Randomization in Clinical Trials: Conclusions and Recommendations

John M. Lachin, ScD

The George Washington University, Department of Statistics/Computer and Information Systems, The Biostatistics Center, Rockville, Maryland

John P. Matts, PhD

Department of Surgery, University of Minnesota, Minneapolis, Minnesota

L.J. Wei, PhD

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan

Abstract: The statistical properties of simple (complete) randomization, permuted-block (or simply blocked) randomization, and the urn adaptive biased-coin randomization are summarized. These procedures are contrasted to covariate adaptive procedures such as minimization and to response adaptive procedures such as the play-the-winner rule. General recommendations are offered regarding the use of complete, permuted-block, or urn randomization. In a large double-masked trial, any of these procedures may be acceptable. For a given trial, the relative merits of each procedure should be carefully weighed in relation to the characteristics of the trial. Important considerations are the size of the trial, overall as well as within the smallest subgroup to be employed in a subgroup-specific analysis, whether or not the trial is to be masked, and the resources needed to perform the proper randomization-based permutational analysis.

KEY WORDS: *Randomization, complete randomization, permuted-block randomization, urn randomization, treatment adaptive, covariate adaptive, response adaptive*

INTRODUCTION

In the preceding articles [1-4] we reviewed the basic properties of various randomization procedures for the assignment of patients to receive one of two treatments designated *a* or *b*. The general properties of randomization procedures were defined and discussed, principally the permutational prop-

erties with respect to all of the possible ways that the treatments could have been assigned to the patients [1]. These include the effects of treatment group imbalances on the statistical power of tests, the impact of predictability of future treatment assignments on selection bias, and the vulnerability of a randomization to covariate imbalances which could lead to accidental bias in the estimate of treatment effect in a linear model. Other central considerations are the dependence of a permutation test on the randomization employed and the differences between a permutation model versus a population model basis for a statistical test. A permutation test requires no assumptions regarding the origin of the study patients or the distribution of their responses. In contrast, a population model assumes that the patients arose (were sampled) at random from a homogeneous (unchanging) population and that their responses follow some common distribution [i.e., are independently and identically distributed (i.i.d.)].

The properties of simple (unrestricted) randomization were then described [2]. The simplest form of randomization is complete randomization, analogous to tossing a fair coin. This was contrasted to the random allocation rule [2] in which a random permutation of $n/2$ out of n patients is selected to receive treatment a and the remainder b. The latter case is rarely applicable to clinical trials because it requires that the sample size n be guaranteed beforehand. The random allocation rule is of interest, however, because it is the simplest possible case of permuted-block randomization (one block of size n), and thus some of the properties of permuted block randomization are obtained as generalizations of those of the random allocation rule.

The subsequent articles then describe the properties of two commonly used and widely studied procedures for restricted randomization—the permuted-block randomization [3] and the urn adaptive biased-coin randomization [4]. In permuted-block randomization, successive blocks of size $2m$ are employed (where m may vary), each block containing a random permutation of m of $2m$ assignments to treatment a, the remainder to b [3]. With the urn randomization, one draws successive assignments from an hypothetical urn containing both a and b balls [4]. The urn contains α balls of each type initially. After each assignment (a or b) is drawn from the urn (and replaced), β balls of the other treatment type (b or a) are added to the urn. The urn design is designated as UD(α, β) and the special case UD(α, 0) is simply complete randomization. Thus, permuted-block randomization forces balance after every $2m$ assignments, whereas the urn randomization UD(α, β) for $\beta > 0$ promotes ongoing balance by altering the probability of assignment in favor of whichever treatment has been assigned least in the past.

CONCLUSIONS

Based on the developments presented in the preceding articles [1–4], the following is a summary of the principal conclusions regarding the properties of these procedures for randomization:

(a) Each of these procedures can be easily implemented using a uniform random number generator on a computer. With a specified seed, the sequence so generated can be replicated and documented. Each of these procedures can also be implemented using a table of random numbers, but this is tedious. Thus, for each procedure, the randomization sequences can be pregenerated.

(b) These randomization procedures can be sharply distinguished according to the pursuant probabilities of an imbalance in the numbers assigned to each treatment. Although power (under a population model) is usually maximized with equal sample fractions, the effects of treatment imbalances on power are trivial unless the imbalances are substantial, on the order or 0.6 or 0.7 to one of the two groups (see Figure 1 in ref. 1). For a large trial with total $n > 200$, such treatment imbalances that might affect power are extremely unlikely with complete randomization (see Figure 1 in ref. 2) and are even less likely with an urn randomization procedure. For a small trial with $n < 100$, such imbalances are more likely with complete randomization. For a trial with $100 < n < 200$, an imbalance of 0.70 vs. 0.30 is very unlikely, but an imbalance of 0.60 to 0.40 can occur with small but nonnegligible probability. With permuted-block randomization, in a trial of any size there are no imbalances provided that all blocks are filled.

Therefore, from a consideration of the likelihood of imbalances and their effects on power, the planned sample size of the trial is important when choosing a randomization procedure. In this context, it is important to note that the probabilities of such imbalances overall in a stratified trial are the same as those in an unstratified trial of the same total size. On the other hand, if the objectives of the trial are to be assessed by separate subgroup analyses within one or more strata, for example, separately among males and separately among females, then the sample size of importance is the smallest among the various separate subgroups or strata to be considered.

(c) In a homogeneous population model, it is assumed that the patients arose (were sampled from) a defined homogeneous (unchanging) population. Under this model, the method of randomization can be ignored in the conduct of a statistical test. However, in ref. 1 it is argued that such a population model is tenuous, at best, in a clinical trial, and that a permutation model provides an alternative basis for conducting a valid statistical test.

Under a permutation model, the observed treatment difference is assessed with reference to all the possible ways (permutations) that the treatments could have been assigned. In this case, the preferred method of analysis is to analyze the way one randomized. The family of linear rank tests provides a simple, direct method for conducting such permutation tests with small or large samples. However, the proper variance of this statistic depends explicitly on the method of randomization employed [1].

(d) Therefore, for a large trial with $n > 200$ overall and within each planned subgroup (if any), simple unrestricted complete randomization is desirable because virtually all simple population model–based tests are equivalent to the corresponding permutation test, thus simplifying the analysis [2]. In this case, virtually all available standard statistical software provides valid statistical tests. Complete randomization also provides optimal bias reduction, especially in an unmasked trial [2].

(e) On the other hand, for a small trial (say $n < 100$ overall or within any principal subgroup or stratum), imbalances that might affect power are more likely with complete randomization. In these cases restricted randomization is desirable.

(f) Under a permutation model it is necessary that the trial be analyzed the way it was randomized in order to yield tests with the proper size (type 1 error probability) [1]. Thus, if a stratified randomization is employed, a like-

stratified analysis should be performed. Also, if restricted randomization is employed, for example, permuted-block or urn design, then the test should employ the proper corresponding permutational variance.

Nevertheless, it is common practice to ignore the randomization in the analysis. If the patients actually arose from a homogeneous population, then ignoring the randomization will have no effect on the size of the test. However, if there is significant heterogeneity in some systematic way among the patients entering the trial, such as a change over time, then ignoring the stratification, or ignoring the type of restricted randomization (permuted-block or urn design) may substantially distort the size of the test [1].

(g) If a permuted-block randomization is employed, the proper permutational analysis requires a blocked or block-stratified analysis. The effect of ignoring the blocks in the analysis is a function of the intrablock correlation [3]. If there is any systematic difference in the characteristics of the patients entering the trial, such as a time trend (i.e., a time heterogeneity) or a difference among strata (e.g., clinics), then the intrablock correlation (R) is expected to be positive. If the blocking is then ignored in the analysis, the value of some common chi-square or F-test statistics ignoring the blocking (say T_I) can be expressed as $T_I = (1 - R)T$, where T is the proper blocked-stratified test statistic value [3]. Thus, a positive intrablock correlation will result in a conservative and less powerful test if the blocking is ignored.

(h) For the urn randomization, the effects of ignoring the randomization are not as easily quantified as with permuted-block randomization. Nevertheless, the urn permutation test variance will also be affected by the presence of any systematic variation of the characteristics of the patients entering the trial, such as a time heterogeneity [4]. Here it is best to always perform the proper permutation test for the analysis of the principal outcomes of the trial.

(i) For a permuted-block randomization, a blocked analysis can be performed using standard software. For most blocked analyses, standard software computes the permutation test, including the block-stratified Mantel–Haenszel test for proportions and the block-stratified log-rank and Wilcoxon tests for lifetimes. In other cases, such as the blocked analysis of variance, standard software computes tests that converge in distribution to the permutation test as the block size increases.

For the permuted-block and the urn randomization, the permutational linear rank test is readily computed [3, 4], but computational software is not widely available.

(j) There remains controversy as to whether the randomization of a trial should be stratified on the basis of prognostic covariates [1]. It has been shown that stratification is principally advantageous for a small trial and has negligible advantages in terms of power or efficiency in a large trial (say $n > 100$). On the other hand, if a stratified randomization were used, one or more strata could be eliminated from the analysis without biasing the analysis of the remaining strata. Accordingly we recommend that the randomization of a multicenter trial always be stratified by clinic, but that further stratification by covariates is unnecessary.

Another consideration is the operational implementation of the randomization. The more complicated the randomization-stratification scheme, the more difficult it is to implement operationally, and the more likely errors are

to occur in the actual randomization of patients, such as where patients are randomized within the wrong stratum. Therefore, the degree of stratification should be as minimal as possible.

Of course, stratification may also be required by the design of the trial, such as where different patient populations (e.g., different disease entities) are to be studied simultaneously, perhaps using different outcome measures for the evaluation of some objectives. In these cases, some of the objectives of the trial will be assessed separately within each stratum (subgroup), as well as in aggregate.

(k) If the randomization is not prestratified by a covariate, then subgroup-specific, post hoc–defined permutational analyses can still be performed with any of these randomization procedures when it can be assumed that the covariate values are statistically independent of the treatment assignments [1–4]. This is a strong but untestable assumption. In the case where separate tests are conducted within each of multiple mutually exclusive strata, these tests are statistically independent for complete randomization and the random allocation rule [2] and for permuted-block randomization [3]. With the urn randomization, however, these stratum-specific tests are correlated [4]. With any of these procedures, a permutational covariate-adjusted test can then be obtained by combining tests over strata [1–4].

(l) Any analysis is complicated when some observations are missing, regardless of whether a population model or a permutation model is assumed. However, with any of these randomization procedures, a permutational analysis of the subset of patients with observed responses can be justified under the assumption of missing-at-random observations [1]. This is also a strong and unverifiable assumption that states that whether a patient's response is observed or missing is statistically independent of the treatment assignments.

With permuted-block randomization, an alternate assumption-free strategy is to analyze only the subset of complete blocks that contain no missing data [3]. Compared to the analysis of the subset of patients with complete data, this approach will probably result in a loss of efficiency. Therefore, if this approach is planned, then the block size should not be so large as to result in a significant loss in efficiency due to incomplete blocks.

(m) Common statistical methods, such as population tests of significance, parameter estimates, confidence intervals, and regression models, can only be applied under population model assumptions [1]. Such analyses are more relevant than permutation tests in terms of attempting to generalize the observed results to some more general hypothetical patient population. The frame of reference for such generalizations, however, is the sample of patients actually studied. In this regard, extensive tabulations of baseline patient characteristics should be performed, both marginally and jointly, so as to provide a complete description of the "population" of patients studied.

(n) Randomization procedures can also be distinguished with respect to the susceptibility to experimental biases in the estimate of treatment effect due to the predictability of the randomization sequences [1]. Among all the procedures considered, complete randomization provides maximum protection against such biases.

In an unmasked trial, the Blackwell–Hodges model assesses the potential for selection bias due to the possibility of correctly guessing future assign-

ments when the investigator guesses which treatment has been assigned least often in the past [1]. With complete randomization, the potential for this bias is eliminated. Among the restricted randomization procedures considered, the urn randomization has the least potential for selection bias [4] while the permuted-block randomization has the highest [3] (see Figure 1 in ref. 4).

With the random allocation rule, the potential for selection bias increases monotonically as n increases [2]. Thus, with the permuted-block design this bias increases as a function of the block size, and the use of random block sizes does not in any way protect against this bias [3]. However, if this model is altered to only allow for selection bias when future assignments can be guessed with certainty, then withholding knowledge of the possible blocking scheme from the investigators will eliminate this bias [3]. Otherwise, the use of random block sizes affords some protection. In any event, selection bias can be eliminated entirely by randomizing patients in a block as each block is filled rather than as patients arrive.

With the urn randomization, since the probability of treatment assignment approaches 1/2 as n increases, the potential for selection bias approaches that of complete randomization as n increases.

(o) Efron's model for accidental bias assesses the susceptibility of a randomization to bias in the estimator of treatment effect in a linear model when important covariates have been ignored [1]. For each of these randomization procedures, the estimate is unbiased but the variance of the estimate, and thus the susceptibility to bias, depends on the nature of the randomization. This is important because such bias arises when there is a covariate imbalance between treatment groups for the ignored covariates.

For all of the randomization procedures considered, the variance of the bias, and thus the susceptibility to accidental bias in the estimate of treatment effect, vanishes asymptotically [1]. However, for given finite n, the variance of the bias is minimized with complete randomization and it is minimally greater with the random allocation rule [2]. For the urn randomization, as n increases, the variance approaches that of complete randomization [4]. With permuted-block randomization the variance is slightly greater than that of these other procedures. Thus, the susceptibility to this bias is slightly greater for permuted-block randomization than for the other procedures considered [3]. However, for all practical purposes, the differences are minuscule for large n (say $n > 200$).

COVARIATE- AND RESPONSE-ADAPTIVE PROCEDURES

In the preceding articles, we have only considered the case of restricted randomization procedures that account for imbalances in the prior numbers of assignments to each of the two treatments, what are termed "treatment adaptive" procedures. These are permuted-block randomization [3] and the urn adaptive biased-coin design [4]. We have not considered those treatment assignment procedures that account for the covariate values of the prior patients entered into the trial (covariate adaptive), nor those that account for the prior patient responses (response adaptive) [5–7].

Among covariate and response adaptive procedures, a distinction must

be made between those that are deterministic and those that are probabilistic. In a deterministic procedure, the treatment assignment of the first patient is determined by chance, but thereafter the assignment of each future patient is predetermined based either on that patient's covariate values, or on the responses of the prior patients entered. Such deterministic methods do not entail use of chance or randomization in the designation of the treatment assignments, and thus cannot be evaluated with respect to the permutational properties of the treatment assignment rule itself. For such methods, however, a valid unbiased analysis can be based on the assumption that the patients in the trial arose from a homogeneous population (i.e., using a population model).

On the other hand, probabilistic adaptive procedures entail modification of the probability of assignment to treatment a (and b) based on a patient's covariate values or on the prior patients' responses. Here, randomization does play a role and thus it is possible, though tedious, to evaluate the permutational properties of the treatment assignment rule.

Well-known covariate adaptive procedures are minimization [8], which is deterministic, and the Pocock–Simon procedure [9], which is probabilistic. Here the treatment assignment of the next patient, or the probability of assignment, is determined so as to minimize a measure of overall covariate imbalance when that patient's covariate values are considered. For minimization, it has been shown that the statistical analysis (under the assumption of a population model) must incorporate adjustments for the covariates employed in the design in order to yield tests of proper size [10, 11].

The play-the-winner rule is an example of a response-adaptive procedure. The original formulation of Zelen [12] is deterministic, whereas the modification of Wei and Durham [13] is probabilistic. Here the treatment assignment of the next patient, or the probability of assignment, is determined such that the patient will receive the treatment that appears to be most favorable based on the responses of the prior patients entered. For the deterministic procedure, the trial can be analyzed under the assumption that the patients arose at random from a homogeneous population (the population model). For the probabilistic procedure, Wei [14] describes the permutational distribution for a single test of a binary response when that response is also used as the basis for the adaptive treatment assignment procedure. However, the permutation test has not been described for the analysis of other outcome variables that were not used as the basis for adjustment of the treatment assignment probabilities.

Therefore, for the most part, these covariate- and response-adaptive procedures require population model assumptions for a valid analysis, and the permutational properties of the probabilistic procedures have not been thoroughly studied. Further, these procedures are tedious to implement operationally, especially in a stratified trial or a masked trial, in part because the sequences of treatment assignments cannot be pregenerated.

On the other hand, permutational properties of treatment adaptive procedures have been widely studied and the family of permutation tests based on these procedures provides a variety of assumption-free significance tests. Also, these procedures can be used to pregenerate the randomization se-

quences, thus simplifying operational implementation in a trial. Therefore, we feel that treatment-adaptive restricted randomization procedures are preferred over covariate- or response-adaptive procedures.

RECOMMENDATIONS

Based on the above, a series of recommendations are offered as a guide to choosing between complete (simple, unrestricted) randomization and either the permuted-block or urn (restricted) randomization procedures. In many cases, we feel that any of these procedures will be acceptable and that the choice of a randomization procedure will depend upon one's evaluation of the relative properties of each procedure. Therefore, rather than present an algorithm as to when one versus another procedure should be employed (which we feel would be unwise to attempt), we have tried to summarize those conditions under which each procedure might be appropriate. For each procedure, the relevant paragraphs in the conclusions section are cited.

Complete Randomization

Complete randomization provides optimal protection against various experimental biases (n, o). For a small trial, permutation tests are easily performed on a computer. For a large trial, virtually all simple common statistical tests are equivalent to the corresponding permutation test (d). However, complete randomization also allows the greatest probability of an imbalance in the numbers assigned to each treatment (b). For a small trial, one must weigh the relative need for balance against the possibility of bias when deciding whether or not to use complete randomization. For a large trial, balance is not that important (b), in which case complete randomization is attractive due to its other favorable properties.

Permuted-Block

The permuted-block procedure is appropriate when there is a strong need, real or perceived, for both periodic and final balance in the numbers of patients assigned to each treatment. The permuted-block procedure, however, is vulnerable to experimental biases (n, o), especially in an unmasked trial, and proper analyses require blocking if any intrablock correlation is present (g).

Central considerations in the use of the permuted-block design are the choice of block size(s), choice of fixed versus random blocks, and whether the trial is to be masked. The block size(s) should not be too large: otherwise a moderate proportion of patients may be members of an incomplete last block within one or more strata (when stratification is employed). This is undesirable because incomplete blocks complicate the analysis when an intrablock correlation exists.

If the trial is masked, even the smallest block size, 2, can safely be employed. However, when the number of allocations to each treatment group is not masked, the choice of block size(s) and/or block types must be weighed against the potential for selection bias. If one adopts the Blackwell–Hodges model for selection bias [1, 3], small block sizes, especially of size 2, should be avoided in favor of larger block sizes. Further, the use of random or variable

block sizes does not in any way protect against this type of bias [3]. However, in ref. 3 it is shown that the use of random block sizes does provide considerable protection against selection bias if bias is viewed as only possible when the investigator can guess with certainty.

Therefore, in an unmasked trial, permuted block randomization should be used with caution. In such a case, the use of a fixed block size of 2 should especially be avoided unless the patients are randomized 2 at a time rather than as they arrive. In fact, with any block size, the only way to eliminate selection bias with permuted-block randomization in an unmasked trial is to randomize patients in blocks as each block of patients is recruited rather than randomizing patients individually as they arrive [3].

It is also recommended that the intrablock correlation should be examined for all principal outcome variables. If intrablock correlation exists, it is necessary that the analysis be blocked in order to provide tests of the proper size (g). Permutational subgroup analyses can also be performed in the case of either pre- or poststratified randomization (k).

Urn Design

Likewise, the urn design is appropriate when there is a desire to promote but not guarantee periodic and final balance in the numbers of patients assigned to each treatment. The urn design is far less susceptible to experimental biases than is the permuted-block procedure (n, o); however, it also requires special analyses in order to obtain tests of proper size (h).

The urn design promotes balance during the early stages of a trial, where it is usually most important, but then approaches complete randomization as the size of the trial increases. This makes the urn design particularly attractive when the size of the trial is uncertain a priori. For example, in a stratified trial one usually does not have foreknowledge of the size of each stratum. Another example is a trial that may be terminated early due to sequential monitoring of treatment effects.

For the urn design, $UD(\alpha, \beta)$, the choice of α depends on how closely it is desired that the design resemble complete randomization in the early stages of the trial. On the other hand, the choice of β depends on how strongly it is desired that balance be maintained. For example, $UD(10, 1)$ closely resembles complete randomization, whereas $UD(0, 10)$ will be highly balanced initially and will slowly approach complete randomization only as n becomes very large. The $UD(0, 1)$ design usually has adequate balancing properties while still being less susceptible to bias than the permuted-block or biased-coin designs.

Proper permutational tests are required (h) that may be easily programmed on a computer. Such tests also may be performed in the case of either pre- or poststratification by a covariate (k).

Therefore, the urn design is acceptable in a large unmasked trial because its susceptibility to experimental biases approaches that of complete randomization as n increases (n, o). However, in a very small trial (say $n < 10$ within any stratum), the urn design has a vulnerability to selection bias as high as that of the permuted-block or biased-coin designs.

As is the case with all of the procedures considered, the urn design is not

recommended for an unmasked trial with very small total or subgroup sizes (< 10). In all other cases, the urn design has highly favorable properties.

The Small Unmasked Trial

Of all the cases considered, the most problematic is the small ($n < 10$) unmasked trial. In this case, the best approach might be to randomize in small blocks as each block is filled, rather than as patients arise. If this is not feasible, then none of the above procedures is wholly satisfactory and each will have significant disadvantages.

REFERENCES

1. Lachin JM: Statistical properties of randomization in clinical trials. Controlled Clin Trials 9:289–311, 1988

2. Lachin JM: Properties of simple randomization in clinical trials. Controlled Clin Trials 9:312–326, 1988

3. Matts JP, Lachin JM: Properties of permuted-block randomization in clinical trials. Controlled Clin Trials 9:327–344, 1988

4. Wei LJ, Lachin JM: Properties of the urn randomization in clinical trials. Controlled Clin Trials 9:345–364, 1988

5. Pocock SJ: Allocation of patients to treatment in clinical trials. Biometrics 35:183–197, 1979

6. Simon R: Restricted randomization designs in clinical trials. Biometrics 35:502–512, 1979

7. Kalish LA and Begg CB: Treatment allocation methods in clinical trials: A review. Stat Med 4:129–144, 1985

8. Taves DR: Minimization: A new method of assigning patients to treatment and control groups. Clin Pharmacol Ther 15:443–453, 1974

9. Pocock SJ, Simon R: Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31:103–115, 1975

10. Forsythe AB, Stitt FW: Randomization or minimization in the treatment assignment of patient trials: Validity and power of tests. Technical Report 28: Health Sciences Computing Facility, University of California, Los Angeles, 1977

11. Birkett NJ: Adaptive allocation in randomized controlled trials. Controlled Clin Trials 6:146–155, 1985

12. Zelen M: Play the winner rule and the controlled clinical trial. J Am Stat Assoc 64:131–146, 1969

13. Wei LJ, Durham S: The randomized play-the-winner rule in medical trials. J Am Stat Assoc 73:840–843, 1978

14. Wei LJ: Exact two-sample permutation tests based on the randomized play-the-winner rule. Biometrika 1988, in press.