Contemporary Clinical Trials 40 (2015) 15–25

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Simulation study for evaluating the performance of response-adaptive randomization

Yining Du [a,b], Xuan Wang [c], J. Jack Lee [a,*]

[a] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030, USA
[b] Division of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler Street, Houston, TX 77030, USA
[c] Baylor Institute for Immunology Research, Baylor Research Institute, Baylor Health Care System, Dallas, TX 75246, USA

ARTICLE INFO

Article history:
Received 18 June 2014
Received in revised form 31 October 2014
Accepted 1 November 2014
Available online 11 November 2014

Keywords:
Allocation probability
Bayesian adaptive design
Efficacy early stopping
Operating characteristics
Patient horizon

ABSTRACT

A response-adaptive randomization (RAR) design refers to the method in which the probability of treatment assignment changes according to how well the treatments are performing in the trial. Holding the promise of treating more patients with the better treatments, RARs have been successfully implemented in clinical trials. We compared equal randomization (ER) with three RARs: Bayesian adaptive randomization, sequential maximum likelihood, and sequential posterior mean. We fixed the total number of patients, considered as patient horizon, but varied the number of patients in the trial. Among the designs, we compared the proportion of patients assigned to the superior arm, overall response rate, statistical power, and total patients enrolled in the trial with and without adding an efficacy early stopping rule. Without early stopping, ER is preferred when the number of patients beyond the trial is much larger than the number of patients in the trial. RAR is favored for large treatment difference or when the number of patients beyond the trial is small. With early stopping, the difference between these two types of designs was reduced. By carefully choosing the design parameters, both RAR and ER methods can achieve the desirable statistical properties. Within three RAR methods, we recommend SPM considering the larger proportion in the better arm and higher overall response rate than BAR and similar power and trial size with ER. The ultimate choice of RAR or ER methods depends on the investigator's preference, the trade-off between group ethics and individual ethics, and logistic considerations in the trial conduct, etc.

1. Introduction

When multiple treatment arms are involved in clinical trials, randomization is commonly applied to avoid allocation bias and to yield a valid statistical inference. It also has ethical implications regarding treatment efficacy. In the beginning of a trial, the equipoise principle implies that treatment effects are equal among the treatment arms. Hence, equal randomization (ER) which randomizes patients equally among treatment arms can be justified. However, as the information accrues in the trial, the treatment effects may no longer be equal among treatments. Response-adaptive randomization (RAR) dynamically assigns patients to treatments based on the accumulating clinical responses, where 'response' generically refers to treatment outcomes. One of the RAR's appealing features is that it can assign more patients to better treatments based on available data.

In two-arm trials, RAR assigns more patients to the superior treatment and exposes fewer patients to the inferior treatment. The challenge is that we know very little about the relative treatment effect initially. As a trial proceeds, data in the trial give RAR higher probability of assigning more patients to the superior treatment. The relative effectiveness of treatments

Abbreviations: RAR, response-adaptive randomization; ER, equal randomization; BAR, Bayesian adaptive randomization; SML, sequential maximum likelihood; SPM, sequential posterior mean.

* Corresponding author. Tel.: +1 713 794 4158.
E-mail address: jjlee@mdanderson.org (J. Jack Lee).

http://dx.doi.org/10.1016/j.cct.2014.11.006
1551-7144/

can be estimated by some sequential estimation rule on a meaningful outcome measure. An important issue is that many of the RAR procedures may not be optimal based on the specific measures [1]. Since they are derived heuristically, we need to define criteria for comparing and evaluating them. A properly chosen RAR method can strike a balance in maximizing the statistical power for testing the treatment efficacy (group ethics) and giving each patient the best treatment (individual ethics). Key desirable features of the RAR designs include retaining randomizations advantage (eliminate bias), maintaining the required statistical power, and treating more patients with superior treatments. The play-the-winner design [2] was a deterministic precursor of RAR. Other approaches use urn models [3,4], and consider multi-arm trials and covariates [5,6], extend RAR to trials with delayed response [7,8], and compare RAR for frequentist group sequential binary response trials [9].

We use the setting of a binary outcome, but RAR can be generalized to other settings. We evaluate only treatment efficacy and assume comparability in safety, cost, and feasibility. We compare ER to three RAR methods: Bayesian adaptive randomization (BAR), derived from [10]; sequential maximum likelihood (SML), with randomization probability calculated based on frequentist sequential maximum likelihood estimators (Section 10.4.1 in [4]); and sequential posterior mean (SPM), which replaces the role of the sequential maximum likelihood in the allocation probability.

The use of RAR versus ER is yet debated in statistical and clinical trial communities [11,12]. Traditionally, the primary goal of clinical trials is to determine better treatment options to benefit future patients. Providing the best possible treatment to the patients enrolled in the trial is rarely the main purpose of the trial. This view, however, has major shortcomings. For patients enrolled in clinical trials, will there be a single patient who does not want to receive the best possible treatment? Although equal randomization can be well justified based on the equipoise principle in the beginning of the trial, as the information accrues, RAR can be applied to assign more patients to better treatments based on available data. Two recent advances in medicine compel us to critically examine the traditional paradigm. (1) New treatment modalities are changing rapidly. Even if a trial is positive and the standard practice is changed, its impact may not be long lasting because new treatment modalities could change again. (2) In the genomic era, considering all possible markers such as mutations from deep sequencing (next generation sequencing), copy number variation, mRNA expression, microRNA expression, and protein expression, no two patients are alike. This N-of-1 trial concept challenges the traditional paradigm that patients are homogeneous and the results are generalizable to all alike patients in the future. Therefore, we take one step further to evaluate designs which can strike a balance between the individual ethics and the group ethics. The concept of the patient horizon [13] is used to evaluate such balance. The patient horizon is the total number of patients with the particular disease that is relevant to the treatments being investigated. It includes patients enrolled in the trial and the future patient population beyond the trial who will benefit from what is learned in the trial. For some diseases the future patient population can be quite large compared to the size of the trial. In that setting, making the right decision about the treatments studied (which depends on the power of the test) is more important. On the other hand, in the case of some rare cancers, most of the patients with the disease of interest will be enrolled in the trial and there will be few future patients beyond the trial. Under that scenario, giving the best treatment possible to each patient in the trial is ethically the right thing to do. We fix the patient horizon size but varying the trial size, then, compare the proportion of patients assigned to the superior treatment, overall response rate, and statistical power between RAR and ER designs. We evaluate the operating characteristics for trials without and with early stopping rules for futility and efficacy.

2. Methods

2.1. Bayesian adaptive randomization method

Bayesian adaptive randomization (BAR) is an allocation scheme based on the posterior probability of one treatment being more effective than the other(s). This posterior probability can be obtained from a binomial likelihood and a beta prior for a binary outcome:

$$x_{ij} \sim Bin(1, \theta_i), \quad \theta_i \sim Beta(\alpha_i, \beta_i) \tag{1}$$

where x_{ij} is the outcome of patient j on treatment i and θ_i is the probability of a response for treatment i. We consider the setting of two treatments $i \in \{1, 2\}$. The probability that treatment 2 is better than treatment 1 is given by $Pr(\theta_2 > \theta_1|x)$. Because of the consistency of the posterior, this probability approaches either 0 or 1 as more data are collected for both treatments unless the two treatments are exactly the same. The basic idea of BAR can be traced to [10], although a good part of his paper was devoted to the computation of $Pr(\theta_1 > \theta_2|x)$, which was then a big hurdle. A detailed review on BAR is provided by Thall and Wathen [14], and an application of BAR in the development of targeted agent can be found in [15]. The main goal of BAR is to assign more patients to the better treatment arms with higher probability. A patient is adaptively randomized to treatment i with a probability of $\Pr(\theta_i > \theta_j|data)$. However, the probability $\Pr(\theta_i > \theta_j|data)$ can be highly variable in the beginning of the trial when the number of patients is small. By adding a tuning parameter, the randomization probability can be stabilized

$$\rho(\lambda) = \frac{Pr(\theta_2 > \theta_1|x)^\lambda}{Pr(\theta_1 > \theta_2|x)^\lambda + Pr(\theta_2 > \theta_1|x)^\lambda} \tag{2}$$

where λ is a tuning parameter and $\lambda \geq 0$. An introduction about the tuning parameter could be found in [16]. Furthermore, [14] recommended using the tuning parameter $n/2N$ instead of a fixed value. Note that $\rho(1) = Pr(\theta_2 > \theta_1|x)$, $\rho(0) = \frac{1}{2}$ and $\rho(\infty)$ behave like the play-the-winner rule. Therefore, λ controls the level of imbalance in the allocation probability.

As is evident, BAR may lead to an extreme preference for a certain treatment arm. One way to avoid such extreme allocation probability is to set bounds on the allocation probability; thus, it does not converge to 0 or 1. For example, we may constrain the randomization probability to be bounded within 0.05 to 0.95, or 0.1 to 0.9 [17] to allow for continued randomization to both arms to gather information for further assessment of the treatment effects.