Statistica Sinica 4(1994), 219-231

INFLUENCE OF SELECTION BIAS ON TYPE I ERROR RATE UNDER RANDOM PERMUTED BLOCK DESIGNS

Michael Proschan

National Heart, Lung, and Blood Institute

Abstract: A model for selection bias in a large, single blind clinical trial is presented. The actual Type I error rate is evaluated, and this is used to quantify the degree of selection bias under random permuted block designs. The approach utilizes results from the theory of random walks to show rigorously that when the total number of patients is fixed and there is only one investigator, the least bias occurs when there is a single block (random allocation). Even under random allocation, however, the bias does not become negligible as the number of patients becomes large. It is also shown that if the total number of patients and blocks is fixed, the bias is maximized when the blocks are all the same size. On the other hand, if there are two or more investigators, each aware only of his own assignments and each attempting to bias the results, the bias appears to be minimized when the investigators enter the same number of patients.

Key words and phrases: Random allocation, random permuted blocks, returns to the origin of a constrained or unconstrained random walk, selection bias.

1. Introduction

In many clinical trials, patients enter sequentially, and must be treated as they arrive. It is desirable to ensure an equal number of patients in the treatment (T) and control (C) groups at several times throughout the trial. This is because there may be time trends, causing a possible unintentional bias if a disproportionate number of early patients were assigned to the control group, for example. Random permuted block randomization is often used to avoid this problem. If the trial is not double blind, an investigator who knew that random permuted blocks were being utilized could keep track of past assignments and figure out to which group the next eligible patient is more likely to be assigned. Such knowledge could affect the investigator's eligibility determination of the next patient. This is called selection bias.

Previous papers on randomization designs have evaluated selection bias in terms of the expected number of correct treatment assignment guesses by the investigator (Blackwell and Hodges (1957), Efron (1971), Matts and Lachin (1988))

or in terms of risk (Stigler (1969), Wei (1978)). An alternative measure of selection bias is the actual Type I error rate. This paper uses such a measure to quantify the degree of selection bias under different random permuted block designs. A simple model is proposed wherein the investigators, using their power to veto prospective patients, *act* according to their guesses (see 1.2). The model allows explicit calculation of the Type I error rate when the investigator attempts to bias the outcome.

1.1. Model

Suppose that treatment is intended to reduce the value of a continuous variable X. For simplicity, assume that X is normally distributed with known variance σ^2. These assumptions of normality and known variance will be relaxed for some of the results. With n observations in each group we base a one-tailed test of the effectiveness of the treatment on the statistic $Z = D/(2n\sigma^2)^{1/2}$, where

$$D = \left(\sum_{\text{control}} - \sum_{\text{treatment}} \right) (X_i). \tag{1.1}$$

Since Z is a constant multiple of D when σ is known, we may consider D to be the test statistic. The approximate one-tailed rejection region is $D \geq (2n\sigma^2)^{1/2} z_\alpha$. The biasing policy may be stated succinctly as follows: The investigator counts the numbers N_C and N_T of patients assigned to the control and treatment groups respectively thus far. He then chooses a patient with expected response:

$$\begin{cases} \mu - \eta & \text{if } N_T < N_C; \\ \mu & \text{if } N_T = N_C; \\ \mu + \eta & \text{if } N_T > N_C. \end{cases} \tag{1.2}$$

This is the setup of Blackwell and Hodges (1957), who then consider the expected number of correct guesses.

Assume throughout this paper that *there is no real treatment difference.*

2. One Block of Size $2k$ (Random Allocation)

A permuted block design with only one block is called random allocation. In this case the total sample size, $2n$, is $2k$, the block size. It is useful to consider the constrained random walk associated with the treatment assignments. With P_i denoting patient i, let

$$\xi_i = \begin{cases} -1 & \text{if } P_i \text{ is assigned to C} \\ +1 & \text{if } P_i \text{ is assigned to T,} \end{cases}$$

and $S_j = \sum_{i=1}^{j} \xi_i$. Let

$$\tau = \min\{i : 2 \leq i \leq 2k : S_i = 0\}$$

be the time of the first return to the origin of the random walk. Note that τ is well defined since we know that $S_{2k} = 0$. Throughout the paper N_{2k} will refer to the number of returns of a symmetric, unconstrained random walk up to and including time $2k$, and N'_{2k} will refer to the number of returns of a symmetric random walk constrained to have a return at time $2k$. Thus any statement about the distribution of N'_{2k} concerns the conditional distribution of N_{2k} given that $S_{2k} = 0$.

The X response of the first patient, P_1, will have mean μ. Assuming P_1 is assigned to T, the investigator will guess C next and choose the second patient with expected response $\mu+\eta$, and he will continue choosing patients with expected response $\mu + \eta$ until the number of C's equals the number of T's. Let

$$D_\tau = \left(\sum_{i \leq \tau, P_i \in C} - \sum_{i \leq \tau, P_i \in T} \right) (X_i).$$

Then by pairing one of the C's among $P_2, \ldots, P_\tau$ with P_1 we can write D_τ as $\eta + U + \sum_{j=1}^{\tau/2-1} V_j$, where U has mean 0 and variance $2\sigma^2$, and V_j are i.i.d. with mean 0 and variance $2\sigma^2$ and independent of U. The same representation holds if the first patient is assigned to C.

Once patient τ is reached, the entire process is repeated. Given that $N'_{2k} = r$,

$$D = r\eta + \sum_{i=1}^{r} U_i + \sum_{j=1}^{k-r} V_j, \tag{2.1}$$

where the U's and V's are jointly independent with 0 means and common variance $2\sigma^2$. Since we assume the response is normally distributed, then conditional on $N'_{2k} = r$, D has a normal distribution with mean $r\eta$ and variance $2k\sigma^2$.

We see that the distribution of the test statistic under (1.2) depends only on the number of returns of the constrained random walk associated with the treatment assignments. The greater the number of returns, the stochastically larger the test statistic becomes.

To find the unconditional distribution of D we must find the distribution of N'_{2k}.

It is shown in Proschan (1991) that

$$\Pr(N'_{2k} > j) = \binom{2k-j}{k}(k-j)2^{j+1}\left\{\binom{2k}{k}(2k-j)\right\}^{-1} \tag{2.2}$$

and therefore that

$$\Pr(N'_{2k} = j) = \binom{2k-j}{k} j 2^j \left\{ \binom{2k}{k} (2k-j) \right\}^{-1}. \tag{2.3}$$

It is also shown that

$$\lim_{k\to\infty} \Pr(N'_{2k}/(2k)^{1/2} > x) \to \exp(-x^2/2). \tag{2.4}$$

Thus the survival distribution of the test statistic is

$$\Pr(D/(2k\sigma^2)^{1/2} > \lambda) = 1 - E\Phi\Big(\lambda - N'_{2k}\eta/(2k\sigma^2)^{1/2}\Big). \tag{2.5}$$

Asymptotically,

$$\begin{aligned} & \Pr\Big(D/(2k\sigma^2)^{1/2} > \lambda\Big) \\ \to\ & 1 - \int_0^\infty \Phi(\lambda - \gamma x) x e^{-x^2/2} dx \\ =\ & 1 - \Phi(\lambda) + \frac{\gamma}{\sqrt{1+\gamma^2}} \exp\left(\frac{-\lambda^2}{2(1+\gamma^2)}\right) \Phi\left(\frac{\gamma\lambda}{\sqrt{1+\gamma^2}}\right), \end{aligned} \tag{2.6}$$

where $\gamma = \eta/\sigma$ = the selection effect.

The probabilities corresponding to different sample sizes and selection effects are given in Table 1 for $\alpha = .05$ and $\alpha = .01$. The degree of bias does not seem to change much with different total sample sizes of 50 or more. Even with a very small selection effect of .1 standard deviations the investigator is able to increase the probability of a Type I error from .05 to nearly .065. A selection effect of .3 standard deviations allows him to double the probability of a Type I error if $\alpha = .05$, and asymptotically to nearly triple it if $\alpha = .01$.

Table 1. Probability of a significant result, one investigator

	$\alpha = .05$			$\alpha = .01$		
$\gamma = \eta/\sigma$	$2k = 50$	$2k = 100$	$2k = \infty$	$2k = 50$	$2k = 100$	$2k = \infty$
.10	.063	.063	.065	.013	.014	.014
.20	.079	.080	.083	.018	.019	.020
.30	.098	.100	.107	.024	.025	.028
.40	.121	.124	.134	.033	.034	.039
.50	.147	.152	.166	.044	.046	.053

The asymptotic result (2.6) actually holds without the assumption of normally distributed responses. To see this, note that from (2.1),

$$D/(2k\sigma^2)^{1/2} = N'_{2k}\eta/(2k\sigma^2)^{1/2} + \frac{N'_{2k}}{(2k\sigma^2)^{1/2}}\left((1/N'_{2k})\sum_{i=1}^{N'_{2k}} U_i\right) + \frac{\{1 - N'_{2k}/k\}^{1/2}}{[2(k - N'_{2k})\sigma^2]^{1/2}}\sum_{i=1}^{k-N'_{2k}} V_i. \tag{2.7}$$

The corresponding U and V_j's if the first patient following a return is assigned to T do not necessarily have the same distribution as when the first patient is assigned to C. They do, however, have the same means and variances. The second term in (2.7) converges a.s. to 0 by the strong law of large numbers. Since $k - N'_{2k} \to \infty$ a.s. as $k \to \infty$, the third term converges in distribution to a standard normal deviate. Moreover the first and third terms are clearly asymptotically independent; so, by (2.4), $\Pr(D/(2k\sigma^2)^{1/2} > \lambda)$ tends to the right side of (2.6).

Up to now the variance has been assumed known, for simplicity. If it is unknown, then one would use the usual pooled estimate from the treatment and control groups. Within each group, there will be a mixture of observations with means μ, $\mu+\eta$, and $\mu-\eta$. The proportion of observations with mean μ becomes negligible as k becomes large. The proportion of observations with mean $\mu+\eta$ is approximately the proportion of time the constrained random walk spends above 0. This proportion converges in distribution to $U = \mu\{t : W^0(t) \geq 0\}$, where μ denotes Lebesgue measure and W^0 is a Brownian Bridge process. U is known to be uniformly distributed on $[0,1]$ (see Billingsley (1968, p.85)). Given that $U = u$, the pooled variance estimator will approach $\sigma^2 + 4\eta^2 u(1-u)$. This attains its maximum value of $\eta^2 + \sigma^2$ when $u = 1/2$, so the asymptotic survival distribution of D when the sample estimate is used in place of σ is at least as large as (2.6) with λ replaced by $\lambda\sqrt{1+\gamma^2}$. This makes very little difference if γ is small.

3. More Than One Block

The biasing policy (1.2) is also effective when there are $m > 1$ blocks. It makes no difference whether the block sizes are equal, and whether they are fixed or random. According to Pocock (1983), "Also, it is advisable not to inform clinicians that blocking is being used and especially they should not know the block size." However, policy (1.2) works well regardless of whether the investigator knows the block size. For any treatment allocation strategy which forces more balance than complete randomization, policy (1.2) is effective. This applies to all

permuted block randomization, whether the block sizes are fixed or random, and to Efron's (1971) biased coin design.

It is intuitively clear that of all random permuted block designs with fixed total sample size, the one minimizing bias is a single block (random allocation), and that the more one forces balance, the greater the selection bias will be. This will be formalized through Theorem 3.1 below.

Recall, first, that random variable Y is said to be stochastically larger than random variable X if $\Pr(X > z) \leq \Pr(Y > z)$ for all z. This is much stronger than the expectation of Y exceeding the expectation of X. In fact it can be shown that if Y is stochastically larger than X then $E\phi(X) \leq E\phi(Y)$ for every increasing function ϕ for which the expectations exist (Marshall and Olkin (1979, p.481)).

Theorem 3.1. *Suppose a study consists of a combination of random permuted blocks of various sizes, with total sample size $2n$. If the investigator adopts policy (1.2), the test statistic D (as well as the number of correct treatment assignment guesses) becomes stochastically smaller each time a pair of blocks is collapsed into one. In particular, the design making D stochastically smallest is a single block design (random allocation).*

Proof. It suffices to show that the number N'_{2k} of returns to the origin of the random allocation design with sample size $2k$ is stochastically smaller than the total number $N'_{2j} + N'_{2k-2j}$ of returns when the sample is split into two blocks of respective sizes $2j$ and $2k - 2j$. Let S_i denote a symmetric random walk, and condition upon $S_{2k} = 0$. We further condition on the value of S_{2j}, and show that for any given value of S_{2j}, N'_{2k} is stochastically smaller than $N'_{2j} + N'_{2k-2j}$. It suffices to show that the number of returns to the origin of a random walk beginning at $S_0 = 2r$ and constrained to be 0 at time $2j$ is stochastically smaller than the number of returns of a random walk beginning at $S_0 = 0$ and constrained to be 0 at time $2j$, for any integer r. Let T_i be the constrained random walk beginning at the origin and U_i be the constrained random walk beginning at $2r$. Without loss of generality assume that $r > 0$ and that $T_1 = 1$. Let τ be the time of the first return to the origin of T_i, and let μ be the time of the first return to the origin of U_i. Independently, generate a realization from each process, and observe both. Then one of the following mutually exclusive events must occur:

$$\begin{aligned} A &= \{\exists\, i^* \leq \min(\tau, \mu) \ni T_{i^*} = U_{i^*}\} \quad \text{or} \\ B &= A^c \cap \{\tau \leq \mu\}. \end{aligned}$$

Given event A, the number of returns for the two processes has the same distribution, and it is easy to show directly using (2.2) that given event B, T_i has a stochastically larger number of returns to the origin than U_i.

In Section 2 we considered the distribution of D under random allocation. Under this scenario the number of returns to the origin is of the order of the square root of the total sample size, so, by expression (2.5) the test statistic converges weakly to a non-degenerate random variable. If more than one block is used and there is a return at $2n$, then (2.5) is still valid with k replaced by n and N'_{2k} replaced by the total number of returns by time $2n$, whether forced by design or occurring by chance. It is therefore clear that *if the number of returns is of the order of the sample size itself instead of its square root, the probability of a Type I error converges to 1.* Thus, rejection of the null hypothesis will be assured with a large enough sample size. This will be the case whenever random permuted blocks with a fixed maximum block size are used, irrespective of whether there is a forced return at $2n$. It is also the case with Efron's (1971) biased coin design. A comparison of different designs for which the number of returns is of the order of the sample size is therefore moot.

A result stronger than Theorem 3.1 undoubtedly holds. It seems reasonable that not only should the test statistic become stochastically larger each time a pair of blocks is collapsed into one, but also each time a pair of blocks is made more nearly equal in size without changing the total size of the two blocks. This leads naturally to the topic of majorization and Schur functions.

For a vector $\boldsymbol{z}$, let $z_{[i]}$ denote the ith largest component of $\boldsymbol{z}$. Recall that the vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ is said to be smaller than $\boldsymbol{y} = (y_1, \ldots, y_n)$ in the majorization ordering if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}$$

for $k = 1, \ldots, n$, with equality when $k = n$. A function $f(\boldsymbol{x})$ is said to be Schur-concave if $f(\boldsymbol{x})$ decreases as $\boldsymbol{x}$ increases in the majorization ordering. The above conjecture may then be rephrased as: With m blocks of respective sizes $b_1, \ldots, b_m$, the probability $\Pr(D > c \mid b_1, \ldots, b_m)$ is a Schur-concave function of $(b_1, \ldots, b_m)$ for each c. Unfortunately, this author has been able to prove this only asymptotically:

Theorem 3.2. *Suppose an investigator follows strategy* (1.2) *in a random permuted block design with m blocks of respective sizes $b_1, \ldots, b_m$, and total sample size $2n$. Fix m and let $n \to \infty$, $b_i \to \infty$ in such a way that $b_i/(b_1 + \cdots + b_m) \to \pi_i > 0$, $1 \leq i \leq m$. Then $\lim_{n\to\infty} \Pr(D/(2n\sigma^2)^{1/2} > c \mid \boldsymbol{\pi})$ is a Schur-concave function of $\boldsymbol{\pi}$ for each c. In particular, D is asymptotically stochastically largest when $\boldsymbol{\pi} = (1/m, \ldots, 1/m)$.*

Proof. It suffices to prove the theorem for $m = 2$, with respective block sizes $b_1 = 2j$ and $b_2 = 2n - 2j$. As usual let N'_{2j} denote the number of returns to the

origin of a symmetric random walk constrained to have a return at $2j$, and let N'_{2j} and N'_{2n-2j} be independent. We need only prove that the limiting distribution of $(N'_{2j} + N'_{2n-2j})/(2n)^{1/2}$ as $j \to \infty$, $n \to \infty$, $j/n \to \lambda$ becomes stochastically larger as λ increases from 0 to 1/2. Now

$$(N'_{2j} + N'_{2n-2j})/(2n)^{1/2} \to \lambda^{1/2}U + (1-\lambda)^{1/2}V,$$

where U and V are independent Weibulls with common survival distribution given by (2.4). It suffices to show that the survival function $\psi(\lambda, t) = \Pr(\lambda^{1/2}U + (1-\lambda)^{1/2}V > t)$ is increasing in λ for $0 < \lambda < 1/2$ for each fixed t.

$$\psi(\lambda, t) = \int_0^{t/(1-\lambda)^{1/2}} \exp\left\{ -\left(\frac{t - (1-\lambda)^{1/2}v}{\lambda^{1/2}}\right)^2 /2 \right\} v \exp(-v^2/2)\, dv + \int_{t/(1-\lambda)^{1/2}}^{\infty} v \exp(-v^2/2)\, dv.$$

Integrating this expression, differentiating with respect to λ and simplifying yields:

$$\frac{\partial \psi}{\partial \lambda} = (2\pi)^{1/2} e^{-t^2/2} \left\{ \phi(v) - \phi(u) + \frac{(v-u)}{2}\Big(\Phi(u) + \Phi(v) - 1\Big) \right\},$$

where ϕ and Φ are the density and distribution function of a standard normal random variable, and $u = (\frac{\lambda}{1-\lambda})^{1/2} t < (\frac{1-\lambda}{\lambda})^{1/2} t = v$ for $\lambda < 1/2$. It therefore suffices to show that $h(s, \delta) = \phi(s) - \phi(s-\delta) + (\delta/2)(\Phi(s-\delta) + \Phi(s) - 1)$ is nonnegative for $0 \le \delta \le s$. It is easily verified that $\partial h/\partial s = \frac{1}{2}(2s - \delta)(\phi(s-\delta) - \phi(s)) \ge 0$, and hence the value of $s \ge \delta$ minimizing $h(s, \delta)$ is $s = \delta$. It is therefore sufficient to verify that $g(\delta) = h(\delta, \delta) = \phi(\delta) - 1/(2\pi)^{1/2} + (\delta/2)(\Phi(\delta) - 1/2) \ge 0$ for all $\delta > 0$. Now $g'(\delta) = (\delta/2)((1/\delta)(\Phi(\delta) - 1/2) - \phi(\delta)) = (\delta/2)(\phi(\delta^*) - \phi(\delta))$ for some $0 \le \delta^* \le \delta$. Since this is nonnegative, g is increasing in $\delta > 0$. The proof is completed by noting that $\lim_{\delta \to 0} g(\delta) = 0$ is nonnegative.

We see, then, that asymptotically the more nearly equal the block sizes the greater the Type I error rate if the investigator is attempting to bias the results of the trial. Equal block sizes are commonly used in practice.

4. More Than One Investigator

It is usually the case in clinical trials that separate blocked randomization is used for each center participating in the trial. It is not uncommon for a center to consist of several different hospitals called satellites. In this case an investigator from one hospital might know the treatment assignments for his hospital, but not those of the other satellites. It is interesting to consider what

bias reduction results when the overall study is a random allocation design, but there are 2 or more investigators, each oblivious to the treatment assignments of the others. To illustrate, consider a random allocation design with $2n$ patients and 2 investigators. Suppose that each investigator, not knowing the treatment assignments of the other, follows strategy (1.2). Let S_i be the random walk representing the state of treatment assignments after patient i, and without loss of generality assume the first investigator selects patients $1, \ldots, 2k$. Let θ_1 be the time of the last return to the origin of S_i at or before time $2k$.

The first investigator will not make as many correct guesses as if a block (random allocation) of size $2k$ were used for him for two reasons:

(1) $S_1, \ldots, S_{2k}$ will have stochastically fewer returns to the origin with the constraint that $S_{2n} = 0$ than it would with the constraint that $S_{2k} = 0$.
(2) There will be a "tail" of more incorrect than correct guesses at times $\theta_1 + 1$, $\ldots, 2k$ unless $S_{2k} = 0$.

For example, Figure 1 shows what might happen with the first investigator entering $2k = 10$ out of the total of $2n = 20$ patients. In that figure $\theta_1 = 4$ and $S_{2k} = 2$. The first investigator may or may not guess correctly for patient 5. He will correctly guess assignments 8 and 9, and incorrectly guess assignments 6, 7, and 10.

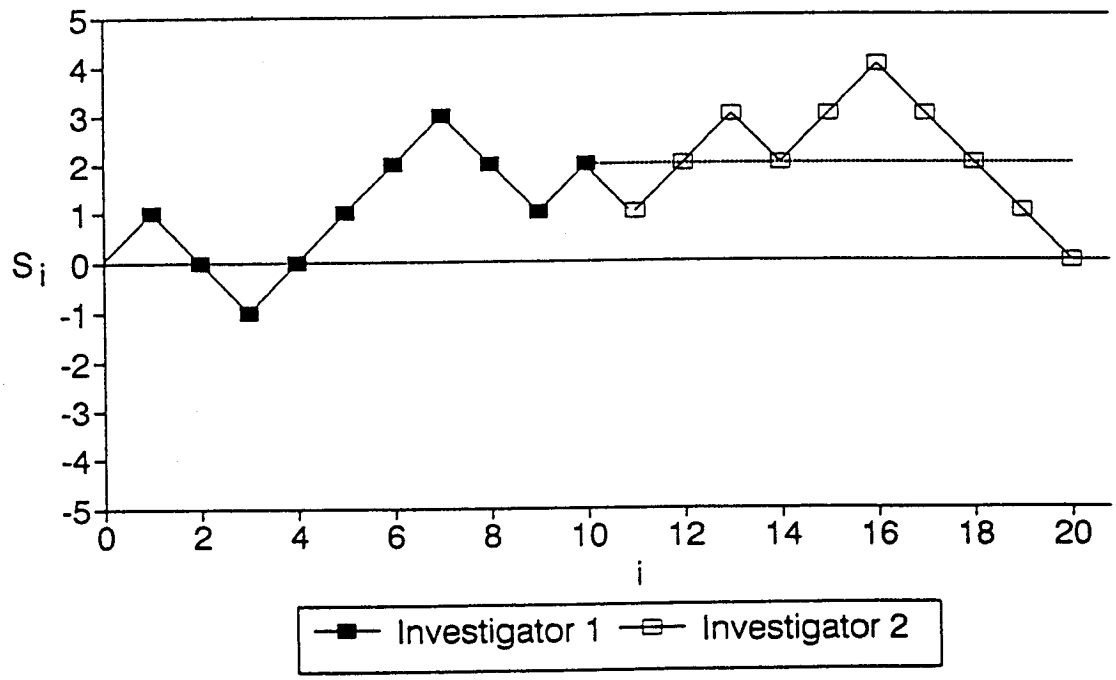

Figure 1. Random walk: 2 investigators
Overall random allocation

How many correct guesses the first investigator makes is seen to depend on the number of returns to the origin up to and including time $2k$, and on S_{2k}. Let N_{2k} be the number of returns to the origin of the random walk up to and including $2k$. We must first determine $\Pr(N_{2k} = \nu_1 \cap S_{2k} = 2j \mid S_{2n} = 0)$. For $j = 0$ one can use formula (2.3) to conclude that $\Pr(N_{2k} = \nu_1 \cap S_{2k} = 0 \mid S_{2n} = 0) = \nu_1 2^{\nu_1} \binom{2k-\nu_1}{k} \binom{2n-2k}{n-k} / [(2k-\nu_1)\binom{2n}{n}]$ for $0 \le \nu_1 \le k$. For $j > 0$,

$$\begin{aligned} &\Pr(N_{2k} = \nu_1 \cap S_{2k} = 2j \mid S_{2n} = 0) \\ &= \sum_{r=1}^{k-j} \Pr(A_{\nu_1,2r} \cap B_{2j,2k-2r} \cap C_{2n-2k,2j}) / \Pr(S_{2n} = 0), \end{aligned}$$

where $A_{\nu_1,2r} = \{\nu_1\text{th return of the random walk occurs at time } 2r\}$, $B_{2j,2k-2r} = \{S_{2r+1} > 0, S_{2r+2} > 0, \ldots, S_{2k} = 2j\}$, and $C_{2n-2k,2j} = \{S_{2n} - S_{2k} = -2j\}$. From Feller (1968) the probability of $A_{\nu_1,2r}$ is $\nu_1 \binom{2r-\nu_1}{r} \Big/ [2^{2r-\nu_1}(2r-\nu_1)]$. Conditioned on $A_{\nu_1,2r}$, the probability that $B_{2j,2k-2r}$ occurs is the probability that a random walk beginning at state 0 is always positive and ends up at state $2j > 0$ after $2k - 2r$ steps. By the ballot theorem (Feller (1968)) this probability equals $j\binom{2k-2r}{k-r+j} / [2^{2k-2r}(k-r)]$. Finally, conditioned on $A_{\nu_1,2r}$ and $B_{2j,2k-2r}$, the probability of $C_{2n-2k,2j}$ is $\binom{2n-2k}{n-k-j} / 2^{2n-2k}$. Putting these all together and simplifying one obtains, for $0 \le \nu_1 \le k - j$,

$$\begin{aligned} &\Pr(N_{2k} = \nu_1 \cap S_{2k} = 2j \mid S_{2n} = 0) \\ &= \begin{cases} \dfrac{\binom{2n-2k}{n-k-j} j 2^{\nu_1}}{\binom{2n}{n}} \left\{ \dfrac{\binom{2k-2\nu_1}{k-\nu_1+j}}{(k-\nu_1)} + \nu_1 \displaystyle\sum_{r=\nu_1+1}^{k-j} \dfrac{\binom{2k-2r}{k-r+j}\binom{2r-\nu_1}{r}}{(2r-\nu_1)(k-r)} \right\} & \text{if } j > 0 \\ \dfrac{\nu_1 2^{\nu_1} \binom{2k-\nu_1}{k} \binom{2n-2k}{n-k}}{(2k-\nu_1)\binom{2n}{n}} & \text{if } j = 0. \end{cases} \end{aligned} \tag{4.1}$$

Now the second investigator will not know the treatment assignments of the first, and consequently will not know that his portion of the constrained random walk is starting at state $2j$. He will act as though the random walk were starting at 0, and will flip a coin to guess whether the first patient will be assigned to T or C. Likewise he will flip a coin each time the random walk returns to the state $2j$. Let N_{2n-2k} be the number of returns of the random walk to the state $2j$ after time $2k$. In Figure 1, $N_{2n-2k} = 3$. In a manner similar to the above one can show that for $0 \le \nu_2 \le n - k - j$,

$$\Pr(N_{2n-2k} = \nu_2 \mid N_{2k} = \nu_1, S_{2k} = 2j, S_{2n} = 0)$$

$$= \begin{cases} \dfrac{j2^{\nu_2}}{\binom{2n-2k}{n-k-j}} \left\{ \dfrac{\binom{2n-2k-2\nu_2}{n-k-\nu_2+j}}{(n-k-\nu_2)} + \nu_2 \displaystyle\sum_{r=\nu_2+1}^{n-k-j} \dfrac{\binom{2n-2k-2r}{n-k-r+j}\binom{2r-\nu_2}{r}}{(2r-\nu_2)(n-k-r)} \right\} & \text{if } j > 0 \\ \nu_2 2^{\nu_2} \dfrac{\binom{2n-2k-\nu_2}{n-k}}{(2n-2k-\nu_2)\binom{2n-2k}{n-k}} & \text{if } j = 0. \end{cases} \tag{4.2}$$

The next step is to find the conditional distribution of D given $S_{2k} = 2j$, $N_{2k} = \nu_1$, and $N_{2n-2k} = \nu_2$. Assume for the moment that $j > 0$. Recall that θ_1 was the last return to the origin at or before time $2k$. Patient $\theta_1 + 1$ will have mean μ, and then patients $\theta_1 + 2, \ldots, 2k$ will all have mean $\mu + \eta$. Thus $(\sum_{\theta_1+1\le i\le 2k, P_i\in C} - \sum_{\theta_1+1\le i\le 2k, P_i\in T})(X_i)$ is normally distributed with mean $(1-2j)(\mu+\eta) - \mu$. Now let θ_2 be the last time at or after patient $2k$ that the random walk returns to state $S_{2k} = 2j$. In Figure 1, $\theta_2 = 18$. $(\sum_{\theta_2+1\le i\le 2n, P_i\in C} - \sum_{\theta_2+1\le i\le 2n, P_i\in T})(X_i)$ is normally distributed with mean $(2j-1)(\mu-\eta)+\mu$. It follows that the conditional distribution of D given $S_{2k} = 2j$, $N_{2k} = \nu_1$, and $N_{2n-2k} = \nu_2$ is normal with variance $2n\sigma^2$ and mean

$$\begin{cases} (\nu_1 + \nu_2 + 2(1-2j))\eta & \text{if } j > 0 \\ (\nu_1 + \nu_2)\eta & \text{if } j = 0. \end{cases} \tag{4.3}$$

If $j < 0$ then (4.1)-(4.3) are valid with j replaced by $|j|$.

Using (4.1)-(4.3) one can obtain the distribution of the test statistic. Table 2 shows the actual Type I error rate if two investigators use strategy (1.2) in an overall random allocation design of total sample size 60. As can be seen, the Type I error rate is lowest when each investigator enters the same number of patients, and this rate is substantially lower than when one investigator enters all patients. For a selection effect of .3 standard deviations the Type I error rate is .099 if one investigator enters all 60 of the patients, and it is .073 if each investigator enters 30 patients. This is not surprising in view of the fact that the constrained random walk tends to be largest in absolute value at the midway point. It is conjectured that with m investigators, each following strategy (1.2) applied to his own treatment assignments, the test statistic is stochastically smallest when the blocks are all the same size.

Table 2. Probability of a significant result, two investigators

$2n = 60$, $\alpha = .05$

$\gamma = \eta/\sigma$	(0,60)	(10,50)	(20,40)	(30,30)
.10	.063	.058	.056	.056
.20	.079	.067	.064	.063
.30	.099	.079	.075	.073
.40	.122	.094	.087	.086
.50	.148	.110	.102	.100

5. Conclusion

It is well recognized that a clinical trial should be double blind, if possible, and that the treatment assignments should be nearly balanced throughout the study. If double blinding is impossible or infeasible, then selection bias can result in serious inflation of the Type I error rate. Moreover this inflation increases as we try to make treatment assignments more balanced. For example, a random allocation design (one large block), which requires balance only at the end of the study and could be quite unbalanced in the interim, was shown to be least susceptible to selection bias. By contrast blocks of size two, which guarantee balance after each pair of patients, yield the largest bias. Furthermore, for a fixed sample size and number of blocks, the bias is maximized when the block sizes are all equal. This is precisely because the more nearly equal the block sizes, the more balanced the treatment assignments tend to be. Even if we sacrifice some treatment balance and use a random allocation design, the bias does not disappear as we increase the sample size. This reinforces the idea that double blinding is critical when possible.

It may be that the overall design is a random allocation but there are two or more investigators, each knowing only his own past treatment assignments. If each is attempting to bias the results, the impact on the Type I error rate is smaller than if there were only one investigator, but it is still noticeable. In this case the bias appears to be minimized when each investigator enters the same number of patients.

Acknowledgement

I would like to thank the referee for many helpful suggestions which made the model more realistic and the paper less verbose.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

Blackwell, D. and Hodges, J. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28**, 449–460.

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

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3rd edition. John Wiley, New York.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

Matts, J. P. and Lachin, J. M. (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials* **9**, 327–344.

Pocock, S. J. (1983). *Clinical Trials: A Practical Approach.* John Wiley, New York.

Proschan, M. A. (1991). A note on Blackwell and Hodges (1957) and Diaconis and Graham (1981). *Ann. Statist.* **19**, 1106-1108.

Stigler, S. M. (1969). The use of random allocation for the control of selection bias. *Biometrika* **56**, 553-560.

Wei, L. J. (1978). On the random allocation design for the control of selection bias in sequential experiments. *Biometrika* **65**, 79–84.

National Heart, Lung, and Blood Institute, Federal Building, Room 2A11, 7550 Wisconsin Avenue, Bethesda, MA 20892, U.S.A.

(Received May 1991; accepted July 1993)