

Chapter 14

Adaptive Randomization for Master Protocols in Precision Medicine



Jianchang Lin, Li An Lin, Veronica Bunn, and Rachael Liu

14.1 Introduction

In the era of precision medicine, especially in the oncology and hematology areas, there have been explosions in knowledge of the molecular profile of disease. With the genomic sequencing becoming more affordable, many tumors can now be classified from a molecular biology perspective, with different treatment options and tailored strategies for patients based on their tumor biomarker status. Under the drug development setting, new generation trials have emerged to target patient selection within any given tumor type based on specific underlying molecular and biologic characteristics: e.g. (1) ‘Basket trials’ usually are focused only on specific molecular aberrations, in several tumor types. (2) ‘Umbrella (or Platform) trials’ focus on drug development targeting several molecular subtypes in one tumor type. (3) ‘Adaptive enrichment strategies’ offer the potential to enrich for patients with a particular molecular feature that is predictive of benefit for the test treatment based on accumulating evidence from the trial. Among them, umbrella, basket and platform trials constitute a new generation of clinical trial design defined as master protocol, which allow for the study of multiple drugs, multiple diseases indications, or both within a single trial. These innovative approaches to clinical drug development have resulted in rapidly revolutionized methodologies, including adaptive randomization (Lin et al. 2016a, b, c), to conduct clinical trials in the setting of biomarkers and targeted therapies, whereas the traditional paradigm of treating

J. Lin (✉) · V. Bunn · R. Liu
Takeda Pharmaceuticals, Cambridge, MA, USA
e-mail: Jianchang.Lin@takeda.com; Veronica.Bunn@takeda.com; Yue.Liu@takeda.com

L. A. Lin (✉)
Merck & Co., Inc., Rahway, NJ, USA
e-mail: li.an.lin@merck.com

a very large number of unselected patients is increasingly less efficient, lacks cost effectiveness and is ethically challenging.

In the past few years, there have been a variety of thought-provoking next generation master protocols conducted multi-institutionally in oncology: specific recognized examples include I-SPY2, BATTLE, NCI MATCH, LUNG-MAP, ALCHEMIST and FOCUS4 (Renfro et al. 2016). As a change from traditional clinical trial design paradigms, statisticians have partnered with clinicians to become fully integrated in these clinical trials and make critical contributions for advancing therapeutic development in this era of molecular medicine. Meanwhile, the new development of immunotherapeutic agents and implementation of next-generation sequencing (NGS) also brings many new and exciting opportunities in the design of biomarker driven trials. From a clinical trial operational perspective, there are some logistical challenges to implementing these innovative designs, e.g. central assay testing, drug supply, multiple institutional collaboration, real time data collection and integrations. However, these additional efforts are all worthwhile given the substantial improvement of efficient medicine development, and most importantly, the benefit of the patients.

In general, the goals of randomized clinical trials are to effectively treat patients and differentiate treatment effects efficiently. On one hand, a clinical trial tries to discriminate the effects of different treatments quickly, so that patients outside of the trial will sooner benefit from the more efficacious treatment. For this purpose, patients' allocation should be (nearly) balanced across the comparative arms. On the other hand, each trial participant should be given the most effective treatment, and patients themselves also hope that they would be assigned to the arm that performs better. This often leads to an unbalanced allocation through adaptive randomization by equipping a better arm with a higher allocation probability (Berry et al. 2010). Therefore, randomized clinical trials need to strike a balance between individual and collective ethics.

During the study planning stage, key components of the protocol such as primary endpoint, key secondary endpoints, clinically meaningful treatment effect difference, and treatment effect variability are pre-specified. Participating investigators and sponsors then collect all data in electronic data capture (EDC) system and perform statistical analyses. The success of the study depends on the accuracy of the original design assumptions or sample size calculation. Adaptive Designs are a way to address uncertainty about design parameters assumptions made during the study planning stage. Adaptive Designs allow a review of accumulating data or patient information during a trial to possibly modify trial characteristics and to promote multiple experimental objectives, while protecting the study from bias and preserving inferential validity of the results. The flexibility can translate into a more efficient drug development process by reducing the number of patients enrolled. This flexibility also increases the probability of success of the trial answering the question of scientific interest (finding a significant treatment effect if one exists or stopping the study as early as possible if no treatment effect exists).

Adaptive Designs have received a great deal of attention in the statistical, pharmaceutical, and regulatory fields. The US Food and Drug Administration (FDA)

released a draft version of the “Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics” in 2010 (U.S. Food and Drug Administration 2010). The guidance defined an adaptive design as ‘a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study.’ The most common adaptive designs used in clinical trials include, but are not limited to, the following types: adaptive randomization design, seamless adaptive phase II/III design, adaptive dose-response design, biomarker adaptive design, adaptive treatment switching design, adaptive-hypothesis design, multiple arm adaptive design, group sequential design, sample size re-estimation design, et al (Kairalla et al. 2012).

14.2 Why Is Adaptive Randomization Important?

The design of any clinical trial starts with formulation of the study objectives. Most clinical trials are naturally multi-objective, and some of these objectives may compete. For example, one objective is to have sufficient power to test the primary study hypothesis, and consequently have sufficient sample size. However, cost considerations may preclude a large sample size, so the twin objectives of maximum power and minimum sample size directly compete. Other objectives may include minimizing exposure of patients to potentially toxic or ineffective treatments, which may compete with having sufficient numbers of patients on each treatment arm to conduct convincing treatment group comparisons. In the case of $K > 2$ treatments, where $(K - 1)$ experimental treatments are to be compared with the placebo group with respect to some primary outcome measure, the primary objective of the trial may be testing an overall hypothesis of homogeneity among the treatment effects, and a secondary objective may be performing all pairwise comparisons among the $(K - 1)$ experimental treatments versus placebo. Investigators may have an unequal interest in such comparisons. In addition to statistical aspects of a clinical trial design, there may be a strong desire to minimize exposure of patients to the less successful (or more harmful) treatment arms. Clearly, in these examples it is very difficult to find a single design criterion that would adequately describe all the objectives. Many of these objectives depend on model parameters that are unknown at the beginning of the trial. It is useful, and indeed sometimes imperative, to use accruing data during the trial to adaptively redesign the trial to achieve these objectives. These design considerations must be achieved without sacrificing the hallmark of the carefully conducted clinical trials—randomization—which protects the study from bias.

Once the study objectives are formally quantified and ranked in the order of their importance, the experimental design problem is to find a design that accommodates several selected design criteria. Frequently, the treatment allocations are unbalanced across treatment groups, and they depend on model parameters that are unknown a priori and must be calibrated through simulation. Adaptive randomization uses

accruing information in the trial to update randomization probabilities to target the allocation criteria. Hu and Rosenberger (2006) classify adaptive randomization into four major types:

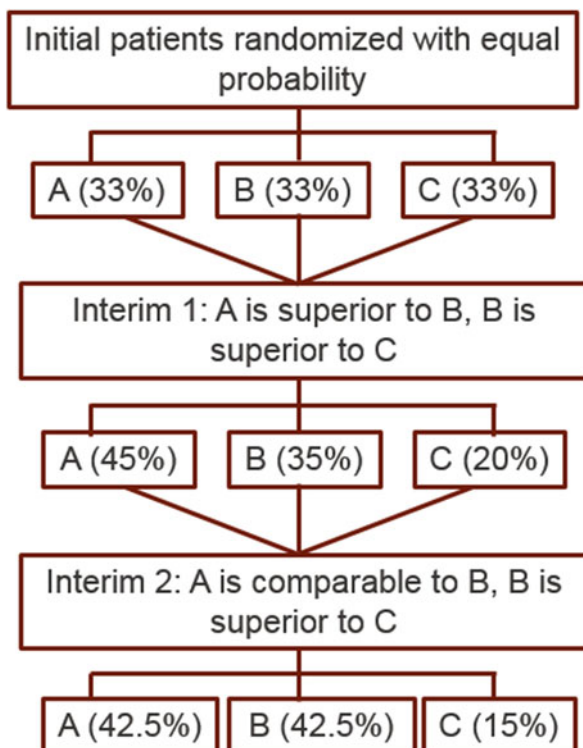
- *Restricted randomization*: a randomization procedure that uses past treatment assignments to select the probability of future treatment assignments, with the objective to balance numbers of subjects across treatment groups.
- *Covariate-adaptive randomization*: a randomization procedure that uses past treatment assignments and patient covariate values to select the probability of future treatment assignments, with the objective to balance treatment assignments within covariate profiles.
- *Response-adaptive randomization*: a randomization procedure that uses past treatment assignments and patient responses to select the probability of future treatment assignments, with the objective to maximize power or minimize expected treatment failures.
- *Covariate-adjusted response adaptive (CARA) randomization*: a combination of covariate-adaptive and response-adaptive randomization procedures.

A typical example of master protocol to screen three experimental treatments, A, B and C simultaneously is illustrated in Fig. 14.1. All patients recruited for the first stage of the trial are randomized to the treatment arms with equal probability. At each interim analysis, we update the Bayesian model used for setting the randomization probabilities. The proportion of patients that are randomized to better performing arms increases, and decreases to arms that are performing poorly.

14.3 Frequentist and Bayesian Approaches for Adaptive Randomization

The commonly used statistical approach to design and analyze clinical trials and other medical experiments is frequentist, while a Bayesian method provides an alternative approach. The Bayesian approach can be applied separately from frequentist methodology, as a supplement to it, or as a tool for designing efficient clinical trials that have good frequentist properties. The two approaches have rather different philosophies, although both use probability and deal with empirical evidence. Practitioners exposed to traditional, frequentist statistical methods appear to have been drawn to Bayesian approaches for three reasons (Ning and Huang 2010; Rosenberger et al. 2012; Thall and Wathen 2007; Yin et al. 2012; Yin 2013). One is that Bayesian approaches implemented with the majority of their informative content coming from the current available data, and not prior information, typically have good frequentist properties (e.g., low mean squared error (MSE) in repeated use). Second, these methods are now easily implemented in WINBUGS, OpenBUGS and other available MCMC software packages. These offer a convenient approach to hierarchical or random effect modeling, as regularly used in longitudinal data, frailty

Fig. 14.1 Overview of adaptive randomization design



model, spatial data, time series data, and a wide variety of other settings featuring interdependent data. Third, practitioners are attracted to the increased levels of flexibility and adaptivity offered by the Bayesian approach which allows for early stopping for efficacy, toxicity, and futility, as well as facilitates a straightforward solution to a great many other advanced problems such as dosing selection, adaptive randomization, equivalence testing, and others.

Flexibility is the major difference between Bayesian and frequentist methods, in both design and analysis. In the Bayesian approach, experiments can be altered midcourse, disparate sources of information can be combined, and expert opinion can play a role in inferences. An important property of Bayesian design is that it can utilize prior information and Bayesian updating while still maintaining good frequentist properties (power and Type I error). Another major difference is that the Bayesian approach can be decision-oriented, with experimental designs tailored to maximize objective functions, such as company profits or overall public health benefit. Overall, designing a clinical trial is a decision problem, such as therapy selection, resource allocation, early stop etc., which involves costs and benefits consideration. In the Bayesian approach, these costs and benefits can be well assessed for all possible scenarios of future observations. However, frequentism fits naturally with the regulatory “gate-keeping” role, through its insistence on

procedures that perform well in the long run regardless of the true state of nature. And indeed, frequentist operating characteristics (Type I and II error, power) are still very important to the FDA and other regulators.

14.4 Response-Adaptive Randomization

Response-adaptive randomization is one of the most important adaptive trial designs, in which the randomization ratio of patients assigned to the experimental treatment arm versus the control treatment arm changes overtime from 1:1 to a higher proportion of patients assigned to the arm that is performing better (Yuan et al. 2011). It is very attractive when ethical considerations or concerns make it potentially undesirable to have an equal number of patients assigned to each treatment arm. For the purpose of simplicity, suppose the trial objective is to compare treatments A and B. Patients are enrolled in sequential groups of size $\{N_j\}$, $j = 1, \dots, J$, where N_j is the sample size of group j . When planning the trial, researchers typically have limited prior information regarding the superiority or effectiveness of the experimental treatment arms. Therefore, at the beginning stage of the trial, for the first j groups, e.g. $j = 1$, patients are equally allocated to two treatments. The responses observed from these patients are utilized to update the allocation probability for subsequent groups of patients.

Let p_A be the response rate of treatment A and p_B be the response rate of treatment B. We set N to be the maximum sample size allowed for the trial and N_A (N_B) to be the maximum number of patient assigned to treatment A (B). We assign the first N_1 patients equally to two treatments (A, B) and observe the response Y_k ($k = A, B$). Assign p_k a noninformative prior of $\text{beta}(\alpha_k, \beta_k)$. If among n_k subjects treated in arm k , we observe y_k responses, then

$$Y_k \sim \text{binomial}(n_k, p_k) \quad (14.1)$$

and the posterior distribution of p_k is

$$p_k \mid \text{data} \sim \text{beta}(\alpha_k + x_k, \beta_k + n_k - x_k) \quad (14.2)$$

During the trial, we could continuously update the Bayesian posterior distribution of p_k , and allocate the next N_j patients to the k th treatment arm according to the posterior probability that treatment k is superior to all other treatment arms

$$\pi_k = \Pr(p_k = \max\{p_l, 1 \leq l \leq K\} \mid \text{data}) \quad (14.3)$$

One of the advantages of a Bayesian approach to inference is the increased flexibility to include sequential stopping compared to the more restrictive requirements of a classical group sequential approach in terms of number of interim analysis, stopping rules, etc. Noninformative stopping rules are irrelevant for Bayesian

inference. In other words, posterior inference remains unchanged regardless of why the trial was stopped. Several designs make use of this feature of Bayesian inference to introduce early stopping for futility and/or for efficacy.

- *Futility*: if $\Pr(p_k < p.min|data) > \theta_u$, where $p.min$ denotes the clinical minimum response rate, that is, there is strong evidence that treatment k is inferior to the clinical minimum response rate, we drop treatment arm k .
- *Superiority*: if $\Pr(p_k > p.target|data) > \theta_l$, where $p.target$ denotes the target response rate, that is, there is strong evidence that treatment k is superior to prespecified response rate, we terminate the trial early and claim the treatment k is *promise*.

At the end of the trial, if $\Pr(p_k > p.min|data) > \theta_l$, then treatment k is selected as the superior treatment. Otherwise, the trial is inconclusive. To achieve desirable operating characteristics (type 1 error and power), we use simulations to calibrate the pre-specified cut-off points θ_u , θ_l , and θ_t .

We conducted simulations to show the procedure for design parameters calibration. The patient allocation probability is determined by algorithm (14.3). The minimum allocation probability is 10% to ensure a reasonable probability of randomizing patients to each arm. The minimum clinical response rate ($p.min$) is 0.2 and the target response rate ($p.target$) is 0.4. In this trial, we set maximum sample sizes of 90 and maximum sample size of 30 per treatment arm. We equally assigned the first 15 patients to three treatments (A, B, or C) and started using the adaptive randomization at the 16th patient. The sequential group size is set as 10, so that the early stopping rule and allocation probability updating will act after 10 new patient's responses cumulated. Although the design allows continuous monitoring after every patient's response outcome becomes available, from the operational and computational point of view, it's more convenient to monitor the trial for early termination with a cohort size of 10. A total of 5,000 independent simulations were performed for each configuration.

In the first stage, we set $\theta_u = \theta_l = 1$, so that the trial would not be terminated early, to determine the threshold values of θ_t . we performed a series of simulation studies with different values of θ_t and compared the corresponding type 1 error rates and powers. Table 14.1 shows the simulation results. Similarly, we can obtain a set of values of θ_t that reached the desired power. The value of θ_t with type 1 error (defined as the selection probability of Arm A) close to 5% and a desired power (defined as the selection probability of Arms B or C) will be selected for the next stage selection.

Table 14.1 Type 1 error rates and power, without early termination

Arm (response rate)	θ_t			
	0.9	0.91	0.92	0.93
A (0.2)	0.07	0.065	0.056	0.049
B (0.4)	0.842	0.838	0.832	0.825
C (0.6)	0.998	0.994	0.992	0.989

Table 14.2 Type 1 error rates and power, with early termination

		θ_l			
θ_u	Arm	0.92	0.93	0.94	0.95
0.85	A	0.073	0.071	0.064	0.066
	B	0.828	0.819	0.84	0.817
	C	0.991	0.996	0.989	0.992
0.86	A	0.082	0.071	0.065	0.057
	B	0.821	0.824	0.814	0.839
	C	0.993	0.99	0.989	0.986
0.87	A	0.08	0.07	0.067	0.057
	B	0.822	0.819	0.838	0.845
	C	0.996	0.994	0.994	0.993
0.88	A	0.078	0.07	0.069	0.053
	B	0.843	0.847	0.82	0.801
	C	0.996	0.993	0.995	0.996
0.89	A	0.079	0.072	0.069	0.048
	B	0.852	0.832	0.845	0.819
	C	0.991	0.994	0.997	0.994
0.9	A	0.069	0.063	0.062	0.048
	B	0.831	0.83	0.821	0.826
	C	0.997	0.989	0.992	0.994

In the second stage, fixing $\theta_t = 0.92$, we followed the similar procedure to calibrate (θ_u, θ_l) , which determine the early termination of a trial due to equivalence or superiority respectively. Note that θ_l has to be greater or equal to θ_t because the decision criteria must be tighter during the trial than at the end of trial. Our goal is still to maintain a treatment-wise type 1 error rate of 5% or lower and to achieve desired power when the trial can terminate early (Table 14.2).

Alternatively, we can set $\theta_t = \theta_l$ which means that we will not relax the decision criteria at the end of the trial. Extensive simulation for various scenarios should be carried out to ensure controlled type 1 error and satisfied power for all possible situations in real trial (Table 14.3).

Suppose the trial require 0.1 type 1 error and at least 0.85 power for treatment B and 0.99 power for treatment C, we chose the design parameters as $\theta_t = \theta_l = 0.89$ and $\theta_u = 0.9$. The operation characteristics is list in Table 14.4.

14.5 Response-Adaptive Randomization for Survival Outcomes

The response-adaptive randomization design with binary outcomes is commonly used in clinical trial where “success” is defined as the desired (or undesired) event occurring within (or beyond) a clinically relevant time. Given that patients

Table 14.3 Type 1 error rates and power with $\theta_l = \theta_l$

		θ_u					
$\theta_l = \theta_l$	Arm	0.8	0.82	0.84	0.86	0.88	0.9
0.85	A	0.142	0.127	0.127	0.126	0.115	0.105
	B	0.871	0.869	0.866	0.864	0.857	0.857
	C	0.992	0.994	0.991	0.995	0.993	0.994
0.87	A	0.121	0.121	0.117	0.108	0.1	0.085
	B	0.876	0.877	0.87	0.87	0.869	0.852
	C	0.996	0.994	0.992	0.994	0.996	0.986
0.89	A	0.109	0.102	0.093	0.091	0.071	0.075
	B	0.855	0.861	0.849	0.861	0.847	0.857
	C	0.992	0.994	0.995	0.988	0.996	0.995
0.91	A	0.097	0.082	0.08	0.078	0.075	0.077
	B	0.83	0.848	0.848	0.84	0.849	0.825
	C	0.993	0.994	0.996	0.987	0.989	0.988
0.93	A	0.095	0.074	0.071	0.071	0.064	0.06
	B	0.797	0.809	0.835	0.833	0.817	0.799
	C	0.994	0.99	0.991	0.996	0.99	0.992
0.95	A	0.065	0.042	0.039	0.039	0.036	0.025
	B	0.784	0.792	0.775	0.764	0.79	0.778
	C	0.988	0.995	0.989	0.989	0.986	0.994

Table 14.4 Operation characteristics with $\theta_l = \theta_l = 0.89$ and $\theta_u = 0.9$

Arm	Response rate	Pr (selected early)	Pr (stopped early)	# patients (2.5%, 97.5%)
A	0.2	0.012	0.386	24.15 (6, 35)
B	0.4	0.496	0.077	27.72 (6, 37)
C	0.6	0.827	0.005	16.45 (7, 32)

enter a trial sequentially, only a fraction of patients will have sufficient follow-up during interim analysis. This results in a loss of information as it is unclear how patients without sufficient follow-up should be handled. Adaptive designs for survival trials have been proposed for this type of trial. However, current practice generally assumes the event times follow a pre-specified parametric distribution. In this section, we adopt a nonparametric model of survival outcome which is robust to model event time distribution, and then apply it to response-adaptive design. The operating characteristics of the proposed design along with parametric design are compared by simulation studies, including their robustness properties with respect to model misspecifications.

Patients are enrolled in sequential groups of size $\{N_j\}$, $j = 1, \dots, J$, where N_j is the sample size of the sequential group j . Typically, before conducting the trial, researchers have little prior information regarding the superiority of the treatment arms. Therefore, initially, for the first j' groups, e.g. $j' = 1$, patients are allocated to K treatment arms with an equal probability $1/K$. As patients accrue, the number of current patients increases. Let T_i be the event time for patient i , τ be the clinically

relevant time and $\theta = \Pr(T > \tau)$ be the probability of interest. For example, a trial is conducted to assess the progression-free survival probability at 9 months. Let $N(s)$ denote the current number of patients who have been accrued and treated at a given time s during the trial. Without censoring, θ can be modeled by binomial model where the likelihood function evaluated at time s is

$$L(\text{data}|\theta) = \prod_{i=1}^{N(s)} \theta^{I(T_i > \tau)} (1 - \theta)^{I(T_i \leq \tau)} \tag{14.4}$$

However, censoring is unavoidable in clinical practice. As patients enter into the trial sequentially, the follow-up time for certain patients may be less than τ when we evaluate θ at any calendar time s . Other reasons for censoring may include patient dropout, failure to measure the outcome of interest, etc. If we ignore the censored patients, substantial information will be lost. Cheung and Chappell (2000) introduced a simple model for dose-finding trial. Later, Cheung and Thall (2002) adopted this model to continuous monitoring for phase II clinical trials. With censoring, the likelihood function (14.4) can be rewritten as

$$L(\text{data}|\theta) = \prod_{i=1}^{N(s)} \Pr\{T_i \leq \min(x_i, \tau)\}^{Y(x_i)} \Pr\{T_i > \min(x_i, \tau)\}^{1-Y(x_i)} \tag{14.5}$$

where $x_i = \min(c_i, t_i)$ is the observed event time, c_i is the censoring time, and $Y(x_i) = I\{T_i \leq \min(x_i, \tau)\}$ is the censoring indicator function.

Furthermore, the parameter θ will be plugged into the likelihood function through probability transformation. Let $t = \min(x_i, \tau)$,

$$\begin{aligned} \Pr(T_i \leq t) &= \Pr(T_i \leq t, T_i \leq \tau) + \Pr(T_i \leq t, T_i > \tau) \\ &= \Pr(T_i \leq t|T_i \leq \tau) \Pr(T_i \leq \tau) + \Pr(T_i \leq t|T_i > \tau) \Pr(T_i > \tau) \\ &= w(t) (1 - \theta) \end{aligned} \tag{14.6}$$

where $w(t) = \Pr(T_i \leq t|T_i \leq \tau)$, is a weight function

Finally, we can obtain a working likelihood with unbiased estimation of $w(t)$.

$$L(\text{data}|\theta) = \prod_{i=1}^{N(t)} \tilde{w}(x_i) (1 - \theta)^{Y(x_i)} \left\{ 1 - \tilde{w}(x_i) (1 - \theta) \right\}^{1-Y(x_i)} \tag{14.7}$$

Theorem *if $\tilde{w}(x_i)$ converges almost surely to $w(x_i)$ for all i as $N(s) \rightarrow \infty$, then $\hat{\theta} = \text{argmax}L(\text{data}|\theta)$ is strongly consistent for true survival probability θ .*

Cheung and Chappell (2000) assumed the nuisance parameter $\tilde{w}(x_i)$ as a linear function $\tilde{w}(x_i) = x_i/\tau$. Ji and Bekele (2009) show that these estimated weights are based on strong assumption of linearity and independence, and may lead to

biased results when the assumptions are violated. We propose to estimate $\tilde{w}(x_i)$ with Kaplan–Meier (KM) estimation of $\tilde{S}(x_i)$, where

$$\tilde{w}(x_i) = \frac{1 - \tilde{S}(x_i)}{1 - \tilde{S}(\tau)}$$

It's easy to show that $\tilde{w}(x_i)$ is an unbiased estimation of $w(x_i)$.

Assign a noninformative prior of beta (α, β), we can obtain posterior distribution of θ . However, the posterior distribution is not available in closed form and standard integral approximations can perform poorly. Without knowing the exact posterior distribution, we can easily draw random MCMC samples and obtain posterior estimation using standard MCMC software packages.

Under model (14.7), the survival probability evaluated at time τ is used as a conventional measure of treatment efficacy. However, such a survival probability at time τ ignores the entire path of survival curve. One of the interests in a clinical trial is the estimation of the difference between survival probability for the treatment groups at several points in time. As shown in Fig. 14.2, the survival curve under treatment B declines faster than that under treatment A, although both treatments have the same survival probability at time τ . In a renal cancer trial, this indicates that patients under treatment B would experience disease progression much faster than those under treatment A. Because delayed disease progression typically leads to a better quality of life, treatment A would be preferred in this situation (Ning and Huang 2010). Another example is showed in Fig. 14.3. The survival curves are almost identical between two treatments before time 20. If we compare the survival probability between two treatments at the time before 20, the treatment effect is inconclusive. To provide a comprehensive measure of efficacy by accounting for the shape of the survival curve, we propose to evaluate survival probability at several points in time. Let θ_{kj} be the survival probability at time τ_j for treatment k where $j=1, \dots, J$. The treatment allocation probability for treatment k is defined as,

$$\pi_k = \sum_{j=1}^J w_j \Pr(\theta_{kj} = \max\{\theta_{lj}, 1 \leq l \leq J\} | \text{data})$$

where w_j is the prespecified weight. Currently, we use equal weight with $w_j = 1/J$.

During the trial, we continuously monitor posterior probability of π_k . When the efficacy of π_k is lower than the prespecified lower limit p_l , then the treatment arm k will be terminated early due to futility. When π_k is higher than p_u , the treatment arm k will be selected as promising treatment. At the end of the trial, if π_k is higher than p_t , then treatment k is selected as the superior treatment. Otherwise, the trial is inconclusive. In practice, the values of p_l , p_u , and p_t are chosen by simulation studies to achieve desirable operating characteristics for the trial.

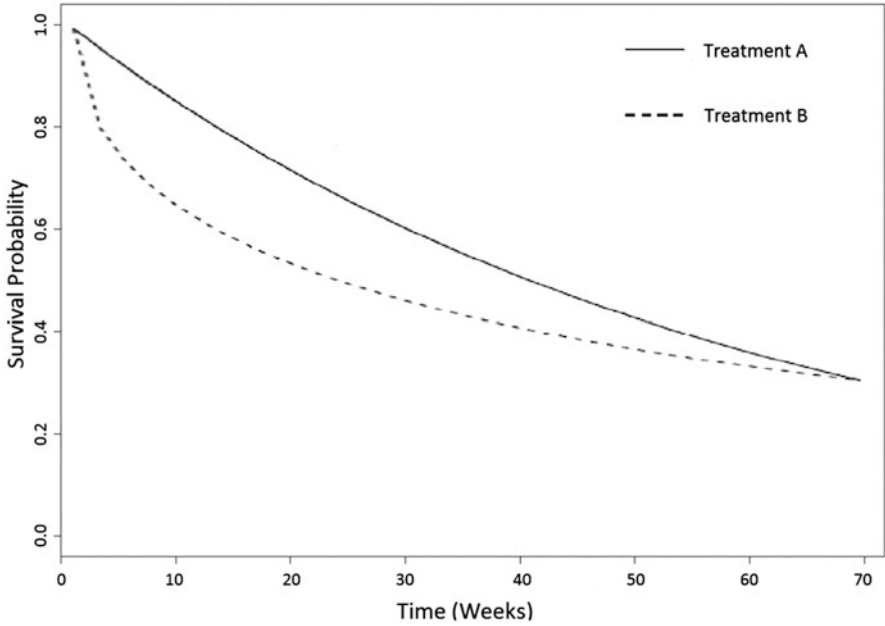


Fig. 14.2 Survival curves of the time to disease progression, where the two survival curves have the same survival probability at the follow-up time $\tau = 70$ weeks, but different areas under the survival curves until τ

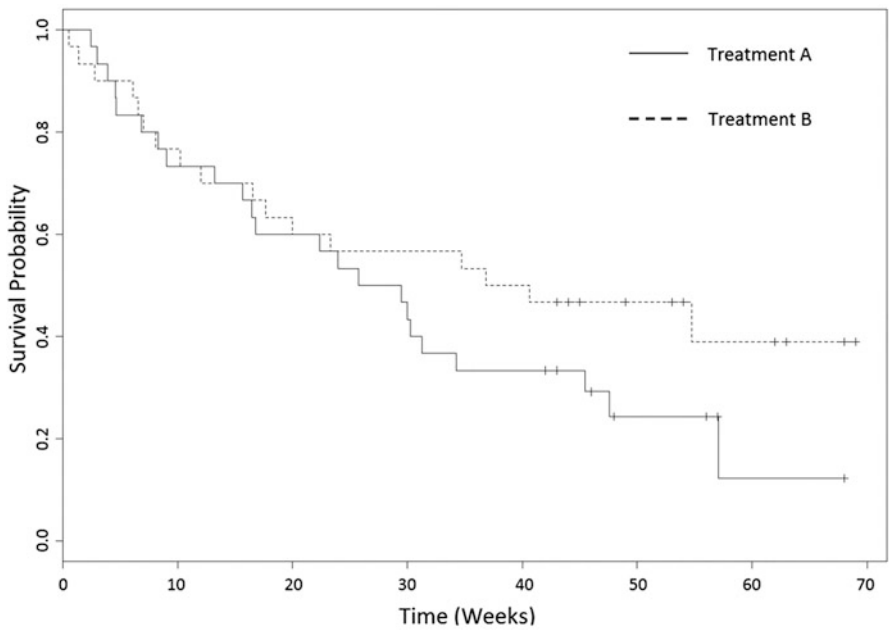


Fig. 14.3 Survival curves of the time to disease progression, where the two survival curves have the similar survival probability before week 20, but gradually show difference as time increase

We simulate a single arm trial where the event times follow a Weibull distribution with $\alpha = 2$ and $\lambda = 50$, where α is the shape parameter and λ is the scale parameter of the distribution. Patients enter the trial sequentially with accrual rate of one per week. At week 50, we stop enrolling the patients and continue to follow the trial for additional 30 weeks. The parameter of interest is $\theta = \Pr (T > 40)$.

The purpose of this simulation study is to compare the performance of estimation with different methods and to show whether the estimation at different trial monitoring time is consistent. Four estimation methods will be evaluated, including the proposed method, the true parametric method (estimate $S(x_i)$ by Weibull distribution), the misspecified parametric method (estimate $S(x_i)$ by exponential distribution), and the original method ($\hat{w}(x_i) = x_i/\tau$). Trial monitoring starts at week 40 and continues until the end of study. Figure 14.4 shows the estimated θ at different monitoring times. The results show that both the true parametric method and proposed method provide unbiased estimation over monitoring time while the original method and misspecified parametric method give large bias. It should be noted that the original method gives small bias at the end of trial because the number of censored observations (e.g. due to treatment ongoing) decreased as follow-up time increased. In Fig. 14.5, we present the coverage probability along the monitoring times. The figure shows that the proposed method and true parametric method provide constant coverage probability over the monitoring time which is close to the nominal value of 95%. In contrast, the original method and misspecified parametric method both give low coverage probability.

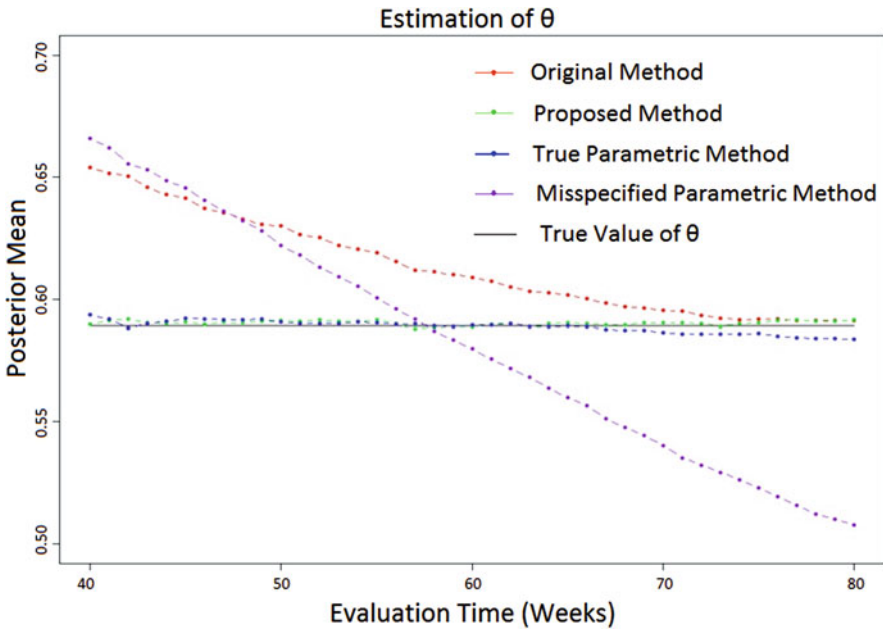


Fig. 14.4 Estimated θ with different methods

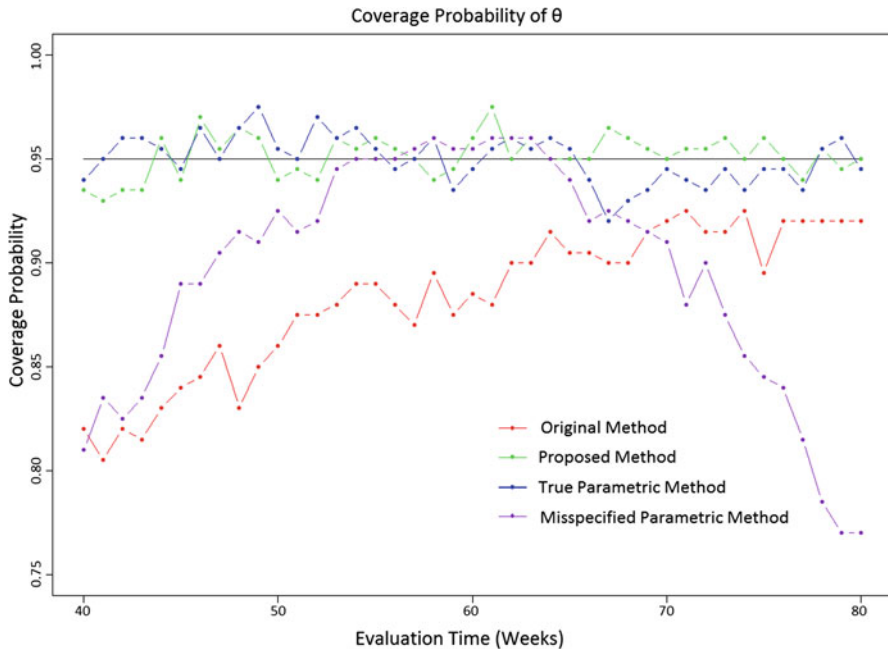


Fig. 14.5 Coverage probability of θ with different methods

We conducted a second set of simulations to evaluate the performance of the proposed adaptive randomization design under various clinical scenarios (1000 simulations per scenario). For the simulations, we set the accrual rate to two patients per week. The maximum number of patients is 120. After the initial 60 weeks of enrollment time, there is an additional follow-up period of 40 weeks. The event times are simulated from a Weibull distribution with $\alpha = 1$ in scenario I and $\alpha = 0.5$ in scenario II. We assigned the first 30 patients equally to two arms (A or B) and started using the adaptive randomization at the 31st patient. The randomization probability was evaluated every 5 weeks. The proposed design will be compared with different estimation methods for the weight function $w(t)$: proposed method, parametric method (estimate $S(x_i)$ by exponential distribution), and original method ($\tilde{w}(x_i) = x_i/\tau$).

Table 14.5 shows the simulation results from scenario I, without early termination ($p_u = 1, p_l = 0$). For each method, we list the average number of patients (with percentage of total patients in the trial) assigned to each treatment arm, and the chance of a treatment being selected as promising. When comparing the parametric method, the proposed method provides comparable operational characteristic where both designs assign more patients to more promising treatment (69% for proposed design and 70.3% for parametric design) and both designs provide the sample level of power (0.978 for proposed design and 0.979 for parametric design). The original method achieves lower power than both the proposed method and parametric method.

Table 14.5 Simulation result for scenario I

Arm	λ	Proposed method ($p_t = 0.955$)		Exponential method ($p_t = 0.99$)		Original method ($p_t = 0.995$)	
		# of patients	Pr (select)	# of patients	Pr (select)	# of patients	Pr (select)
A	40	26.3 (31%)	0.005	24.72 (29.6%)	0.003	36.5 (36.1%)	0.003
B	100	58.6 (69%)	0.978	61.36 (70.3%)	0.979	64.5 (63.9%)	0.749
		84.9		90.08		101.0	

Table 14.6 shows simulation results for scenario II, without early termination ($p_u = 1, p_l = 0$). In the presence of event time distribution misspecification, the parametric method provides lower power than the proposed method (0.836 vs 0.647). In addition, the proposed method assigns more patients to the more promising treatment. Once again, the original method has lower power than the other two methods.

Table 14.6 Simulation result for scenario II

Arm	λ	Proposed method ($p_t = 0.965$)		Exponential method ($p_t = 0.995$)		Original method ($p_t = 0.995$)	
		# of patients	Pr (select)	# of patients	Pr (select)	# of patients	Pr (select)
A	50	27.8 (28.5%)	0.005	32.48 (32.2%)	0.0003	35.6 (33.8%)	0.001
B	200	69.6 (71.5%)	0.836	68.34 (67.8%)	0.647	69.8 (66.2%)	0.51
		97.4		100.82		105.4	

14.6 Case Studies

14.6.1 Investigation of Serial Studies to Predict Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY 2)

I-SPY 2 is an adaptive phase II clinical trial that pairs oncologic therapies with biomarkers for women with advanced breast cancer. The goal is to identify improved treatment regimens for patient’s subsets based on molecular characteristics (biomarkers) of their disease (Barker et al. 2009).

The trial (Fig. 14.6) is initialized with two standard-of-care arms, and five treatment arms. Each treatment is tested on a minimum of 20 patients, and a maximum of 120 patients. Patient’s biomarkers are determined at enrollment, and patients are randomized to treatment arms based on their biomarker signature. Bayesian methods of adaptive randomization are used to achieve a higher probability of efficacy. Thus, treatments that perform well within a biomarker subgroup will have an increased probability of being assigned to patients with that biomarker.

Treatments will be dropped for futility if they show a low Bayesian predictive probability of being more effective than the standard of care with any biomarker. Treatment regimens that show a high Bayesian predictive probability of being more effective than the standard of care will stop for efficacy at interim time-points. These treatments will advance (with their corresponding biomarkers) to phase III trials. Depending on the patient accrual rate, new drugs can be added to the trial as other drugs are discontinued for either futility or efficacy.

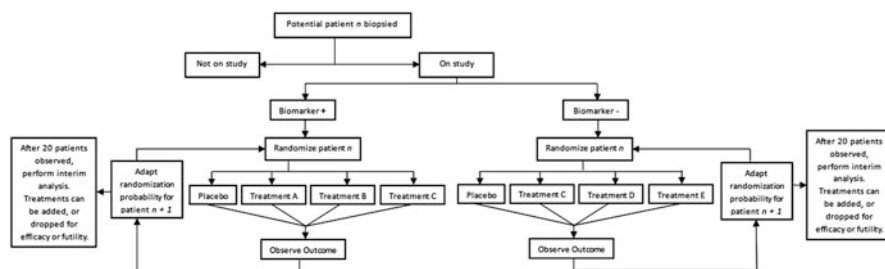


Fig. 14.6 I-SPY 2 trial

As of March 2017, 12 experimental treatment arms have been explored. Five agents, after showing promise within their biomarker groups, advanced to further studies and others are in queue for entry. A new I-SPY 3 master protocol is under planning to provide further evidence of effectiveness for agents successfully graduating from I-SPY2.

14.6.2 Gastric Cancer Umbrella Design for an Investigational Agent

This is an open-label, multicenter, phase 1b study of an investigational agent in combination with regimen A, regimen B, paclitaxel, or docetaxel in adult patients with locally advanced and metastatic gastric or gastroesophageal adenocarcinoma (Fig. 14.7). The study consists of a dose escalation phase (Part 1) and a dose expansion phase (Part 2). In Part 2, this study uses equal and adaptive randomization.

Any patient who enters Part 2 of the study is screened to determine whether their tumor tissue is positive for EBV (approximately 9% of patients with gastric cancer). An estimated 28 patients who are EBV-positive are assigned to treatment with regimen A in combination with the investigational agent (Cohort A). Patients who are EBV-negative initially are randomized equally to 1 of the other treatment cohorts, 5 patients per group: investigational agent + regimen B (Cohort B), investigational agent + paclitaxel (Cohort C), or investigational agent + docetaxel (Cohort D). These patients' data are assessed using a proportional weighted clinical utility function (allocating specific weights for complete response [CR], partial

response [PR], stable disease [SD], and progressive disease [PD]). New patients are then randomized to treatment according to an adaptive randomization algorithm, which incorporates a weighted clinical utility function. The resulting probability is continually updated per accumulating data on the associations between the response rate and Bayesian stopping rules.

Adaptive randomization increases the opportunity for each patient to receive the most effective experimental treatment possible based on posterior probabilities. Up to an additional 25 patients may be enrolled in each treatment regimen. Based on simulation results, the sample size for Part 2 (umbrella portion) of the study may be between 61 and 90 patients.

Overall response rate is used as the efficacy benchmark. Target effect size of 25% (0.25) and an undesirable effect size of 10% (0.1) are chosen based on clinical judgment. Early stopping rules are prespecified if there is a clear signal of efficacy or lack of efficacy. The stopping rules are as follows:

1. achieve maximum sample size of each arm (30 patients);
2. stop an arm if posterior probability $\Pr(\text{response rate [RR]} > 0.25/\text{Data}) > 80\%$ and $\Pr(\text{RR} > 0.10/\text{Data}) > 90\%$;
3. suspend accrual to an arm if $\Pr(\text{RR} \leq 0.10/\text{Data}) > 80\%$.

The treatment arm(s) is/are chosen in relation to the efficacy bar prespecified (target and undesirable); therefore, it is possible to select multiple treatment arms per this study design.

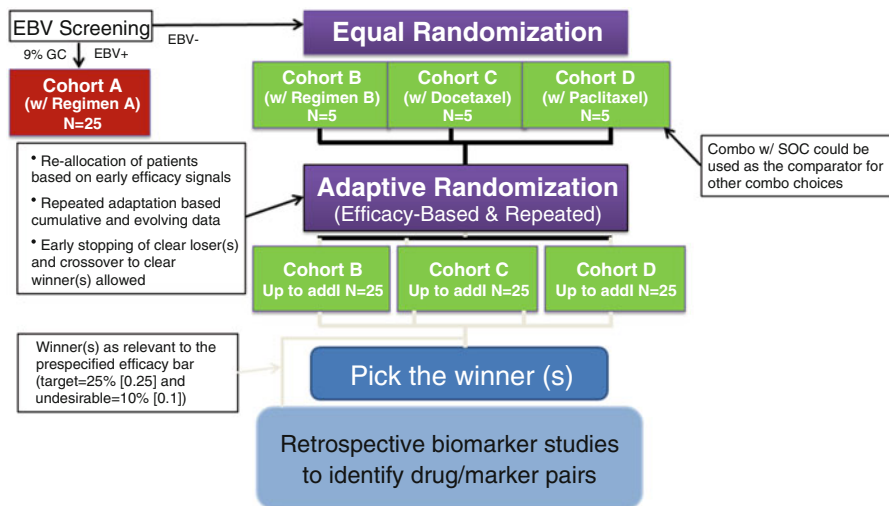


Fig. 14.7 Gastric cancer umbrella design

14.7 Discussion

With the closer collaborations between government, academia and industry, as well as the need to increase the probability of success of drug development across varied therapeutic areas, there are significant growing uses of innovative adaptive designs in master protocols, including the umbrella or platform trials, to screen multiple drugs simultaneously (Woodcock and LaVange 2017). Though different master protocols come with different sizes and settings, they share many common features, e.g. additional planning from the beginning of trial design, coordination between different stakeholders and increasingly sophisticated infrastructures for the research effects. Adaptive randomization is becoming a critical component and statistical methodology under these settings. While response-adaptive randomization procedures are not appropriate in clinical trials with a limited recruitment period and/or outcomes that occur after a long follow up, there is no reason why response-adaptive randomization cannot be used in clinical trials with moderately delayed response. Sequential estimates and allocation probabilities can be updated as data become available. For ease of implementation, updates can also be made after groups of patients have responded, rather than individually. From a practical perspective, there is no logistical difficulty in incorporating delayed responses into the response-adaptive randomization procedure, provided some responses become available during the recruitment and randomization period.

We have developed a Bayesian response-adaptive randomization design for survival trial. A nonparametric survival model is applied to estimate the survival probability at a clinically relevant time. The proposed design provides comparable operational characteristics as true parametric design. When the event time distribution is misspecified, the proposed design performs better than parametric one. The proposed design can be extended to Response-Adaptive Covariate-Adjusted Randomization (RACA) design when we need to control important prognostics among treatment arms (Lin et al. 2016a, b, c). Another potential approach of updating treatment allocation probability could be based on the restricted mean survival time. The benefits of adaptive randomization for survival trial depend on the distributions of event times and patient accrual rate as well as on the adaptive design under consideration (Case and Morgan 2003). If there are short-term responses that are quickly available and predictive of long-term survival, we can use those short-term responses to “speed up” adaptive randomization for survival trial (Huang et al. 2009).

A major criticism of response-adaptive randomization is that, despite stringent eligibility criteria, there may be a drift in patient characteristics over time. Using covariate-adjusted response-adaptive randomization can be a solution to this problem if the underlying covariates causing the heterogeneity are known in advance. This may not cause issues with large sample sizes since the randomization automatically balances prognostic factors among treatment groups asymptotically. For clinical trials with small or moderate sample sizes, the impact from the imbalance of the prognostic factors can be substantial when using response-

adaptive randomization designs, and thus causes difficulties to the interpretation after randomization. Thus, it is encouraged to have a randomization procedure that could also actively balance the covariate across treatment arms. Consequently, such design can help balance patient characteristics between different treatment arms, and thereby control the inflated type I error rates that occur in response-adaptive randomization (Lin et al. 2016a, b, c; Lin and Bunn 2017).

References

- Barker, A., Sigman, C.C., Kelloff, G.J., Hylton, N.M., Berry, D.A., Esserman, L.J.: I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin. Pharmacol. Ther.* **86**(1), 97–100 (2009)
- Berry, S.M., Carlin, B.P., Lee, J.J., Muller, P.: *Bayesian Adaptive Methods for Clinical Trials*. CRC Press, Boca Raton (2010)
- Case, L.D., Morgan, T.M.: Design of phase II cancer trials evaluating survival probabilities. *BMC Med. Res. Methodol.* **3**(1), 6 (2003)
- Cheung, Y.K., Chappell, R.: Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics.* **56**(4), 1177–1182 (2000)
- Cheung, Y.K., Thall, P.F.: Monitoring the rates of composite events with censored data in phase II clinical trials. *Biometrics.* **58**(1), 89–97 (2002)
- Hu, F., Rosenberger, W.F.: Introduction. In: *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, Hoboken, NJ (2006)
- Huang, X., Ning, J., Li, Y., Estey, E., Issa, J.P., Berry, D.A.: Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Stat. Med.* **28**(12), 1680–1689 (2009)
- Ji, Y., Bekele, B.N.: Adaptive randomization for multi-arm comparative clinical trials based on joint efficacy/toxicity outcomes. *Biometrics.* **65**(3), 876–884 (2009)
- Kairalla, J.A., Coffey, C.S., Thomann, M.A., Muller, K.E.: Adaptive trial designs: A review of barriers and opportunities. *Trials.* **13**(1), 145 (2012)
- Lin, J., Bunn, V.: Comparison of multi-arm multi-stage design and adaptive randomization in platform clinical trials. *Contemp. Clin. Trials.* **54**, 48–59 (2017)
- Lin, J., Lin, L., Sankoh, S.: A general overview of adaptive randomization design for clinical trials. *J. Biom. Biostat.* **7**, 294 (2016a)
- Lin, J., Lin, L., Sankoh, S.: A Bayesian response-adaptive covariate-adjusted randomization design for clinical trials. *J. Biom. Biostat.* **7**, 287 (2016b)
- Lin, J., Lin, L., Sankoh, S.: A phase II trial design with Bayesian adaptive covariate-adjusted randomization. In: *Statistical Applications from Clinical Trials and Personalized Medicine to Finance and Business Analytics*, ICSA Book Series in Statistics, pp. 61–73. Springer, Cham (2016c)
- Ning, J., Huang, X.: Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. *Stat. Med.* **29**(17), 1761–1768 (2010)
- Renfro, L.A., Mallick, H., An, M.-W., Sargent, D.J., Mandrekar, S.J.: Clinical trial designs incorporating predictive biomarkers. *Cancer Treat. Rev.* **43**, 74–82 (2016). <https://doi.org/10.1016/j.ctrv.2015.12.008>
- Rosenberger, W.F., Sverdlov, O., Hu, F.: Adaptive randomization for clinical trials. *J. Biopharm. Stat.* **22**(4), 719–736 (2012)
- Thall, P.F., Wathen, J.K.: Practical Bayesian adaptive randomisation in clinical trials. *Eur. J. Cancer.* **43**(5), 859–866 (2007)
- U.S. Food and Drug Administration. *Guidance for industry: Adaptive design clinical trials for drugs and biologics*. Draft guidance (2010)

- Woodcock, J., LaVange, L.: Master protocols to study multiple therapies, multiple diseases, or both. *N. Engl. J. Med.* **377**, 62–70 (2017)
- Yin, G.: *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods*, vol. 876. Wiley, Hoboken (2013)
- Yin, G., Chen, N., Jack Lee, J.: Phase II trial design with Bayesian adaptive randomization and predictive probability. *J. R. Stat. Soc.: Ser. C: Appl. Stat.* **61**(2), 219–235 (2012)
- Yuan, Y., Huang, X., Liu, S.: A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial. *Stat. Med.* **30**(11), 1218–1229 (2011)