-1} \textstyle\sum_{k=1}^{n-1} |2p(D_k/k) - 1|. \tag{2.12}$$

The asymptotic distribution of U_n is given by the following:

THEOREM 2.

$$n^{1/2}(1 + 2\rho)^{1/2}U_n \to_d \rho \int_0^1 \left| \frac{Z(t)}{t} \right| dt$$

where Z is as in Theorem 1.

PROOF. First define

$$U_n^* = n^{-1} \sum_{k=1}^{n-1} \rho \,|\, D_k/k \,|.$$

Now $n^{1/2}E\,|\,U_n - U_n^*\,| <$ constant $n^{-1/2} \sum_{k=1}^{n} E(D_k^2/k^2) \to 0$ as $n \to \infty$, so it suffices to prove the result with U_n^* replacing U_n. Rewrite

$$n^{1/2}(1 + 2\rho)^{1/2}U_n^* = \rho \sum_{k=1}^{n-1} k^{-1}\,|\,Z_n(k/n)\,|.$$

If we replace Z_n by Z then the right-hand side converges almost surely to $\rho \int_0^1 |\,Z(t)/t\,|\,dt$ because Z is almost surely a continuous function. Therefore it suffices to prove

$$\sum_{k=1}^{n} k^{-1}\,|\,Z_n(k/n) - Z(k/n)\,| \to_p 0,$$

whenever Z_n and Z satisfy (2.10). However it follows immediately that the result holds provided the sum is restricted to $k \ge n\varepsilon$, for any fixed $\varepsilon > 0$, so it suffices to show that ε may be chosen so as to make both $\sum_{k<n\varepsilon} k^{-1}\,|\,Z_n(k/n)\,|$ and $\sum_{k<n\varepsilon} k^{-1}\,|\,Z(k/n)\,|$ arbitrarily small in probability, uniformly in n. By Lemma 1,

$$\begin{aligned} E \sum_{k<n\varepsilon} k^{-1}\,|\,Z_n(k/n)\,| &= (1 + 2\rho)^{1/2}n^{-1/2} \sum_{k<n\varepsilon} E\,|\,D_k/k\,| \\ &\le (1 + 2\rho)^{1/2}n^{-1/2} \sum_{k<n\varepsilon} k^{-1/2} < \text{const. } \varepsilon^{1/2}. \end{aligned}$$

This suffices to give the result for Z_n, and the same argument also works for Z. This proves the theorem.

The exact distribution of the limit may be found from Kac's formula (Ito and McKean, 1965), after rewriting in terms of ξ. We note, however, that the mean is easily calculated, since $E\,|\,Z(t)\,| = (2t/\pi)^{1/2}$ and hence

$$E\left\{\int_0^1 \left|\,\frac{Z(t)}{t}\,\right| dt\right\} = 2\left(\frac{2}{\pi}\right)^{1/2}.$$

Therefore we have

$$\text{mean selection bias} \approx 2\rho\{2/(n\pi(1 + 2\rho))\}^{1/2}.$$

This result improves on that of Wei because it gives an explicit expression for the bias. In particular, it is possible to evaluate the trade-off between the imbalance of the design and selection bias, so enabling the experimenter, in principle at least, to choose ρ to get the best trade-off.

3. Biased coin designs for $R > 2$ treatments. Biased coin designs for an arbitrary number of treatments have been proposed by Wei (1978b) and by Atkinson (1982). Wei's procedure is a generalization of his procedure for two treatments and is based on the assumption that it is desired to balance the experiment so that each of the R treatments receives a proportion R^{-1} of patients. Atkinson's procedure is based on the concept of D_A optimality and allows for the possibility that the limiting optimal design is asymmetric. Therefore we propose a procedure which generalises Wei's procedure to achieve a limiting design measure $(\xi_1, \cdots, \xi_R)$, where $\xi_r \geq 0$ for $1 \leq r \leq R$ and $\sum \xi_r = 1$. In other words, the limiting proportion of patients assigned to treatment r is required to be ξ_r. We assume $(\xi_1, \cdots, \xi_R)$ are known at the start of the experiment.

Suppose, after n patients have been allocated, the number of patients on treatment r is $D_{n,r}$. Define $\delta_{n+1,r}$ to be 1 if the $(n + 1)$st patient is allocated to treatment r, 0 otherwise. Let $\mathscr{F}_n$ be the sigma field generated by $\{D_{k,r}, 1 \leq k \leq n, 1 \leq r \leq R\}$. Our allocation rule is of the form

$$P\{\delta_{n+1,r} = 1 \mid \mathscr{F}_n\} = p_r(n^{-1}D_{n,1}, \cdots, n^{-1}D_{n,R}).$$

Define $\Omega = \{(y_1, \cdots, y_R): y_r \geq 0, \sum y_r = 1\}$ and suppose $\mathbf{p} = (p_1, \cdots, p_R)$ is a function from an open neighborhood of Ω into Ω satisfying:

A1 $\mathbf{p}$ is twice continuously differentiable with uniformly bounded second derivatives;

A2 if $y_r \geq \xi_r$ then $p_r(\mathbf{y}) \leq \xi_r$.

Let $B = (b_{rq})$ be the matrix of first order partial derivatives defined by

$$b_{rq} = (\partial p_r/\partial y_q)(\xi_1, \cdots, \xi_R).$$

Condition A2, which says that the design always works towards the desired limiting proportions $\xi_1, \cdots, \xi_R$, imposes quite strong restrictions on the matrix B. First, it is obvious that $b_{rr} \leq 0$ and that $\sum_r b_{rq} = 0$ for all q. Secondly,

$$p_r(\xi_1 + \delta, \xi_2 - \delta, \xi_3, \cdots, \xi_R) = \xi_r + \delta b_{r1} - \delta b_{r2} + O(\delta^2).$$

By A2 this is at most ξ_r if $r > 2$, irrespective of the sign of δ. Hence $b_{r1} = b_{r2}$ for $r > 2$. By extension, for each r we have b_{rq} equal to a constant, say $c_r \geq 0$, for $q \neq r$. It follows next that $b_{rr} = c_r - \sum_q c_g$. Denote $\sum_q c_q$ by ρ; it will soon be seen that this ρ has the same interpretation as in the two-treatment problem. Finally we have

$$b_{rq} = \begin{cases} c_r, & r \neq q \\ c_r - \rho, & r = q. \end{cases}$$

Note that B has two distinct eigenvalues, namely 0 with multiplicity 1 and $-\rho$ with multiplicity $R - 1$. Any vector $\mathbf{u} = (u_r)$ such that $\sum c_r u_r = 0$ is a left eigenvector with eigenvalue $-\rho$. The constant c_r may be thought of as a weighting factor which measures how strongly it is desired to balance the rth treatment. This interpretation, however, is misleading to some extent, as it will be seen that the limiting properties of the design depend only on ρ.

REMARK. It is essential that $\mathbf{p}$ be defined on an open neighbourhood of Ω, and has values in Ω throughout that neighbourhood, for these relations on the partial derivatives to be correct. I am grateful to Profs. L. J. Wei and R. T. Smythe for correspondence on this point.

We now state the main result:

THEOREM 3. *Define for* $n \geq 1, 1 \leq r \leq R, 0 \leq t \leq 1,$

$$\bar{D}_{n,r} = D_{n,r} - n\xi_r,$$

$$Z_{n,r}(t) = n^{-1/2}(1 + 2\rho)^{1/2}[\bar{D}_{k,r} - (t - k/n)(\delta_{k+1,r} - \xi_r)], \quad k/n \leq t < (k+1)/n.$$

Then $\{(Z_{n,1}(t), \cdots, Z_{n,R}(t)), 0 \leq t \leq 1\}$ *converges weakly on* C_ρ^R *to a limiting process* $\{(Z_1(t), \cdots, Z_R(t)), 0 \leq t \leq 1\}$ *which is a zero mean Gaussian process with continuous sample paths and covariance function*

$$E\{Z_r(s)Z_q(t)\} = \begin{cases} (\xi_r - \xi_r^2)s^{1+\rho}t^{-\rho}, & r = q \\ -\xi_r\xi_q s^{1+\rho}t^{-\rho}, & r \neq q, \end{cases} \tag{3.1}$$

whenever $0 \leq s \leq t \leq 1$.

PROOF. Let $\mathbf{u}^T = (u_1, \cdots, u_R)$ be an arbitrary R-dimensional vector. We shall show that $\Sigma\, u_r Z_{n,r}(t)$ converges to a zero mean Gaussian process with continuous sample paths and the same covariance function as $\Sigma\, u_r Z_r(t)$. This is sufficient, since u is arbitrary.

We begin with two lemmas. We use the notation $x_+ = \max(x, 0)$.

LEMMA 3. *Given* $x_1, \cdots, x_R$ *with* $\sum x_r = 0$,

$$\sum x_r^2 \leq R \sum (x_r)_+^2.$$

PROOF. Suppose $x_1 > 0, \cdots, x_m > 0, x_{m+1} \leq 0, \cdots, x_R \leq 0$, where $1 \leq m \leq R - 1$. If $\sum_1^R (x_r)_+^2 = \sum_1^m x_r^2 = c$ then $\sum_1^m x_r \leq (mc)^{1/2}$. So $-\sum_{m+1}^R x_r \leq (mc)^{1/2}$, $\sum_{m+1}^R x_r^2 \leq mc$. Thus $\sum_1^R x_r^2 \leq (m+1)c \leq Rc$.

LEMMA 4. $\sum_r E\{\bar{D}_{n,r}^2\} < R^2 n$.

PROOF. We start with the identity,

$$\bar{D}_{n+1,r}^2 = \bar{D}_{n,r}^2 + 2\bar{D}_{n,r}(\delta_{n+1,r} - \xi_r)^2 + (\delta_{n+1,r} - \xi_r)^2.$$

Then, by considering separately the cases $\bar{D}_{n,r} \geq 0$, $\bar{D}_{n,r} < 0$ it follows that

$$(\bar{D}_{n+1,r})_+^2 \leq (\bar{D}_{n,r})_+^2 + 2(\bar{D}_{n,r})_+(\delta_{n+1,r} - \xi_r) + (\delta_{n+1,r} - \xi_r)^2.$$

Take expected values, noting that $(\delta_{n+1,r} - \xi_r)^2 \leq 1$ while $E\{(\delta_{n+1,r} - \xi_r)(\bar{D}_{n,r})_+\} \leq 0$ by A2. Hence $E\{(\bar{D}_{n+1,r})_+^2\} \leq E\{(\bar{D}_{n,r})_+^2\} + 1$ and so $E\{(\bar{D}_{n,r})_+^2\} \leq n$. The result now follows from Lemma 3.

From Lemma 4, it follows at once that

$$\sum_r E|\bar{D}_{n,r}| = O(n^{1/2})$$

and that $n^{-1}\bar{D}_{n,r} \to_p 0$ for each r.

From now on let $Y_n = \sum_r u_r \bar{D}_{n,r}$. Then

$$\begin{aligned} E\{Y_{n+1} \mid \mathscr{F}_n\} &= Y_n + E\{\textstyle\sum_r u_r(\delta_{n+1,r} - \xi_r) \mid \mathscr{F}_n\} \\ &= Y_n + \textstyle\sum_r u_r\{p_r(n^{-1}D_{n,1}, \cdots, n^{-1}D_{n,R}) - \xi_r\} \\ &= Y_n + \textstyle\sum_r u_r \sum_q b_{rq} n^{-1}\bar{D}_{n,q} \\ &\quad + \textstyle\sum_r u_r\{p_r(n^{-1}D_{n,1}, \cdots, n^{-1}D_{n,R}) - \xi_r - \sum_q b_{rq}n^{-1}\bar{D}_{n,q}\} \\ &= \alpha_n Y_n + \beta_n \end{aligned}$$

where $\alpha_n = 1 - \rho/n$ and $E \mid \beta_n \mid$ is uniformly $O(n^{-1})$, by the boundedness of second derivatives and Lemma 3.

The proof now follows closely that of Theorem 1. For $n > n_0 > \rho$, let

$$A_n = \textstyle\prod_{k=n_0}^{n-1} \alpha_k^{-1}, \quad B_n = \sum_{k=n_0}^{n-1} A_{k+1}\beta_k, \quad M_n = A_n Y_n - B_n.$$

Then $(M_n, \mathscr{F}_n)$ is a martingale. If $X_n = M_n - M_{n-1}$, then

$$\begin{aligned} A_n^{-1} X_n &= \textstyle\sum_r u_r(\delta_{n,r} - \xi_r) + \rho Y_{n-1}/(n-1) - \beta_{n-1} \\ &= \textstyle\sum_r u_r(\delta_{n,r} - \xi_r) + \varepsilon_n \end{aligned}$$

where ε_n is uniformly bounded and tends to zero in probability. Then

$$\begin{aligned} A_n^{-2} X_n^2 &= \textstyle\sum_r \sum_q u_r u_q(\delta_{n,r} - \xi_r)(\delta_{n,q} - \xi_q) + \varepsilon_n \\ &= \textstyle\sum_r u_r^2 \delta_{n,r} + \sum_r \sum_q u_r u_q(\xi_r \xi_q - 2\xi_r \delta_{n,q}) + \varepsilon_n. \end{aligned}$$

Thus $E\{A_n^{-2}X_n^2\} \to \sum_r u_r^2 \xi_r - (\sum_r u_r \xi_r)^2 = \bar{u}$ (say). We may now write

$$n^{-2\rho-1} \textstyle\sum_{k \le n} X_k^2 = n^{-2\rho-1} \sum_{k\le n} A_k^2 \bar{u} + n^{-2\rho-1} \sum_{k \le n} A_k^2(A_k^{-2}X_k^2 - \bar{u}).$$

Since $A_k \sim A_0 k^\rho$ the first term converges to $A_0^2(1 + 2\rho)^{-1}\bar{u}$. The second term is

$$n^{-2\rho-1} \textstyle\sum_{k\le n} A_k^2 \{\sum_r u_r^2(\delta_{k,r} - \xi_r) - 2 \sum_r \sum_q u_r u_q \xi_r(\delta_{k,q} - \xi_q) + \varepsilon_k\}.$$

We claim that this tends to zero both in mean and in probability. First, we know $\varepsilon_n \to_p 0$ and by the dominated convergence theorem $E \mid \varepsilon_n \mid \to 0$ also. Hence $n^{-2\rho-1} \sum_{k \le n} A_k^2 E \mid \varepsilon_k \mid \to 0$. For the rest, it will suffice to show

$$n^{-2\rho-1} E \mid \textstyle\sum_{k\le n} A_k^2(\delta_{k,r} - \xi_r) \mid \to 0.$$

But

$$\begin{aligned} \textstyle\sum_{k\le n} A_k^2(\delta_{k,r} - \xi_r) &= \textstyle\sum_{k \le n} A_k^2(\bar{D}_{k,r} - \bar{D}_{k-1,r}) \\ &= A_n^2 \bar{D}_{n,r} - \textstyle\sum_{k=1}^{n-1} \bar{D}_{k,r}(A_{k+1}^2 - A_k^2). \end{aligned}$$

We know already that $n^{-2\rho-1} A_n^2 E \mid D_{n,r} \mid \to 0$, and it is easily verified that $A_{k+1}^2 - A_k^2 = O(k^{2\rho-1})$ so that

$$\begin{aligned} n^{-2\rho-1} \textstyle\sum_{k=1}^{n-1} (A_{k+1}^2 - A_k^2)\, E \mid D_{k,r} \mid &\le \text{constant } n^{-2\rho-1} \textstyle\sum_{k=1}^{n-1} k^{2\rho-1}k^{1/2} \\ &\to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

This proves our claim.

Thus we deduce

$$n^{-2\rho-1} \sum_{k\le n} X_k^2 \to_p A_0^2(1 + 2\rho)^{-1}\bar{u}, \quad n^{-2\rho-1} E\{\sum_{k\le n} X_k^2\} \to A_0^2(1 + 2\rho)^{-1}\bar{u}.$$

The remainder of the proof follows that of Theorem 1. We conclude that $\sum u_r Z_{n,r}(t)$ converges on C_ρ to a zero mean Gaussian process with continuous sample paths and covariance function $\bar{u}s^{1+\rho}t^{-\rho}$, $0 \le s \le t \le 1$. But this is precisely the covariance function of $\sum u_r Z_r$ when $Z_1, \cdots, Z_R$ have the covariances (3.1). With this the proof of Theorem 3 is complete.

EXAMPLE. Atkinson (1982, Appendix) derived an allocation procedure for an experiment in which it is desired to estimate all contrasts among R treatments. This procedure corresponds to choosing

$$p_r(y) = (y_r^{-1} - 1)/(\sum y_q^{-1} - R).$$

We claim this satisfies A1 and A2 with $\xi_r = R^{-1}$ for all r. It is a little tricky to show A2 in this case. However, if y_1 (say) is fixed and $y_2, \cdots, y_R$ allowed to vary then $\sum y_r^{-1} \ge y_1^{-1} + (R - 1)^2(1 - y_1)^{-1}$ and so

$$p_1(\mathbf{y}) \le (y_1^{-1} - 1)/(y_1^{-1} + (R - 1)^2(1 - y_1)^{-1} - R).$$

It may now be checked that the right side is at most R^{-1} whenever $y_1 \ge R^{-1}$. Consequently A2 is satisfied and the result of Theorem 3 holds with $\rho = R/(R - 1)$.

REMARK. It seems natural to extend condition A2 to require also $P_r(\mathbf{y}) \ge \xi_r$ when $y_r \le \xi_r$ but this is false for Atkinson's procedure. This makes Lemma 4 harder than would otherwise be the case.

3.1 *Lack of balance.* The quantities $\{\bar{D}_{n,r}\}$ measure the imbalance in the design. Theorem 3 shows that their variances are asymptotically proportional to $(1 + 2\rho)^{-1}$. Thus the dependence on ρ is exactly the same as in the two-treatment problem.

Specific measures of imbalance can be treated in the same way. Wei (1978b) takes $\sum_r D_{n,r}^{-1}$ as a measure of imbalance, pointing out that it is equivalent to the design robustness measure of Box and Draper (1975). The optimal value $n^{-1}R^2$ is obtained when $D_{n,r} = nR^{-1}$ for all r (so $\xi_r = R^1$ here). Now

$$\begin{aligned} n \sum_r D_{n,r}^{-1} &= n \sum_r \{nR^{-1} + n^{1/2}(1 + 2\rho)^{1/2}Z_{n,r}(1)\}^{-1} \\ &= R \sum_r \{1 + n^{-1/2}(1 + 2\rho)^{-1/2}RZ_{n,r}(1)\}^{-1} \\ &= R^2\{1 + n^{-1}(1 + 2\rho)^{-1}R \sum_r Z_{n,r}^2(1) + o(n^{-1})\}. \end{aligned}$$

But $\sum_r Z_{n,r}^2(1)$ converges to $\sum_r Z_r^2(1)$ which has mean $(R - 1)/R$, so

$$nE\{n \sum_r D_{n,r}^{-1} - R^2\} \to (1 + 2\rho)^{-1}R^2(R - 1).$$

Here, letting $\rho \to \infty$ gives the limiting case of perfect balance, while $\rho = 0$ corresponds to perfect randomisation.

3.2 *Selection bias.* Suppose an experimenter is trying to guess the sequence of patient assignments. A reasonable measure of selection bias is

$$n^{-1} \sum_{k=1}^{n-1} \max_r p_r(k^{-1}D_{k,1}, \cdots, k^{-1}D_{k,R}).$$

Wei (1978b) shows that for his procedures this quantity tends to its optimum value R^{-1} as $n \to \infty$. We shall obtain a stronger result using Theorem 3. It is implicit in this calculation that $\xi_1 = \cdots = \xi_R = R^{-1}$.

Neglecting higher order terms in the Taylor expansion we have

$$\begin{aligned} p_r(n^{-1}D_{n,1}, \cdots, n^{-1}D_{n,R}) &= R^{-1} + \sum_q b_{rq} n^{-1}\bar{D}_{n,q} \\ &= R^{-1} + \sum_q b_{rq} n^{-1/2}(1 + 2\rho)^{-1/2} Z_{n,q}(1) \\ &= R^{-1} - n^{-1/2}\rho(1 + 2\rho)^{-1/2} Z_{n,r}(1). \end{aligned}$$

Hence

$$E\{n^{1/2}(\max_r p_r(n^{-1}D_{n,1}, \cdots, n^{-1}D_{n,R}) - R^{-1})\} \to \rho(1 + 2\rho)^{-1/2} E\{\max_r Z_r(1)\}.$$

To evaluate this expression, suppose $U_1, \cdots, U_R$ are independent $N(0, 1)$ and define

$$V_r = R^{-1/2}(U_r - R^{-1} \sum_q U_q), \quad 1 \le r \le R.$$

It is readily verified that $(V_1, \cdots, V_R)$ have the same joint distribution as $(Z_1(1), \cdots, Z_R(1))$. Consequently

$$E\{\max_r Z_r(1)\} = E\{\max_r V_r\} = R^{-1/2} E\{\max_r U_r\}.$$

For the calculation of $E\{\max_r U_r\}$, see David (1981), Section 3.2.

The main point of this calculation is that the selection bias, as a function of ρ, is proportional to $\rho(1 + 2\rho)^{-1/2}$, exactly as in the two-treatment problem.

It should be possible to derive the asymptotic distribution of the selection bias, along the lines of Theorem 2, but we have not attempted this.

4. An additive covariate model for two treatments. In this section we study the properties of a procedure for two treatments in which explicit allowance is made for prognostic factors and the need to balance with respect to the prognostic factors as well as the main treatment effects. The procedure is closely related to recent proposals of Begg and Iglewicz (1980) and Atkinson (1982).

Consider the linear model

$$E\{y_n\} = \alpha\delta_n + \sum_{j=1}^{p} z_{nj}\beta_j$$

where α is the treatment effect, δ_n is +1 or −1 according as the nth patient is assigned to the first treatment or the second treatment.

$\mathbf{z}_n = (z_{n1}, \cdots, z_{np})^T$ is a $p \times 1$ vector of covariates and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ is a vector of nuisance parameters. The model may be written in matrix notation as

$$E\{\mathbf{Y}_n\} = X_n\begin{pmatrix}\alpha\\ \boldsymbol{\beta}\end{pmatrix}$$

where $\mathbf{Y}_n = (y_1, \cdots, y_n)^T$ and X_n is the design matrix

$$X_n = (\Delta_n Z_n)$$

where $\Delta_n = (\delta_1, \cdots, \delta_n)^T$ is $n \times 1$ and

$$Z_n = \begin{bmatrix} \mathbf{z}_1^T \\ \vdots \\ \mathbf{z}_n^T \end{bmatrix}$$

is the $n \times p$ matrix of prognostic factors. Note that there is no overall constant in this model, but one may be introduced by assuming $z_{n,1} = 1$ for all n. In most of the discussion we shall assume this to be the case.

We shall assume that both $X_n^T X_n$ and $Z_n^T Z_n$ are invertible. The standard least squares estimator of α is

$$\hat{\alpha} = n^{-1} \cdot \frac{\Delta_n^T \mathbf{Y}_n - \Delta_n^T Z_n (Z_n^T Z_n)^{-1} Z_n^T \mathbf{Y}_n}{1 - n^{-1} \Delta_n^T Z_n (Z_n^T Z_n)^{-1} Z_n^T \Delta_n} \tag{4.1}$$

and its variance, under the assumption of uncorrelated errors with mean zero and common variance σ^2, is given by

$$\operatorname{var} \hat{\alpha} = n^{-1}\sigma^2 \{1 - n^{-1} \Delta_n^T Z_n (Z_n^T Z_n)^{-1} Z_n^T \Delta_n\}^{-1}. \tag{4.2}$$

Suppose a new patient arrives with covariate vector $\mathbf{z}_{n+1}$. It may be shown that the procedure of Atkinson (1982) is equivalent to setting $\delta_{n+1} = \pm 1$ with probabilities proportional to

$$\{1 \mp \mathbf{z}_{n+1}^T (Z_n^T Z_n)^{-1} Z_n^T \Delta_n\}^2.$$

Begg and Iglewicz arrive at a rather similar conclusion by a different route. They assume $Z_n^T Z_n \approx nI$ and then base their allocation rule on $z_{n+1}^T Z_n^T \Delta_n$. More precisely, their rule (in our notation) is the deterministic rule

$$\delta_{n+1} = \begin{cases} -1 & \text{if } \mathbf{z}_{n+1}^T Z_n^T \Delta_n > 0, \\ +1 & \text{if } \mathbf{z}_{n+1}^T Z_n^T \Delta_n < 0. \end{cases}$$

An obvious generalization of these procedures is to choose $\delta_{n+1} = +1$ with probability either

$$\phi(\mathbf{z}_{n+1}^T (Z_n^T Z_n)^{-1} Z_n^T \Delta_n) \quad \text{or} \quad \phi(n^{-1} \mathbf{z}_{n+1}^T Z_n^T \Delta_n)$$

for some nonincreasing function ϕ. (We use ϕ here, rather than p, because we have used p to denote the number of prognostic factors.)

The procedure we analyze here is a compromise between these. We assume that $\{\mathbf{z}_n, n \ge 1\}$ are independent, identically distributed random vectors with $E\{\mathbf{z}_n \mathbf{z}_n^T\} = Q$, where Q is nonsingular, and all third moments of z_n are finite. The procedure is:

$$P\{\delta_{n+1} = 1 \mid \Delta_n, \mathbf{z}_{n+1}, Z_n\} = \phi(n^{-1} \mathbf{z}_{n+1}^T Q^{-1} Z_n^T \Delta_n) \tag{4.3}$$

where ϕ is nonincreasing and satisfies $\phi(x) + \phi(-x) = 1$. We assume ϕ is twice continuously differentiable with bounded second derivative and let $\rho = -2\phi'(0)$. Define the $p \times 1$ vector $\mathbf{D}_n$ to be $Z_n^T \Delta_n = \sum_{k=1}^n \mathbf{z}_k \delta_k$, and let $D_{n,j}$ be the jth component of $\mathbf{D}_n$. Let

$$Z_{n,j}(t) = n^{-1/2}(1 + 2\rho)^{1/2}[D_{k,j} + (t - k/n)(D_{k+1,j} - D_{k,j})],$$

$$k/n \le t < (k+1)/n.$$

THEOREM 4. *The process* $\{(Z_{n,1}(t), \cdots, Z_{n,p}(t)), 0 \le t \le 1\}$ *converges weakly on* C_ρ^p *to a limiting process* $\{(Z(t), \cdots, Z_p(t)), 0 \le t \le 1\}$ *which is a zero mean Gaussian process with continuous sample paths and covariance function*

$$E\{Z_i(s)Z_j(t)\} = q_{ij} s^{1+\rho} t^{-\rho}, \quad 0 \le s \le t \le 1$$

where q_{ij} $(1 \le i \le p, 1 \le j \le p)$ *is the generic element of* Q.

PROOF. As usual the proof starts with a lemma which essentially ensures that eventual balance is achieved.

LEMMA 5. $E\{\mathbf{D}_n^T Q^{-1} \mathbf{D}_n\} \le np$.

PROOF.

$$\begin{aligned} E\{&\mathbf{D}_{n+1}^T Q^{-1} \mathbf{D}_{n+1} \mid \Delta_n, Z_n, \mathbf{z}_{n+1}\} \\ &= \mathbf{D}_n^T Q^{-1} \mathbf{D}_n + 2\mathbf{D}_n^T Q^{-1} \mathbf{z}_{n+1}\{2\phi(n^{-1}\mathbf{z}_{n+1}^T Q^{-1} \mathbf{D}_n) - 1\} \\ &\quad + \delta_{n+1}^2 \mathbf{z}_{n+1}^T Q^{-1} \mathbf{z}_{n+1} \\ &\le \mathbf{D}_n^T Q^{-1} \mathbf{D}_n + \mathbf{z}_{n+1}^T Q^{-1} \mathbf{z}_{n+1}, \end{aligned}$$

using the fact that

$$\mathbf{D}_n^T Q^{-1} \mathbf{z}_{n+1} \quad \text{and} \quad 2\phi(n^{-1}\mathbf{z}_{n+1}^T Q^{-1} \mathbf{D}_n) - 1$$

are necessarily of opposite sign. Now take expected values, noting that

$$E\{\mathbf{z}_{n+1}^T Q^{-1} \mathbf{z}_{n+1}\} = E\{\operatorname{tr} \mathbf{z}_{n+1} \mathbf{z}_{n+1}^T Q^{-1}\} = \operatorname{tr} QQ^{-1} = p.$$

Thus

$$E\{\mathbf{D}_{n+1}^T Q^{-1} \mathbf{D}_{n+1}\} \le E\{\mathbf{D}_n^T Q^{-1} \mathbf{D}_n\} + p$$

from which the result follows.

Note that Lemma 5 effectively means that all the components of $\mathbf{D}_n$ have variance of at most a multiple of n. For $\mathbf{D}_n^T Q^{-1} \mathbf{D}_n$ is invariant under orthogonal transformations of the $\mathbf{z}$'s; there exists an orthogonal transformation with respect to which Q is diagonal, and $E\{\mathbf{D}_n^T Q^{-1} \mathbf{D}_n\}$ is just the sum of the variances of the principal components, which must therefore all be of $O(n)$.

Now we turn to the main part of the proof. Define $\mathscr{F}_n$ to be the sigma field generated by $\delta_1, \cdots, \delta_n$ and $\mathbf{z}_1, \cdots, \mathbf{z}_n$.

$$\begin{aligned} E\{\delta_{n+1} \mid \mathscr{F}_n, \mathbf{z}_{n+1}\} &= 2\phi(n^{-1}\mathbf{z}_{n+1}^T Q^{-1}\mathbf{D}_n) - 1 \\ &= -\rho n^{-1}\mathbf{z}_{n+1}^T Q^{-1}\mathbf{D}_n \\ &\quad + \{2\phi(n^{-1}\mathbf{z}_{n+1}^T Q^{-1}\mathbf{D}_n) - 1 - 2\phi'(0) n^{-1}\mathbf{z}_{n+1}^T Q^{-1}\mathbf{D}_n\}. \end{aligned}$$

The part in brackets is at most a constant times $(n^{-1}\mathbf{z}_{n+1}^T Q^{-1}\mathbf{D}_n)^2$.

Let $\mathbf{u}$ be an arbitrary p-vector. Then

$$E\{\delta_{n+1}\mathbf{u}^T\mathbf{z}_{n+1} \mid \mathscr{F}_n\} = -\rho n^{-1}\mathbf{u}^T E\{\mathbf{z}_{n+1}\mathbf{z}_{n+1}^T\} Q^{-1}\mathbf{D}_n + \beta_n$$

where

$$\beta_n = O(\mathbf{u}^T\mathbf{z}_{n+1}(n^{-1}\mathbf{z}_{n+1}^T Q^{-1}\mathbf{D}_n)^2).$$

Expanding this expression and using the facts that (i) $\mathbf{z}_{n+1}$ is independent of $\mathbf{D}_n$, (ii) the third moments of $\mathbf{z}_{n+1}$ are all finite and (iii) the components of $\mathbf{D}_n$ have variance of $O(n)$, by the remark following Lemma 5, we conclude that $E|\beta_n| = O(n^{-1})$. Defining $V_n = \mathbf{u}^T\mathbf{D}_n$, $\alpha_n = 1 - \rho/n$ we then have

$$E\{V_{n+1} \mid \mathscr{F}_n\} = \alpha_n V_n + \beta_n.$$

Following the proof of Theorem 1, for $n > n_0 > \rho$ let

$$A_n = \prod_{k=m_0}^{n-1} \alpha_k^{-1}, \quad B_n = \sum_{k=n_0}^{n-1} A_{k+1}\beta_k, \quad M_n = A_n V_n - B_n, \quad X_n = M_n - M_{n-1}.$$

Then $\{M_n\}$ is a martingale with respect to $\{\mathscr{F}_n\}$, we have $A_n \sim A_0 n^\rho$, and the Martingale differences $\{X_n\}$ satisfy

$$A_n^{-1}X_n = \delta_n\mathbf{u}^T\mathbf{z}_n + \rho V_{n-1}/(n-1) - \beta_{n-1} = \delta_n\mathbf{u}^T\mathbf{z}_n + o_p(1).$$

Thus $E\{A_n^{-2}X_n^2\} = \mathbf{u}^T Q\mathbf{u} + o(1)$ and by arguments similar to those in the proof of Theorem 4 it follows that

$$n^{-2\rho-1} \sum_{k\le n} X_k^2 \to_p A_0^2(1 + 2\rho)^{-1}\mathbf{u}^T Q\mathbf{u},$$

$$n^{-2\rho-1} E\{\sum_{k\le n} X_k^2\} \to A_0^2(1 + 2\rho)^{-1}\mathbf{u}^T Q\mathbf{u}.$$

The remainder of the proof follows that of Theorem 1, and leads to the following conclusion. Defining Z_n by

$$Z_n(t) = n^{-1/2}(1 + 2\rho)^{1/2}[V_k + (t - k/n)\delta_{k+1}\mathbf{u}^T\mathbf{z}_{k+1}], \quad k/n \le t < (k+1)/n,$$

Z_n converges weakly on C_p to a zero-mean Gaussian process Z with continuous sample paths and covariance function

$$E\{Z(s)Z(t)\} = (\mathbf{u}^T Q\mathbf{u}) s^{1+\rho} t^{-\rho}, \quad 0 \le s \le t \le 1.$$

But this is just the covariance function of $\sum u_i Z_i(t)$, where $Z_1, \cdots, Z_p$ are as in the statement of the theorem. Since the vector $\mathbf{u}$ was arbitrary, the result is proved.

4.1 *Lack of balance.* These designs are constructed so as to achieve balance with respect to not only the main treatment effect but also the various prognostic factors. There are a number of measures of imbalance. One natural criterion is to consider the effect of imbalance on the variance of α, given by (4.2). The variance is increased, compared with the case of "perfect balance" for which $\mathbf{D}_n = Z_n^T \Delta_n = \mathbf{0}$, by a fraction $n^{-1}\mathbf{D}_n^T(Z_n^T Z_n)^{-1}\mathbf{D}_n$. But

$$E\{\mathbf{D}_n^T(Z_n^T Z_n)^{-1}\mathbf{D}_n\} \sim n^{-1}E\{\mathbf{D}_n^T Q^{-1}\mathbf{D}_n\} = n^{-1}E\{\operatorname{tr} \mathbf{D}_n\mathbf{D}_n^T Q^{-1}\}$$
$$\rightarrow \operatorname{tr} QQ^{-1}(1 + 2\rho)^{-1} = p(1 + 2\rho)^{-1}.$$

Hence

$$\operatorname{var} \hat{\alpha} \approx n^{-1}\sigma^2\{1 + n^{-1}p(1 + 2\rho)^{-1}\}.$$

Note the dependence on p as well as on ρ.

4.2 *Selection bias.* As in Section 2.2, a suitable measure of selection bias is

$$U_n = n^{-1} \textstyle\sum_{k=1}^{n-1} \mid 2\phi(k^{-1}\mathbf{z}_{k+1}^T Q^{-1} Z_k^T \Delta_k) - 1 \mid.$$

This is asymptotically equivalent to

$$U_n^* = n^{-1} \textstyle\sum_{k=1}^{n-1} \rho k^{-1} \mid \mathbf{z}_{k+1}^T Q^{-1} Z_k^T \Delta_k \mid.$$

We shall only calculate the asymptotic mean of U_n^* and shall not attempt to find its whole distribution. For large k, conditionally on $\mathbf{z}_{k+1}$, the distribution of $\mathbf{z}_{k+1}^T Q^{-1} Z_k^T \Delta_k$ is approximately normal with mean zero and variance

$$E\{\mathbf{z}_{k+1}^T Q^{-1}\mathbf{D}_k\mathbf{D}_k^T Q^{-1}\mathbf{z}_{k+1} \mid \mathbf{z}_{k+1}\} \approx k(1 + 2\rho)^{-1}\mathbf{z}_{k+1}^T Q^{-1}\mathbf{z}_{k+1}.$$

Thus

$$E\{\mid z_{k+1}^T Q^{-1} Z_k^{-1} \Delta_k \mid \mid z_{k+1}\} \approx [2k\mathbf{z}_{k+1}^T Q^{-1}\mathbf{z}_{k+1}/\pi(1 + 2\rho)]^{1/2}.$$

Substituting in the formula for U_n^* and performing the summation yields eventually:

$$\text{mean selection bias} \approx 2\rho\{2/(n\pi(1 + 2\rho))\}^{1/2}E\{(\mathbf{z}^T Q^{-1}\mathbf{z})^{1/2}\}.$$

Here $\mathbf{z}$ denotes an arbitrary member of the random sequence $\{\mathbf{z}_n\}$. The dependence on ρ and n are the same as for the situation of Section 2.2. The dependence on $\mathbf{z}$ depends on the distribution of $\mathbf{z}^T Q^{-1}\mathbf{z}$. As an example, suppose $z_{n,1} = 1$ for all n and the remaining covariates have a normal distribution with mean zero. By applying a transformation, these covariates may be taken to be independent with common variance 1 (so that $Q = I$). Then $z^T Q^{-1} z = 1 + \chi_{p-1}^2$ and

$$E\{(z^T Q^{-1} z)^{1/2}\} = [2^{(p-1)/2}\Gamma((p - 1)/2)]^{-1} \int_0^\infty (1 + x)^{1/2} x^{(p-3)/2} e^{-x/2}\, dx.$$

The author is not aware of any simple expression for the right-hand side, but it is asymptotically $(2p)^{1/2}$ as $p \rightarrow \infty$, which again emphasises the dependence on p.

5. Concluding remarks. The methods of this paper apply to a wide class of designs, but that does not include Efron's biased coin design. The analysis of

Efron (1971) is complete in itself, but the development of corresponding results for the procedure of Efron (1980) is an open question.

We have concentrated in this paper on the asymptotics rather than on applications of the results. A fuller discussion of applications and of other aspects of the procedures is given in Smith (1984).

REFERENCES

ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–68.

BEGG, C. B. and IGLEWICZ, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics* **36** 81–90.

BLACKWELL, D. and HODGES, H. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28** 449–460.

BOX, G. E. P. and DRAPER, N. L. (1975). Robust designs. *Biometrika* **62** 347–352.

DAVID, H. A. (1981). *Order Statistics*, 2nd ed. Wiley, New York.

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

EFRON, B. (1980). Randomizing and balancing a complicated sequential experiment. *Biostatistics Casebook* (ed. R. G. Miller) 19–30. Wiley, New York.

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

ITO, K. and MCKEAN, H. P. (1965). *Diffusion Processes and Their Sample Paths.* Springer-Verlag, Berlin.

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

SMITH, R. L. (1984). Sequential treatment allocation using biased coin designs. *J. Roy. Statist. Soc. B.*, to appear.

WEI, L. J. (1978a). The adaptive biased coin design for sequential experiments. *Ann. Statist.* **6** 92–100.

WEI, L. J. (1978b). A class of treatment assignment rules for sequential experiments. *Commun. Statist. Theor. Meth.* **A7** 285–295.

WICHURA, M. J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist.* **41** 284–291.

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY
HUXLEY BUILDING
QUEEN'S GATE, LONDON SW7 2BZ
ENGLAND