A Comparative Study of Restricted Randomization Methods in Clinical Trials

Myung-Hoe Huh*

ABSTRACT

In clinical trials subjects are available sequentially and must be assigned to treatments immediately. Completely randomized procedure for the allocation of treatments to each subject may result in severe imbalance among the number of subjects in treatment groups, especially for small experiments or interim analyses of large experiments. In this study, restricted randomization methods such as biased coin designs (Efron, 1971), permuted block design, and truncated binomial design are compared to the completely randomized design in the presence of selection and/or accidental bias by Monte-Carlo simulations.

1. Introduction

When the experimental units are available at the same time for the experiments, it is possible to balance the number of units in treatments groups. Completely randomized design in such cases is one way of achieving perfectly balanced experiments. In clinical trials, however, subjects (experimental units) arrive sequentially for the experiment and must be assigned to treatments immediately. As a consequence, perfectly balancing plans such as Student sandwich plan $TCCTTCCT...$ (Efron, 1971) have some determinicity (lack of randomness) which may bring in biases, where 'T' for the treatment and 'C' for the control. On the other hand, completely randomized design (CRD), in which each subject has equal chance of being assigned to a particular treatment, may result in severe imbalance among the number of subjects in treatment groups, especially for small experiments or interim analyses of large experiments.

Efron (1971) proposed the biased coin design (BCD) as a compromise of perfect balancing and complete randomization. In BCD, higher chance is given to the treatment group which has fewer number of subjects so far. Put it more formally, suppose that the number of treatments is two (control and treatment), and let n_c and n_T be the number of subjects in the control and the treatment group respectively after assigning $n = n_c + n_T$ subjects. Then the next subject is assigned to the control with the probability

$$0.5, \quad \text{if } n_c = n_T,$$
$$p, \quad \text{if } n_c < n_T,$$
$$1-p, \quad \text{if } n_c > n_T,$$

where $p \geq 0.5$. The above design is abbreviated as BCD(p), where p is called the biasing parameter. Note that BCD (1/2) is identical with CRD. BCD (2/3) is advocated in Efron (1971) without any discussion and, later, BCD (3/4) is used in Efron (1980). The Northern California Oncology Group (NCOG) is using a variant of BCD (3/4) in all clinical trials (Hannigan and Brown, 1982).

Permuted block design (PBD) is another alternative in sequential experiments. In the case of two treatments (control and treatment), the experiment is divided into blocks of even length, say $2b$, and within each block b subjects are randomly assigned to the control and the other b subjects to the treatment. Therefore near the end of each block there is some determinicity in the sense that the experimenter knows which treatment will be assigned to the next subject before it is available. As an example, for the case $2b = 10$, PBD may yield a sequence

$$TCCCTCTCTT.$$

In this case, after allocating the eighth subject the remaining subjects are to be assigned to the treatment 'T'. In any case assignment of the last subject in each block is pre-determined.

Another competing plan using blocks is the truncated binomial design (TBD). In each block of legnth $2b$, subjects are assigned to treatments (control and treatment) in a completely random manner until either control or treatment gets its quota b. Then the following subjects belonging to that block are to be assigned to the other treatment.

Experiments with restricted randomization procedures can be biased from the two sources, as mentioned by Efron (1971). First, they are not free from selection bias,

which is due to the prejudice of experimenter if he/she knows with certain probability which treatment the next subject is to be assigned. Second, they are not free from accidental bias, which is due to nuisance factors systematically affecting subjects.

In this study, restricted randomization methods such as BCD (2/3), BCD (3/4), PBD, TBD are compared to CRD in the presence of selection and/or accidental bias by Monte-Carlo methods for two treatment sequential experiments without stratification.

In practice, however, subjects are stratified according to their characteristics before the treatment assignment without few exceptions (Simon, 1979). Even though this study is not directly concerned with stratified clinical trials, the author believes that the conclusions can be extended to such cases since a clinical trial with several strata can be regarded as a collection of clinical trials with single stratum.

2. Comparisons in the Presence of Selection Bias

Suppose that the experimenter believes the treatment is superior to the control and that, as a consequence, the response is upgraded by the amount $B(-B)$ if he/she knows with certainty that the next subject will be assigned to the treatment (control). If the next patient is going to be assigned to the treatment with probability p_{NEXT} and if the experimenter is rational, the selection bias is reflected on the response of the next subject by the amount

$$B \cdot p_{\text{NEXT}} + (-B)(1 - p_{\text{NEXT}}) = B \cdot (2p_{\text{NEXT}} - 1).$$

Therefore, CRD is free from the selection bias since p_{NEXT} is always equal to 1/2. On the other hand, BCD's, PBD, TBD are not so.

Comparisons among various randomization plans are made as follows. Let y_{ij} be the jth response of treatment i ($i = C$ or T). Assume that y_{ij} are independent and that

$$E y_{ij} = \mu_i + B \cdot (2p_{\text{NEXT}} - 1) \equiv \mu_i + \beta_{ij}, \ \text{Var } y_{ij} = \sigma^2,$$

where μ_i is the expected respose of treatment i in the absence of selection bias. The object of the experiment is to estimate the difference $\mu_T - \mu_C$.

After allocating n subjects, suppose that we have n_C controls and n_T treatments ($n = n_C + n_T$). Then

$$\text{Var}(\bar{y}_T - \bar{y}_C) = \sigma^2(1/n_C + 1/n_T) \tag{2.1}$$

where $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. Note that (2.1) achieves the minimum for fixed n when n_C is

close to n_T as nearly as possible. Hence CRD suffers more from imbalancing than restricted randomization plans. On the other hand. CRD is free from selection bias which can be written as

$$E(\bar{y}_T - \bar{y}_c) - (\mu_T - \mu_c) = \sum_j \beta_{Tj}/n_T - \sum_j \beta_{cj}/n_c.$$

Therefore, mean squared error (MSE) criterion

$$\text{MSE}(\bar{y}_T - \bar{y}_c) = \sigma^2[1/n_c + 1/n_T] + \left[\sum_j \beta_{Tj}/n_T - \sum_j \beta_{cj}/n_c\right]^2$$

is considered to compare various restricted randomization plans. Monte-Carlo evaluation proceeded roughly as follows. First, we simulated a number of sequential trials (100 for various restricted randomization designs and 332 for CRD, which will be explained shortly) with each randomization method for the trial size n and the magnitude of bias B. The MSE's are computed accordingly for each trial, finally yielding the average MSE. Second, for each restricted randomization design and fixed n and B, we calculate "percent MSE" which is by definition the ratio of its average MSE over the average MSE of CRD. One advantage of using "percent MSE" for comparing various restricted randomization designs with CRD is that "percent MSE" depends only on the magnitude σ^2/B^2 while the MSE for restricted randomization designs depends on both σ^2 and B. (Note that MSE is the sum of two terms: the first term is proportional to σ^2 and the second term to B^2.) Therefore σ^2 is set to be equal to 1 and non-negative values of B are tried without loss of generality.

Note that the same average MSE of CRD can be used for comparison with other designs for various sizes (eleven cases) of B, since CRD is free from selection bias. Hence, more precision is desirable for the MSE of CRD than for the MSE of respective restricted randomization designs. In computing the average MSE of CRD, 332 $(= \sqrt{11} \times 100)$ simulation results are combined, while 100 simulations are used for other designs.

To limit the number of different levels of B, let's try to give a meaning to the magnitude of B, say 0.5. If the next subject is to be assigned to the treatment (control) with probability one the would-be responses take the value around $\mu_T + 0.5$ ($\mu_c - 0.5$), which corresponds to the 69th (31st) percentile for the normal distribution with mean μ_T (μ_c) and unit variance. Therefore the selection bias displaces the 50th percentile to the right or left sides. See Table 1 for the amount of displacements in terms of percentiles with different values of B.

Table 1. Location of the 50th percentiles under the selection bias of magnitude
B in the reference normal distribution

B	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
Treatment	50	54	58	62	66	69	73	76	79	82	84
Control	50	46	42	38	34	31	27	24	21	18	16

In author's opinion, selection bias larger than 0.7 is rather extreme and of little practical importance. However, Monte-Carlo simulations are performed for eleven values of B between 0.0 and 1.0 in steps of 0.1. The last three cases ($B=0.8$, 0.9, and 1.0) are considered to see what can be said under such rather extreme cases. For the size n of sequential trials, three values 10, 20 and 50 are tried. Trial size $n=$ 10 and/or 20 might be appropriate for comparisons of small experiments or interim analyses of large experiments. Large trial size $n=50$ is considered in the hope that it might suggest the asymptotic behaviors.

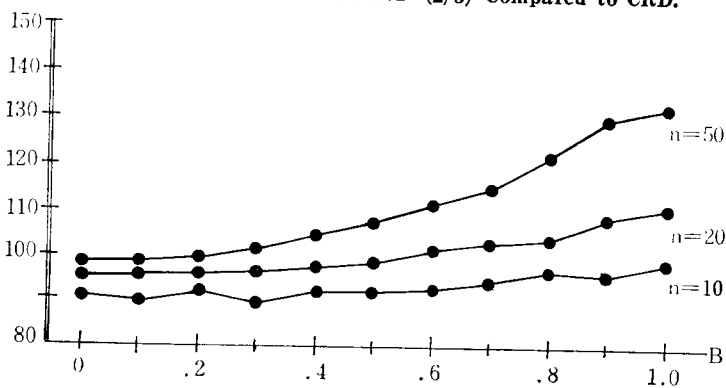**Figure 2.1 Percent MSE of BCD (2/3) Compared to CRD.**
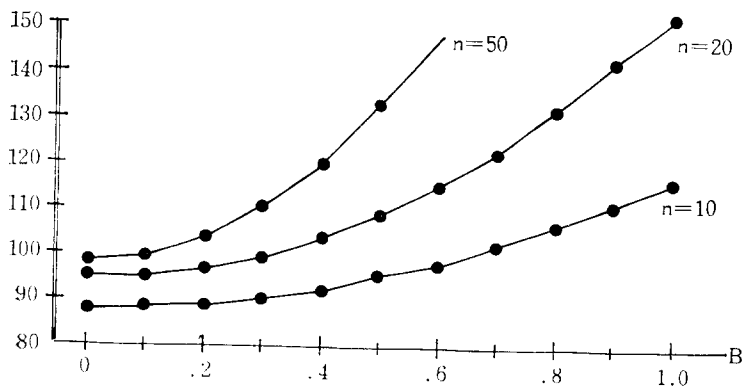

Figure 2.2 Percent MSE of BCD (3/4) Compared to CRD.

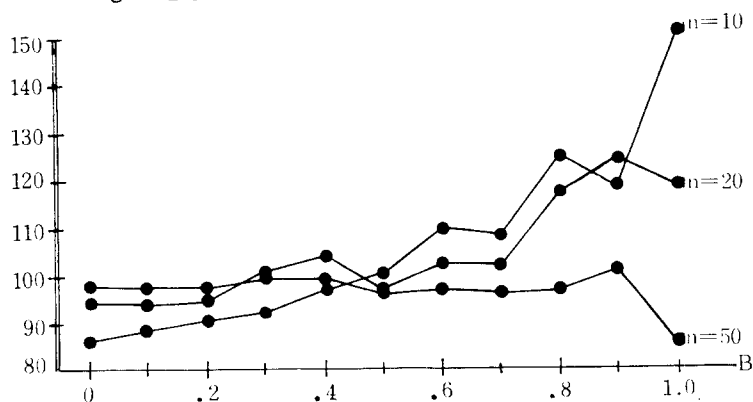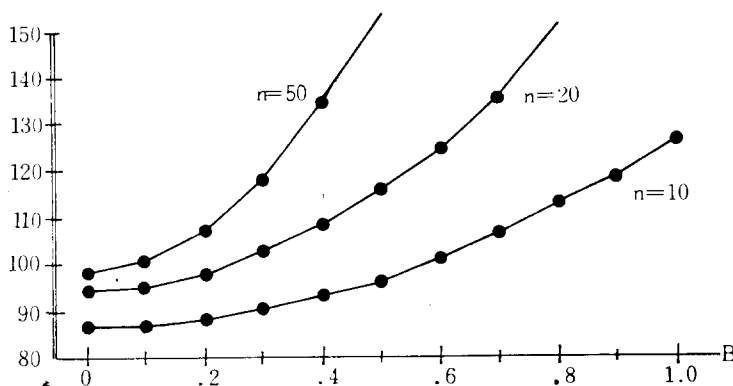**Figure 2.3 Percent MSE of PBD Compared to CRD.**

Figure 2.4 Percent MSE of TBD Compared to CRD.

See Figures 2.1—2.4 for the "percent MSE" of various restricted randomization designs compared to the CRD. By the reasons stated before, "percent MSE" curve is symmetrical around the vertical axis $B=0$.

In Figure 2.1, BCD (2/3) is preferred to CRD on the "whole" interval $[0, 1]$ of B for $n=10$, on $[0, 0.5]$ for $n=20$, and on $[0, 0.2]$ for $n=50$.

In Figure 2.2, BCD (3/4) is preferred to CRD on the interval $[0, 0.65]$ of B for $n=10$, on $[0, 0.3]$ for $n=20$, and on $[0, 0.1]$ for $n=50$.

In Figure 2.3, PBD (with block size $2b=10$) is preferred to CRD on the interval $[0, 0.6]$ of B for $n=10$, on $[0, 0.25]$ for $n=20$, and on $[0, 0.1]$ for $n=50$.

In Figure 2.4, TBD is preferred to CRD on the interval $[0, 0.4]$ of B for $n=10$, on $[0, 0.25]$ for $n=20$, and on $[0, 0.1]$ for $n=50$.

Note that all restricted randomization methods considered here have "good" performances compared to CRD for "small" absolute values of B and "small" trial size n.

Also note that the intervals of B on which such dominance over CRD appears are nested, for each n, in a sequence of (from the smallest to the largest) TBD, PBD, BCD (3/4), BCD (2/3). In other words, BCD (2/3) dominates CRD whenever any other designs do. On the other hand, BCD (3/4) has smaller "percent MSE" than BCD (2/3) for B between -0.4 and 0.4.

In summary, BCD (2/3) is generally recommendable for small experiments. However, BCD (3/4) is more suitable when relative magnitude of bias B can be controlled to be small.

3. Comparisons in the Presence of Accidental Bias

It is not possible to incorporate all possible types of accidental biases into a Monte-Carlo study. In this section, restricted randomization designs are compared to CRD in the presence of two typical types of accidental biases. First, biases are assumed to appear alternating:

$$+B \ -B \ +B \ -B \ +B \ -B \ +B \ -B \ +B \ -B\cdots\cdots \qquad (3.1)$$

Second, biases are assumed to appear alternating with groups of five:

$$+B \ +B \ +B \ +B \ +B \ -B \ -B \ -B \ -B \ -B\cdots\cdots \qquad (3.2)$$

Monte-Carlo study was carried out in the same way as in Section 2 and showed that all of the restricted randomization designs considered, BCD (2/3), BCD (3/4), PBD, and TBD, have smaller or nearly the same average MSE compared to CRD under the accidental biases of types (3.1), and (3.2) except for TBD under (3.2). See Figure 3.1 for the "percent MSE" of TBD compared to CRD under the accidental bias of type

Figure 3.1 Percent MSE of TBD Compared to CRD under the Accidental Bias of Type (3.2).
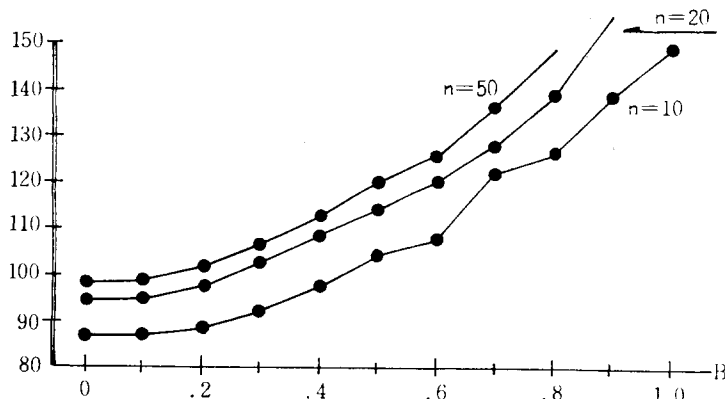
(3.2). In computing average MSE of CRD for each n, 332 simulation results are combined while 100 simulations are for other designs. It is designed for the analogical presentation of results to that in Section 2. (In computing average MSE of CRD for $n=10$ at $B=0.0$, (0.1), 0.2, 0.4, (0.6), 0.7, and 0.9, one (two) out of 332 Monte-Carlo trials which yielded perfect imbalance ($n_C=0$ or $n_T=0$) is (are) excluded).

In Figure 3.1, we can see that TBD suffers much from the accidental bias of type (3.2) for $B \geq 0.6$ and $n=10$ or 20. However, such phenomenon diappears for the case $n=50$ simply because the period of (3.2), 10, is much smaller than n. Thus it is apparent that TBD is not strong against the accidental biases with long period.

4. Conclusion

In the presence of moderate selection and/or accidental bias, it is shown by Monte-Carlo study that BCD (2/3) is generally recommendable among several restricted randomization methods in clinical trials without stratification. However, if one can validly assume that the selection bias is pretty small, more restricted biased coin design such as BCD (3/4) is more suitable.

REFERENCES

(1) Efron, B. (1971) Forcing a sequential experiment to be balanced. *Biometrika, 58*, 403—417.
(2) Efron, B. (1980) Randomizing and balancing a complicated sequential experiment. *Biostatistics Casebook* (edited by R.G. Miller, Jr. et al.), Wiley and Sons, Inc., New York, 19—30.
(3) Hannigan, J.F., Jr., and Brown, B.W., Jr. (1982) Adaptive randomization biased coin-design: Experience in a cooperative group clinical trial. Technical Paper #74, Division of Biostatistics, Stanford University, Stanford, CA.
(4) Simon, R. (1979) Restricted randomization designs in clinical trials. *Biometrics, 35*, 503—512.