Properties of Simple Randomization in Clinical Trials

John M. Lachin, ScD

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

ABSTRACT: This article presents the properties of complete randomization (e.g., coin toss) and of the random allocation rule (random permutation of \( n/2 \) of \( n \) elements). The latter is principally used in cases where the total sample size \( n \) is known exactly a priori. The likelihood of treatment imbalances is readily computed and is shown to be negligible for large trials \( (n > 200) \), regardless of whether a stratified randomization is used. It is shown that substantial treatment imbalances are extremely unlikely in large trials, and therefore there is likely to be no substantial effect on power.

The large-sample permutational distribution of the family of linear rank tests is presented for complete randomization unconditionally and conditionally, and for the random allocation rule. Asymptotically the three are equivalent to the distribution of these tests under a sampling-based population model. Permutation tests are also presented for a stratified analysis within one or more subgroups of patients defined post hoc on the basis of a covariate. This provides a basis for analysis when some patients’ responses are assumed to be missing-at-random.

Using the Blackwell–Hodges model, it is shown that complete randomization eliminates the potential for selection bias, but that the random allocation rule yields a substantial potential for selection bias in an unmasked trial. Finally, the Efron model for accidental bias is used to assess the potential for bias in the estimate of treatment effect due to covariate imbalance. Asymptotically, this probability approaches zero for complete randomization and for the random allocation rule. However, for finite \( n \), complete randomization minimizes the probability of accidental bias, whereas this probability is slightly higher with a random allocation rule.

It is concluded that complete randomization has merit in large clinical trials.

KEY WORDS: Randomization, complete binomial, random allocation rule, permutation tests, stratified analysis, missing data, selection bias, accidental bias

INTRODUCTION

Many procedures have been proposed for the random assignment of treatments to patients in clinical trials [1–4]. In the simplest cases, no restrictions are placed on the nature of the randomization sequence, apart perhaps from prespecification of the total sample size and the sample size in each group.
Simple Randomization

where feasible. Such procedures for unrestricted randomization are commonly referred to as simple randomization.

Lachin [1] describes some of the properties by which randomization procedures have been evaluated. In this article, the properties of simple randomization are described as a yardstick against which the properties of restricted randomization procedures can be compared. Likewise, Matts and Lachin [5] describe the properties of permuted-block randomization, and Wei and Lachin [6] describe those of urn adaptive biased-coin randomization. The factors considered are the probabilities of treatment imbalances and their potential effects on power, the permutational distribution of linear rank tests, the potential for selection bias, and the potential for accidental bias. Lachin [1] describes the models by which these properties are assessed. Only the case of two treatment groups (a and b) is considered, where approximately equal sample sizes are desired.

SIMPLE RANDOMIZATION

In order to describe "simple" randomization, two different cases must be distinguished. In the first case, the total sample size \( n \) and the sample sizes in each group are prespecified exactly and are under the direct control of the investigator. In this case the usual simple randomization procedure is the random allocation rule [7, 8], whereby a randomly chosen subset of \( n/2 \) out of \( n \) is assigned to group a, the remainder to group b. Thus, the sample sizes in each group, say \( n_a \) and \( n_b \), each equal to 1/2.

In the second case, which is most common in large clinical trials, a target sample size is established but the final sample size is not known with certainty. In this case, the usual simple randomization procedure is complete randomization [7], analogous to tossing a fair coin. Here the sample sizes in each group are binomially distributed random variables.

For complete randomization, the marginal and conditional probability of assignment is a constant \( 1/2 \) for all assignments. That is, \( p_j = \Pr(a) = 1/2 \) marginally for the \( j \)th assignment, and the conditional probabilities are such that \( p_{j'} = 1/2 \) for all \( j' < j \), even if all prior assignments were to one of the two groups.

For the random allocation rule, the marginal probability of assignment is also \( 1/2 \) for all assignments. However, the conditional probability of assignment at the \( j \)th step given prior assignments is not always \( 1/2 \). Specifically, let \( n_a = n_b = m = n/2 \) and let \( n_{aj} \) and \( n_{bj} \) denote the numbers assigned to \( a \) and \( b \) after \( j \) assignments. For the random allocation rule, \( p_{j' < j} = (m - 1)/(n - 1) \) if \( j \) and \( j' \) have the same assignment, \( p_{j' < j} = m/(n - 1) \) otherwise. Also, \( p_{j' < j} = (m - n_{a,j-1})/(n - j + 1) \). Clearly \( p_{j' < j} = 1/2 \) only if the numbers of prior assignments \( n_{a,j-1} \) and \( n_{b,j-1} \) are equal. For example, suppose \( n = 10 \) (\( n_a = n_b = m = 5 \)) and the first two assignments are \( a \) and \( a \). Then, the unconditional (marginal) probability of an \( a \) on the third assignment is \( 1/2 \) over the set of all possible permutations, but within the particular permutation chosen, the conditional probability of an \( a \) on the third assignment given the past assignments is \( p_{j=3} = 3/8 \).
TREATMENT IMBALANCES

With a random allocation rule \( n_a = m \) provided that \( n \) patients are randomized. If \( j < n \) patients are randomized, then the probability of an imbalance of \( n_{aj} \) to \( n_{bj} \) or worse, can be obtained from Fisher's exact test for the resulting \( 2 \times 2 \) table with cells \( n_{aj}, n_{bj}, m - n_{aj}, \) and \( m - n_{bj} \). Asymptotically, this probability can be assessed by the ordinary chi-square test for the \( 2 \times 2 \) table.

With complete randomization, after \( n \) (or \( j \)) assignments, \( n_a \) (or \( n_d \)) follows the binomial distribution with \( p = \frac{1}{2} \) and sample size \( n \) (or \( j \)). The specific probability of a treatment imbalance of \( n_a \) to \( n_b \) can readily be calculated using the large sample approximation to the binomial, where \( n_a \) is approximately normally distributed with mean \((0.5)n = m\) and variance \((0.5)^2n = n/4\). If we denote the larger of the two sample fractions as \( q_u = \max(n_a, n_b)/n \), then \( \Pr[q_u > r] \approx 2\Phi(-Z_r) \), where \( \Phi(\cdot) \) is the standard normal cumulative distribution [e.g., \( \Phi(1.645) = 0.95 \)] and \( Z_r = 2(r - 0.5)\sqrt{n} \). Therefore, the probability of a stated treatment imbalance \( r \) decreases as the sample size increases.

This property is displayed in Figures 1 and 2. Figure 1 presents the probability of obtaining a treatment imbalance (larger sample fraction) of \( q_u \geq r \) for \( r = 0.55, 0.60, \) and 0.70. Figure 1 shows that there is a probability of less than 0.05 of an imbalance of 0.70 or more for a sample size of 30 or more; of an imbalance of 0.60 or more for a sample size of 100 or more; and of an imbalance of 0.55 or more for a sample size of 400 or more. Figure 2 shows those degrees of imbalance \( q_u \geq r \) that would occur with probabilities 0.01, 0.05, or 0.10 as sample size increases. This shows that severe imbalances become extremely unlikely with increasing sample sizes.

Now consider the case of a stratified randomization with \( K \) strata, each of size \( n_k \) (1 \( \leq k \leq K \)). By the normal approximation to the binomial, the number assigned to treatment \( a \) in the \( k \)th strata, \( n_{ak} \), is asymptotically normally distributed with mean \( \mu_k = n_k/2 \) and variance \( \sigma_k^2 = n_k/4 \). Therefore, in aggregate overall \( K \) strata, \( n_a = \Sigma n_{ak} \) is also asymptotically normally distributed with mean \( \mu = \Sigma \mu_k = n/2 \) and variance \( \sigma^2 = \Sigma \sigma_k^2 = n/4 \). Therefore the probability of an aggregate treatment imbalance in a stratified trial is the same as that in an unstratified trial of the same total sample size.

The principal concern is that a treatment imbalance may affect the power of a statistical test. It has been shown that power is nontrivially reduced only if the treatment imbalance is on the order of \( q_u = 0.70 \) or greater [1]. However, from Figures 1 and 2 it is clear that even for a small-sized study, the probability of an imbalance this great is extremely remote. Thus, there is little support for the widespread concern that complete randomization is likely to result in treatment imbalances that could have a substantial effect on statistical power.

PERMUTATION TESTS

As described in ref. 1, there is a fundamental difference between the use of a population model and a randomization model as a basis for statistical tests. Permutation tests make no assumptions regarding the origin of the patient samples or the distribution of the measurements. However, the variance of permutation tests depends on the precise nature of the randomization
Figure 1

Probability of treatment imbalance for complete randomization as a function of sample size (n), for imbalances

\[ \text{max}(n, m) \in \{0.5, 0.6, 0.7, 0.8\} \]

Sample Size (n)

Probability of Imbalance

Simple Randomization
Figure 2  Treatment imbalances $\max(n_a, n_b)/n \geq r$, which occur with tail probability 0.01, 0.05, or 0.10 for complete randomization as a function of sample size ($n$).
procedure employed. Table 1 in ref. 1 presents an example of an exact permutation test based on a complete randomization.

For large samples, Lachin [1] describes the family of permutational linear rank tests. This test is based on the set of \( n \) scores \( \{ c_j \} \) that are some function of the patient responses \( \{ Y_j \} \) as \( c_j = f(Y_1, \ldots, Y_n), j = 1, \ldots, n \), with mean score \( \bar{c} \). The treatment assignments are designated as \( \tau_j = 1 \) if \( a \), and \( = 0 \) if \( b \), with marginal expectation \( E(\tau_j) \). The scores \( \{ c_j \} \) are treated as fixed constants while the \( \{ \tau_j \} \) are random with a distribution dictated by the randomization procedure employed. This distribution is defined over the reference set of size \( \Omega \) of all possible permutations, either conditionally given \( n_a \) and \( n_b \) fixed (\( \Omega_\| \)), or unconditionally (\( \Omega_n \)). The linear rank statistic with centered scores (i.e., mean-corrected) is then defined as

\[
W = \frac{\sum (c_j - \bar{c})(\tau_j - E(\tau_j))}{\sqrt{\text{var} \sum (c_j - \bar{c})(\tau_j - E(\tau_j))}} \sqrt{\lambda} = \frac{S}{\sqrt{\lambda}} \tag{1}
\]

which is asymptotically distributed as standard normal under the null hypothesis \( H_0 \) that there is no difference between treatments \( a \) and \( b \). When used with the appropriate scores [1], this test yields the equivalent of the chi-square test for a \( 2 \times 2 \) table, the Wilcoxon rank sum test, and the log-rank and Wilcoxon tests for survival data, among many others.

Using complete randomization, the \( \{ \tau_j \} \) are independent Bernoulli variables. Thus, unconditionally (\( \Omega_n = 2^n \)), \( E(\tau_j) = \frac{1}{2} \) and the variance is of the form

\[
V = \sum (c_j - \bar{c})^2 \text{var}(\tau_j) = \sum (c_j - \bar{c})^2/4. \tag{2}
\]

Therefore, from eq. (6) of ref. 1, the unconditional complete randomization permutation test is equivalent to the test obtained under a population model. If one conditions on the sample sizes \( (\Omega_c = nC_{n_a}) \), then \( E(\tau_j) = q_a = n_a/n \), and the variance is

\[
V = \frac{n_a(n - n_a)}{n(n - 1)} \sum (c_j - \bar{c})^2 \tag{3}
\]

(see the Appendix). This is the expression for the permutational variance of the linear rank test that is usually presented in reference texts [9,10].

Using a random allocation rule such that \( n_a = n/2 = m \), then \( E(\tau_j) = \frac{1}{2} \), and, from eq. (3), the variance is

\[
V = \frac{m}{2(2m - 1)} \sum (c_j - \bar{c})^2. \tag{4}
\]

Asymptotically, \( n_a \to m \) in eq. (3), and \( m/(4m - 2) \to 1/4 \) in eq. (4), in which case these variances are equivalent. For moderate sample sizes, however, the variances are different. This demonstrates that the proper permutational variance depends upon the nature of the randomization and the corresponding reference set.

In addition, if one adopts a permutational basis for inference, it is also necessary to "analyze the way you randomize" in order to conduct a large-
sample test with proper size (type 1 error probability). For a stratified randomization, the proper analysis is likewise stratified, such as a stratified linear rank test [1] or a Mantel–Haenszel analysis of 2 × 2 tables. Such a stratified test is likely to be different from that obtained by a simple aggregate analysis ignoring strata. For example, a stratified randomization using a random allocation rule within strata is the same as a permuted-block randomization. In this case, Matts and Lachin [5] show that a conservative test is likely to result if the blocks (strata) are ignored in the analysis, that is, if an unstratified analysis is conducted.

**POST HOC STRATIFIED (SUBGROUP) ANALYSES AND MISSING DATA**

**Subgroup Analyses**

The permutation tests described above apply to the analysis of the total collection of \( n \) patients randomized into the trial. Often one may also wish to perform a test separately among those patients who are members of a subgroup defined post hoc on the basis of a covariate, that is, a subgroup or stratum within which a stratified randomization was not performed. In order to perform a valid analysis among the subset of patients within such a subgroup, it is sufficient to assume that the covariate values among these \( n \) patients are statistically independent of the randomly assigned treatment indicators \( \{ \tau_j \} \). Clearly this assumption is satisfied for any baseline covariate when there is no potential for selection bias. Under this assumption, a post hoc stratified linear rank test can be obtained as follows. Details are presented in the Appendix.

For each patient \( (1 \leq j \leq n) \) let \( v_j \) be an indicator for whether that patient is a member of the designated subgroup of interest. For example, for an analysis among males, \( v_j = 1 \) if the \( j \)th patient is a male, \( v_j = 0 \) if female. The pattern of subgroup membership is then represented by the vector of subgroup indicators \( v = (v_1, \ldots, v_n) \). For the collection of \( n \) patients, the rank scores are obtained as some function of the responses among those who are members of the subgroup as \( c_j = f(v_1, Y_1, \ldots, v_n, Y_n) \), where \( c_j \) is undefined if \( v_j = 0 \). The mean score among those in the subgroup is \( \bar{c} = \frac{\sum v_j c_j}{n'} \), where \( n' = \sum v_j \) is the size of the subgroup.

If we now condition on the pattern of subgroup indicators \( v \) that are assumed to be independent of the treatment assignments \( \{ \tau_j \} \), then the linear rank statistic can be written as \( W = S/(V^{1/2}) \), where

\[
S = \sum v_j (c_j - \bar{c}) (\tau_j - E(\tau_j | v))
\]

and where \( E(\tau_j | v) \) is the probability of assignment to \( a \) within the subgroup. For an unconditional test with reference to \( \Omega_a \), under the covariate-treatment independence assumption, \( E(\tau_j | v) = 1/2 \). For a test conditional on the subgroup sample sizes, \( E(\tau_j | v) = n'/n' \), where \( n'_a = \sum v_j \tau_j \) and \( n'_b = n' - n'_a \). In either case, \( E(S) = 0 \) under the null hypothesis of no treatment effect.

With complete randomization, since \( p_j = 1/2 \) for all assignments, it readily follows that the unconditional permutational variance (not conditional on the sample sizes \( n'_a \) and \( n'_b \)) is simply

\[
V = \sum v_j (c_j - \bar{c})^2/4.
\]
In the Appendix it is also shown that the conditional permutational variance (conditional on the sample sizes $n'_i$ and $n'_j$) is

$$V = \frac{n'_i(n'_i - n'_j)}{n'(n' - 1)} \sum_i n_i (c_i - \bar{c})^2. \quad (7)$$

For the random allocation rule, the probabilities of assignment within a given sequence are a function of the prior assignments. However, because the sample sizes are fixed a priori as $n/2$ to each group, each permutation within the conditional reference set is equiprobable. Likewise, conditional on the sample sizes within a subgroup ($n'_i$ and $n'_j$), each permutation within the reduced conditional reference set for that subgroup is also equiprobable. Thus, in the Appendix it is shown that eq. (7) also applies in this case.

It may also be desired to perform separate tests within each of the multiple strata defined on the basis on a covariate, such as analyses both among males and among females. With complete randomization, since $p = 1/2$ independently for all treatment assignments, it follows that these multiple stratified tests are statistically independent. In the Appendix it is also shown that these tests are likewise independent with the random allocation rule. Thus, it is possible to perform a poststratified covariate-adjusted combined test of treatment effect as though stratified randomization had been performed as described in ref. 1.

**Missing Data**

Up to this point we have assumed that all patients have complete data, namely all patients' responses were observed. Usually, however, some patients' responses will be missing. Therefore, it is desirable to be able to perform an analysis using only that subset of patients for whom responses were observed. In this case, the statistical considerations are the same as for a subgroup analysis since the subset of patients with observed data is exactly such a post hoc defined subgroup.

For a complete-data subgroup analysis, the subgroup indicators are defined as $v_j = 1$ if the $j$th patient's response was observed, $v_j = 0$ if missing. As described in ref. 1, a valid analysis then can be performed under the assumption of missing-at-random observations, that is, where the missingness indicators $\{v_j\}$ can be assumed to be statistically independent of the treatment indicators $\{\tau_j\}$. However, this is a strong assumption which must be carefully considered before performing such a complete-data subgroup analysis, whether on the basis of a population model or a permutational model [1].

**SELECTION BIAS**

In an unmasked study, a difference (bias) in the composition of the treatment groups can be introduced by biases in patient selection. Whether such biases will be allowed to operate is a function of the predictability of treatment assignments within the randomization sequence employed. Lachin [1] describes the Blackwell–Hodges [7] model by which the potential for selection bias of a randomization procedure can be assessed. This potential is quantified by the expected bias factor $E(F)$, which is the expected number of correct
guesses for treatments a (α) and b (β) less the number expected by chance, $E(F) = (\alpha + \beta) - m$. An example is shown in Table 2 of ref. 1.

For complete randomization the expected number of correct guesses is simply $m$, in which case $E(F) = 0$ for all $n$. For the random allocation rule, however, there is substantial potential for selection bias in an unmasked study. In this case, the successive cumulative numbers of assignments to a and b constitute a nondecreasing random walk from the point (0,0) to $(m, m)$. Thus, the random walk makes $m$ steps away from the diagonal and $m$ steps toward the diagonal. At each step towards the diagonal, the experimenter can guess correctly, yielding $m$ correct guesses. In addition, the experimenter is expected to guess correctly half the times that the random walk actually lands on the diagonal (say $d$). Thus $E(F) = E(d)/2$. Blackwell and Hodges [7] show that

$$E(d) = \frac{2^{2m}}{2mC_m} - 1,$$

(8)

where $\lim_{m \to \infty} E(d) = 0.886 m^{1/2}$. Therefore,

$$E(F) = \frac{2^{2m-1}}{2mC_m} - \frac{1}{2},$$

(9)

and $\lim_{m \to \infty} E(F) = 0.443 m^{1/2}$.

Figure 3 presents the expected bias factor as a function of increasing sample size for the random allocation rule. This shows that the potential selection bias is substantial for a large study.

However, it should be noted that the potential for selection bias is completely eliminated if a random allocation rule is used to randomize the entire sample of $n$ patients simultaneously, rather than individually as they arrive. Unfortunately, this is rarely practical.

ACCIDENTAL BIAS—COVARIATE IMBALANCES

As described by Lachin [1], Efron [11] presented a model for the assessment of accidental bias in the estimate ($\hat{\alpha}^*$) of treatment effect (α) in a linear model in which one or more covariates are ignored. Here, $T_j = +1$ to designate treatment a, $T_j = -1$ to designate treatment b, and $X_j$ is the omitted covariate, $j = 1, \ldots, n$. Asymptotically, the bias is such that

$$n \text{ var}(\hat{\alpha}^* - \alpha) = \frac{\beta}{n} \Sigma_i T_i X_i = \bar{X}_a - \bar{X}_b,$$

$$\text{var}(\hat{\alpha}^* - \alpha) = \text{var}[\Sigma_i T_i X_i ](\beta/n)^2,$$

(10)

where $\beta$ is the coefficient for $X$ in an simple linear model. It follows that $\text{var}(\hat{\alpha}^* - \alpha)$ is proportional to $X' \Sigma X$, where $X' = (X_1, X_2, \ldots, X_n)$ is the vector of covariate values, $X' X = 1$, $\Sigma$ is the covariance matrix of the vector $T' = (T_1, T_2, \ldots, T_n)$. The vulnerability of a randomization procedure to
Figure 3  Selection bias factor $E(F)$ for the random allocation rule as a function of sample size ($n$).
accidental bias is then determined by the largest eigenvalue $\lambda$ of $\Sigma_T$ [1]. The probability that the bias will exceed some value is then bounded by

$$\Pr[|\hat{\alpha}^* - \alpha| \geq \xi] \leq 2\Phi(-Z_\xi), \quad (11)$$

where $\Phi(\cdot)$ is the standard normal c.d.f., $Z_\xi = \xi[\max \text{var}(\hat{\alpha}^* - \alpha)]^{-1/2}$; and $\max \text{var}(\hat{\alpha}^* - \alpha) = (\beta/n)^2\lambda$.

For complete randomization, $\Sigma_T = I_n$ and $\lambda = 1$. This is the minimum value $\lambda$ for any randomization procedure. Therefore, complete randomization yields estimators that have bounded variance for all $n$. Further, since the variance vanishes asymptotically, the probability $\to 0$ that the bias will exceed any value $\xi$.

For the random allocation rule, Efron [11] showed that $\Sigma_{T(j,j')} = 1$ and $\Sigma_{T}\Sigma_{T}^{*} = -(2m - 1)^{-1}$. It then follows that $\lambda = 1 + (2m - 1)^{-1}$, and, for finite $n = 2m$, the random allocation rule has a slightly larger upper bound $\lambda$ on the variance of the estimate than is provided by complete randomization. Therefore, there is a greater likelihood of a covariate imbalance. Asymptotically, $\lambda \to 1$ and var($\hat{\alpha}^* - \alpha$) vanishes, but it does not converge to zero as rapidly as for complete randomization.

For example, let $n = 100$ and let $X$ be a moderately prognostic covariate with $\beta = 1$. This would represent a substantial association between $X$ and $Y$ if $Y$ were standardized to have var($Y$) = 1. With complete randomization, var($\hat{\alpha}^* - \alpha$) $\leq (1/100)^2$, whereas for the random allocation rule var($\hat{\alpha}^* - \alpha$) $\leq (1/100)[1 + (1/99)]$. For a slight bias of $\xi = 0.02$, both procedures yield a probability that is $\leq 0.0456$; $Z_{0.02} = 2.0$ for complete randomization, and $Z_{0.02} = 1.98997$ for the random allocation rule. For such sample sizes, there is virtually no difference between the two methods.

**CONCLUSIONS**

Based on the above review, the following conclusions can be reached with respect to complete randomization and the random allocation rule.

**Complete Randomization**

As $n$ increases, the probability of a meaningful treatment imbalance approaches 0, such an imbalance being extremely unlikely for total sample sizes of 200 or more, with or without stratification. Statistical power is usually maximized for equal sample sizes, but is reduced only by extreme imbalances that have negligible probability for large sample sizes. For sample sizes under 200, the probabilities of imbalances and the potential effects on power should be evaluated. In such cases, it may be feasible to use a random allocation rule. Alternately, a restricted randomization procedure might be considered.

Asymptotically the unconditional permutational variance equals the conditional permutational variance, which usually equals the population model variance. Thus, special analyses are not required to consider the permutational distribution of statistical tests. Virtually all statistical analyses yield tests of proper size with the expected levels of power. However, if a stratified randomization is performed, a similarly stratified analysis is required to yield a test of the proper size.
Post hoc covariate-stratified permutation tests are easily performed and yield tests within strata that are statistically independent. This provides a method for analysis when some patients' responses can be assumed to be missing-at-random. In an unmasked study, selection bias is completely eliminated. Also, an estimate of treatment effect from a linear model–based analysis following complete randomization is asymptotically free of accidental bias, and, therefore, is free of covariate imbalances.

Thus, if one can accept a possibility of a minor imbalance in treatment assignments, such as 55% of assignments to one of the two groups, then complete randomization has much to offer.

Random Allocation Rule

The simplest form of restricted randomization is the random allocation rule wherein the total sample size is known a priori. If \( n = 2m \) patients are then actually recruited, there is no treatment imbalance and \( n_a = m \). Any analysis that then conditions on the sample sizes, such as Fisher's exact test, yields a test of proper size. Asymptotically, the conditional permutational variance for such tests usually equals the population model variance, and, thus, as with complete binomial randomization, special analyses are not required to consider the permutational distribution of statistical tests.

As with complete randomization, post hoc covariate-stratified permutation tests are easily performed and yield tests within strata that are statistically independent. This provides a method for analysis when some patients' responses can be assumed to be missing-at-random. In this case the permutation test is equivalent to the conditional complete randomization test and is asymptotically equivalent to that under a population model.

However, in an unmasked study with staggered patient entry, there is substantial potential for selection bias. In this case, selection bias can only be eliminated if the set of \( n \) patients are randomized concurrently as a block, rather than as they arrive.

Also, an estimate of treatment effect from a model-based analysis is unbiased, but with larger mean square error than provided by complete randomization. Thus, there is a greater likelihood of a covariate imbalance with the random allocation rule than with complete randomization, but the difference is trivial for large sample sizes.

The work was partially supported by the Lupus Nephritis Collaborative Study under grant RO1-AM-27769 from the National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK), and by the Study of ACE Inhibition in Diabetic Nephropathy under grant RO1-DK-39826 from the NIDDK and a grant from E.R. Squibb and Sons, Inc.

REFERENCES


APPENDIX

Permutational Variances

Here we present derivations of the permutational variances presented in eqs. (2)–(4) for the case of complete data (no missing observations).

For complete randomization, $\tau_i$ is a Bernoulli variable with $E(\tau_i) = 1/2$ and $\text{var}(\tau_i) = 1/4$. Thus the unconditional variance of eq. (2) results directly.

The conditional variance of eq. (3) is obtained as follows. Expanding eq. (1) using $E(\tau_i) = q_a = n_a/n$ yields

$$V = \sum (c_j - \bar{c})^2 E(\tau_i - q_a)^2 + \sum_{j \neq j'} (c_j - \bar{c})(c_{j'} - \bar{c}) E[(\tau_i - q_a)(\tau_{i'} - q_a)],$$

$$V = \sum (c_j - \bar{c})^2 q_a(1 - q_a) + \sum_{j \neq j'} (c_j - \bar{c})(c_{j'} - \bar{c}) [E(\tau_i\tau_{i'}) - q_a^2].$$

There are $n(n-1)$ possible pairs $j \neq j'$, of which $\tau_j = \tau_{j'} = 1$ for $n_a(n_a - 1)$ pairs. Therefore, $E(\tau_i\tau_{i'}) = [n_a(n_a - 1)]/[n(n-1)]$. Substituting into the above, and noting that

$$\sum (c_j - \bar{c})(c_{j'} - \bar{c}) = \left[\sum (c_j - \bar{c})\right]^2 - \sum (c_j - \bar{c})^2$$

$$= -\sum (c_j - \bar{c})^2,$$

eq. (3) results.

The conditional variance in eq. (4) for the random allocation rule is similarly obtained. By definition, $q_a = 1/2$. Since $n = 2m$,

$$E(\tau_i\tau_{i'}) = \frac{m(m-1)}{n(n-1)} = \frac{m - 1}{2(2m - 1)}.$$ (A3)

Substituting into eq. (A2), eq. (4) results.
Post Hoc Stratified Permutation Tests

As in the text, the vector of subgroup membership indicators is \( \nu = (\nu_1, \ldots, \nu_n) \), where \( \nu_j = 1 \) if the \( j \)th patient is a member of the designated subgroup and \( \nu_j = 0 \) otherwise. Among these \( n \) patients, the subgroup indicators \( \{\nu_j\} \) are assumed to be mutually statistically independent of the randomly assigned treatment indicators \( \{\tau_j\} \). The rank scores are then defined as some function of the responses among members of the subgroup \( c_j = f(\nu_1, Y_1, \ldots, \nu_n, Y_n) \), with mean \( \bar{c} = (\Sigma_j \nu_j c_j)/n' \), where \( n' = \Sigma_j \nu_j \) is the size of the subgroup. Conditioning on the observed pattern of subgroup membership, \( \nu \), the linear rank statistic can be written as \( W = S/(V^{1/2}) \), where \( S \) is given in eq. (5).

Under complete randomization or the random allocation rule, conditional on the subgroup sample sizes, \( E(\tau_j|\nu) = n'_a/n' \), where \( n'_a = \Sigma \nu_j \tau_j \) is the number of those in the subgroup assigned to treatment \( a \), and \( n'_b = n' - n'_a \) is the number assigned to \( b \). With complete randomization unconditionally, \( E(\tau_j|\nu) = E(\tau_j) = 1/2 \). In each case, under the null hypothesis of no treatment effect, it is readily shown that \( E(S) = 0 \).

Now consider the variance \( V \). With complete randomization, since \( \tau_j = 1/2 \) independently for all assignments, \( \tau_j \) is a Bernoulli variable with \( E(\tau_j) = 1/2 \) and \( \text{var}(\tau_j) = 1/4 \) for all \( j \) where \( \nu_j = 1 \). It follows, therefore, that the unconditional permutational variance (not conditional on the sample sizes \( n'_a \) and \( n'_b \)) is simply as given in eq. (6).

To obtain the conditional permutational variance (conditional on the sample sizes \( n'_a \) and \( n'_b \)), the variance of \( S \) in eq. (5) can be written as

\[
V = \Sigma_j \nu_j (c_j - \bar{c})^2 \text{var}(\tau_j|\nu) + \sum_{j \neq j'} \nu_j \nu_j' (c_j - \bar{c})(c_j' - \bar{c})[E(\tau_j \tau_j'|\nu) - E(\tau_j|\nu)^2].
\]

Conditioning on the pattern of observed responses \( \nu \), it follows that \( \text{var}(\tau_j|\nu) = q'_a (1 - q'_a) \), where \( q'_a = n'_a/n' \). Likewise, \( E(\tau_j \tau_j'|\nu) = (n'_a (n'_a - 1))/(n' (n' - 1)) \). Using the same operations applied to eq. (A2) above, the conditional variance in eq. (7) is obtained.

For the random allocation rule, since it is assumed that subgroup membership is at random, if we condition on the pattern of subgroup membership \( \nu \), then all possible permutations of \( n' \) out of \( n \) patients are equally likely, as are all possible permutations of \( n'_a \) out of \( n' \) assignments to treatment \( a \). Therefore, the permutational variance is also given by eq. (7).

Now consider that separate tests are performed within each of multiple mutually exclusive strata defined on the basis of a covariate. Without any loss of generality, consider the case of two strata with membership indicators \( \nu_{1j} \) and \( \nu_{2j} \) to indicate membership in stratum 1 (\( \nu_{1j} = 1, \nu_{2j} = 0 \)) or in stratum 2 (\( \nu_{1j} = 0, \nu_{2j} = 1 \)) for the \( j \)th patient. Note that \( \nu_{1j} + \nu_{2j} = 1 \) and that \( \nu_{1j} \nu_{2j} = 0 \) for all \( j \). As before, for the \( n \) patients, the rank scores may be defined as some function of the responses among members of the subgroup, most generally as

\[
c_j = \nu_{1j} f(\nu_{11}, Y_1, \ldots, \nu_{1n}, Y_n) + \nu_{2j} f(\nu_{21}, Y_1, \ldots, \nu_{2n}, Y_n).
\]
For the lth subgroup (l = 1, 2), the total sample size is \( n'_i = \sum_j v_{ij} \), of whom \( n'_{ia} = \sum_j v_{ij} \tau_j \) are assigned to treatment \( a \) and the remaining \( n'_{ib} \) to \( b \). The subgroup mean score is \( \bar{c}_i = (\sum_j v_{ij} \bar{c}_j) / n'_i \).

For the lth subgroup (l = 1, 2), the linear rank statistic can be expressed as

\[
S_l = \sum_j v_{ij} (c_j - \bar{c})(\tau_j - E(\tau_j|v)),
\]

where \( \bar{c} = v_{1j}\bar{c}_1 + v_{2j}\bar{c}_2 \) and \( E(\tau_j|v) = v_{1j}E(\tau_j|v_1) + v_{2j}E(\tau_j|v_2) \) are the expectations of \( c_j \) and \( \tau_j \), respectively, depending upon the subgroup to which the jth patient belongs. For each subgroup, the conditional variance is likewise obtained from eq. (7) using the vector of indicators for the lth subgroup. We now show that the covariance between the rank statistics \( S_1 \) and \( S_2 \) is zero, and therefore, that the statistics are independent.

Let \( \tau \) be the vector of values \( [\tau_j - v_{1j}E(\tau_j|v_1)] - v_{2j}E(\tau_j|v_2) \) and let \( \Psi \) be the covariance matrix of \( \tau \). For complete randomization, then over the unconditional reference set, \( E(\tau_j|v_i) = 1/2 \) and \( \Psi \) is the identity matrix of size \( n \). It follows that the stratum-specific tests are independent in this case.

For complete randomization over the conditional reference set, or for the random allocation rule, the terms \( E(\tau_j|v_i) \) can be expressed as \( v_{ij}q_{ia} \), where \( q_{ia} = n'_{ia}/n'_i \). The matrix \( \Psi \) then has diagonal elements

\[
\psi_{jj} = v_{1j}q_{ia}(1 - q_{ia}) + v_{2j}q_{2a}(1 - q_{2a})
\]

and off diagonal elements \( (j \neq j') \)

\[
\psi_{jj'} = \left[ \frac{n'_{ia}(n'_{ia} - 1)}{n'_i(n'_i - 1)} - q_{ia}^2 \right] v_{1j}v_{1j'}
+ \left[ \frac{n'_{2a}(n'_{2a} - 1)}{n'_i(n'_i - 1)} - q_{2a}^2 \right] v_{2j}v_{2j'}
+ (v_{1j}v_{1j'} + v_{2j}v_{2j'}) \left[ \frac{n'_{ia}n'_{2a}}{n(n - 1)} - q_{ia}q_{2a} \right].
\]

Therefore, the covariance is obtained as

\[
\text{cov}(S_1, S_2) = \sum_j v_{1j}v_{2j}(c_j - \bar{c})^2 \psi_{jj}
+ \sum_{j \neq j'} v_{1j}v_{2j'}(c_j - \bar{c})(c_{j'} - \bar{c})\psi_{jj'}.
\]

Clearly, the first term is zero since \( v_{1j}v_{2j} = 0 \) for all \( j \). The second term is also zero since \( \sum_{j'} v_{2j'}(c_{j'} - \bar{c}) = 0 \) from the definition of \( \bar{c} \). Therefore, \( \text{cov}(S_1, S_2) = 0 \) and the stratum-specific statistics are independent.