9

Covariate-Adaptive Randomization

In the previous chapters, we have assumed that a set of s strata is defined on the basis of one or more covariates and a separate randomization is performed within each stratum. An entirely different approach would be to determine the treatment assignment of a new subject so as to minimize the covariate imbalances within treatment groups. This approach has been called by various names in the literature: *adaptive stratification*, *dynamic allocation*, or *minimization*. When randomization is involved in the procedure, we prefer the term *covariate-adaptive randomization*.

9.1 EARLY WORK

Some of the early work in this area appeared in the 1970s in papers by some of the great researchers in clinical trials: Zelen, Pocock, Simon, Taves, and Wei. It appears that, with the exception of Wei (1978), which references the Pocock and Simon (1975) paper, these authors were unaware of the other authors' work, despite the similarity in intentions.[1]

Except for Zelen's approach, all of these procedures attempt to induce balance between treatment groups with respect to covariate margins and are therefore called *marginal approaches*. One can immediately see the benefit of such an approach if there are a large number of stratification variables. For example, if there are 8 binary covariates, there will be 256 strata, while the number of covariate margins is only 16. The marginal approach is not concerned with treatment balance in each stratum, but only across each covariate separately.

[1] All three papers were published in 1974 or 1975, so it is difficult to establish priority, but the submission dates put Taves at February 1973, Pocock and Simon at November 1973, and Zelen at November 1973. Both Zelen and Pocock were at SUNY Buffalo at the time.

Randomization in Clinical Trials: Theory and Practice, Second Edition.
William F. Rosenberger and John M. Lachin.

9.1.1 Zelen's rule

Zelen's (1974) rule uses a preassigned randomization sequence (which could be generated by complete randomization or some restricted randomization design) ignoring strata. Let $N_{ik}(n)$ be the number of patients in stratum $i = 1, \ldots, S$ on treatment $k = 1, 2$ $(1 = A, 2 = B)$. When patient $n + 1$ in stratum i is ready to be randomized, one computes $D_i(n) = N_{i1}(n) - N_{i2}(n)$. For an integer c, if $|D_i(n)| < c$, then the patient is randomized according to schedule; otherwise, the patient receives the opposite treatment. Zelen proposes $c = 2, 3$, or 4. He also proposes randomizing the value of c for each new patient.

9.1.2 The Pocock–Simon procedure

Pocock and Simon (1975) proposed a randomized version of Zelen's rule. Let $N_{ijk}(n), i = 1, \ldots, I, j = 1, \ldots, n_i, k = 1, 2$ $(1 = A, 2 = B)$, be the number of patients in category j of covariate i on treatment k after n patients have been randomized. (In our previous notation, $\prod_{i=1}^{I} n_i = S$ is the total number of strata in the trial.) Suppose the $(n + 1)$th patient to be randomized is a member of categories $r_1, \ldots, r_I$ of covariates $1, \ldots, I$. Again, we define D to be a difference metric; in this case, let $D_i(n) = N_{ir_i1}(n) - N_{ir_i2}(n)$. We then take a sum over weighted strata defined by $D(n) = \sum_{i=1}^{I} w_i D_i(n)$, where w_i are weights chosen depending on which covariates are deemed of greater importance. If $D(n)$ is less than 0, then the weighted difference measure indicates that B has been favored thus far for that set, $r_1, \ldots, r_I$, of strata and the patient $n + 1$ should be assigned with higher probability to treatment A, and vice versa, if $D(n)$ is greater than 0. Pocock and Simon suggest biasing a coin with

$$p = \frac{c^* + 1}{3} \tag{9.1}$$

and implementing the following rule: if $D(n) < 0$, assign the next patient to treatment A with probability p; if $D(n) > 0$, assign the next patient to treatment A with probability $1 - p$; and if $D(n) = 0$, assign the next patient to treatment A with probability $1/2$, where $c^* \in [1/2, 1]$.

Note that if $c^* = 1$, we have a rule very similar to Efron's biased coin design in Section 3.5. If $c^* = 2$, we have the deterministic *minimization method* proposed by Taves (1974). Many other rules could be considered, all derivatives of Zelen's rule and Taves's minimization method with a biased coin twist to give added randomization; Efron (1980) describes one such rule and applies it to a clinical trial in ovarian cancer.

Pocock and Simon generalize their covariate-adaptive randomization procedure for more than two treatments by considering a general metric $D_i^k(n), k = 1, \ldots, K$, which could be the standard deviation of the $N_{ir_ik}(n)$'s, and a weighted sum $D^k(n) = \sum_{i=1}^{I} w_i D_i^k(n)$. The D^k's are then ordered from smallest to largest, and a corresponding set of probabilities $p_1 \geq p_2 \geq \ldots \geq p_K$ is determined such that $\sum_{k=1}^{K} p_k = 1$. The value $p_k, k = 1, \ldots, K$, is then the probability that patient $n + 1$ with strata $r_1, \ldots, r_I$

will be assigned to treatment k. Pocock and Simon suggest the following functional form:

$$p_k = c^* - \frac{2(Kc^* - 1)}{K(K+1)}k,\ k = 1, \ldots, K. \tag{9.2}$$

Note that (9.2) reduces to (9.1) for $K = 2$.

9.1.3 Example: Adjuvant chemotherapy for locally invasive bladder cancer

In a multicenter clinical trial of patients with locally invasive bladder cancer (Stadler *et al.*, 1993), who have undergone a radical cystectomy with lymph node dissection, and whose tumors demonstrate p53 abnormalities, 190 patients were to be randomized to either adjuvant chemotherapy (95 patients) or routine follow-up (95 patients) after surgery. The primary outcome was time to recurrence. Because of the large number of stratification variables, it was decided that the Pocock–Simon procedure would be used (see Section 9.1.2). Stratification variables were age (dichotomized at 65 years), stage (dichotomous), grade (dichotomous), and p21 status (dichotomous). The value of p in (9.1) was set to 0.75 ($c^* = 1.25$). The trial was stopped for futility after 114 patients were randomized. All baseline covariates included in the randomization procedure were balanced, with a maximum imbalance between treatment groups of 6 in the over 65 years old category.

9.1.4 Wei's marginal urn design

Wei (1978) described the use of an urn model for covariate-adaptive randomization. When the number of covariates is such that the resulting number of strata is large and the stratum sizes are small, using a separate urn in each stratum can result in imbalances in treatment assignments within strata. Let n_i be the number of categories for the ith of I covariates considered jointly, such that there are $s = \prod_{i=1}^{I} n_i$ unique strata. Instead of using s urns, one for each unique stratum, Wei proposed using an urn for each category of each covariate, for a total of $\sum_{i=1}^{I} n_i$ urns. For a given new subject with covariate values $r_1, \ldots, r_I$, the treatment group imbalances within each of the corresponding urns are examined. The one with the greatest imbalance is used to generate the treatment assignment. A ball from that urn is chosen and then replaced. Then β balls representing the opposite treatment are added to the urns corresponding to that patient's covariate values. Wei called this a *marginal urn*.

9.1.5 Is marginal balance sufficient?

Marginal balance may seem to be inadequate compared with balance in each stratum, but this is not necessarily so. Wei (1978), in a remarkably cogent short proof, shows that, if there is no interaction among the covariates nor between the treatment effect and covariates, marginal balance is sufficient. Consider a linear model where Y is

the response of a patient to treatment $k = 1, \ldots, K$ that has covariate category r_i for factor $i, i = 1, \ldots, I$. We can write the standard ANOVA model as

$$Y = \mu + \sum_{i=1}^{I} \beta_{ij_i} + t_k + \epsilon, \tag{9.3}$$

where t_k and β_{ij_i} are treatment and covariate category effects, μ is a constant, and ϵ is a $N(0, \sigma^2)$ error term, with the usual constraint that $\sum_{j=1}^{n_i} \beta_{ij} = 0, i = 1, \ldots, I$.

Suppose we wish to estimate a treatment contrast $\sum_{k=1}^{K} c_k t_k$ for a K-dimensional contrast vector $\boldsymbol{c}$. If we have equal numbers of patients in each stratum, then the best linear unbiased estimator is $\sum_{k=1}^{K} c_k \bar{Y}_k$, where $\bar{Y}_k$ is the observed mean of Y in treatment group k. If we do not have equal numbers in each stratum, then the mean squared error of $\sum_{k=1}^{K} c_k \bar{Y}_k$ can be computed as

$$\text{MSE}\left(\sum_{k=1}^{K} c_k \bar{Y}_k\right) = \sum_{k=1}^{K} \frac{c_k^2 \sigma^2}{n_k} + \left\{\sum_{k=1}^{K} c_k \left(\sum_{i=1}^{I} \sum_{j_i=1}^{n_i} \frac{n_{ij_ik}}{n_k} \beta_{ij_i}\right)\right\}^2, \tag{9.4}$$

where n_k is the number of patients assigned to treatment k and n_{ij_ik} is the number of patients assigned to k in category j_i of covariate i. We can see that if n_{ij_ik}/n_k is constant for $k = 1, \ldots, K$, which means equal allocation proportions within covariate margins across treatment arms, then the bias term is 0. Hence, marginal balance is sufficient to ensure that the treatment contrast estimator is unbiased.

If we have interactions, either among covariates or among covariates and treatment, then one needs to have balance within all covariate-by-treatment strata. In the next section, we describe some covariate-adaptive procedures that attempt to do this.

9.1.6 Is randomization necessary?

Taves's (1974) minimization procedure is not randomized, and there was never any intention in the original paper of using any form of randomization. Rather minimization was proposed as an alternative to randomization. Hidden in the discussion of that paper is a very cogent heuristic argument about why minimization can never result in a distribution of unknown covariates that is worse than from a randomized clinical trial.

> Consider an important but neglected or unknown variate: either it does or does not have some degree of correlation with the variates used for minimizing. If it does have some correlation, either positive or negative, it will tend to be distributed like the variates with which it is correlated. If the variate is not correlated it will be placed in the two groups independently of anything done to the minimized variates, i.e., placed randomly. To have the full range of a random distribution it must be possible to have extremely skewed combinations. Since minimizing works against these skewed combinations theoretically–and practically, … , the odds for skewed combinations with minimization will never be greater than with randomization.

Aickin (2001) has argued that randomization is not needed in covariate-adaptive procedures because the covariates themselves are random, leading to randomness in the treatment assignments. In addition, for large multicenter clinical trials with centralized randomization and minimization, he believes the likelihood of correctly guessing the treatment assignment of the next patient would be minimal, thus eliminating concerns about selection bias. Thirty years after proposing minimization, Taves (2004) still supports the view that randomization is not necessary, writing:

> I hope the day is not too far distant when we look back on the current belief that randomization is essential to good clinical trial design and realize that it was … "credulous idolatry".

Berger (2010) and Senn believe that selection bias and predictability are the major concerns with deterministic minimization. Barbáchano, Coad, and Robinson (2008) show that, under deterministic minimization, scant information about the last treatment assignment can lead to more than 60 percent correct guesses, and this increases to 80 percent if the last two assignments are known. Berger (2010) states:

> … The idea behind minimization is brilliant, but its failure is the flip side of the same coin, and cannot be separated from its benefit. The claim to fame is that subjects are allocated not randomly but rather deterministically, so as to minimize an imbalance function (hence the name). It is this, the heart and soul of the method (and not some tangential aspect), that leads to its downfall . . . . If the odds of each treatment come close enough to 50% so as to truly take a bite out of selection bias, then the baby is lost with the bath water, and minimization no longer does what it purports to do. There is simply no way around this.

Senn (2013) writes:

> … The advantage of minimisation compared with randomisation is a potential gain in efficiency. However, it is also known that the gain is very small and dependent on a number of assumptions. The disadvantages include an increase in predictability with its attendant disadvantages (loss of blinding vulnerability to manipulation) and increased uncertainty about precision itself. My judgement is that this particular game is not worth the candle, and I strongly prefer randomisation to minimisation.

9.2 MORE RECENT COVARIATE-ADAPTIVE RANDOMIZATION PROCEDURES

In this section, we review more recent covariate-adaptive randomization procedures. We first describe methods to balance within strata, marginally, and overall for a set of discrete strata. We then describe methods that treat covariates as continuous.

9.2.1 Balancing within strata

One of the criticisms of the methods described thus far based on marginal balance across covariates is that, within individual strata, imbalances can result. While this

has been shown to not be problematic if there are no interactions, some authors have developed techniques to balance within strata as well as marginally. Signorini *et al.* (1993) developed *dynamic balanced randomization* based on Soares and Wu's big stick design. First, a hierarchical ordering of the stratification structure is defined. For example, a clinical trial with two covariates, gender and clinic, say, the hierarchy may be gender within clinic ($i = 1$), then clinic ($i = 2$), then the overall trial ($i = 3$). Each of these I hierarchical stratification levels has an associated imbalance intolerance, $b_i, i = 1, \ldots, I$. Randomization is conducted with probability $1/2$ unless one of the imbalance intolerance bounds is reached, where b_1 is examined first, then b_2, until b_I. If an imbalance intolerance bound is attained, then the treatment assigned less often is allocated with probability 1.

In our example with two covariates, and two clinics, if the next patient to be randomized is a female in clinic 2, then the difference between treatment assignments among females in clinic 2 is compared to b_1. If it exceeds the threshold, then the treatment assigned less often so far is assigned with probability 1. Otherwise, one examines the imbalance in clinic 2 and compares it to the threshold. If it exceeds the threshold, then the treatment assigned less often so far is assigned with probability 1. Otherwise, one looks at the treatment imbalance overall in the trial and compares it to the threshold b_3. If it exceeds the threshold, then the treatment assigned less often so far is assigned with probability 1. Otherwise, the patient is randomized with probability $1/2$.

An alternative method is to create a compound distance metric that is a weighted sum of the imbalance within strata, within margins of covariates, and overall (Hu and Hu, 2012). Suppose the $(n + 1)$th patient enters the trial with covariate values $(r_1, \ldots, r_I)$ of covariates $(1, \ldots, I)$. For each treatment k that could be assigned, say $k = A, B$, we compute three sets of difference metrics. Let $D_{\text{stratum},k}(n)$ be the squared difference of patient numbers on treatment A versus treatment B among the first n patients whose covariate profile is $(r_1, \ldots, r_I)$. Let $D_{ik}(n)$ be the squared difference of patient numbers among the first n patients whose ith covariate value is r_i. Let $D_k(n)$ be the squared difference of patient numbers in the trial after n patients. Define a set of $I + 2$ nonnegative weights, $(w_{\text{margin},i}, i = 1, \ldots, I, w_{\text{trial}}, w_{\text{stratum}})$ that sum to 1. We compute an imbalance score, $G_k, k = A, B$ as follows:

$$G_k(n) = \sum_{i=1}^{I} w_{\text{margin},i} D_{ik}(n) + w_{\text{trial}} D_k(n) + w_{\text{stratum}} D_{\text{stratum},k}(n). \tag{9.5}$$

Efron's biased coin design is then used to assign the next treatment, depending on the value of $G_A(n) - G_B(n)$. The $n + 1$th patient is then assigned to treatment A with probability p if $G_A(n) - G_B(n) < 0$, with probability $1 - p$ if $G_A(n) - G_B(n) > 0$, and with probability $1/2$ if $G_A(n) - G_B(n) = 0$, where $p \in (1/2, 1)$.

9.2.2 Balancing with respect to continuous covariates

All of the methods discussed thus far have discretized covariates to create strata. Hu *et al.* (2015) review nine papers that attempt to balance continuous covariates

across treatments. Some of the procedures are designed to induce comparability with respect to means, medians, variances, or ranks. Hu, *et al.* mention that the ultimate goal would be to balance the actual distribution of the covariates across treatments. One method of doing this is to define a metric based on a comparison of the areas under the empirical distribution function among treatments. Another possible metric compares the kernel density estimators across treatments. All of these procedures rely on the same technique of defining a distance metric and then performing a biased coin randomization to favor the treatment that will promote greater balance. While feasible in practice, it is not clear that such procedures actually perform better than covariate-adaptive randomization on discretized covariates.

An alternative approach when there are continuous (and/or categorical) covariates is to find the treatment that maximizes the efficiency of estimating the treatment effect in a covariate-adjusted regression model, as described in the next section.

9.3 OPTIMAL DESIGN BASED ON A LINEAR MODEL

The covariate-adaptive randomization procedures in the preceding sections are arbitrary in the sense that they are developed intuitively rather than based on some optimal criterion. While one can simulate these procedures for different parameter values and find appropriate designs to fit certain criteria, none of the designs has been shown to be optimal. Instead of concerns about balance of treatment assignments across strata, one can take an entirely different approach and find an allocation rule that minimizes the variance of the estimated treatment effect in the presence of covariates. Such a rule would necessarily require the specification of a model linking the covariates and the treatment effect.

Begg and Iglewicz (1980) and Atkinson (1982) select a standard linear regression model. Here we follow Atkinson's development. We begin with the classical regression model, given by

$$E(Y_i) = \boldsymbol{x}_i'\boldsymbol{\beta}, i = 1, \ldots, n,$$

where the Y_i's are independent responses with $\text{Var}(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}$ and $\boldsymbol{x}_i$ includes a treatment indicator and selected covariates of interest. Then $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, where $\boldsymbol{X}'\boldsymbol{X}$ is the $p \times p$ dispersion matrix from n observations.

For the construction of an optimal design, we wish to find the n points of experimentation at which some function is optimized (in our case we will be finding the optimal sequence of n treatment assignments). The dispersion matrix evaluated at these n points is given by $\boldsymbol{M}(\xi_n) = \boldsymbol{X}'\boldsymbol{X}/n$, where ξ_n is the n-point design. It is convenient, instead of thinking of n points, to formulate the problem in terms of a measure ξ (which is, in this case, a frequency distribution) over a design region Ξ.

Since an important goal of clinical trials is to estimate a treatment effect, possibly adjusting for important covariates, Atkinson formulates the optimal design problem as a design that minimizes, in some sense, the variance of $\boldsymbol{A}'\boldsymbol{\beta}$, where $\boldsymbol{A}$ is an $s \times p$

matrix of contrasts, $s < p$. One possible criterion is Sibson's (1974) D_A-optimality that maximizes

$$|\boldsymbol{A}'\boldsymbol{M}^{-1}(\xi)\boldsymbol{A}|^{-1}. \tag{9.6}$$

Other criteria could also be applied. Atkinson compares the D_A criterion to standard *D-optimality*, which maximizes the log determinant of $\boldsymbol{M}$. Ball, Smith, and Verdinelli (1993) investigate the *Bayesian D-optimality criterion*, where a Bayesian prior distribution is assumed for $\boldsymbol{\beta}$, and the procedure maximizes the expectation (with respect to the prior distribution) of the log determinant of $\boldsymbol{M}$.

For any multivariable optimization problem, we compute the directional derivative of the criterion. In the case of the D_A criterion in (9.6), we can derive the directional (Frechet) derivative as

$$d_A(\boldsymbol{x}, \xi) = \boldsymbol{x}'\boldsymbol{M}^{-1}(\xi)\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{M}^{-1}(\xi)\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{M}^{-1}(\xi)\boldsymbol{x}.$$

By the classical equivalence theorem of Kiefer and Wolfowitz (1960), the optimal design ξ^* that maximizes the criterion (9.6) then satisfies the following equations:

$$\sup_{\boldsymbol{x}\in\Xi} d_A(\boldsymbol{x}, \xi) \leq \boldsymbol{s}, \forall \xi \in \Xi \tag{9.7}$$

and

$$\sup_{\boldsymbol{x}\in\Xi} d_A(\boldsymbol{x}, \xi^*) = s \tag{9.8}$$

(Kiefer and Wolfowitz (1960); see Atkinson and Donev (1992) for further details.)

In the model with covariates, we have

$$E(Y) = x_1\beta_1 + \boldsymbol{x}_2'\boldsymbol{\beta}_2,$$

where x_1 is the treatment indicator vector and $\boldsymbol{x}_2$ is a vector of important covariates. In this case, if we are interested in estimating the treatment effect in the presence of covariates, $\boldsymbol{A}' = [\boldsymbol{A}_1' : 0]$ with $\boldsymbol{A}_1$ identifying the treatment differences. This formulation can be simplified with two treatments, but the optimal design that satisfies (9.7) and (9.8) must be determined numerically.

Such a design is optimal for estimating linear contrasts of β, but the solution will provide only an allocation ratio to each of K treatments, without incorporating the sequential nature of a clinical trial. Assume that n patients have already been allocated, and the resulting n-point design is given by ξ_n. Atkinson proposes a sequential design that allocates the $(n + 1)$th patient to the treatment $k = 1, \ldots, K$ for which $d_A(k, \xi_n)$ is a maximum. However, this design is deterministic.

In order to randomize the allocation, Atkinson suggests biasing a coin with probabilities

$$p_k = \frac{d_A(k, \xi_n)}{\sum_{k=1}^{K} d_A(k, \xi_n)} \tag{9.9}$$

and allocating to treatment k with the corresponding probability. With two treatments, $k = 1, 2$ ($1 = A, 2 = B$), we have $s = 1, \boldsymbol{A}' = [-1, 1]$, and the probability of assigning treatment A is given by

$$p = \frac{d_A(1, \xi_n)}{d_A(1, \xi_n) + d_A(2, \xi_n)}. \tag{9.10}$$

With no covariates in the model, the model becomes $E(Y) = \beta_k, k = 1, 2$, and the equations in (9.7) and (9.8) can be solved analytically. We can write (9.10) as

$$p = \frac{\{N_B(n)\}^2}{\{N_A(n)\}^2 + \{N_B(n)\}^2}, \tag{9.11}$$

where $N_A(n)$ and $N_B(n)$ are the numbers of patients assigned to treatments A and B, respectively, through n patients (Problem 4.4). Note that this is the design in equation (3.25) with $\rho = 2$. In a similar vein, Ball, Smith, and Verdinelli (1993) and Atkinson (1998) investigate Bayesian optimality criteria.

While Atkinson's (1982) original approach relies on a homoscedastic linear model. In principle, one could extend these results to any generalized linear models, such as logistic regression or parametric survival models, but the optimal solution will depend on the unknown parameters of the model. There are three options for dealing with this problem. One can substitute a best guess of the parameter values to obtain a *locally optimal design*. Secondly, one can take a Bayesian approach by putting a prior distribution on the unknown parameters and obtain a *Bayesian optimal design*. Finally, one could substitute sequentially computed estimates of those parameters based on the response data accrued thus far. Such models are called *covariate-adjusted response-adaptive* (CARA) designs and will be discussed in the next chapter. (The topic is dealt with more extensively in Chapter 9 of Hu and Rosenberger (2006)).

9.4 THE TRADE-OFF AMONG BALANCE, EFFICIENCY, AND ETHICS

There is no doubt that minimization and covariate-adaptive randomization tend to balance a set of known covariates marginally across treatments and can thus improve the cosmetic appearance of comparisons of baseline characteristics (often found in Table 1 of clinical trials reports). This has been shown in countless simulation papers that will not be described here. But is balance on covariates really essential in the

valid comparison of treatments in a clinical trial, or is its primary importance only cosmetic? This has been the subject of much controversy. Certainly if the patient cohorts on different treatment arms have vastly different characteristics, the trial can be less convincing. While covariate-adjusted treatment effects can be estimated using post-hoc regression analysis, many would argue that heterogeneity should be eliminated entirely, through either stratification or covariate-adaptive randomization. The same arguments used for and against stratification, given in Chapter 7, apply to covariate-adaptive randomization.

Rosenberger and Sverdlov (2008) describe the problem in terms of a trade-off among competing objectives of the clinical trials, shown in Figure 9.1. As Atkinson (1982) showed, in the event that the treatment effect arises from a homogeneous linear regression model, the optimal design for efficient estimation of the treatment contrast is one that balances marginally over covariates. However, this model is not applicable for nonlinear models with heterogeneous variances, such as logistic regression or survival models (e.g., Begg and Kalish, 1984; Kalish and Harrington, 1988). In these cases, balancing marginally over covariates can actually reduce the efficiency of estimation. These issues have been hotly debated in the literature; see Atkinson (1999) and the ensuing discussion.

The third component in Figure 9.1 that has yet to be mentioned is ethical considerations. While balance may be more efficient in limited cases, balanced allocation can lead to more patients being assigned to the inferior treatment. This may be problematic, particularly in trials where the outcome is grave. We will discuss this issue further in the next chapter, along with randomization methods that may mitigate the problem. Rosenberger and Sverdlov (2008) give several examples where balance in treatment numbers and balance marginally over covariates are less efficient and less ethically appealing than imbalances. They agree with Senn that "… cosmetic balance, while psychologically reassuring, should not be the goal if power or efficiency is lost in the process of forcing balance."

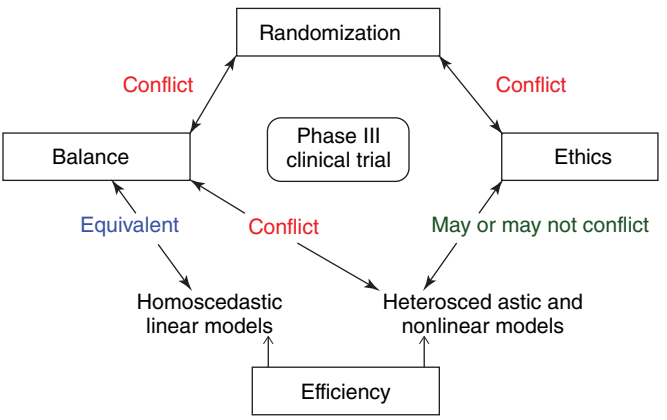

Fig. 9.1 *Multiple objectives of a phase III clinical trial. Source: Rosenberger and Sverdlov, 2008, p. 411. Reproduced with permission of Institute of Mathematical Statistics.*

9.5 INFERENCE FOR COVARIATE-ADAPTIVE RANDOMIZATION

In this section, we describe techniques to analyze data following covariate-adaptive randomization. We begin with model-based inference and then address randomization-based inference.

9.5.1 Model-based inference

Shao, Yu, and Zhong (2010) derive a sufficient condition for a hypothesis test to be valid under covariate-adaptive randomization when a model is specified among the responses and a set of covariates. Let $(\mathcal{T}, \mathcal{Y}, \mathcal{X})$ be the set of treatment assignments, responses, and covariates, respectively, for n patients. A hypothesis test S is *valid* for a significance level α if, under H_0,

$$\lim_{n\to\infty} P_{\mathcal{Y},\mathcal{T}}(|S| > c_\alpha|\mathcal{X}) \leq \alpha, \tag{9.12}$$

with equality holding for at least some cases, and $P_{\mathcal{Y},\mathcal{T}}(\cdot|\mathcal{X})$ denotes a probability calculated with respect to the conditional distribution of $(\mathcal{Y}, \mathcal{T})$, given $\mathcal{X}$. If theallocation is fixed, so that $\mathcal{T}$ is nonrandom, and a regression model between $\mathcal{Y}$ and $\mathcal{X}$ is correctly specified, then

$$\lim_{n\to\infty} P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}) \leq \alpha, \tag{9.13}$$

so that S is valid under any deterministic allocation. For example, if $E(Y_i|X_i)$ is linear in X_i, then the one-way analysis of covariance test is valid under any fixed allocation, provided $E(Y_i|X_i)$ is correctly specified. If we now incorporate randomization, an application of the dominated convergence theorem gives the equality

$$\lim_{n\to\infty} P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}, \mathcal{T}) = E_{\mathcal{T}}\left\{\lim_{n\to\infty} P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}, \mathcal{T})|\,\mathcal{X}\right\}.$$

Hence, a sufficient condition for a test S satisfying (9.13) to be valid is

$$P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}, \mathcal{T}) = P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}), \tag{9.14}$$

or that $\mathcal{T}$ and $\mathcal{Y}$ are conditionally independent, given $\mathcal{X}$. This condition holds under complete randomization and under any restricted randomization procedure. However, it does not necessarily hold under covariate-adaptiverandomization.

However, under covariate-adaptive randomization, if the set of covariates used in the design, $\mathcal{Z}$ is a function of $\mathcal{X}$, the test is still valid. This is because

$$\begin{aligned} P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}, \mathcal{T}) &= P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}, \mathcal{Z}, \mathcal{T}) \\ &= P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}, \mathcal{Z}) \\ &= P_{\mathcal{Y}}(|S| > c_\alpha|\mathcal{X}), \end{aligned}$$

as $\mathcal{T}$ and $\mathcal{Y}$ are conditionally independent, given $(\mathcal{X}, \mathcal{Z})$. Hence, for valid inference following covariate-adaptive randomization, the covariates $\mathcal{Z}$ must be a function of the model covariates $\mathcal{X}$ and the model must be correctly specified. Forsythe (1987) among others have recommended that all covariates used in the covariate-adaptive randomization should also be used in the analysis, and hypothesis tests will be valid if this is followed, provided the model for analysis is correctly specified.

This result is sufficient, but not necessary. A number of authors have shown simulation evidence that hypothesis tests can be conservative if $\mathcal{Z}$ is left out of the analysis (e.g., Birkett, 1985; Forsythe, 1987; Barbáchano and Coad, 2013). Shao, Yu, and Zhong (2010) give a theoretical justification of this phenomenon. Consider the following linear model linking $\mathcal{Y}, \mathcal{T}$, and $\mathcal{Z}$:

$$Y_{ij} = \mu_j + \beta Z_i + \epsilon_{ij},\ i = 1, \ldots, n, j = A, B,$$

where the Z_i's are i.i.d. with finite second moment, μ_j and β are unknown parameters, and ϵ_{ij} are i.i.d random errors with mean zero, variance σ^2, independent of the Z_i's. They show that, unless $\beta = 0$ in the model, the t-test under covariate-adaptive randomization is conservative, due to inducedcovariance between the sample means $\bar{Y}_A$ and $\bar{Y}_B$; that is, there exists a constant α_0 such that, under H_0,

$$\lim_{n\to\infty} P_{\mathcal{Y},\mathcal{T}}(|S| > c_\alpha) \leq \alpha_0 < \alpha.$$

Senn (2004) gives a more philosophical view of why the treatment effect should be adjusted for the covariates used in the design:

> If an investigator uses [covariate-adaptive randomization], she or he is honour bound, in my opinion, as a very minimum, to adjust for the factors used to balance, since the fact that they are being used to balance is an implicit declaration that they have prognostic value. In the case of a linear model the standard error quoted will generally be too small if this is not done ....

In conclusion, valid inference under a population model can only be attained under covariate-adaptive randomization if (1) a covariate-adjusted analysis is used, adjusting for the variables used for the randomization and (2) the model is correctly specified. Usual population-based inference methods valid under fixed, complete, or restricted randomization will be conservative otherwise.

9.5.2 Randomization-based inference

Randomization-based inference is more appropriate than population-based inference (see Chapter 6), and the technique of re-randomization for the Pocock–Simon method was described by Simon (1979, p. 508):

> It is possible, though cumbersome, to perform the appropriate randomization test generated by a nondeterministic adaptive stratification design. One assumes that the patient responses, covariate values, and sequence of patient arrivals are all

> fixed. One then simulates on a computer the assignment of treatments to patients using the [Pocock–Simon procedure] and the treatment assignment probabilities actually employed. Replication of the simulation generates the approximate null distribution of the test statistic adopted, and the significance level. One need not make the questionable assumption that the sequence of patient arrivals is random.

Computation of a re-randomization test, while fixing the covariate values and responses, is no longer "cumbersome" due to today's computational resources, and the establishment of a randomization p-value follows the procedure described in Section 6.9. However, the randomization distribution is completely determined by the random sequences generated by the biased coin; with deterministic minimization, there is no randomization distribution (cf., Kuznetsova, 2010). Hasegawa and Tango (2009) show that a randomization test of the difference of means is equivalent to the analysis of covariance when there is marginal balance in covariates. In that case, the randomization test is more powerful and less conservative than the t-test.

Proschan, Brittain, and Kammerman (2011) and Simon and Simon (2011) show that the re-randomization test preserves the type I error rate in the event of time trends. While, in principle, unequal allocation can be targeted using covariate-adaptive randomization, randomization-based hypothesis tests can be subject to a serious loss of power compared to tests under normal theory in a population model (cf., Proschan, Brittain, and Kammerman, 2011; Kuznetsova and Tymofyeyev, 2013).

9.6 CONCLUSIONS

Covariate-adaptive randomization can be an effective technique to force balance marginally on a set of known covariates. Whether marginal balance is essential in most clinical trials has not been resolved. In many cases, balance may be cosmetic and may not improve efficiency of estimation, particularly in the context of heteroscedastic, nonlinear models. Another approach is based on the theory of optimal designs; while more complicated to implement, it attempts to maximize the efficiency of estimation of the treatment effect. In any event, we favor those procedures that are randomized, to mitigate selection bias and to provide a basis for inference. These include the Pocock–Simon procedure and its variants, as well as Atkinson's biased coin procedure in (9.9). If such a procedure is used, it is important to adjust for the covariates in any population-based analysis. However, the advantages of randomization-based inference in such a context should be clear: the type I error rate is preserved in the presence of time trends, and valid inference does not depend on a correctly specified model.

There is an important logistical distinction between restricted and stratified randomization and covariate-adaptive randomization. For restricted and stratified randomization, the entire randomization sequence can be generated in advance, which aids in prepackaging treatments. For covariate-adaptive randomization, the randomization of each patient is not available until the patient's covariate profile is known.

9.7 PROBLEMS

9.1 Consider a two-arm clinical trial in which there are three baseline covariates over which balance is sought: (Z_1, Z_2, Z_3) representing gender, age, and cholesterol level. Assume that $Z_{ij}, i = 1, \ldots, n, j = 1, 2, 3$ are independently distributed as Bernoulli$(1/2)$, discrete uniform $[18, 75]$, and normal$(\mu = 200, \sigma = 30)$, respectively. Furthermore, assume that Z_2 and Z_3 are discretized into 4 and 3 levels, respectively:

$$\tilde{Z}_2 = \begin{cases} 1 \text{ (young)}, & \text{if } Z_2 \in [18, 30]; \\ 2 \text{ (older young)}, & \text{if } Z_2 \in [31, 45]; \\ 3 \text{ (middle age)}, & \text{if } Z_2 \in [46, 59]; \\ 4 \text{ (elderly)}, & \text{if } Z_2 \in [60, 75], \end{cases}$$

and

$$\tilde{Z}_3 = \begin{cases} 1 \text{ (normal level)}, & \text{if } Z_3 < 200; \\ 2 \text{ (borderline risk)}, & \text{if } Z_3 \in [200, 240); \\ 3 \text{ (high risk)}, & \text{if } Z_3 \geq 240. \end{cases}$$

Conduct a simulation study to describe the operating characteristics of the Pocock–Simon procedure in terms of balance (overall and within-covariate levels) and randomness for various levels of p.

9.2 Consider the model in (9.3).

a. Show that, if there are equal numbers in each stratum, the best linear unbiased estimator (BLUE) of $\sum_{k=1}^{K} c_k t_k$, for a K-dimensional contrast vector $\boldsymbol{c}$, is $\sum_{k=1}^{K} c_k \bar{Y}_k$, where $\bar{Y}_k$ is the observed mean of Y in treatment group k.
b. Derive equation (9.4).

9.3 Consider the covariate-adaptive randomization procedure based on the imbalance score, described in (9.5). Suppose there are $J = 2$ covariates: clinic (1, 2, 3) and gender (male, female), and 100 patients have already been randomized according to the following table:

Gender	Treatment	Clinic 1	Clinic 2	Clinic 3
Male	*A*	11	7	6
	B	11	7	7
Female	*A*	9	9	8
	B	11	8	7

Assume that weights are $(w_{\text{trial}}, w_{\text{margin},1}, w_{\text{margin},2}, w_{\text{stratum}} = (0.2, 0.2, 0.2, 0.4)$, and $p = 0.85$. Compute the probability that patient 101, a female in clinic 1, will be randomized to treatment *A*. (Hu *et al.*, 2015.)

9.4 [*Maximum entropy constrained balance*, Klotz (1978)] Consider a trial with $K \geq 2$ treatments. Let $B_i \geq 0$ be some measure of "overall covariate imbalance" from assigning the new patient to treatment i and $\mathbf{B} = (B_1, \ldots, B_K)$. Let $\mathbf{P} = (P_1, \ldots, P_K)$ be a vector of randomization probabilities to treatments $1, \ldots, K$. The expected covariate imbalance is

$$E(\mathbf{B}) = \sum_{i=1}^{K} B_i P_i.$$

One measure of randomness of treatment assignments is *entropy*, defined as

$$H(\mathbf{P}) = -\sum_{i=1}^{K} P_i \log \; P_i.$$

Consider the following optimization problem:

$$\begin{array}{ll} \max\limits_{\boldsymbol{P}} & H(\boldsymbol{P}) \\ \text{s.t.} & E(\boldsymbol{B}) \leq \eta B_{(1)} + (1-\eta)\bar{B} \\ \text{and} & 0 \leq P_i \leq 1, \; \sum\limits_{i=1}^{K} P_i = 1. \end{array} \tag{9.15}$$

In (9.15), $B_{(1)} = \min_{1 \leq i \leq K} B_i$, $\bar{B} = K^{-1} \sum_{i=1}^{K} B_i$, and $0 \leq \eta \leq 1$ is a user-specified constant that determines trade-off between complete randomization ($\eta = 0$) and strict balance ($\eta = 1$).

Using Lagrange multiplier optimization, show that the optimal randomization probabilities are as follows:

$$P_i = \frac{e^{-\lambda B_i}}{\sum_{k=1}^{K} e^{-\lambda B_k}}, \quad i = 1, \ldots, K,$$

where $\lambda = \lambda(\eta)$ is chosen such that $\sum_{i=1}^{K} B_i P_i = \eta B_{(1)} + (1-\eta)\bar{B}$.

9.5 [*Minimum quadratic distance constrained balance*, Titterington (1983)] Let $\boldsymbol{P} = (P_1, \ldots, P_K)$ and $\boldsymbol{P}_B = (K^{-1}, \ldots, K^{-1})$. The quadratic distance between $\boldsymbol{P}$ and $\boldsymbol{P}_B$ is

$$\delta(\boldsymbol{P}, \boldsymbol{P}_B) = \sum_{i=1}^{K} (P_i - K^{-1})^2.$$

Consider the following optimization problem:

$$\begin{array}{ll} \min\limits_{\boldsymbol{P}} & \delta(\boldsymbol{P}, \boldsymbol{P}_B) \\ \text{s.t.} & E(\mathbf{B}) \leq \eta B_{(1)} + (1-\eta)\bar{B} \\ \text{and} & 0 \leq P_i \leq 1, \; \sum\limits_{i=1}^{K} P_i = 1. \end{array} \tag{9.16}$$

a. Using Lagrange multiplier optimization, show that the optimal randomization probabilities are as follows:

$$P_i = \frac{1}{K} + \eta \frac{(B_{(1)} - \bar{B})}{\sum_{k=1}^{K} (B_k - \bar{B})^2}, i = 1, \ldots, K. \tag{9.17}$$

b. Show that when $K = 2$, (9.17) simplifies to

$$P_1 = (1 + \eta)/2,\ P_2 = (1 - \eta)/2 \quad \text{if } B_1 < B_2;$$
$$P_1 = (1 - \eta)/2,\ P_2 = (1 + \eta)/2 \quad \text{if } B_1 > B_2,$$

which is equivalent to Efron's biased coin design with coin bias $(1 + \eta)/2$.

9.6 Derive equation (9.11).

9.7 Consider a classical linear regression model with constant variance:

$$\boldsymbol{Y}_n = \boldsymbol{Z}_n \boldsymbol{\beta} + \alpha \boldsymbol{t}_n + \epsilon,\ \ \epsilon \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$

where $\boldsymbol{Y}_n$ is a vector of responses from n patients in the trial, α is the treatment effect, $\boldsymbol{\beta}$ is a $p \times 1$ vector of covariate effects, $\boldsymbol{t}_n$ is a $n \times 1$ vector of treatment assignments (1 for A; -1 for B), and $\boldsymbol{Z}_n$ is a $n \times p$ matrix of covariate values (including $\boldsymbol{1}$, the intercept). Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})'$ and $\boldsymbol{X}_n = (\boldsymbol{Z}_n, \boldsymbol{t}_n)$.

a. Verify that the BLUE of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}_n' \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n' \boldsymbol{Y}_n$ with $\text{Var}(\hat{\boldsymbol{\theta}}) = \sigma^2 (\boldsymbol{X}_n' \boldsymbol{X}_n)^{-1}$.
b. Show that $\text{Var}(\hat{\alpha}) = \sigma^2 \{n - \boldsymbol{t}_n' \boldsymbol{Z}_n (\boldsymbol{Z}_n' \boldsymbol{Z}_n)^{-1} \boldsymbol{Z}_n' \boldsymbol{t}_n\}^{-1}$ and $\text{Var}(\hat{\alpha})$ is minimized when $\boldsymbol{t}_n' \boldsymbol{Z}_n = \boldsymbol{0}$, which means that $\boldsymbol{t}_n$ is orthogonal to the columns of $\boldsymbol{Z}_n$ ($\boldsymbol{t}_n' \boldsymbol{1} = 0$ means balance in treatment totals, and $\boldsymbol{t}_n' \boldsymbol{z}_k$ means balance over the kth covariate).

9.8 Consider a linear regression model

$$Y = \mu_A t + \mu_B (1 - t) + \beta z + \epsilon,$$

where $t = 1(0)$ if treatment is $A(B)$, $z = 1(0)$, if a subject is male (female), and $\epsilon \sim N(0, 1)$. Based on a sample of n subjects, let n_{1k} be the number of males assigned to treatment $k = A, B$ and n_{0k} be the number of females assigned to treatment $k = A, B$ such that $n_A = n_{0A} + n_{1A}$, $n_B = n_{0B} + n_{1B}$, and $n = n_A + n_B$.

a. Show that for model (9.18),

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} n_A & 0 & n_{1A} \\ 0 & n_B & n_{1B} \\ n_{1A} & n_{1B} & n_1 \end{pmatrix},$$

b. For Atkinson's (1982) approach, we take the directional derivative

$$d(\boldsymbol{x}_{n+1}, \xi_n) = \boldsymbol{x}'_{n+1}\boldsymbol{M}^{-1}(\xi_n)\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{M}^{-1}(\xi_n)\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{M}^{-1}(\xi_n)\boldsymbol{x}_{n+1},$$

where $\boldsymbol{M}^{-1}(\xi_n) = n(\boldsymbol{X}'\boldsymbol{X})^{-1}$, $\boldsymbol{A} = (1, -1, 0)$ and

$$\boldsymbol{x}'_{n+1} = \begin{cases} (1, 0, z), & \text{if treatment} = A; \\ (0, 1, z), & \text{if treatment} = B, \end{cases}$$

where $z = 1$ if $(n + 1)$th subject is male and $z = 0$ if the subject is female. Show that

$$d(\boldsymbol{x}_{n+1}, \xi_n) = \begin{cases} cn_{1B}^2, & \text{if} \boldsymbol{x}'_{n+1} = (1, 0, 1) \text{ (treatment} = A, \text{ male)}; \\ cn_{1A}^2, & \text{if} \boldsymbol{x}'_{n+1} = (0, 1, 1) \text{ (treatment} = B, \text{ male)}; \\ cn_{0A}^2, & \text{if} \boldsymbol{x}'_{n+1} = (1, 0, 0) \text{ (treatment} = A, \text{ female)}; \\ cn_{0B}^2, & \text{if} \boldsymbol{x}'_{n+1} = (0, 1, 0) \text{ (treatment} = B, \text{ female)}, \end{cases}$$

where c depends on n_{ik}, $i = 0, 1$, $k = A, B$.

c. Hence, show that Atkinson's randomization procedure in this case is:

$$\phi_{n+1} = \begin{cases} \frac{n_{1B}^2}{n_{1A}^2 + n_{1B}^2}, & \text{if the } (n + 1)\text{th patient is male;} \\ \frac{n_{0B}^2}{n_{0A}^2 + n_{0B}^2}, & \text{if the } (n + 1)\text{th patient is female.} \end{cases}$$

9.8 REFERENCES

Aickin, M. (2001). Randomization, balance and the validity and efficiency of design-adaptive allocation methods. *Journal of Statistical Planning and Inference* **94** 97–119.

Atkinson, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67.

Atkinson, A. C. (1998). Bayesian and other biased-coin designs for sequential clinical trials. *Tatra Mountains Mathematical Publications* **17** 133–139.

Atkinson, A. C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statistics in Medicine* **18** 1741–1752.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Clarendon Press, Oxford.

Ball, F. G., Smith, A. F. M., and Verdinelli, I. (1993). Biased coin designs with Bayesian bias. *Journal of Statistical Planning and Inference* **34** 403–421.

Barbáchano, Y. and Coad, D. S. (2013). Inference following designs which adjust for imbalances in prognostic factors. *Clinical Trials* **10** 540–551.

Barbáchano, Y., Coad, D. S., and Robinson, D. R. (2008). Predictability of designs which adjust for imbalances in prognostic factors. *Journal of Statistical Planning and Inference* **138** 756–767.

Begg, C. B. and Iglewicz, B. (1980). A treatment allocation procedure for sequential clinical trials. *Biometrics* **36** 81–90.

Begg, C. B. and Kalish, L. (1984). Treatment allocation in sequential clinical trials: the logistic model. *Biometrics* **40** 409–420.

BERGER, V. W. (2010). Minimization, by its nature, precludes allocation concealment, and invited selection bias. *Contemporary Clinical Trials* **31** 406–(letter to the editor).

BIRKETT, N. J. (1985). Adaptive allocation in randomized clinical trials. *Controlled Clinical Trials* **6** 146–155.

EFRON, B. (1980). Randomizing and balancing a complicated sequential experiment. In *Biostatistics Casebook* (MILLER, R. G., EFRON, B., BROWN, B. W., AND MOSES, L. E., eds.). John Wiley & Sons, Inc., New York, pp. 19–30.

FORSYTHE, A. B. (1987). Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics and Data Analysis* **5** 193–200.

HASEGAWA, T. AND TANGO, T. (2009). Permutation test following covariate-adaptive randomization in randomized clinical trials. *Journal of Biopharmaceutical Statistics* **19** 106–119.

HU, Y. AND HU, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Annals of Statistics* **40** 1794–1815.

HU, F., HU, Y., MA, Z., AND ROSENBERGER, W. F. (2015). Adaptive randomization for balancing over covariates. *WIRES Statistical Computing* **6** 288–303.

HU, F. AND ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley & Sons, Inc., New York.

KALISH, L. A. AND HARRINGTON, D. P. (1988). Efficiency of balanced treatment allocation for survival analysis. *Biometrics* **44** 815–821.

KIEFER, J. AND WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12** 363–366.

KLOTZ, J. H. (1978). Maximum entropy constrained balance randomization for clinical trials. *Biometrics* **34** 283–287.

KUZNETSOVA, O. M. (2010). On the second role of the random element in minimization. *Contemporary Clinical Trials* **31** 587–588.

KUZNETSOVA, O. M. AND TYMOFYEYEV, Y. (2013). Shift in re-randomization distribution with conditional randomization test. *Pharmaceutical Statistics* **12** 82–91.

POCOCK, S. J. AND SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** 103–115.

PROSCHAN, M., BRITTAIN, E., AND KAMMERMAN, L. (2011). Minimize the use of minimization with unequal allocation. *Biometrics* **67** 1135–1141.

ROSENBERGER, W. F. AND SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statistical Science* **23** 404–419.

SENN, S. (2004). Controversies concerning randomization and additivity in clinical trials. *Statistics in Medicine* **23** 3729–3753.

SENN, S. (2013). Seven myths of randomisation in clinical trials. *Statistics in Medicine* **32** 1439–1450.

SHAO, J., YU, X., AND ZHONG, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* **97** 347–360.

SIBSON, R. (1974). *D*-optimality and duality. In *Progress in Statistics* (GANI, J., SARKADI, K., AND VINCZE, J., eds.). North-Holland, Amsterdam.

SIGNORINI, D. F., LEUNG, O., SIMES, R. J., BELLER, F., AND GEBSKI, V. J. (1993). Dynamic balanced randomization for clinical trials. *Statistics in Medicine* **12** 2343–2350.

SIMON, R. (1979). Restricted randomization designs in clinical trials. *Biometrics* **35** 503–512.

SIMON, R. AND SIMON, N. R. (2011). Using randomization tests to preserve type I error with response adaptive and covariate adaptive randomization. *Statistics and Probability Letters* **81** 767–772.

Stadler, W. M., Lerner, S. P., Groshen, S., Stein, J. P., Shi, S.-R., Raghavan, D., Esrig, D., Steinberg, G., Wood, D., Klotz, L., Hall, C., Skinner, D. G., and Cote, R. J. (2011). Phase III study of molecularly targeted adjuvant therapy in locally advanced urothelial cancer of the bladder based on p53 status. *Journal of Clinical Oncology* **29** 3443–3449.

Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics* **15** 443–453.

Taves, D. R. (2004). Faulty assumptions in Atkinson's criteria for clinical trial design. *Journal of the Royal Statistical Society A* **167** 179–180 (letter to the editor).

Titterington, D. M. (1983). On constrained balance randomization for clinical trials. *Biometrics* **39** 1083–1086.

Wei, L. J. (1978). An application of an urn model to the design of sequential controlled clinical trials. *Journal of the American Statistical Association* **73** 559–563.

Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases* **28** 365–375.