Dynamic randomization and a randomization model for clinical trials data

Lee D. Kaiser*†

Randomization models are useful in supporting the validity of linear model analyses applied to data from a clinical trial that employed randomization via permuted blocks. Here, a randomization model for clinical trials data with arbitrary randomization methodology is developed, with treatment effect estimators and standard error estimators valid from a randomization perspective. A central limit theorem for the treatment effect estimator is also derived. As with permuted-blocks randomization, a typical linear model analysis provides results similar to the randomization model results when, roughly, unit effects display no pattern over time. A key requirement for the randomization inference is that the unconditional probability that any patient receives active treatment is constant across patients; when this probability condition is violated, the treatment effect estimator is biased from a randomization perspective. Most randomization methods for balanced, 1 to 1, treatment allocation satisfy this condition. However, many dynamic randomization methods for planned unbalanced treatment allocation, like 2 to 1, do not satisfy this constant probability condition, and these methods should be avoided. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: linear model; unit-treatment additivity; permuted-blocks randomization; biased-coin randomization; minimization; randomization test

1. Introduction

A theory for estimation in a linear model, specifically accounting for the randomization used in a clinical trial will be developed. A particular goal is to compare the inferences from the usual population-based linear model with those from the randomization model. A further goal is to compare randomization model inference that follows from permuted-blocks randomization (PBR) with those from dynamic randomization methods. Here, ‘dynamic randomization’ refers to methodology in which the probability of assignment of a given patient to experimental treatment is a function of the patient’s stratification variables and the stratification variables and treatment assignments of previously randomized patients. Other terminology has been used, such as adaptive randomization and minimization.

Randomization of treatments to patients in clinical trials has perhaps most commonly been performed with PBR, often within strata defined by crossings of prognostic factors and/or investigational sites. Permuted blocks are used instead of ‘complete randomization’, which assigns the next patient to experimental treatment essentially by a fair coin toss, to ensure that the number of patients assigned to each treatment group remains similar within strata, regardless of the number of patients ultimately enrolled in the strata [1]. Efron [2] introduced biased-coin randomization, which assigns the next patient to experimental treatment with a biased coin if the counts of patients previously assigned to the two treatments are imbalanced. Pocock and Simon [3] generalized the method to include stratification variables, where the probability of assignment to experimental treatment depends on the balance of treatment counts across the margins of the stratification variables. There have subsequently been many proposals for other dynamic randomization procedures [4–7].
A theory for estimation and testing exists for permutation-based randomizations paired with a linear model, where the linear model includes the randomization’s stratification variables in addition to a treatment term, and leads to unbiased estimators of treatment effects and unbiased estimators of the variance of the treatment effect estimator [8–11]. The unbiasedness is over the distribution of the observations induced by repeated randomizations of treatments to fixed experimental units, under the assumption of unit-treatment additivity with fixed unit and treatment effects. There have been many theoretical developments for randomizations that are not of the permutation type, but these have mostly focused on randomization tests rather than estimation.

- Smythe and Wei [12] derived the asymptotic distribution of a linear rank test of the null hypothesis of equality of two treatments when the assignment of treatments to patients is according to an urn-based randomization scheme.
- Smith [13] summarized treatment allocation methods developed to date. For treatment allocation probabilities that are functions of \((n_1 - n_0)/(n_0 + n_1)\), where the \(n_1\)’s are the numbers of patients previously assigned to the two treatments, Smith developed an approximation to a randomization test based on the difference in sample means. Smith also developed similar approximations when treatment assignment probabilities are a particular function of stratification variables of the current and previous patients and developed asymptotic approximations to randomization tests.
- Wei et al. [14] proposed a biased-coin scheme for the allocation of \(k\) treatments in proportion \(\xi_1, \ldots, \xi_k\), where \(\xi_1 + \ldots + \xi_k = 1\), and generalized the work of Smythe and Wei [12] to \(k > 2\).
- Shao et al. [15] proposed a theory for clinical trials with dynamic randomization, but the results were developed on an assumption of a model that relates the dependent variable to the variables used in constructing hypothesis tests.
- Rosenkranz [16] explored the sampling properties of a difference in sample means under various randomization schemes and noted that a \(t\)-test can be very conservative with biased-coin randomization.

Regardless of the randomization method used, it is common in clinical trials data analysis to apply a statistical test that is based on assumptions of random sampling from a distribution. In general, these tests behave as follows for biased-coin and urn-based randomizations [2, 13, 16–18].

- If there is low-frequency variation in the responses of patients over time of entry into the trial, then a population-based test will tend to result in a larger \(p\)-value than the corresponding randomization test result. With this pattern of variation, patients enrolled closer together in time tend to respond more similarly than patients enrolled farther apart in time.
- If there is high-frequency variation in responses over enrollment time, then a population-based test will tend to yield a smaller \(p\)-value than the corresponding randomization test result. In this case, patients enrolled closer together in time tend to respond less similarly than patients enrolled farther apart in time. This is generally noted to be less likely than the scenario above.
- Finally, when responses appear to come from a homogeneous, time-independent model, then the two test results tend to be similar.

These same qualitative comparisons apply for PBR when the blocking in time is ignored in the analysis [1], in which case the variation in patient responses is quantified by the intrablock correlation.

Motivation for the present work on estimation and testing under a randomization model comes from a ‘points to consider’ document on adjustment for baseline covariates [19] that states ‘… techniques of dynamic allocation such as minimization are sometimes used to achieve balance across several factors simultaneously. Even if deterministic schemes are avoided, such methods remain highly controversial. Thus, applicants are strongly advised to avoid such methods.’ Furthermore, Halpern and Brown [18] noted that dynamic randomization schemes ‘can lead to complex randomization distributions and hence to complex analyses about which little is currently known, and which, in fact, may present intractable difficulties.’ With a randomization-based theory for estimation and testing applicable across all randomization methods, perhaps regulatory authorities will be more accepting of dynamic randomization methods.

Section 2 contains the theory for randomization-based estimation in a linear model with unit-treatment additivity. Section 3 evaluates the estimation of the variance of the treatment effect estimator through a population-based linear model, in comparison with the randomization-based variance estimator. Section 4 compares the estimation results with randomization test results. Sections 5 through 7 explore
randomization-based inference when there is planned unbalanced treatment allocation, like two experimental to one control. A key requirement, often not met with proposed dynamic randomization methods for such unbalanced allocation, is that the probability that a patient receives experimental treatment should be constant across patients. The paper finishes with a discussion and conclusion.

2. The randomization model

Suppose \( n \) patients are in the trial, and they are assigned one of two treatments: experimental or control. The probability that a patient receives experimental treatment is a function of this patient’s stratification variables and the stratification variables and treatment assignments of previous patients. No assumptions are made on this function just yet. Let \( \delta \) be the \( n \times 1 \) vector that indicates assignment to experimental treatment: \( \delta_i = 0 \) if patient \( i \) receives control and \( 1 \) if patient \( i \) receives experimental treatment. Assume unit-treatment additivity, so that \( y \), the \( n \times 1 \) vector of observations on the patients, is

\[
y = y_0 + \delta \tau.
\]

for unknown constant unit effects \( y_0 \) and a scalar treatment effect \( \tau \). A vector of independent and identically distributed random variables could be added as an error term to the model, but this only complicates the development without changing the conclusions.

Let \( X \) be an \( n \times q \) matrix of covariates measured on the patients. Typically, some or all of the stratification variables used in the randomization will be included in \( X \), but this is not required in what follows. Also, it will typically be thought that the variables in \( X \) are prognostic of patient outcome, but there is no assumed relationship between the unit effects in \( y_0 \) and the covariates in \( X \). The only requirement is that \( 1 \), an \( n \times 1 \) vector of ones, is in the column space of \( X \) and, without loss of generality, \( X \) has full column rank. As a simple example, \( X \) could have two columns, with the first column containing \( 1 \)s and the second column denoting Eastern Cooperative Oncology Group performance status of \( 1 \) or more. An analysis of covariance example has \( X \) with a column of \( 1 \)s and a column for a continuous covariate.

We want to estimate \( \tau \) after adjustment for \( X \), and so use the normal equations

\[
egin{pmatrix}
X'X & X\delta \\
\delta'X & \delta\delta
\end{pmatrix}
\begin{pmatrix}
\beta \\
\tau
\end{pmatrix} =
\begin{pmatrix}
X' y \\
\delta' y
\end{pmatrix}
\]

where \( \beta \) is a vector of nuisance parameters. With simple matrix algebra,

\[
\hat{\tau} = \frac{\delta' My}{\delta' M \delta}, \text{ where } M = I - X(X'X)^{-1}X'.
\]

It is worthwhile to explicitly state what is fixed and what is random in this set-up, because this will provide the relative frequency basis for the inferences. The patients have entered the study in a given order, and while they could conceptually have entered in a different order, it is not clear what a reasonable probability model should be for these other potential patient entry orders. Also, would the patients’ unit effects be different if they entered in a different order, and how should these effects be modeled? Would the patients’ baseline covariates differ? Thus, patient entry order, the unknown unit effects, and patient covariates are considered fixed. What does clearly vary, and in a way that we can model, is the treatment assignment to patients, as captured in \( \delta \). The variability in the observations \( y \) then comes only from the randomization mechanism, and the conceptual repetitions of the study arise from independent applications of the randomization method to these patients as entered with their fixed covariates and unit effects. This topic will be addressed further in the discussion section.

Up to this point, no assumptions have been made on the randomization algorithm, and the only one needed to make progress is that \( E\delta = p1 \), with \( p \) a constant: \( 1 \) to \( 1 \) randomization would have \( p = 0.5 \), \( 2 \) to \( 1 \) randomization would have \( p = 2/3 \), and so on. The problems with randomization methods with \( E\delta \neq p1 \) are addressed in Section 6.

This restriction to \( E\delta = p1 \) means that the unconditional probability that each patient receives experimental treatment is constant, even though the probability that a patient receives experimental treatment, conditional on the treatment allocations of previous patients, will often be different from \( p \) to more closely attain treatment balance, either overall or within stratification variable groupings. Note that it is not enough that \( E\delta \) averages to the desired fraction, like \( (1/n)1'E\delta = 1/2 \). With \( 1 \) to \( 1 \) treatment allocation, the \( E\delta = (1/2)1 \) condition is easily satisfied because of the usual symmetry present in
randomization algorithms, while dynamic randomization algorithms for unbalanced treatment allocations often do not maintain a constant $E \delta_i$ value [20]. Examples of this are given in Section 6.1 and a potential solution for certain cases is given in Section 6.

Note that $\bar{r}$ is a function of $\delta' M \delta$, which does not depend on unknown parameters and so is an ancillary statistic. In simple cases, as in Table I, $\delta' M \delta$ is a function of sample sizes within the strata, and most would prefer to condition on observed sample sizes [21]. As shown in Appendix B, $\delta' M \delta/n$ converges in probability to $p(1 – p)$ under mild conditions, so for large samples, inference that is conditional on $\delta' M \delta$ and unconditional inference should be similar. However, for small samples, conditional inference may be preferred, so both approaches are developed in the next sections.

### 2.1. Unconditional inference

Appendix A.1 develops approximations to the mean and variance of $E \delta = p \mathbf{1}$ and unit-treatment additivity and without conditioning on a given $\delta' M \delta$ value. First, $\bar{r}$ is an approximately unbiased estimator of $\tau$. Second,

$$\text{Var}(\bar{r}) \approx \frac{y' M M y_0}{(tr M V)^2}$$

(1)

where $\text{tr}(\cdot)$ denotes the trace of a matrix and $V = \text{Var}(\delta)$ is the variance–covariance matrix of $\delta$. $\text{Var}(\bar{r})$ is a function of the residuals in the regression of $y_0$ on $X$, so good use of covariance adjustment reduces the variance, which is as it should be. It is not so clear what the role of the randomization method is in the reduction of variance, because it is tied up in the $V$ matrix. As an example, however, especially in a small study with a stratified analysis, a stratified permuted-blocks randomization is better than unstratified permuted-blocks randomization [22, 23].

Next, it is shown in Appendix A.1 that

$$\text{Var}(\bar{r}) = \frac{y' M M y - \bar{r}^2 \delta' M M \delta}{(tr M V)(\delta' M \delta - \delta' M M \delta)}$$

(2)

is an approximately unbiased estimator of the Var($\bar{r}$) approximation of Equation (1). Finally, under mild conditions it is shown in Appendix B that $(\bar{r} – \tau)/\sqrt{\text{Var}(\bar{r})}$ has asymptotically a standard normal distribution, so that tests and confidence intervals on $\tau$ can be determined in the usual way.

In a common class of randomization methods, $\bar{r}$ is unbiased, and not just approximately unbiased. Define a symmetrical 1 to 1 randomization method as one with $P(\delta = \delta_0) = P(\delta = 1 - \delta_0)$ for each treatment assignment sequence $\delta_0$. In words, the probability of obtaining any treatment sequence equals the probability of that sequence with the treatments switched between experimental and control. Most 1 to 1 randomization methods treat the experimental and control arms symmetrically and satisfy this condition. For such randomization methods, the unbiasedness of $\bar{r}$ is shown in Appendix A.1.

The diagonal elements of $V$ are $p(1 – p)$ and the off-diagonal elements are $E(\delta_i \delta_j) – p^2$, where $E(\delta_i \delta_j)$ is the probability that patients $i$ and $j$ both receive experimental treatment. Analytical expressions for the first two moments of counts of assignments to treatments for urn-based randomizations [24] and for biased-coin randomizations [25] are available. For some cases, $E(\delta_i \delta_j)$ can be determined algorithmically by stepping through all assignment probabilities [26]. Alternatively, a straightforward

**Table I. $\delta' M \delta$ for three examples of the $X$ matrix.**

| $X$ | $\delta' M \delta$ |
|-----|-------------------|
| Intercept term only | $(1/n_0 + 1/n_1)^{-1}$, where $n_j$ is the number of patients on treatment $j$ |
| Two columns denoting membership in one of two strata, such as_low ECOG status vs. high status | $(1/n_{10} + 1/n_{11})^{-1} + (1/n_{20} + 1/n_{21})^{-1}$, where $n_{ij}$ is the number of patients on treatment $j$ in stratum $i$ |
| Intercept and continuous covariate as in an ANCOVA | $(1/n_0 + 1/n_1)^{-1} \{1 – (1/n_0 + 1/n_1)^{-1}(x_0 – x_1)^2 / \Sigma (x_{ij} – x)^2 \}$, where $n_j$ is the number of patients on treatment $j$, $x_{ij}$ is the covariate value of patient $j$ on treatment $i$ |
solution for any randomization method is to rerandomize treatments to the patients in the trial according to the trial’s algorithm millions of times and estimate \( \mathbf{V} \) with the sample variance–covariance matrix of the resulting \( \delta \)s.

2.2. Conditional inference

For symmetrical 1 to 1 randomizations as defined above, inference conditional on a \( \delta' \mathbf{M} \delta = c \) value is developed in Appendix B.1, where first it is shown that \( E(\hat{\tau} \mid \delta' \mathbf{M} \delta = c) = \tau \). Furthermore, letting \( \text{Var}_{c}(\hat{\tau}) = \text{Var}(\hat{\tau} \mid \delta' \mathbf{M} \delta = c) \), it is shown that

\[
\text{Var}_{c}(\hat{\tau}) = \frac{y' \mathbf{MV}_{c} \mathbf{M} y}{(\delta' \mathbf{M} \delta)^2}
\]

where \( \mathbf{V}_{c} = \text{Var}(\delta \mid \delta' \mathbf{M} \delta = c) \). Finally,

\[
\text{Var}_{c}(\hat{\tau}) = \frac{y' \mathbf{MV}_{c} \mathbf{M} y - \hat{\tau}^2 \text{tr}\mathbf{MV}_{c} \mathbf{M} c}{(\delta' \mathbf{M} \delta)^2 - \text{tr}\mathbf{MV}_{c} \mathbf{M} c}
\]

is a conditionally unbiased estimator of \( \text{Var}_{c}(\hat{\tau}) \).

Because most 1 to 1 randomizations are symmetrical, conditional inference is straightforward, as outlined above. For planned unbalanced allocation, however, even when \( E \delta = p \mathbf{1} \), \( E(\hat{\delta}_i \mid \delta' \mathbf{M} \delta = c) \) is typically not constant over \( i \). For example, consider PBR in blocks of 6 in a 2 to 1 ratio to 10 patients, with \( \mathbf{M} = \mathbf{I} - \mathbf{J}/12 \). Unconditionally, \( E \delta_i = 2/3 \). However, from Table I, if \( \delta' \mathbf{M} \delta = (1/8 + 1/2)^{-1} = 1.6 \), then \( E(\hat{\delta}_i \mid \delta' \mathbf{M} \delta = 1.6) = 2/3 \) for \( 1 \leq i \leq 6 \), but \( E(\hat{\delta}_i \mid \delta' \mathbf{M} \delta = 1.6) = 1 \) for \( 7 \leq i \leq 10 \).

The variation in the conditional probability of assignment to experimental treatment leads to conditional bias in \( \hat{\tau} \). Let \( p_c = E(\delta \mid \delta' \mathbf{M} \delta = c) \), then

\[
E \left( \frac{\delta' \mathbf{M} \delta}{\delta' \mathbf{M} \delta} \mid \delta' \mathbf{M} \delta = c \right) = \tau \frac{p' \mathbf{M} y}{c}
\]

and the conditional bias is proportional to the covariance between \( p_c \) and \( \mathbf{M} y \). The problem of conditional inference for planned unbalanced allocations with \( E \delta = p \mathbf{1} \) is illustrated in Section 5 and a potential solution is raised in the conclusion section.

3. Comparison of standard error estimators — linear model versus randomization model

Regardless of the randomization method used, it is common to analyze trial data with a population-based linear model

\[
y = \mathbf{X} \mathbf{\beta} + \delta \tau + \varepsilon. \tag{3}
\]

The elements of \( \varepsilon \) are independent and identically distributed with zero mean and variance \( \sigma^2 \), \( \delta \) is considered fixed, and \( \mathbf{X} \) often includes stratification factors that are balanced by the randomization [27]. It is common, however, to ignore in the model the balancing in time that randomizations typically achieve.

By construction, the treatment effect estimator is the same under the randomization model as under model (3); however, the variance estimator from Equation (2) will generally differ from the variance estimator derived from model (3). How similar are these estimators from a randomization perspective? The estimators will undoubtedly be different in any particular trial, but it is informative to consider their expectations.

The linear model estimate of the variance of \( \hat{\tau} \) is

\[
\text{Var}_{\text{lm}}(\hat{\tau}) = y' \left( \mathbf{I} - (\mathbf{X}' \delta) \left( \mathbf{X}' \mathbf{X} \delta' \mathbf{X} \delta \right)^{-1} \left( \mathbf{X}' \delta' \delta \right) \right) \frac{1}{(\text{trM} - 1) \delta' \mathbf{M} \delta} = \frac{y' \mathbf{M} y}{(\text{trM} - 1) \delta' \mathbf{M} \delta} - \frac{\hat{\tau}^2}{\text{trM} - 1}. \tag{4}
\]
What is the expectation of $\text{Var}_{\text{lm}}(\hat{\tau})$ from a randomization perspective? Again, the numerator and denominator both contain random variables, but approximate by simply taking expectations individually of the numerator and denominator to obtain

$$E_r\text{Var}_{\text{lm}}(\hat{\tau}) \approx \frac{y_0'y_0 + \tau^2 \text{tr} M}{(\text{tr} M - 1) \text{tr} M} - \frac{\text{Var}_r(\hat{\tau})}{\text{tr} M - 1}$$

where the subscript $r$ on $E_r$ and $\text{Var}_r$ denotes expectation and variance over the randomization method used. For what unit effects $y_0$ might $E_r\text{Var}_{\text{lm}}(\hat{\tau})$ be similar to, less than, or greater than $\text{Var}_r(\hat{\tau})$? Some guidance is given in the next two sections.

### 3.1. Linear model versus randomization model standard error estimators when no pattern is expected for unit effects

Suppose that apart from dependence on $X$, the unit effects $y_0$ display no further systematic variation. A reasonable model for this is $y_0 = X\beta + e$, where $e$ is a vector of independent and identically distributed random variables with expectation zero and finite variance $\sigma^2$. It is easy to see that the expectation of both $E_r\text{Var}_{\text{lm}}(\hat{\tau})$ and $\text{Var}_r(\hat{\tau})$ equals $\sigma^2/\text{tr} M$.

Therefore, if no discernable pattern in unit effects is expected, apart from variation owing to dependence on $X$, then the usual linear model estimate of the estimated treatment effect variance estimates a quantity similar to the proper randomization variance. This is true regardless of the type of randomization, as long as $E \delta = p I$.

Note also that if complete randomization were used, with $V = p(1 - p) I$, then the randomization model variance estimator in (2) equals the linear model estimator in (4), illustrating the equivalence of randomization-based and population-based inference with complete randomization [28].

### 3.2. Linear model versus randomization model standard error estimators when unit effects have systematic variation

The difference between the expectation of the linear model variance estimator in Equation (5) and the randomization variance in Equation (1) is proportional to $y_0'y_0/\text{tr} M - y_0'\text{tr} M y_0/\text{tr} M$. These quadratic forms have been evaluated in the context of accidental bias by Efron [2] and Smith [13] for biased-coin randomization.

Efron defined accidental bias as the expectation, over the randomization used, of the squared bias of $\hat{\tau}$ calculated under an assumed regression model for the observations, where the regression model includes a covariate $z$ not accounted for in the treatment effect estimator. This squared bias is proportional to $z'\text{tr} M z$, and therefore the work of these authors applies to the comparison of variance estimators considered here. The applicable conclusion for biased-coin randomization is that when unit effect residuals, after regression on $X$, exhibit low-frequency variation or a smooth trend from patient to patient, then the linear model estimate of variance will tend to be larger than the randomization model estimate. With high-frequency variation, the comparison of the variances is reversed.

Urn randomization is asymptotically free of accidental bias [17] so that for very large samples, the linear model variance estimate is likely to be similar to the randomization-based variance. For smaller samples, the control of sample size differences achieved by urn randomization is likely to lead to comparisons of variances similar to those for biased-coin randomization.

This parallels the situation with PBR where the variance estimate ignoring the blocking in time will generally be greater than the estimate accounting for this blocking when there is low-frequency variation or a smooth time trend in the unit effect residuals, as quantified by the intrablock correlation [1].

### 4. Randomization tests

The null hypothesis that treatment has absolutely no effect in any patient can be tested with a randomization test. Under this hypothesis, the vector of observations $y$ equals $y_0$, and the observed $\hat{\tau}$ is just one of the possible treatment effect estimates that could be obtained by the random assignment of treatment labels to these patients. Is the observed $\hat{\tau}$ unusually large or small in this distribution of possible $\hat{\tau}$ values?
A randomization test is implemented by performing the randomization repeatedly with the patients enrolled in their fixed order and determining the resulting $\hat{\tau}$ values. For an alternative hypothesis of positive $\tau$, the randomization test $p$-value is estimated as the fraction of the $\hat{\tau}$ values that are greater than the observed $\hat{\tau}$; for a two-sided alternative, compare $|\hat{\tau}|$ to $|\bar{\hat{\tau}}|$. As with estimation, the randomization test can be performed conditionally on $\delta^T\mathbf{M}\delta = c$ or unconditionally. For the conditional randomization test, only those randomizations with $\delta^T\mathbf{M}\delta = c$, or suitably close to it, are used for the generation of the $\hat{\tau}$ values.

What are the first two moments of this randomization distribution of $\hat{\tau}$? With $E\delta = \rho\mathbf{1}$, from Section 2.1, unconditionally $E\hat{\tau} = 0$ and

$$\text{Var}(\hat{\tau}) \approx \frac{y^T\text{MVMy}}{(\text{trMV})^2} \quad (6)$$

Under the asymptotic normality conditions at the beginning of Appendix B, an approximation to the randomization test would be to compare $\hat{\tau}/\sqrt{\text{Var}(\hat{\tau})}$ with standard normal quantiles. This test statistic is similar to the $Z$-statistic obtained from use of SE ($\hat{\tau}$) from Equation (2), except that a ‘sum of squares for treatment’ is removed in (2) relative to the variance in the randomization distribution. Through use of approximations as in Appendix A.1, the expectation of Equation (6) exceeds that of Equation (2) by approximately $\frac{\tau^2}{4}\text{trMV}^2/(\text{trMV})^2$, which is small for small $\tau$ or large $n$.

Thus, the randomization test inference will necessarily be similar to the inference that comes from simple use of $\hat{\tau}$ and its estimated standard error. As above, however, if the estimated standard error is obtained from a linear model based on $X$, then the randomization test need not agree with the linear model results. From Section 3.1, when the unit effects display no trend apart from dependence on $X$ then the randomization inference should be similar to that from linear model inference. Likewise, from Section 3.2, when unit effects display a low-frequency trend, then the randomization test $p$-value would be expected to be smaller than the $p$-value from the linear model, with the reverse order when the unit effects display high-frequency variation. This agrees with the summary of randomization tests given in the introduction.

5. Conditional inference with planned unbalanced treatment allocation ratio

Section 2.2 noted that planned unbalanced treatment allocation with $E\delta = \rho\mathbf{1}$ yields an approximately unbiased treatment effect estimator unconditionally, but the estimator is typically biased conditioned on a $\delta^T\mathbf{M}\delta$ value. This section illustrates by example the potential for bias with conditional inference and shows that the problem is applicable to PBR and to dynamic randomization.

Consider the randomization of two treatments to 38 patients in a 2 to 1 ratio, with patients categorized by two binary stratification variables. The patient counts in the four cells defined by these two variables are respectively 10, 8, 11, and 9 for cells (0,0), (0,1), (1,0), and (1,1), with the first value of the pair indicating the Stratum 1 level and the second value indicating the Stratum 2 level. The order of entry of these patients to the simulated trial was determined at random and then held fixed.

Two randomization methods are evaluated. The first is stratified PBR, with 2 to 1 allocation in blocks of six applied independently within each of the four cells. The second is a sequential method [5, 29] that followed the recommendation at the end of Section 6.1 to perform a symmetrical 1 to 1 to 1 randomization with three treatments, with two treatment arms subsequently combined (see Appendix C for details). The matrix $X$ includes an intercept and main effects for the two strata. Each randomization method was replicated 10 million times to determine the $V_r$ matrices for this population. The methods were subsequently replicated one million times to determine the conditional properties of the estimators.

For each randomization method, the expected value of the ancillary statistic $\delta^T\mathbf{M}\delta$ is 8.3 and approximately 90% of the $\delta^T\mathbf{M}\delta$ values are between 7.7 and 8.9; smaller values represent randomizations that generally exceed the 2 to 1 ratio and larger values represent randomizations with closer to 1 to 1 allocation. Figure 1 illustrates the fluctuation across the patients in the empirically estimated $E(\delta_i|\delta^T\mathbf{M}\delta = c)$ for selected values of $c$, grouped to the nearest 0.2. For stratified PBR, $E(\delta_i|\delta^T\mathbf{M}\delta = c)$ is exactly 2/3 for earlier patients who belong to complete blocks, and similarly for the sequential method, $E(\delta_i|\delta^T\mathbf{M}\delta = c)$ is approximately 2/3 for early patients. However, for patients in incomplete blocks with PBR and for the later patients with the sequential method, the conditional probability of assignment to active treatment diverges from 2/3, with $E(\delta_i|\delta^T\mathbf{M}\delta = c)$ generally above 2/3 for smaller values of $c$ and below 2/3 for larger values of $c$. 
From Section 2.2, the conditional bias is proportional to the covariance between $p_{c, \delta} = E(\delta | \delta'M\delta = c)$ and the residuals of the unit effects after regression on $X$. Suppose here that unit effects are related to order of patient entry via a simple linear regression with a positive slope and coefficient of determination $R^2$, and let $E_{lm}$ denote the expectation with respect to this model. Through simple algebra, the expectation of the conditional bias normalized by $\sqrt{(E_{lm} \text{Var}_{c}(\hat{\tau}))}$ is

$$E_{lm}E\left(\hat{\tau} | \delta'M\delta = c\right) - \tau = \frac{p_{c}'Mz}{\sqrt{\{(\text{tr}MV_c)(1 - R^2)/R^2 + z'MV_cMz\}}}$$

where $z$ is a unit-length vector proportional to patient randomization order. This normalized bias is displayed in Figure 2 for each randomization method for $R^2 = 0.5$ and $X = 1$. The normalized bias is similar between randomization methods and is positive for smaller values of $\delta'M\delta$ and negative for larger values of $\delta'M\delta$. At the smallest and largest values of $\delta'M\delta$ displayed the normalized bias is approximately 10% in absolute value.

Thus, for planned unbalanced treatment allocation with $E\delta = p1$, randomization inference conditional on $\delta'M\delta$ can be problematic, and this problem is present with both dynamic randomization and PBR. Because this problem arises in stratified PBR only for the patients in the last potentially incomplete block in each stratum, the bias is likely negligible in large trials. Likewise, for dynamic randomization, although based on limited empirical evaluations, $p_{c}$ appears to diverge from the desired fraction only for the latest patients randomized in a trial, so that the conditional bias for large trials would also seem to be negligible.

The next sections shift to the problem of unconditional bias when $E\delta \neq p1$ and illustrate that this problem is common with planned unbalanced treatment allocation.
6. Estimation and randomization tests when $E\delta \neq p1$

When $E\delta = p \neq p1$, then $\hat{\tau}$ is biased. Using a first-order approximation, the bias is

$$E \left( \frac{\delta' My_0}{\delta' \hat{M} \delta} \right) \approx \frac{E\delta' My_0}{E\delta' \hat{M}\delta} = \frac{p' My_0}{\text{tr}(MV) + p' Mp}$$

The bias is a function of the correlation between $p$ and the residuals from the regression of $y_0$ on $X$.

For the randomization test when $E\delta \neq p1$, the randomization distribution of the usual estimator $\hat{\tau}$, is not centered at zero, but at $p0_{My}/(\text{tr}(MV) + p' Mp)$, approximately. This nonzero expectation does not affect the Type I error rate by construction of the randomization test, although the power of the test may be affected [30].

A notable example of $E\delta \neq p1$ is contained in the documents of the Endocrinologic and Metabolic Drugs Advisory Committee meeting held October 21, 2008 [30, 31]. In this case, the linear model inference indicated a statistically significant treatment effect, while the randomization test did not, resulting in $p = 0.035$ versus $p = 0.06$ for the analysis of one endpoint and $p = 0.046$ versus $p = 0.15$ for another endpoint.

Because the bias is a function of $p' My_0$, a simple method to obtain an approximately unbiased estimator of $\tau$ is to augment $X$ with the vector $p$, use $[X p]/c141$ in the normal equations, and proceed as before with estimation of $\hat{\tau}$ and the standard error. A randomization test would also proceed as before. The asymptotic normality results of Appendix B were developed with $E\delta \neq p1$, so it is uncertain whether asymptotic normality applies in this case. Even if it did, the best solution is not to use randomization methods with $E\delta \neq p1$ in the first place.

6.1. Dynamic randomization with planned unbalanced treatment allocation ratio

Consider randomizations for a trial with 2 to 1 randomization of experimental to control treatment. $P(\delta_i = 1) = 2/3$ for PBR applied in blocks of multiples of three. What about biased-coin randomization? A natural attempt is to use a biased coin whenever $n_c/2$ differs too much from $n_e$, where $n_c$ and $n_e$ denote the number of patients previously assigned to experimental and control. Suppose a threshold of 1 or more is used, with $P(\delta_i = 1) = 0.9$ for too many previous control arm assignments, $P(\delta_i = 1) = 0.2$ for too many previous experimental arm assignments, and $P(\delta_i = 1) = 2/3$ otherwise. This procedure satisfies conditions C1 and C2 in [13] and the probabilities satisfy the recent recommendation of Han et al. [6]. With the first patient’s assignment to experimental treatment determined with probability 2/3, it is easy to calculate the $P(\delta_i = 1)$ values for $i = 2, 3, 4,$ and 5 as 0.744, 0.467, 0.674, and 0.761. Stepping through the probabilities algorithmically yields $P(\delta_{20} = 1) = 0.775$. The probabilities do not converge to 0.667, although the average of these 20 assignment probabilities is close to 0.667.

The situation is worse with the marginal balance metric [6]. In this simple case, the biased coin is applied to the treatment that yields the smaller value of $|(n_c + 1)/2 - n_c|$ versus $|n_e/2 - n_c - 1|$. 

Copyright © 2012 John Wiley & Sons, Ltd. 

Statist. Med. 2012, 31 3858–3873
If the former is smaller, then experimental treatment is assigned with probability 0.9, and if the latter is smaller, then experimental treatment is assigned with probability 0.2. With equal values and also for the first patient, experimental treatment is assigned with probability 2/3. \( P(\delta_1 = 1) \) values for \( i = 2, 3, 4, \) and 5 are 0.433, 0.807, 0.881, and 0.342. Stepping through the probabilities algorithmically yields \( P(\delta_20 = 1) = 0.340. \)

The generalization of 1 to 1 urn randomization to unbalanced allocation in Rosenberger and Lachin [32] is incorrect. For example, with 2 to 1 allocation, and starting with 2 E balls and 1 C ball, one application of their generalization would add 2 E balls if control treatment is assigned, with 1 C ball added if experimental treatment is assigned. The first patient is assigned experimental treatment with probability 2/3, and \( P(\delta_1 = 1) \) values for \( i = 2, 3, 4, \) and 5 are 0.6, 0.590, 0.589, and 0.588. These probabilities converge to 0.586. A fix in this simple case is to add 4 E balls, rather than 2 E balls, if control treatment is assigned, with other aspects of the method unchanged. Then the \( P(\delta_1 = 1) \) values for \( i = 2, 3, 4, \) and 5 are 0.620, 0.641, 0.656, and 0.662 and the probabilities converge to 2/3. In large samples, urn sampling behaves like complete randomization [17] so that deviation from the target allocation ratio is likely to be greater with urn sampling than with other methods.

A solution for \( r_1 : r_2 \) treatment allocation is to perform randomization for balanced allocation to \( r_1 + r_2 \) treatment arms, and then combine \( r_1 \) of these arms for the experimental arm assignment and \( r_2 \) of these arms for control. As long as the randomization method treats the arms symmetrically then \( E\delta = p\mathbf{1} \) will hold, and unconditional inference can be carried out as developed above.

### 7. Discussion

The estimation of \( \tau \) involves only \( \delta, M, \) and \( y, \) and the estimator \( \hat{\tau} \) is approximately unbiased as long as \( E\delta = p\mathbf{1} \) and unit-treatment additivity holds. This approximate unbiasedness is true regardless of the relationship between \( y_0 \) and \( X \) and any other properties of the randomization algorithm. Therefore, at least from the point of view of unbiased estimation of \( \tau, \) there is no natural determination of \( X \) based on the randomization used, such as if the randomization is stratified on a particular factor, then that factor should be included in \( X. \) However, from the point of view of \( \text{Var}(\hat{\tau}), \) covariates that are related to the unit effects should be included in \( X, \) as smaller residuals from a regression of \( y_0 \) on \( X \) reduce \( \text{Var}(\hat{\tau}). \) It is unclear how manipulation of the randomization scheme influences \( \text{Var}(\hat{\tau}), \) but it is likely that randomization methods that seek to balance the covariates in \( X \) between treatment groups are somewhat better than those that do not, at least for small studies. Finally, for a proper randomization-based variance estimator, both \( X \) and the randomization method, through \( V, \) should be accounted for. On this point, the advice to ‘analyze as you randomize’ should be followed.

Variability in the observations in the randomization model arises only from the random assignment of treatments to the patients in the order they entered the trial. This order and the patients’ baseline covariates determine the distribution of the treatment assignment vector \( \delta \) and subsequent properties of the treatment effect estimator. In the discussion of randomization tests, this fixing of patient order is common [12, 13, 17, 26, 33]. A reviewer asked whether study conclusions can only be applied to these patients with the same order of entry? Now, it is common [34, 35] to consider that statistical inference in a randomized trial applies to the actual patients in the trial, with inference to the broader population of similar patients being either nonstatistical or based on assumptions that trial data follow specified distributions. In answer to the reviewer’s question, it seems necessary to go a bit further and to say that the statistical inference from the randomization model applies to the actual patients in the trial, with inference to the broader population somewhat different circumstances, but to make statistical statements about such a treatment effect would require further assumptions.

One generalization of the model of Section 2 is to drop the assumption of unit-treatment additivity. Let \( \tau_i \) be the difference in response if patient \( i \) were given experimental treatment versus the patient’s response with control treatment. The estimator \( \hat{\tau} \) in Section 2 is in general a biased estimator of \( \tau, \) defined as the average of the \( \tau_i \) values across the \( n \) patients. Using the first-order approximation to the expectation of a ratio of random variables as in Section 2.1, the bias is a function of a contrast in the diagonal elements of the matrix \( VM. \) These diagonal elements will often be nearly constant, in which case the bias will be small. Gadbury [36] showed for permutation-based randomizations that the usual estimator of \( \text{Var}(\hat{\tau}) \) is positively biased. If a similar result applied to the variance estimator in Equation (2) for the
more general randomizations considered here, then the randomization model estimators would yield, for example, nearly unbiased estimates of the average treatment effect, along with a conservative standard error estimate and conservative hypothesis tests on $\gamma$.

Another generalization is to more than two treatments. This appears to be relatively straightforward for a continuous endpoint with unit-treatment additivity through the introduction of a separate $\delta$ vector for each of $t - 1$ treatments.

It would be interesting to see whether randomization model inference could be extended from the continuous endpoints considered here to binary responses and time-to-event endpoints. Certainly, randomization tests of a null hypothesis of no treatment effect in any patient are available for such endpoints, but treatment effect estimation from a randomization perspective does not seem to be available.

In addition to the analysis of the whole trial, it is important to explore treatment effects within subgroups of patients. If $E\delta = p1$ and there is unit-treatment additivity within the subgroup, then a linear model estimate within the subgroup of this common treatment effect as in Section 2 is approximately unbiased, regardless of the covariates included in such an estimate. An estimate of the standard error of this estimator could come from Equation (2), with $V$ determined only from this subset of patients. As in Section 3.1, if unit effects within this subset are expected to display no systematic variation other than dependence on the covariates, then the linear model estimate of the standard error should be a reasonable approximation to the randomization-based standard error estimate.

8. Conclusion

For symmetrical 1 to 1 randomization, inference conditional on $\delta' M \delta$ and unconditional inference are both straightforward, and a randomization test will necessarily be similar to a test result obtained from $\hat{\gamma}$ and its estimated variance, which is justified by a central limit theorem. With planned unbalanced treatment allocation, Proshcan et al. [30] argued to ‘minimize the use of minimization.’ However, it is only with $E\delta = p1$ that unconditional inference is problematic. Unfortunately, as illustrated in Section 6.1, many dynamic randomization methods for planned unbalanced allocation result in $E\delta \neq p1$, and care must be exercised to achieve $E\delta = p1$, in which case unconditional inference is also straightforward. Although likely only relevant for smaller trials, with $E\delta = p1$ and planned unbalanced allocation, conditional inference will typically lead to bias in $\hat{\gamma}$. This conditional bias is also present with PBR when the blocking in time is ignored in the analysis. A potential repair to $\hat{\gamma}$ is to include, as suggested in Section 6 for unconditional inference, $p_c = E(\delta | \delta' M \delta = c)$ as a covariate in $X$ and proceed with inference based on $V_c = \text{Var}(\delta | \delta' M \delta = c)$. This possible repair needs further evaluation.

An additional similarity between PBR and dynamic randomization methods that attempt to balance treatment counts is that inference from a population-based model will likely be similar to a randomization-based analysis when unit effects, or rather the residuals from the regression of unit effects on $X$, are essentially sampled from a homogenous population. Furthermore, for both PBR and dynamic methods, when unit effect residuals display low-frequency variation or a general trend related to enrollment time, then the estimated variance of $\hat{\gamma}$ from a population-based model will tend to overestimate the randomization-based variance, with the opposite order in variance terms for high-frequency variation.

In our company’s experience, the use of dynamic randomization has never been an issue in disqualifying a positive study. However, at the time of protocol review, we often receive comments from health authority review staff on our randomization method and trial analysis methods. It is very common that we receive a request for a randomization test for the main trial endpoints. I believe we would not receive such comments and requests if we routinely used PBR. With the results of this manuscript, owing to the similarity between randomization inference for dynamic randomization and PBR, it seems inconsistent to single out dynamic methods for the performance of a randomization test. Given that the blocking in time achieved by PBR and dynamic methods is routinely ignored in analyses, one should either (1) be satisfied that these analyses are likely to match a randomization analysis or to be conservative or (2) ask for a randomization test for any randomization method other than complete randomization.

Appendix A.1

Approximations to the unconditional moments of $\hat{\gamma}$ will be developed here under the assumptions of $E\delta = p1$ and unit-treatment additivity. Recall that $X$ is an $n \times q$ full column rank matrix with $I$ in its
column space and that $M = I - X(X'X)^{-1}X$. First, note that

$$\hat{\tau} = \frac{\delta'M(y_0 + \delta\tau)}{\delta'M\delta} = \tau + \frac{\delta'My_0}{\delta'M\delta}$$

so that the moments of $\hat{\tau}$ can be found from the moments of $\delta'My_0/\delta'M\delta$. Approximate this ratio with a first-degree Taylor series expansion about the means of the numerator and denominator to obtain the approximation $\delta'My_0/\delta'M\delta$. Therefore, $\hat{\tau}$ is approximately unbiased.

For the variance of $\hat{\tau}$,

$$\text{Var} \left( \frac{\delta'My_0}{\delta'M\delta} \right) \approx \text{Var} \left( \frac{\delta'My_0}{\text{trMV}} \right) = \frac{y_0^2 \text{MV} \text{My}_0}{(\text{trMV})^2} = \text{Vår} (\hat{\tau})$$

For the estimation of $\text{Var} (\hat{\tau})$, note that

$$Ey'MVMy = E (y_0 + \delta\tau)'MVM (y_0 + \delta\tau) = y'MVMy_0 + \tau^2 E\delta'MVM\delta = (\text{trMV})^2 \text{Vår} (\hat{\tau}) + \tau^2 \text{tr}(MVM)$$

Define

$$\text{Vår}_0 (\hat{\tau}) = \frac{y'MVMy - \hat{\tau}^2 \text{trMV} MVM}{(\text{trMV})^2 - \text{trMV} MVM}$$

It is easy to see that

$$E\text{Vår}_0 (\hat{\tau}) = \text{Vår} (\hat{\tau}) + \left( \tau^2 - (E\hat{\tau})^2 \right) \frac{\text{trMV} MVM}{(\text{trMV})^2 - \text{trMV} MVM}$$

Because $E\hat{\tau} \approx \tau$, $\text{Vår}_0 (\hat{\tau})$ is an approximately unbiased estimator of $\text{Vår} (\hat{\tau})$.

A problem with $\text{Vår}_0 (\hat{\tau})$, inherent in its unconditional nature, is that it averages over variation in $\delta'M\delta$. A potentially better estimator that explicitly depends on $\delta'M\delta$ can be found by requiring that the variance estimator match the linear model estimator of Equation (4) in Section 3.0 when complete randomization with $V = p(1 - p)I$ is used. With this requirement

$$\text{Vår} (\hat{\tau}) = \frac{y'MVMVy - \hat{\tau}^2 \delta'MVM\delta}{\delta'M\delta \text{trMV} - \delta'MVM\delta}$$

Approximating the expectation of the ratio by the ratio of the expectations and approximating $E \left( \hat{\tau}^2 \delta'MVM\delta \right)$ by the product of the expectations, $\text{Vår} (\hat{\tau})$ is also an approximately unbiased estimator of $\text{Var} (\hat{\tau})$.

Appendix A.2 shows for symmetrical 1 to 1 randomization methods, defined in Section 2.1, that $\hat{\tau}$ is unbiased conditional on $\delta'M\delta = c$. The estimator $\hat{\tau}$ is therefore unconditionally unbiased for such randomization methods.

**Appendix A.2**

For the properties of $\hat{\tau}$ conditional on $\delta'M\delta = c$, it is first necessary to evaluate $E(\delta | \delta'M\delta = c)$. In general, this conditional expectation is not proportional to the vector $I$ even with $E(\delta) = pI$. An example for PBR in blocks of six for treatments in a 2 to 1 ratio is given in Section 2.2.

In spite of potential bias issues with conditional inference with planned unbalanced treatment allocation, $\hat{\tau}$ is conditionally unbiased with symmetrical 1 to 1 randomization, defined in Section 2.1, where $P(\delta = \delta_0) = P(\delta = 1 - \delta_0)$. To show this, note first that if $\delta'M\delta = c$ then $(1 - \delta)'M(1 - \delta) = c$. 
Thus, let \( \{ \delta_{ij} \} \) denote all treatment assignment sequences with \( \delta_{ij}'M\delta_{ij} = c \), where \( \delta_{i2} = 1 - \delta_{i1} \). Also, let \( p_{ij} \) denote the probability of obtaining the sequence \( \delta_{ij} \). Because of symmetry, \( p_{11} = p_{12} \). By definition,

\[
E(\delta | \delta'BM = c) = \frac{\sum_{ij} p_{ij} \delta_{ij}}{P(\delta'BM = c)} = \frac{\sum_{i} p_{1i}(\delta_{i1} + (1 - \delta_{i1}))}{P(\delta_{i1}'M\delta_{ij} = c)} = (1/2)1
\]

Hence, \( E(\hat{\tau} | \delta'BM = c) = \tau \) for all \( c \).

Let \( \text{Var}_c(\hat{\tau}) \) denote \( \text{Var}(\hat{\tau} | \delta'BM = c) \). Continuing with symmetrical 1 to 1 randomization, it is easy to see that

\[
\text{Var}_c(\hat{\tau}) = \frac{\nu_o'MV_cMy_0}{(\delta'BM)^2}
\]

where \( V_c = \text{Var}(\delta | \delta'BM = c) \). Following the approach in Appendix A.1,

\[
\text{Var}_{c} (\hat{\tau}) = \frac{\nu MV_cMy - \hat{\tau}^2 trMV_cMV_c}{(\delta'BM)^2 - trMV_cMV_c}
\]

is a conditionally unbiased estimator of \( \text{Var}_c(\hat{\tau}) \).

**Appendix B**

Here, a proof is sketched of the asymptotic normality of \( \hat{\tau} \) under the randomization model with \( E\delta = p1 \). As in Section 2, \( \hat{\tau} \) is adjusted for \( X \), which has full column rank \( q \) and \( 1 \) in its column space.

Intuitively, for reasonable randomization algorithms, treatment assignments for very large \( n \) have almost no dependence on treatment assignments for small \( n \), so central limit theorems for sequences of random variables that satisfy a ‘strong mixing’ condition [37,38] should be helpful. However, the strong mixing conditions are difficult to verify for randomization algorithms.

The approach taken here, then, is to approximate the sequence \( \delta_1, \delta_2, \ldots \) with an asymptotically equivalent triangular array \( \{ \delta_{n1}, \delta_{n2}, \ldots \delta_{nn} \} \). The original randomization scheme is applied within the array, except that treatment assignments are based on only the most recent \( m \) patients, with treatment assignments in earlier patients ignored. This makes the triangular array \( m \)-dependent [38], and \( m \) will be allowed to increase with \( n \). Note that permuted-blocks randomization generates a sequence of \( \delta \) that is already \( m \)-dependent, where \( m \) is the block size.

Because \( \hat{\tau} = \tau + \delta'BM_0/\delta'BM \) it suffices to show the asymptotic normality of \( \sqrt{n}\delta'M_0/\delta'BM \). It will first be shown that \( \delta'BM_0/\sqrt{n} \) is asymptotically normal and then that \( \delta'BM/\sqrt{n} \rightarrow_p p(1 - p) \), where \( \rightarrow_p \) denotes convergence in probability.

As before, let \( V = \text{Var}(\delta) \), with \( (i, j) \) element \( v_{ij} \). Make the following assumptions:

(A1) \( \nu_0'MVM_0/n \rightarrow v > 0 \) as \( n \rightarrow \infty \).

(A2) Let \( (a_{n1}, \ldots, a_{nn})' = a_n = My_0 \) and assume there is a finite \( A \) with \( |a_{ni}| < A \) for all \( n \) and \( i \).

(A3) For all \( i, \Sigma_j |v_{ij}| \rightarrow B_i \) as \( n \rightarrow \infty \), and there is a \( B \) with \( B_i < B \) for all \( i \).

Let \( \delta_n = (\delta_{n1}, \delta_{n2}, \ldots \delta_{nn})' \), the vector of the \( m \)-dependent array elements, where \( m \) is selected as the smallest integer less than or equal to \( n^{0.5 - \epsilon} \) for arbitrarily small \( \epsilon > 0 \).

Now apply Berk’s theorem [39]. His condition (i) on the uniform boundedness of \( E|a_{n1}\delta_{n1}|^{2 + r} \) applies for arbitrarily large \( r \) owing to Assumption A2. His condition (ii) on the uniform boundedness of \( \text{Var}(a_{n1}\delta_{n1} + \ldots + a_{nn}\delta_{nn})/(j - i + 1) \) for \( 1 \leq i < j \leq n \) is satisfied because of both Assumptions A2 and A3. His condition (iii) on the limit of \( \text{Var}(a_{n1}\delta_{n1} + \ldots + a_{nn}\delta_{nn})/n \) follows from Assumption A1. Finally, his condition (iv) on \( \lim_{n \rightarrow \infty} n^{(2 + 2/r)/r} = 0 \) is satisfied for \( r > (1 - \epsilon)/\epsilon \), where \( \epsilon \) was given above in the construction of the \( m \)-dependent array \( \delta_n \). It follows that \( (a_{n1}\delta_{n1} + \ldots + a_{nn}\delta_{nn})/\sqrt{n} \) converges in distribution to \( N(0, v) \).

Next, assume

(A4) the maximum eigenvalue of \( V \), denoted \( \lambda_{\text{max}} \), is \( o(n) \).
Using the inequality \( \|V\| \leq \lambda_{\text{max}} \) for any unit-length vector \( \mathbf{l} \), it is then straightforward to show that \( \delta' \mathbf{M}\delta / n \to_p p(1 - p) \). First, let \( \mathbf{M} = \mathbf{I} - \mathbf{L}\mathbf{L}' \), where \( \mathbf{L}/\mathbf{L} = \mathbf{l}_q \). Let \( \mathbf{l}_i \) denote the \( i^{th} \) column of \( \mathbf{L} \), where without loss of generality, \( \mathbf{l}_1 = 1 / \sqrt{n} \). Then

\[
\frac{\delta' \mathbf{M} \delta}{n} = \frac{\delta' (\mathbf{I} - \mathbf{J}/n) \delta}{n} + \sum_{i \geq 1} (\mathbf{l}_i' \delta)^2 / n
\]

Now, \( \delta' (\mathbf{I} - \mathbf{J}/n) \delta / n = \delta' \mathbf{1}/n - (\delta' \mathbf{1})^2 / n \to_p p(1 - p) \), because \( E(\delta' \mathbf{1}/n) = 0 \) and \( \text{Var}(\delta' \mathbf{1}/n) = (\mathbf{1}' \mathbf{V} \mathbf{1}/n)/n \leq \lambda_{\text{max}} \to 0 \). The result follows because for each \( i = 2, \ldots, q \), \( E(\mathbf{l}_i' \delta / \sqrt{n}) = 0 \) and \( \text{Var}(\mathbf{l}_i' \delta / \sqrt{n}) \leq \lambda_{\text{max}} \to 0 \). Thus, \( \sqrt{n} (\hat{\tau} - \bar{\tau}) \) converges in distribution to \( N(0, v/(p(1 - p))^2) \).

Because \( v \) is unknown, practical application of the asymptotic normality of \( \hat{\tau} \) for tests and confidence intervals requires a consistent estimator of \( v/(p(1 - p))^2 \). It will now be shown that \( \text{Var}(\hat{\tau}) \) of Equation (A1.3) provides one solution, but a further assumption is needed.

(A5) \( \lambda_{\text{max}} \), the maximum eigenvalue of \( \mathbf{V} \), is \( o(\sqrt{n}) \). This is a stronger version of Assumption A4.

Now,

\[
n \text{Var}(\hat{\tau}) = \frac{y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0 / n - 2 \tau y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0 / n + (\bar{\tau}^2 - \bar{\tau}^2) \delta' \mathbf{M} \delta / n}{(\delta' \mathbf{M} \delta / n)(\mathbf{V} \mathbf{M} \delta / n) - \delta' \mathbf{M} \delta / n^2}
\]

The pieces of this expression will be handled in turn, starting with the denominator. As shown above, \( \delta' \mathbf{M} \delta / n \to_p p(1 - p) \). Next, with \( \mathbf{M} = \mathbf{I} - \mathbf{L}\mathbf{L}' \) as above, \( \text{tr} \mathbf{M} \delta / n = (\mathbf{V} - \text{V} \mathbf{V} / n) / n \). The trace of \( \mathbf{V} \) is \( n(1 - p) \) and \( \text{tr} \mathbf{V} / n \) is bounded between 0 and \( q \lambda_{\text{max}} / n \), which converges to zero by A5. Thus, \( (\delta' \mathbf{M} \delta / n)(\mathbf{V} \mathbf{M} \delta / n) \to_p p(1 - p)^2 \). Continuing with the denominator, \( \delta' \mathbf{M} \delta / n^2 = (\delta' \mathbf{M} \delta / \sqrt{n})^2 \leq (\lambda_{\text{max}} / n)(\delta' \mathbf{M} \delta / n) \to_p 0 \) from both A5 and the convergence in probability of \( \delta' \mathbf{M} \delta / n \).

For the numerator of \( n \text{Var}(\hat{\tau}) \), \( y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0 / n \to v \) by Assumption A1. Next, \( E(y_0' \mathbf{V} \mathbf{M} \mathbf{y}_0) = 0 \) and it will be shown that \( \text{Var}(y_0' \mathbf{V} \mathbf{M} \mathbf{y}_0) \to 0 \), implying that \( y_0' \mathbf{V} \mathbf{M} \mathbf{y}_0 / n \to_p 0 \). The following repeatedly uses the inequality \( \|V\| \leq \lambda_{\text{max}} \) for any unit-length vector \( \mathbf{l} \) and that \( x' \mathbf{M} \mathbf{x} \leq x' \mathbf{x} \) for any \( \mathbf{x} \).

\[
\text{Var}(y_0' \mathbf{V} \mathbf{M} \mathbf{y}_0) = \frac{y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0 \cdot y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0}{n^2 y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0} \leq \frac{\lambda_{\text{max}}^2 y_0' \mathbf{V} \mathbf{M} \mathbf{V} \mathbf{y}_0}{n^2}
\]

which converges to 0 because of Assumptions A1 and A5. For the final term in the numerator of the expression, \( \sqrt{n} (\bar{\tau}^2 - \bar{\tau}^2) \) converges in distribution to a central normal distribution, so in absolute value

\[
|\bar{\tau}^2 - \bar{\tau}^2| \delta' \mathbf{M} \delta / n = \frac{|\sqrt{n} (\bar{\tau}^2 - \bar{\tau}^2) | \delta' \mathbf{M} \delta / \sqrt{n}}{\delta' \mathbf{M} \delta / n}\leq |\sqrt{n} (\bar{\tau}^2 - \bar{\tau}^2)| \frac{\lambda_{\text{max}} \delta' \mathbf{M} \delta / n}{n}\]

which converges in probability to 0, completing the proof that \( (\hat{\tau} - \bar{\tau}) / \sqrt{\text{Var}(\hat{\tau})} \) converges to a standard normal distribution.

Appendix C

The dynamic randomization method of Section 5 that randomizes patients in a 2 to 1 ratio proceeds as follows. Determine the stratification variable values \( s_1 \) and \( s_2 \) for the current patient to be randomized and determine \( n_{10} \), \( n_{11} \), and \( n_{12} \) for previously randomized patients, where, for example, \( n_{11} \) equals the number of the previously randomized patients with Stratum 1 level equal to the level \( s_1 \) of the current patient who were randomized to treatment level 1. Likewise, determine \( n_{20} \), \( n_{21} \), and \( n_{22} \). The assignment probabilities for the current patient are determined by the balance in these counts first for Stratum 1 and then for Stratum 2. If \( \max(n_{1j}) - \min(n_{1j}) \geq 1 \), then the treatment assignment probabilities depend on the pattern of the \( n_{10} \), \( n_{11} \), and \( n_{12} \) counts as follows:

- With three unique values, the probabilities are, respectively, 0.75, 0.15, and 0.10 for the treatments associated with the minimum, intermediate, and maximum treatment counts \( n_{1j} \).
- With two values at the minimum count and one at the maximum, the probabilities are 0.40 for each of the treatments associated with the minimum count and 0.2 for the treatment associated with the maximum count.
- With one value at the minimum count and two values at the maximum, the probabilities are 0.12 for each of the treatments associated with the maximum count and 0.76 for the treatment associated with the minimum count.

If \( \max(n_{1j}) - \min(n_{1j}) = 0 \) then a similar check on Stratum 2 counts is performed, with unequal treatment assignment probabilities as above if there is imbalance. If there is balance of counts for Stratum 2, then the treatment for the current patient is selected from 0, 1, and 2 with equal probability.

To convert from 1 to 1 to 1 randomization to 2 to 1 randomization, patients randomized to treatment 0 are given control treatment and patients randomized to treatments 1 or 2 are given experimental.

The probabilities in the above algorithm were selected to match the conditional treatment assignment probabilities of the stratified PBR in blocks of six in the following sense. With complete randomization, the probability of assignment to experimental treatment is \( 2/3 \) regardless of previous treatment assignments, so the deviation from the target unconditional probability of \( 2/3 \) is zero. At the other extreme is a deterministic scheme where a fixed one-third of the patients are given control treatment and a fixed two-thirds are given experimental, so the squared deviation from the target unconditional probability of \( 2/3 \) is \( (0 - 2/3)^2/3 + 2(1 - 2/3)^2/3 = 0.22 \).

In the example of Section 5 the mean squared deviation as estimated from the simulation is approximately 0.05 for both the stratified PBR and the sequential method described above.

**Acknowledgements**

I thank the reviewers for their helpful comments that led to substantial improvements in this manuscript.

**References**

1. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials* 1988; 9:327–344. DOI: 10.1016/0197-2456(88)90047-5.
2. Efron B. Forcing a sequential experiment to be balanced. *Biometrika* 1971; 58:403–417. DOI: 10.1093/biomet/58.3.403.
3. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975; 31:103–115.
4. Wei LJ. The application of an urn model to the design of sequential controlled trials. *Journal of the American Statistical Association* 1978; 73:559–563. DOI: 10.1080/01621459.1978.10480054.
5. Borm GF, Hoogendoorn EH, den Heijer M, Zielhuis GA. Sequential balancing: A simple method for treatment allocation in clinical trials. *Contemporary Clinical Trials* 2005; 26:637–645. DOI: 10.1016/j.cct.2005.09.002.
6. Han B, Enas NH, McEntegert D. Randomization by minimization for unbalanced treatment allocation. *Statistics in Medicine* 2009; 28:3329–3346.
7. Su Z. Balancing multiple baseline characteristics in randomized clinical trials. *Contemporary Clinical Trials* 2011; 32:547–550. DOI: 10.1016/j.cct.2011.03.004.
8. Kempthorne O. *The Design and Analysis of Experiments*. Krieger: NY, 1975.
9. Scheffe H. *The Analysis of Variance*. Wiley: NY, 1959.
10. Nelder JA. The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proceedings of the Royal Society of London A* 1965; 283:147–162. DOI: 10.1098/rspa.1965.0012.
11. Nelder JA. The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. *Proceedings of the Royal Society of London A* 1965; 283:63–178. DOI: 10.1098/rspa.1965.0013.
12. Smythe RT, Wei LJ. Significance tests with restricted randomization design. *Biometrika* 1983; 70:496–500. DOI: 10.1093/biomet/70.2.496.
13. Smith RL. Sequential treatment allocation using biased coin designs. *Journal of the Royal Statistical Society, B* 1984; 46:519–543.
14. Wei LJ, Smythe RT, Smith RL. K-treatment comparisons with restricted randomization rules in clinical trials. *The Annals of Statistics* 1986; 14:265–274. DOI: 10.1214/aos/1176349854.
15. Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* 2010; 97:347–360. DOI: 10.1093/biomet/asp014.
16. Rosenkranz GK. The impact of randomization on the analysis of clinical trials. *Statistics in Medicine* 2011; 30:3475–3487. DOI: 10.1002/sim.4376.
17. Wei LJ, Lachin JM. Properties of the urn randomization in clinical trials. *Controlled Clinical Trials* 1988; 9:345–364. DOI: 10.1016/0197-2456(88)90048-7.
18. Halpern J, Brown BW Jr. Sequential treatment allocation procedures in clinical trials—with particular attention to the analysis of results for the biased coin design. *Statistics in Medicine* 1986; 5:211–229. DOI: 10.1002/sim.4780050303.
19. Committee for Proprietary Medicinal Products. Points to consider on adjustment for baseline covariates. *Statistics in Medicine* 2004; 23:701–709. DOI: 10.1002/sim.1647.
20. Kuznetsova OM, Tyomofyeyev Y. Brick tunnel randomization for unequal allocation to two or more treatment groups. *Statistics in Medicine* 2011; 30:812–824. DOI: 10.1002/sim.4167.

21. Cox DR. A remark on randomization in clinical trials. *Utilitas Mathematica* 1982; 21:245–252.

22. McHugh R, Mats J. Post-stratification in the randomized clinical trial. *Biometrics* 1983; 39:217–225.

23. McEntegert DJ. The pursuit of balance using stratified and dynamic randomization techniques: An overview. *Drug Information Journal* 2003; 37:293–308.

24. Oden NL, McIntosh MJ. Exact moments and probabilities for Wei’s urn randomization model. *Statistics & Probability Letters* 2006; 76:1694–1700. DOI: 10.1016/j.spl.2006.04.012.

25. Markaryan T, Rosenberger WF. Exact properties of Efron’s biased coin randomization procedure. *The Annals of Statistics* 2010; 38:1546–1567. DOI: 10.1214/09-AOS758.

26. Hollander M, Peña E. Nonparametric tests under restricted treatment-assignment rules. *Journal of the American Statistical Association* 1988; 83:1144–1151. DOI: 10.1080/01621459.1988.10478712.

27. International Conference on Harmonisation: Guidance on statistical principles for clinical trials. *Federal Register* 1998; 63(179):49583–49598. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073137.pdf (Accessed 2 May 2012).

28. Lachin JM. Properties of simple randomization in clinical trials. *Controlled Clinical Trials* 1988; 9:312–326. DOI: 10.1016/0197-2456(88)90046-3.

29. Herrtier S, Gebshi V, Pillai P. Dynamic balancing randomization in controlled clinical trials. *Statistics in Medicine* 2005; 24:3729–3741. DOI: 10.1002/sim.2421.

30. Proschan M, Brittain E, Kammerman L. Minimize the use of minimization with unequal allocation. *Biometrics* 2011; 67:1135–1141. DOI: 10.1111/j.1541-0420.2010.01545.x.

31. Endocrinologic and Metabolic Drugs Advisory Committee, October 21, 2008 (Meeting ID: 2008-4389). Available from: http://www.fda.gov/ohrms/dockets/ac/cder08.html#EndocrinologicMetabolic (Accessed 2 May 2012).

32. Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*. Wiley: New York, 2002.

33. Simon R. Restricted randomization designs in clinical trials. *Biometrics* 1979; 35:503–512.

34. Kempthorne O. Why randomize? *Journal of Statistical Planning and Inference* 1977; 1:1–25. DOI: 10.1016/0378-3758(77)90002-7.

35. Lachin JM. Statistical properties of randomization in clinical trials. *Controlled Clinical Trials* 1988; 9:289–311. DOI: 10.1016/0197-2456(88)90045-1.

36. Gadbury GL. Randomization inference and bias in standard errors. *The American Statistician* 2001; 55:310–313. DOI: 10.1198/0003130013175327268.

37. Rosenblatt M. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences USA* 1956; 42:43–47.

38. Serfling RJ. Contributions to central limit theory for dependent variables. *The Annals of Mathematical Statistics* 1968; 39:1158–1175. DOI: 10.1214/aoms/1176698240.

39. Berk KN. A central limit theorem for m-dependent random variables with unbounded m. *The Annals of Probability* 1974; 1:352–354. DOI: 10.1214/aop/1176996992.