

A General Overview of Adaptive Randomization Design for Clinical Trials

Jianchang Lin^{1*}, Li-An Lin² and Serap Sankoh¹

¹Takeda Pharmaceutical Company Limited, Cambridge, MA, USA

²Merck Research Laboratories, Whitehouse Station, New Jersey, USA

*Corresponding author: Jianchang Lin, Takeda Pharmaceutical Company Limited, Cambridge, MA, USA, Tel: +1-617-679-7000; E-mail: jianchang.lin@takeda.com

Rec date: March 25, 2016; Acc date: April 14, 2016; Pub date: April 21, 2016

Copyright: © 2016 Lin J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

With the recent release of FDA draft guidance (2010), adaptive designs, including adaptive randomization (e.g. response-adaptive (RA) randomization) has become popular in clinical trials because of its advantages of flexibility and efficiency gains, which also have the significant ethical advantage of assigning fewer patients to treatment arms with inferior outcomes. In this paper, we presented a general overview of adaptive randomization designs for clinical trials, including Bayesian and frequentist approaches as well as response-adaptive randomization. Examples were used to demonstrate the procedure for design parameters calibration and operating characteristics. Both advantages and disadvantages of adaptive randomization were discussed in the summary from practical perspective of clinical trials.

Keywords: Adaptive design; Clinical trials; Bayesian adaptive randomization

Introduction

The general goals of randomized clinical trials are to treat patients effectively and differentiate treatment effects efficiently. On one hand, a clinical trial tries to discriminate the effects of different treatments quickly, so that more patients outside of the trial would benefit from the more efficacious treatment sooner. For this purpose, patients' allocation should be (nearly) balanced across the comparative arms. On the other hand, each trial participant should be treated the most effectively, 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 [1]. Therefore, randomized clinical trials need to strike a balance between individual and collective ethics.

With the recent release of FDA draft guidance (2010), adaptive designs, including adaptive randomization (e.g. response-adaptive (RA) randomization) has become popular in clinical trials because of its advantages of flexibility and efficiency gains, which also have the significant ethical advantage of assigning fewer patients to treatment arms with inferior outcomes. In this paper, we presented a general overview of adaptive randomization designs for clinical trials, including Bayesian and frequentist approaches as well as response-adaptive randomization. Examples were used to demonstrate the procedure for design parameters calibration and operating characteristics. Both advantages and disadvantages of adaptive randomization were discussed in the summary from practical perspective of clinical trials.

ADs 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 [2]. 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 adaptive design, group sequential design, sample size re-estimation design, etc. (Figure 1).

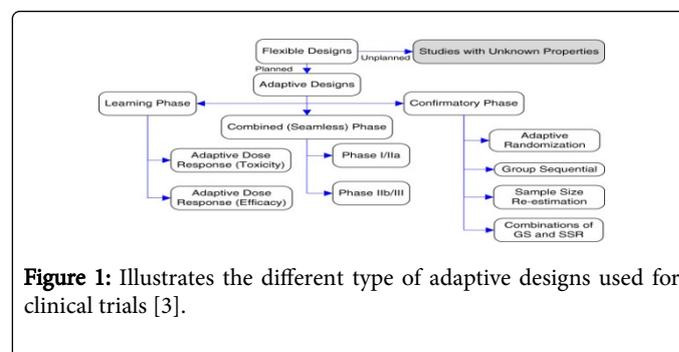


Figure 1: Illustrates the different type of adaptive designs used for clinical trials [3].

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 treatment, 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 course of 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, an 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 [4] 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 and minimize expected treatment failures.
- Covariate-adjusted response adaptive (CARA) randomization: a combination of covariate-adaptive and response-adaptive randomization procedures.

Frequentist and Bayesian Approach for Adaptive Randomization

The standard statistical approach to designing and analyzing clinical trials and other medical experiments is frequentist. A primary purpose of this report is to describe an alternative approach called the Bayesian approach. The eponym originates from a mathematical theorem derived by Thomas Bayes (1763), an English clergyman who lived from 1702 to 1761. Bayes' theorem plays a fundamental role in the inferential and calculational aspects of the Bayesian 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 deal with empirical evidence and both use probability.

Practitioners exposed in traditional, frequentist statistical methods appear to have been drawn to Bayesian approaches for three reasons [5-9]. 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 as now easily implemented in WINBUGS, Open BUGS and other available MCMC software packages now offer the convenient approach to hierarchical or random effect modeling, as regularly used in longitudinal data, frailty model, spatial data, time series data, and a wide variety of other settings featuring interdependent data. Third, practitioners are attracted by the greater flexible and adaptive features of 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 method, in both design and analysis. In the Bayesian approach, experiments can be altered in 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. Designing a clinical trial is a decision problem. Drawing a conclusion from a trial, such as recommending a therapy is a decision problem. Allocating resources among R&D projects is a decision problem. When to stop device development is a decision problem. There are costs and benefits involved in every such problem. In the Bayesian approach these costs and benefits can be assessed for each possible sequence of future observations.

All of this is not to say that the frequentist approach to clinical trials is totally without merit. 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.

Response-Adaptive Randomization

Response-adaptive randomization is one of the most important adaptive trial design in which the randomization ratio of patients assigned to the experimental treatment arm versus the control treatment arm changes from 1:1 over time to randomly assigning a higher proportion of patients to the arm that is doing better [10]. It is very attractive when ethical considerations or concerns make it potentially undesirable to have an equal number of patients assigned to each treatment arms. Suppose the trial objective is to compare treatment 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 the sequential group j . Typically, before planning the trial, researchers 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 allocated to two treatment arms with an equal probability of 0.5. The response information observed from these patients then can be used to update the allocation probability in subsequent coming 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

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

and the posterior distribution of p_k is

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

During the trial conduct, 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.

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

One of the advantages of a Bayesian approach to inference is the increased flexibility to include sequential stopping compared to the more restrictive requirements under a classical approach. Noninformative stopping rules are irrelevant for Bayesian inference. In other words, posterior inference remains unchanged no matter 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 $\Pr(p_k < p.\text{min} | \text{data}) > \theta_u$, where $p.\text{min}$ 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 $\Pr(p_k > p.\text{target} | \text{data}) > \theta_l$, where $p.\text{target}$ 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 $\Pr(p_k > p.\text{min} | \text{data}) > \theta_b$, 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 .

Example

We conducted simulations to show the procedure for design parameters calibration. The patient allocation probability is determined by algorithm [3]. The minimum allocation probability is 10% to ensure a reasonable probability of randomizing patients to each arm. The minimum clinical response rate ($p.\text{min}$) is 0.2 and the target response rate ($p.\text{target}$) 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 next 16th patient. The cohort size is set as 10, so that the early stopping rule and allocation probability updating will act after 10 new patients response 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 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 that close to both type 1 error and power will be selected for the next stage selection (Table 1).

Arm	θ_t			
	0.9	0.91	0.92	0.93
A	0.07	0.065	0.056	0.049
B	0.842	0.838	0.832	0.825
C	0.998	0.994	0.992	0.989

Table 1: Type 1 error rates and power, without early termination.

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 is allowed to terminate early (Table 2).

θ_u	Arm	θ_l			
		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

Table 2: Type 1 error rates and power, with early termination.

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 have to be carried out to ensure controlled type 1 error and satisfied power for all possible situations in real trial (Table 3).

		θ_u					
$\theta_t = \theta_l$	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 3: Type 1 error rates and power with $\theta_t = \theta_l$.

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 4.

Arm	Pr (Success)	Pr (Selected Early)	Pr (Stopped Early)	#Patients (2.5%, 97.5%)
Arm1	0.2	0.012	0.386	24.15 (6, 35)
Arm2	0.4	0.496	0.077	27.72 (6, 37)
Arm4	0.6	0.827	0.005	16.45 (7, 32)

Table 4: Operation characteristics with $\theta_t = \theta_l = 0.89$ and $\theta_u = 0.9$.

Discussion

While response-adaptive randomization procedures are not appropriate in clinical trials with a limited recruitment period and outcomes that occur after all patients have been randomized, there is no reason why response-adaptive randomization cannot be used in clinical trials with moderate delayed response. Sequential estimates and allocation probabilities can be updated as data become available. 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.

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 size since the randomization automatically balances prognostic factors among treatment groups asymptotically. However, 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 [11].

References

- Berry SM, Carlin BP, Lee JJ, Muller P (2010) Bayesian adaptive methods for clinical trials. CRC press.
- Brannath W, Burger HU, Glimm E, Stallard N, Vandemeulebroecke M, et al. (2010) Comments on the draft guidance on "adaptive design clinical trials for drugs and biologics" of the U.S. Food and Drug Administration. J Biopharm Stat 20: 1125-1131.
- Kairalla JA, Coffey CS, Thomann MA, Muller KE (2012) Adaptive trial designs: a review of barriers and opportunities. Trials 13: 145.
- Hu F, Rosenberger WF (2006) Introduction- The theory of response-adaptive randomization in clinical trials. John Wiley & Sons Inc, Hoboken, NJ, USA.
- Ning J, Huang X (2010) Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. Stat Med 29: 1761-1768.
- Rosenberger WF, Sverdlov O, Hu F (2012) Adaptive randomization for clinical trials. J Biopharm Stat 22: 719-736.
- Thall PF1, Wathen JK (2007) Practical Bayesian adaptive randomisation in clinical trials. Eur J Cancer 43: 859-866.
- Yin G, Chen N, Jack Lee J (2012) Phase II trial design with Bayesian adaptive randomization and predictive probability. Journal of the Royal Statistical Society 61: 219-235.
- Yin G (2013) Clinical trial design: Bayesian and frequentist adaptive methods. John Wiley & Sons.
- Yuan Y, Huang X, Liu S (2011) A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial. Statistics in Medicine 30: 1218-1229.
- Lin J, Lin L, Sankoh S (2016) A bayesian response-adaptive covariate-adjusted randomization design for clinical Trials. J Biom Biostat 7: 287.