Optimality in Blocked Response-Adaptive Randomization Designs for Non-Inferiority Trials

Thevaa Chandereng *
Department of Biostatistics, Columbia University
and
Ying Kuen Cheung
Department of Biostatistics, Columbia University

Abstract

Response-adaptive randomization (RAR) designs were introduced due to the captivating characteristics of tilting the randomization probability to assign more subjects to the better-performing treatment group based on interim data. They suffer from the severe flaw of bias and inflated type I error due to potential confounding of treatment with time in the presence of time trends, which are usually unavoidable. Blocked RAR approaches where the subjects are assigned in strata within which the randomization ratio is constant have been proposed to avoid these problems. Both blocked and unblocked RAR strategies are less efficient than balanced 1:1 designs in the usual superiority trial setting. Surprisingly, this may not be true for non-inferiority trials. For such studies, we provide a theoretical framework for optimizing blocked RAR with binary responses using the Cochran-Mantel-Haenszel test. We discuss how the framework is useful in the design of non-inferiority trials using blocked RAR for both additive and multiplicative margins. We evaluate our blocked RAR designs and compare them with the traditional 1:1 design for two treatment arms. We conclude that the blocked RAR design can be more efficient than the traditional 1:1 design for non-inferiority trials. The Cardiovascular Patient Outcomes Research Team (CPORT) study provides a motivating example for the current work.

*email : tc3123@columbia.edu

Keywords: play-the-winner rule, adaptive designs, clinical trials, Neymann allocation

1 Introduction

Adaptive clinical trials are clinical trial designs that allow prospectively planned modifications to one or more aspects of the trial using interim data (Food et al., 2018). Adaptive clinical trial designs have attractive features, including reduction of sample size, early stopping for success and futility, and acceleration of the drug discovery process. RAR designs are adaptive designs that utilize accumulating information to tilt the randomization ratio to assign more subjects to the better-performing treatment arms (Wei and Durham, 1978). Patients enrolled in these trials are treated in the best way possible according to the current interim analysis. Unlike traditional trials, RAR trials are cumbersome due to the effort required to constantly update the allocation probabilities based on patients' outcomes.

Traditional RAR designs where the allocation probabilities are altered on a patient-by-patient basis are flawed. In traditional RAR, confounding of treatment with time induces a potential bias. Time-trends are likely to occur in clinical trials (Karrison et al., 2003; Chappell and Karrison, 2007). In long duration trials, patients' characteristics (which are also known as "patient drift") often change throughout the trial. On the other hand, traditional RAR designs assume that the sequence of patients who arrive for entry into the trial represents samples drawn at random from two homogenous populations. This assumption does not usually hold. Liu and Lee (2015) reported that there were more smokers in the latter part of the trial than the beginning part of the trial in the Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE) trial. Time trends in RAR trials not only bias estimates of treatment differences but can also incorrectly reject a true null hypothesis. The problems are not solved by parametric modeling which in turn relies on further assumptions. Thus standard RAR designs form, in one opinion "... an insidious threat to the most important tool in the clinical research armamentarium (Begg,

2015)." The important tool referred to is, of course, randomization itself.

There are a handful of designs and analyses introduced to address the time-trend issue in RAR (Coad, 1992; Simon and Simon, 2011; Karrison et al., 2003; Thall et al., 2015). Karrison et al. (2003) introduced a blocked RAR approach to tackle this issue. They proposed a block design (group-sequential) where the randomization probability is altered in a block level instead of on a patient-by-patient basis. The final analysis is stratified by blocks to eliminate bias due to patient drift because the randomization probability is kept fixed within in each block. Chandereng and Chappell (2019) examined the operating characteristics of this design under various frequentist and Bayesian designs. They concluded that a small number of blocks have a good tradeoff between efficiency and treating more subjects to the better-performing treatment group. They also warned against the use of a large number of blocks because if a stratum does not randomize subjects to two treatment groups, then the stratum is uninformative. Blocked RAR also speeds up the process of RAR trials since the randomization ratio is modified for a group of patients in the block rather than on a by-patient basis.

There are many RAR designs proposed (Rosenberger et al., 2001; Thall and Wathen, 2007; Eisele, 1994; Wei and Durham, 1978; Zelen, 1969). However, most of the designs have been myopic schemes to randomize more subjects to the better performing treatment arm and reduce the non-trivial risk of assigning more patients to the inferior arm. There are various RAR designs including the urn models and biased coin designs. Some of these designs also have a nice limiting property. However, they are not useful. There are articles that have discussed the optimal allocation for traditional RAR using the non-centrality parameter of the chi-square test for binary responses (Hu and Rosenberger, 2003; Rosenberger et al., 2001). However, it shares with all unblocked RAR designs the fundamental

flaw of bias and inflated type I error in the presence of time trends. On the other hand, the blocked RAR approach provides a good solution.

Non-inferiority (NI) trials are clinical trials that are designed to show if a treatment is not much worse than that of a standard control treatment (McDaniel et al., 2016; Chandereng et al., 2020). The new treatment is considered to offer ancillary benefits in terms of safety, side effects or other elements. Unlike superiority trials, NI trials allow the possibility for the new treatment to be slightly worse than the standard control. NI trials do not restrict the upper bound (the possibility of the new treatment being better than the standard treatment) unlike equivalence trials. A non-inferior margin (Δ) is a predetermined margin that allows a small loss in effect by the new treatment compared to the control treatment (Temple and Ellenberg, 2000). It represents how much worse the new treatment can be compared with the control and still be considered similar.

Denote $\Delta > 0$ for an additive and $\Delta > 1$ for a multiplicative margin. The corresponding non-inferiority trial hypotheses for an additive margin to test the difference in means are

$$\begin{aligned} H_0 : p_C - p_T &\geq \Delta \\ H_A : p_C - p_T &< \Delta, \end{aligned}$$

where p_C and p_T are the proportions of success of the control and experimental groups. The corresponding non-inferiority trial hypotheses for a multiplicative margin to test the differences in means are

$$\begin{aligned} H_0 : \frac{p_C}{p_T} &\geq \Delta \\ H_A : \frac{p_C}{p_T} &< \Delta, \end{aligned}$$

where p_C and p_T are the proportions of success of the control and experimental groups.

RAR designs have been widely used in superiority trials. However, in superiority trials, RAR designs are usually not efficient compared to a traditional 1:1 design. Chandereng et al. (2020) showed that the optimal randomization for non-inferiority trials can be far from 1:1 and can greatly improve efficiency of estimators. The probability of correctly rejecting the null hypothesis when the alternative is true increases as the efficiency of estimators increase.

In Section 2 of the article, we derive the optimal allocation with respect to efficiency and to minimizing the total number of failures for blocked RAR in a non-inferiority trial setting using both additive and multiplicative margins. In the following section, we dicusss how the adaptive design works in a blocked RAR setting and the convergence. In Section 4 we discuss our methods for simulation study and the results before moving on to use the CPORT study of percutaneous revascularization as an example. We also compare these results with the RPW rule and with the equal allocation rule. We focus on binary responses, but the derivation of optimal allocation with respect to efficiency is easily extendable to continuous and survival outcomes.

2 Methods

2.1 Optimal allocation: efficiency

2.1.1 Additive Margin ($\Delta > 0$)

The corresponding non-inferiority trial hypotheses for additive margin, where p_C and p_T are the probabilities of success for the control and treatment groups and Δ is the non-inferiority

margin.

$$H_0 : p_C - p_T \geq \Delta \qquad H_A : p_C - p_T < \Delta.$$

The Wald test for blocked RAR with k blocks,

$$Z^2 = \frac{(p_{C,1} - p_{T,1} - \Delta)^2}{(\frac{p_{C,1}(1-p_{C,1})}{n_{C,1}} + \frac{p_{T,1}(1-p_{T,1})}{n_{T,1}})} + ... + \frac{(p_{C,k} - p_{T,k} - \Delta)^2}{(\frac{p_{C,k}(1-p_{C,k})}{n_{C,k}} + \frac{p_{T,k}(1-p_{T,k})}{n_{T,k}})},$$

where $p_{C,i}$ and $p_{T,i}$ are the probabilities of success for the control and treatment groups in block i , Δ is the non-inferiority margin, and $n_{C,i}$ and $n_{T,i}$ are the numbers of subjects in block i. $Z^2 \sim \chi^2$ with k degrees of freedom.

Instead of maximizing the Z^2 for the sum of all blocks, we can maximize for a block because the functions for all blocks are identical. For block i,

$$\begin{aligned} \phi(i) &= \frac{(p_{C,i} - p_{T,i} - \Delta)^2}{(\frac{p_{C,i}(1-p_{C,i})}{n_{C,i}} + \frac{p_{T,i}(1-p_{T,i})}{n_{T,i}})}, \\ &= \frac{(p_{C,i} - p_{T,i} - \Delta)^2}{(\frac{p_{C,i}(1-p_{C,i})}{h_i * n_i} + \frac{p_{T,i}(1-p_{T,i})}{(1-h_i) * n_i})}, \end{aligned}$$

where $h_i n_i$ and $(1-h_i)n_i$ are the optimal sample sizes allocated to the control and treatment groups.

By optimization, we obtain

$$h_i = \frac{\sqrt{p_{C,i}(1-p_{C,i})}}{\sqrt{p_{C,i}(1-p_{C,i})} + \sqrt{p_{T,i}(1-p_{T,i})}}. \tag{1}$$

2.1.2 Multiplicative margin ($\Delta > 1$)

The corresponding non-inferiority trial hypotheses for multiplicative margin, where p_C and p_T are the probabilities of success for the control and treatment groups and Δ is the non-inferiority margin.

$$H_0 : \frac{p_C}{p_T} \geq \Delta \qquad H_A : \frac{p_C}{p_T} < \Delta, \quad \Delta > 1.$$

The Wald test for blocked RAR with k blocks,

$$Z^2 = \frac{(p_{C,1} - \Delta p_{T,1})^2}{(\frac{p_{C,1}(1-p_{C,1})}{n_{C,1}} + \frac{\Delta^2 p_{T,1}(1-p_{T,1})}{n_{T,1}})} + ... + \frac{(p_{C,k} - \Delta p_{T,k})^2}{(\frac{p_{C,k}(1-p_{C,k})}{n_{C,k}} + \frac{\Delta^2 p_{T,k}(1-p_{T,k})}{n_{T,k}})},$$

where $p_{C,i}$ and $p_{T,i}$ are the probabilities of success for the control and treatment groups in block i , Δ is the non-inferiority margin, and $n_{C,i}$ and $n_{T,i}$ are the numbers of subjects in block i. $Z^2 \sim \chi^2$ with k degrees of freedom.

Instead of maximizing the Z^2 for the sum of all blocks, we can maximize for a block because the functions for all blocks are identical.

$$\begin{aligned} \phi(i) &= \frac{(p_{C,i} - \Delta p_{T,i})^2}{(\frac{p_{C,i}(1-p_{C,i})}{n_{C,i}} + \frac{\Delta^2 p_{T,i}(1-p_{T,i})}{n_{T,i}})}, \\ &= \frac{(p_{C,i} - \Delta p_{T,i})^2}{(\frac{p_{C,i}(1-p_{C,i})}{h*n_i} + \frac{\Delta^2 p_{T,i}(1-p_{T,i})}{(1-h)*n_i})}, \end{aligned}$$

where $h_i n_i$ and $(1-h_i)n_i$ are the optimal sample sizes allocated to the control and treatment groups.

By optimization, we obtain

$$h_i = \frac{\sqrt{p_{C,i}(1-p_{C,i})}}{\sqrt{p_{C,i}(1-p_{C,i})} + \Delta\sqrt{p_{T,i}(1-p_{T,i})}}. \quad (2)$$

2.2 Optimal allocation: minimize the expected number of failures

2.2.1 Additive margin ($\Delta > 0$)

The corresponding non-inferiority trial hypotheses for additive margin, where p_C and p_T are the probabilities of success for the control and treatment groups and Δ is the non-inferiority

margin.

$$H_0 : p_C - p_T \geq \Delta \qquad H_A : p_C - p_T < \Delta.$$

The Wald test for blocked RAR with k blocks,

$$Z^2 = \frac{(p_{C,1} - p_{T,1} - \Delta)^2}{(\frac{p_{C,1}(1-p_{C,1})}{n_{C,1}} + \frac{p_{T,1}(1-p_{T,1})}{n_{T,1}})} + ... + \frac{(p_{C,k} - p_{T,k} - \Delta)^2}{(\frac{p_{C,k}(1-p_{C,k})}{n_{C,k}} + \frac{p_{T,k}(1-p_{T,k})}{n_{T,k}})},$$

where $p_{C,i}$ and $p_{T,i}$ are the probabilities of success for the control and treatment groups in block i , Δ is the non-inferiority margin, and $n_{C,i}$ and $n_{T,i}$ are the numbers of subjects in block i. $Z^2 \sim \chi^2$ with k degrees of freedom.

Similar to the approach in 2.1, instead of minimizing the total number of failures in all the blocks, we can minimize the number of failures in one block because the functions are identical. We wish to find the optimal allocation, $h = \frac{n_{C,i}}{n_{C,i}+n_{T,i}}$, where $n_{C,i}$ and $n_{T,i}$ are the numbers of subjects in block i, which minimizes the expected number of failures, D, for a fixed variance. The equation for the variance of block i,

$$\begin{aligned} k &= \frac{p_{C,i}(1-p_{C,i})}{n_{C,i}} + \frac{p_{T,i}(1-p_{T,i})}{n_{T,i}} \\ k &= \frac{p_{C,i}(1-p_{C,i})}{hn_i} + \frac{p_{T,i}(1-p_{T,i})}{(1-h)n_i} \\ n_i &= \frac{p_{C,i}q_{C,i}}{h_i k} + \frac{p_{T,i}q_{T,i}}{(1-h_i)k}, \end{aligned}$$

where k is a constant and $h_i n_i$ and $(1-h_i)n_i$ are the optimal sample sizes allocated to the control and treatment groups.

$$\begin{aligned} D &= \arg\min_{h_i}\{n_i h_i q_{C,i} + n_i(1-h_i)q_{T,i}\}; \qquad 0 < h_i < 1, \\ &= \arg\min_{h_i}\{n_i(h_i q_{C,i} + (1-h_i)q_{T,i})\} \\ &= \arg\min_{h_i}\left\{ \left(\frac{p_{C,i}q_{C,i}}{h_i k} + \frac{p_{T,i}q_{T,i}}{(1-h_i)k}\right)(h_i q_{C,i} + (1-h_i)q_{T,i})\right\}, \end{aligned}$$

where $q_{C,i} = 1 - p_{C,i}$ and $q_{T,i} = 1 - p_{T,i}$.

$$\begin{aligned}
\frac{\partial D}{\partial h_i} &= \frac{h_i(1-h_i)K\big[(p_{T,i}q_{T,i} - p_{C,i}q_{C,i})(h_i q_{C,i} + (1-h_i)q_{T,i}) +}{K^2(1-h_i)^2h_i^2} \\
&\quad \frac{(q_{C,i} - q_{T,i})((1-h_i)p_{C,i}q_{C,i} + h_i p_{T,i}q_{T,i})\big] -}{K^2(1-h_i)^2h_i^2} \\
&\quad \frac{(1-2h_i)K\big[((1-h_i)p_{C,i}q_{C,i} + h_i p_{T,i}q_{T,i})(h_i q_{C,i} + (1-h_i)q_{T,i})\big]}{K^2(1-h_i)^2h_i^2} \\
&\propto \frac{h_i(1-h_i)\big[2h_i p_{T,i}q_{C,i}q_{T,i} + (1-2h_i)p_{T,i}q_{T,i}^2 - 2(1-h_i)p_{C,i}q_{C,i}q_{T,i} + (1-2h_i)p_{C,i}q_{C,i}^2\big] -}{(1-h_i)^2h_i^2} \\
&\quad \frac{(1-2h_i)\big[h_i(1-h_i)p_{C,i}q_{C,i}^2 + (1-h_i)^2 p_{C,i}q_{C,i}q_{T,i} + h_i^2 p_{T,i}q_{C,i}q_{T,i} + h_i(1-h_i)p_{T,i}q_{T,i}^2\big]}{(1-h_i)^2h_i^2} \\
&= \frac{2h_i(1-h_i)\big[h_i p_{T,i}q_{C,i}q_{T,i} + (1-h_i)p_{C,i}q_{C,i}q_{T,i}\big] - (1-2h_i)\big[h_i^2 p_{T,i}q_{C,i}q_{T,i} + (1-h_i)^2 p_{C,i}q_{C,i}q_{T,i}\big]}{(1-h_i)^2h_i^2} \\
&= \frac{h_i^2 p_{T,i}q_{C,i}q_{T,i} - (1-h_i)^2 p_{C,i}q_{C,i}q_{T,i}}{(1-h_i)^2h_i^2}.
\end{aligned}$$

Setting the numerator $\frac{\partial D}{\partial h_i} = 0$ to obtain the minimum, we get

$$h_i = \frac{\sqrt{p_{C,i}}}{\sqrt{p_{C,i}} + \sqrt{p_{T,i}}}. \tag{3}$$

2.2.2 Multiplicative margin ($\Delta > 1$)

The corresponding non-inferiority trial hypotheses for multiplicative margin, where p_C and p_T are the probabilities of success for the control and treatment groups and Δ is the non-inferiority margin.

$$H_0 : \frac{p_C}{p_T} \geq \Delta \qquad H_A : \frac{p_C}{p_T} < \Delta, \quad \Delta > 1.$$

The Wald test for blocked RAR with k blocks,

$$Z^2 = \frac{(p_{C,1} - \Delta p_{T,1})^2}{(\frac{p_{C,1}(1-p_{C,1})}{n_{C,1}} + \frac{\Delta^2 p_{T,1}(1-p_{T,1})}{n_{T,1}})} + ... + \frac{(p_{C,k} - \Delta p_{T,k})^2}{(\frac{p_{C,k}(1-p_{C,k})}{n_{C,k}} + \frac{\Delta^2 p_{T,k}(1-p_{T,k})}{n_{T,k}})},$$

where $p_{C,i}$ and $p_{T,i}$ are the probabilities of success for the control and treatment groups in block i , Δ is the non-inferiority margin, and $n_{C,i}$ and $n_{T,i}$ are the numbers of subjects in block i. $Z^2 \sim \chi^2$ with k degrees of freedom.

We can minimize the number of failures in one block because the functions are identical. We wish to find the optimal allocation, $h = \frac{n_{C,i}}{n_{C,i}+n_{T,i}}$, where $n_{C,i}$ and $n_{T,i}$ are the numbers of subjects in block i, which minimizes the expected number of failures, D, for a fixed variance. The equation for the variance of block i,

$$\begin{aligned} k &= \frac{p_{C,i}(1-p_{C,i})}{n_{C,i}} + \frac{\Delta^2 p_{T,i}(1-p_{T,i})}{n_{T,i}} \\ k &= \frac{p_{C,i}(1-p_{C,i})}{hn_i} + \frac{\Delta^2 p_{T,i}(1-p_{T,i})}{(1-h)n_i} \\ n_i &= \frac{p_{C,i}q_{C,i}}{h_i k} + \frac{\Delta^2 p_{T,i}q_{T,i}}{(1-h_i)k}, \end{aligned}$$

where k is a constant and $h_i n_i$ and $(1-h_i)n_i$ are the optimal sample sizes allocated to the control and treatment groups.

$$\begin{aligned} D &= \arg\min_{h_i}\{n_i h_i q_{C,i} + n(1-h_i)q_{T,i}\}; \qquad 0 < h_i < 1, \\ &= \arg\min_{h_i}\{n_i(h_i q_{C,i} + (1-h_i)q_{T,i})\} \\ &= \arg\min_{h_i}\left\{ \left(\frac{p_{C,i}q_{C,i}}{h_i k} + \frac{\Delta^2 p_{T,i}q_{T,i}}{(1-h_i)k} \right)(h_i q_{C,i} + (1-h_i)q_{T,i})\right\}, \end{aligned}$$

where $q_{C,i} = 1 - p_{C,i}$ and $q_{T,i} = 1 - p_{T,i}$.

$$
\begin{aligned}
\frac{\partial D}{\partial h_i} &= \frac{h_i(1-h_i)K\big[(\Delta^2 p_{T,i}q_{T,i} - p_{C,i}q_{C,i})(h_i q_{C,i} + (1-h_i)q_{T,i}) +}{K^2(1-h_i)^2h_i^2} \\
&\quad \frac{(q_{C,i} - q_{T,i})((1-h_i)p_{C,i}q_{C,i} + \Delta^2 h_i p_{T,i}q_{T,i})\big] -}{K^2(1-h_i)^2h_i^2} \\
&\quad \frac{(1-2h_i)K\big[((1-h_i)p_{C,i}q_{C,i} + \Delta^2 h_i p_{T,i}q_{T,i})(h_i q_{C,i} + (1-h_i)q_{T,i})\big]}{K^2(1-h_i)^2h_i^2} \\
&\propto \frac{h_i(1-h_i)\big[2\Delta^2 h_i p_{T,i}q_{C,i}q_{T,i} + \Delta^2(1-2h_i)p_{T,i}q_{T,i}^2 - 2(1-h_i)p_{C,i}q_{C,i}q_{T,i} + (1-2h_i)p_{C,i}q_{C,i}^2\big] -}{(1-h_i)^2h_i^2} \\
&\quad \frac{(1-2h_i)\big[h_i(1-h_i)p_{C,i}q_{C,i}^2 + (1-h_i)^2 p_{C,i}q_{C,i}q_{T,i} + \Delta^2 h_i^2 p_{T,i}q_{C,i}q_{T,i} + \Delta^2 h_i(1-h_i)p_{T,i}q_{T,i}^2\big]}{(1-h_i)^2h_i^2} \\
&= \frac{2h_i(1-h_i)\big[\Delta^2 h_i p_{T,i}q_{C,i}q_{T,i} + (1-h_i)p_{C,i}q_{C,i}q_{T,i}\big] -}{(1-h_i)^2h_i^2} \\
&\quad \frac{(1-2h_i)\big[\Delta^2 h_i^2 p_{T,i}q_{C,i}q_{T,i} + (1-h_i)^2 p_{C,i}q_{C,i}q_{T,i}\big]}{(1-h_i)^2h_i^2} \\
&= \frac{\Delta^2 h_i^2 p_{T,i}q_{C,i}q_{T,i} - (1-h_i)^2 p_{C,i}q_{C,i}q_{T,i}}{(1-h_i)^2h_i^2}.
\end{aligned}
$$

Setting the numerator $\frac{\partial D}{\partial h_i} = 0$ to obtain the minimum, we get

$$
h_i = \frac{\sqrt{p_{C,i}}}{\sqrt{p_{C,i}} + \Delta\sqrt{p_{T,i}}}. \tag{4}
$$

3 Blocked designs and the adaptive process

In the blocked RAR design, the randomization probability is altered in a block level instead of on a patient-by-patient basis (Karrison et al., 2003). In the adaptive process, we begin with unknown probabilities of success for the control and treatment groups. Based on the current estimates of probabilities of success, we alter the randomization probabilities. For

instance, using the optimal allocation relative to efficiency in a non-inferiority trial with an additive hypothesis, the allocation probability for the control group in block i is defined as

$$\pi_{i,C} = \frac{\sqrt{\hat{p}_{C,i-1}(1-\hat{p}_{C,i-1})}}{\sqrt{\hat{p}_{C,i-1}(1-\hat{p}_{C,i-1})} + \sqrt{\hat{p}_{T,i-1}(1-\hat{p}_{T,i-1})}},$$

where $\hat{p}_{C,i-1}$ and $\hat{p}_{T,i-1}$ are the estimated proportions of success for the control and treatment groups after block $i-1$.

The blocked RAR for non-inferiority trials with an additive hypothesis and extending the derivation of Melfi and Page (1998) for a non-inferiority trial,

$$\frac{n_C}{n} \to \frac{\sqrt{p_C(1-p_C)}}{\sqrt{p_C(1-p_C)} + \sqrt{p_T(1-p_T)}}, \qquad \text{as} \quad n, k \to \infty,$$

where n_C is the total number of subjects in the control group, n is the total sample size, and k is the number of blocks.

Similarly, with the optimal allocation relative to efficiency in a non-inferiority trial with a multiplicative hypothesis, the allocation probability for the control group in block i is defined as

$$\pi_{i,C} = \frac{\sqrt{\hat{p}_{C,i-1}(1-\hat{p}_{C,i-1})}}{\sqrt{\hat{p}_{C,i-1}(1-\hat{p}_{C,i-1})} + \Delta\sqrt{\hat{p}_{T,i-1}(1-\hat{p}_{T,i-1})}},$$

where $\hat{p}_{C,i-1}$ and $\hat{p}_{T,i-1}$ are the estimated proportions of success for the control and treatment groups after block $i-1$ and Δ is the non-inferiority margin.

The blocked RAR for non-inferiority trials with a multiplicative hypothesis and extending the derivation of Melfi and Page (1998) for a non-inferiority trial,

$$\frac{n_C}{n} \to \frac{\sqrt{p_C(1-p_C)}}{\sqrt{p_C(1-p_C)} + \Delta\sqrt{p_T(1-p_T)}}, \qquad \text{as} \quad n, k \to \infty,$$

where n_C is the total number of subjects in the control group, n is the total sample size, and k is the number of blocks.

The allocation probability for the control group in block i using the optimal allocation relative to minimizing the total number of failures in a non-inferiority trial with an additive hypothesis is defined as

$$\pi_{i,C} = \frac{\sqrt{\hat{p}_{C,i-1}}}{\sqrt{\hat{p}_{C,i-1}} + \sqrt{\hat{p}_{T,i-1}}},$$

where $\hat{p}_{C,i-1}$ and $\hat{p}_{T,i-1}$ are the estimated proportions of success for the control and treatment groups after block $i-1$.

Extending Theorem 2 in Melfi et al. (2001) for a non-inferiority trial,

$$\frac{n_C}{n} \to \frac{\sqrt{p_C}}{\sqrt{p_C} + \sqrt{p_T}}, \qquad \text{as} \quad n, k \to \infty,$$

where n_C is the total number of subjects in the control group, n is the total sample size, and k is the number of blocks.

The allocation probability for the control group in block i using the optimal allocation relative to minimizing the total number of failures in a non-inferiority trial with a multiplicative hypothesis is defined as

$$\pi_{i,C} = \frac{\sqrt{\hat{p}_{C,i-1}}}{\sqrt{\hat{p}_{C,i-1}} + \Delta\sqrt{\hat{p}_{T,i-1}}},$$

where $\hat{p}_{C,i-1}$ and $\hat{p}_{T,i-1}$ are the estimated proportions of success for the control and treatment groups after block $i-1$ and Δ is the non-inferiority margin.

Extending Theorem 2 in Melfi et al. (2001) for a non-inferiority trial,

$$\frac{n_C}{n} \to \frac{\sqrt{p_C}}{\sqrt{p_C} + \Delta\sqrt{p_T}}, \qquad \text{as} \quad n, k \to \infty,$$

where n_C is the total number of subjects in the control group, n is the total sample size, and k is the number of blocks.

4 Simulations and Results

We investigated the effects of various allocation rules with a different number of blocks, $K = 2$ and 4, using 10,000 simulations for both additive and multiplicative hypotheses. We simulated the power (probability of correctly concluding that the treatment group is non-inferior to the control group), the number of failures, and the treatment effect for the allocation rules optimized for the total number of failures (3)(4) and efficiency (1)(2), the randomized play-the-winner (RPW) rule, and equal allocation. The final analysis was done using stratification except for the equal allocation rule. The power indicates the proportion of 10,000 trials that declares the treatment group non-inferior to the control group. The targeted sample size varies depending on the probabilities of success in the control (p_C) and treatment groups (p_T). We obtained the sample size that would yield 0.90 power and 0.05 type I error for equal allocation.

Table 1 shows the power for non-inferiority trials for both additive and multiplicative margins using four different algorithms. The type I error rates are controlled under the nominal level under all the designs. The equal allocation design provides the highest power under all cases due to stratification in the other designs. The blocked RPW rule provides the lowest power in most cases. The differences in power between having two and four blocks for all three adaptive designs are small.

Table 2 displays the means and standard deviations of the total number of failures for non-inferiority trials for both additive and multiplicative margins using four different algorithms. The optimal allocation optimized for the total number of failures has the lowest total number of failures under most the null cases with an additive margin. The optimal allocation optimized for efficiency has a lower total number of failures compared to the equal allocation except when $p_C + p_T > 1$ for all additive margins.

p_C	p_T	Δ	N	E	F_2	F_4	A_2	A_4	R_2	R_4
Additive hypothesis										
0.012	0.008	0.004	25384	0.05	0.04	0.04	0.04	0.04	0.05	0.05
0.05	0.025	0.025	2604	0.04	0.05	0.04	0.05	0.05	0.05	0.05
0.25	0.20	0.05	2570	0.05	0.04	0.05	0.04	0.04	0.04	0.04
0.50	0.40	0.10	858	0.05	0.04	0.04	0.04	0.04	0.05	0.04
0.75	0.65	0.10	644	0.04	0.05	0.06	0.04	0.04	0.05	0.05
0.012	0.012	0.004	25384	0.90	0.82	0.81	0.86	0.85	0.83	0.82
0.05	0.05	0.025	2604	0.91	0.84	0.82	0.85	0.85	0.82	0.82
0.25	0.25	0.05	2570	0.89	0.82	0.82	0.86	0.85	0.82	0.80
0.50	0.50	0.10	858	0.90	0.83	0.83	0.84	0.87	0.81	0.79
0.75	0.75	0.10	644	0.88	0.84	0.82	0.86	0.85	0.80	0.81
Multiplicative hypothesis										
0.012	0.008	1.5	18334	0.06	0.04	0.04	0.04	0.04	0.04	0.03
0.05	0.025	2	1628	0.05	0.05	0.04	0.05	0.04	0.04	0.03
0.25	0.20	1.25	2108	0.04	0.05	0.05	0.05	0.04	0.04	0.04
0.50	0.40	1.25	704	0.06	0.04	0.04	0.04	0.03	0.04	0.03
0.012	0.012	1.5	18334	0.88	0.87	0.86	0.88	0.87	0.83	0.81
0.05	0.05	2	1628	0.89	0.85	0.85	0.86	0.86	0.83	0.82
0.25	0.25	1.25	2108	0.87	0.84	0.85	0.87	0.86	0.82	0.81
0.50	0.50	1.25	704	0.89	0.84	0.84	0.88	0.87	0.83	0.79

Table 1: Power for the equal allocation (E), optimal allocation to minimize the total number of failure for two (F_2) and four (F_4) blocks, optimal allocation to maximize power for two (A_2) and four (A_4) blocks, and the RPW rule for two (R_2) and four (R_4) blocks. 10,000 simulations were done for each case.

p_C	p_T	Δ	N	E	F_2	F_4	A_2	A_4	R_2	R_4
Additive hypothesis										
0.012	0.008	0.004	25384	25130	25128	25126	25128	25126	25130	25130
				(16)	(16)	(16)	(16)	(16)	(16)	(16)
0.05	0.025	0.025	2604	2506	2503	2502	2504	2502	2506	2506
				(10)	(10)	(10)	(10)	(10)	(10)	(10)
0.25	0.2	0.05	2570	1992	1990	1989	1991	1990	1991	1991
				(21)	(21)	(21)	(21)	(21)	(21)	(21)
0.5	0.4	0.1	858	472	471	470	471	471	471	470
				(15)	(15)	(15)	(15)	(14)	(15)	(15)
0.75	0.65	0.1	644	193	193	192	194	194	192	191
				(12)	(12)	(11)	(12)	(12)	(12)	(12)
0.012	0.012	0.004	25384	25079	25079	25079	25079	25080	25079	25079
				(18)	(17)	(17)	(17)	(17)	(17)	(17)
0.05	0.05	0.025	2604	2474	2474	2474	2474	2474	2474	2474
				(11)	(11)	(11)	(11)	(11)	(11)	(11)
0.25	0.25	0.05	2570	1927	1928	1927	1928	1927	1927	1928
				(22)	(22)	(22)	(22)	(22)	(22)	(22)
0.5	0.5	0.1	858	429	429	429	429	429	429	429
				(15)	(15)	(15)	(15)	(15)	(15)	(15)
0.75	0.75	0.1	644	161	161	161	161	161	161	161
				(11)	(11)	(11)	(11)	(11)	(11)	(11)
Multiplicative hypothesis										

0.012	0.008	1.5	18334	18151	18152	18153	18152	18154	18151	18151
				(14)	(13)	(14)	(13)	(13)	(14)	(13)
0.05	0.025	2	1628	1567	1569	1569	1569	1569	1567	1567
				(8)	(8)	(8)	(8)	(8)	(8)	(8)
0.25	0.2	1.25	2108	1634	1635	1636	1635	1636	1633	1633
				(19)	(19)	(19)	(19)	(19)	(19)	(19)
0.5	0.4	1.25	704	387	388	389	389	390	387	386
				(13)	(13)	(13)	(13)	(13)	(13)	(13)
0.012	0.012	1.5	18334	18114	18114	18114	18114	18114	18114	18114
				(15)	(15)	(15)	(15)	(15)	(15)	(15)
0.05	0.05	2	1628	1547	1547	1547	1546	1547	1547	1547
				(9)	(9)	(9)	(9)	(9)	(9)	(9)
0.25	0.25	1.25	2108	1581	1581	1581	1581	1581	1581	1581
				(20)	(20)	(20)	(20)	(20)	(20)	(20)
0.5	0.5	1.25	704	352	352	352	352	352	352	352
				(13)	(13)	(13)	(13)	(13)	(13)	(13)

Table 2: The means and (standard deviations) of the total number of failures for the equal allocation (E), optimal allocation to minimize the total number of failure for two (F_2) and four (F_4) blocks, optimal allocation to maximize power for two (A_2) and four (A_4) blocks, and the RPW rule for two (R_2) and four (R_4) blocks. 10,000 simulations were done for each case.

p_C	p_T	Δ	N	E	F_2	F_4	A_2	A_4	R_2	R_4
Additive hypothesis										
0.012	0.008	0.004	25384	0.0013	0.0012	0.0012	0.0012	0.0013	0.0012	0.0012
0.05	0.025	0.025	2604	0.0075	0.0074	0.0075	0.0074	0.0074	0.0074	0.0074
0.25	0.20	0.05	2570	0.0163	0.0164	0.0164	0.0161	0.0161	0.0166	0.0166
0.50	0.40	0.10	858	0.0339	0.0342	0.0343	0.0337	0.0339	0.0338	0.0338
0.75	0.65	0.10	644	0.0361	0.0359	0.0354	0.0357	0.0363	0.0361	0.0362
0.012	0.012	0.004	25384	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014
0.05	0.05	0.025	2604	0.0087	0.0087	0.0086	0.0085	0.0086	0.0085	0.0086
0.25	0.25	0.05	2570	0.0170	0.0170	0.0171	0.0172	0.0172	0.0170	0.0171
0.50	0.50	0.10	858	0.0345	0.0345	0.0340	0.0344	0.0343	0.0345	0.0344
0.75	0.75	0.10	644	0.0345	0.0342	0.0342	0.0343	0.0340	0.0345	0.0345
Multiplicative hypothesis										
0.012	0.008	1.5	18334	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018
0.05	0.025	2	1628	0.0139	0.0133	0.0138	0.0134	0.0134	0.0138	0.0138
0.25	0.20	1.25	2108	0.0208	0.0203	0.0203	0.0204	0.0202	0.0207	0.0207
0.50	0.40	1.25	704	0.0424	0.0414	0.0422	0.0423	0.0421	0.0424	0.0424
0.012	0.012	1.5	18334	0.0021	0.0020	0.0020	0.0021	0.0020	0.0021	0.0021
0.05	0.05	2	1628	0.0172	0.0167	0.0165	0.0169	0.0169	0.0172	0.0173
0.25	0.25	1.25	2108	0.0211	0.0213	0.0211	0.0210	0.0210	0.0213	0.0214
0.50	0.50	1.25	704	0.0421	0.0425	0.0425	0.0423	0.0429	0.0429	0.0429

Table 3: Standard deviations of the estimated treatment differences for the equal allocation (E), optimal allocation to minimize the total number of failure for two (F_2) and four (F_4) blocks, optimal allocation to maximize power for two (A_2) and four (A_4) blocks, and the RPW rule for two (R_2) and four (R_4) blocks. 10,000 simulations were done for each case.

The total number of failures are similar for both the null and alternative cases under all the designs. Under the optimal allocation optimized for efficiency, the design does not assign more subjects to the better-performing treatment group when $p_C + p_T > 1$ for an additive margin. Similarly, under the multiplicative margin, the optimal allocation optimized for efficiency and the total number of failures do not always assign more subjects to the better-performing treatment arm.

In clinical trials, the main objective is to estimate the treatment effects accurately. Table 3 shows the standard deviations of the estimated treatment differences for non-inferiority trials for both additive and multiplicative margins using four different algorithms. The standard deviations of the estimated treatment differences are lower for most cases under the optimal allocation optimized for efficiency compared to the equal allocation and the RPW rule. Thus, the optimal allocation optimized for efficiency provides a good tradeoff between the power, the total number of failures, and the standard deviation of estimated treatment effects. Under the optimal allocation optimized for efficiency, even if a study is powered for small difference in treatment effects, much can be gained in efficiency if the actual difference is big. However, even under no or small difference in treatment effects, the optimal allocation optimized for efficiency provides a more accurate estimate compared to the equal allocation design.

5 Example: CPORT

The CPORT study provides a motivating example for the current work. Physicians were interested in comparing the performance of percutaneous coronary intervention (PCI) at hospitals with vs. without on-site cardiac surgery in the CPORT study (Aversano et al.,

2012). Patients' availability to receive PCI are often restricted to hospitals with on-site cardiac surgery. Based on historical data, the six weeks rate of mortality (all-cause) in hospitals with on-site cardiac surgery (control group) was estimated to be 0.008. The outcome followup (six weeks) is short to be able to obtain the outcome of the trial for future randomization. The CPORT research team was interested in showing a non-inferiority margin of $p_T - p_C < 0.004$. p_C and p_T denote the proportions of mortality in hospitals with on-site cardiac surgery and hospitals without on-site cardiac surgery ($p_C = 0.008$, $p_T = 0.012$).

CPORT was a multi-center randomized trial with 18,867 patients who were randomized at 3:1 in the treatment group to undergo PCI with 14,149 patients undergoing PCI at a hospital without on-site cardiac surgery and 4,718 patients undergoing PCI at hospitals with on-site cardiac surgery. The six week mortality rate observed was 0.01 at hospitals with on-site surgery and 0.009 at hospitals without on-site surgery. The 95% CI for the difference in six-week mortality rate was (-0.0031, 0.0023) with a p-value of 0.004 for non-inferiority. The results suggest that PCI performed at hospitals without on-site cardiac surgery was non-inferior to PCI performed at hospitals with on-site cardiac surgery with respect to mortality at six weeks and major adverse cardiac events at nine months.

6 Conclusion

In this article, we examined the power, total number of failures, and variance of treatment effects comparing 4 different algorithms (3 adaptive designs and equal allocation) for binary responses in a non-inferiority trial setting. The equal allocation procedure provides the highest power under all the scenarios. The optimal allocation optimized for total number

of failures gives the lowest total number of failures under most of the cases. The optimal allocation optimized for efficiency produces the lowest variance of treatment effects under most of the cases.

In the American Statistical Association (ASA) statement on p-values, Wasserstein and Lazar (2016) highlighted that p-values can only indicate how incompatible the data is with a particular statistical model. In their statement, they also stress the importance of measuring the size of an effect. In the context of clinical trials, proponents of adaptive designs have argued the trade-off between power and total number of failures, however, they need to emphasize the importance of estimating the size of a treatment effect precisely. The optimal allocation optimized for efficiency should be favored to estimate the treatment effects accurately.

An R package (blockRAR), for the RPW rule, the optimal allocation optimized for efficiency, the optimal allocation optimized for total number of failures, and the equal allocation, is implemented in R and released as open source software under the MIT license. The blockRAR package is available at Comprehensive R Archive Network (CRAN) and at `https://thevaachandereng.github.io/blockRAR/`. We used blockRAR version 1.0.4 for all analyses.

References

Aversano, T., C. C. Lemmon, and L. Liu (2012, May). Outcomes of PCI at Hospitals with or without On-Site Cardiac Surgery. *New England Journal of Medicine 366*(19), 1792–1802.

Begg, C. B. (2015, April). Ethical concerns about adaptive randomization. *Clinical Tri-*

als 12(2), 101–101.

Chandereng, T. and R. Chappell (2019, September). Robust Blocked Response-Adaptive Randomization Designs. *arXiv:1904.07758 [stat]*. arXiv: 1904.07758.

Chandereng, T., X. Wei, and R. Chappell (2020). Imbalanced randomization in clinical trials. *Statistics in Medicine 39*(16), 2185–2196.

Chappell, R. and T. Karrison (2007, July). Continuous Bayesian adaptive randomization based on event times with covariates by Cheung et al., Statistics in Medicine 2006; 25:55-70. *Statistics in Medicine 26*(15), 3050–3052; author reply 3052–3054.

Coad, D. S. (1992, April). Data-dependent allocation rules a comparative study of some for bernoulli data. *Journal of Statistical Computation and Simulation 40*(3-4), 219–231. Publisher: Taylor & Francis.

Eisele, J. R. (1994, February). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference 38*(2), 249–261.

Food, U., D. Administration, et al. (2018). Guidance for industry on adaptive designs for clinical trials of drugs and biologics.

Hu, F. and W. F. Rosenberger (2003, September). Optimality, Variability, Power. *Journal of the American Statistical Association 98*(463), 671–678. Publisher: Taylor & Francis.

Karrison, T. G., D. Huo, and R. Chappell (2003). A group sequential, response-adaptive design for randomized clinical trials. *Controlled Clinical Trials 24*(5), 506–522.

Liu, S. and J. J. Lee (2015, September). An overview of the design and conduct of the BATTLE trials. *Chinese Clinical Oncology 4*(3), 33.

McDaniel, L. S., M. Yu, and R. Chappell (2016, April). Sample size under the additive hazards model. *Clinical Trials (London, England) 13*(2), 188–198.

Melfi, V. and C. Page (1998). *Variability in Adaptive Designs for Estimation of Success Probabilities*. Institute of Mathematical Statistics.

Melfi, V. F., C. Page, and M. Geraldes (2001). An adaptive randomized design with application to estimation. *Canadian Journal of Statistics 29*(1), 107–116.

Rosenberger, W. F., N. Stallard, A. Ivanova, C. N. Harper, and M. L. Ricks (2001). Optimal Adaptive Designs for Binary Response Trials. *Biometrics 57*(3), 909–913.

Rosenberger, W. F., A. N. Vidyashankar, and D. K. Agarwal (2001, November). Covariate-adjusted response-adaptive designs for binary response. *Journal of Biopharmaceutical Statistics 11*(4), 227–236.

Simon, R. and N. R. Simon (2011, July). Using randomization tests to preserve type I error with response adaptive and covariate adaptive randomization. *Statistics & Probability Letters 81*(7), 767–772.

Temple, R. and S. S. Ellenberg (2000, September). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Annals of Internal Medicine 133*(6), 455–463.

Thall, P., P. Fox, and J. Wathen (2015, August). Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Annals of Oncology: Official Journal of the European Society for Medical Oncology 26*(8), 1621–1628.

Thall, P. F. and J. K. Wathen (2007, March). Practical Bayesian adaptive randomisation in clinical trials. *European Journal of Cancer 43*(5), 859–866. Publisher: Elsevier.

Wasserstein, R. L. and N. A. Lazar (2016, April). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician 70*(2), 129–133.

Wei, L. J. and S. Durham (1978). The Randomized Play-the-Winner Rule in Medical Trials. *Journal of the American Statistical Association 73*(364), 840–843. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Zelen, M. (1969). Play the Winner Rule and the Controlled Clinical Trial. *Journal of the American Statistical Association 64*(325), 131–146. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].