Ao Yuan*, Qizhai Li, Ming Xiong and Ming T. Tan

Adaptive Design for Staggered-Start Clinical Trial

DOI 10.1515/ijb-2015-0011

Abstract: In phase II and/or III clinical trial study, there are several competing treatments, the goal is to assess the performances of the treatments at the end of the study, the trial design aims to minimize risks to the patients in the trial, according to some given allocation optimality criterion. Recently, a new type of clinical trial, the staggered-start trial has been proposed in some studies, in which different treatments enter the same trial at different times. Some basic questions for this trial are whether optimality can still be kept? under what conditions? and if so how to allocate the the coming patients to treatments to achieve such optimality? Here we propose and study a class of adaptive designs of staggered-start clinical trials, in which for given optimality criterion object, we show that as long as the initial sizes at the beginning of the successive trials are not too large relative to the total sample size, the proposed design can still achieve optimality criterion asymptotically for the allocation proportions as the ordinary trials; if these initial sample sizes have about the same magnitude as the total sample size, full optimality cannot be achieved. The proposed method is simple to use and is illustrated with several examples and a simulation study.

Keywords: adaptive design, allocation proportion, optimality, sequential design, staggered-start

1 Introduction

In phase II and/or III clinical trial, there are several competing treatments to be assessed during the trial process. During the trial, each coming patient is allocated to one of the treatments according to some pre-specified rule, which is based on the current clinical knowledge of the treatment performances of the previously allocated patients up to that time. The aims are to assess the treatment performances well and minimize the the overall treatment losses to the patients. The rule for the allocation is called the design of a clinical trial and the minimal loss is often characterized by some optimality criteria. If the design involves unknown parameter(s) to be estimated and updated along with the trial progression, it is called adaptive design. Optimal adaptive clinical trial designs and their properties have been studied extensively in the literature, for example by Rosenberger and Lachin [1], Eisele [2], Melfi and Page [3], Rosenberger et al. [4], Bai et al. [5], Bai and Hu [6], Zhang et al. [7], Yuan and Chai [8], Zhu and Hu [9], Yuan et al. [10], among others. In all these existing trials, the treatments under investigation enter the trial at the same time. Recently in phase II (or phase III) clinical trials, a new type of trial, the staggered-start clinical trial has been proposed and studied by a number of researchers/experimenters [11–13] etc.). In this trial the treatments enter the trial at different times, as in the studies by Kublin [14], Cummings et al. [15], Hendrix et al. [16], and the experimental studies at the Fred Hutchinson Cancer Research Center. Staggered-start and delayed withdrawal designs were originally proposed by Leber and others as a way of providing evidence of disease modification without relying on biomarkers. In this study patients are randomized to drug or placebo at baseline and after a pre-specified period the placebo group is treated with the active agent. If the staggered-start group “catches up” with the group receiving treatment from the onset of the trial, no disease modification affect has been observed and the agent is deemed to have symptomatic benefits only. If the staggered-start group does not catch up with the group receiving treatment from the onset, then it is concluded that disease modification has occurred [14].

*Corresponding author: Ao Yuan, Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA, E-mail: yuanao@hotmail.com

Qizhai Li, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Ming Xiong, School of Mathematics and Statistics, Central China Normal University, Wuhan 430079, China

Ming T. Tan, Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA

As an example for such trial, in the vaccine trial study [14], vaccine 1 and placebo are in the trial, after 12 months another new treatment vaccine 2 is available and joined the ongoing trial, and a third vaccine joined at a latter time. Thus, three vaccine regimens vs a shared placebo group are investigated in their trials. The trial is aimed to address the question: How should a lag in the availability of a vaccine product be accommodated in the research trial design? There are variety of other reasons for such type of trials, for example, by joining the new treatment(s) and the one(s) already under study, it can save the sample size and costs as compared to running separate trials. Also some times it is desirable to compare the behavior of the newly available treatment(s) to the ongoing one(s) in the same trial study.

In such staggered-start clinical trial there are some basic questions, such as how to allocate the coming patients in a reasonable way? how different this type trial compared to the existing ones in basic properties, will some properties be kept with this new trial and under what conditions? A simple way in this trial is to allocate all the coming new patients to the new treatment(s) and letting the existing one(s) in waiting until all the treatments have roughly the same number of patients, then one can study the treatments like in the standard trials. But this method is known to be undesirable, as the subjects in study is human being, any design should aim at allocating more patients to treatment with best efficacy and minimize the sample size as possible; also the deterministic assignment in the waiting period violates the basic principle of clinical trial. Since the existing treatments have been in the study for relatively long time, they can not be treated as initial observations in relation to the newly added treatments. In this type of trial, we want to know if it is still possible to allocate the coming patients to the treatments to achieve a given optimality criterion, or to minimize the treatment losses as at least in some extent. These pose new problems in clinical trial study.

For standard clinical trial, the adaptive design uses accumulating data to update aspects of the study as it continues without undermining the validity and integrity of the trial [3, 17, 18]; among others). The optimal design is to achieve some targeting objective criteria for the allocation proportions [2, 4, 7, 8, 19, 20]. These methods have various merits in application, but it seems none of the existing methods addressed the case of staggered-start trial.

Here we propose and study a class of adaptive designs for the mentioned type of clinical trials, in which weights are used in the assignment of patients to the treatments to balance the current allocation proportions, in such a way that the allocation proportions of the treatments will asymptotically achieve, under suitable conditions, some given optimality criterion as in the standard trial. This design is applicable to any type of responses and simple to use. Our initial findings for such design are that if the sample sizes of the ongoing treatments are relatively small as compared to those of the newly joined treatments, then asymptotically the trial can achieve the given optimality criterion as in the standard trial with all treatments starting at the same time; otherwise the optimality can only be partially achieved in some sense. In Section 2, we give a brief review of some related commonly used existing methods for standard adaptive clinical trials. Section 3 describes the proposed allocation rule for staggered-start trial; investigates its asymptotic optimality behavior in Section 4, by distinguish two cases: the initial sample sizes are not too large relative to the total sample size, so that asymptotic optimality can be achieved; the initial sample sizes has roughly the same magnitude as the total sample size and asymptotic optimality cannot be achieved. Lastly, in Section 5 we illustrate these methods by some examples and a simulation study. A short discussion is given at the end. The relevant technical proofs are given in the Appendix.

2 Brief review of existing methods for standard clinical trials

Assume a standard clinical trials with k treatments. The patients come to the trial sequentially. Let $\mathbf{r}_n$ be the response (may be multiple) of the n -th patient under the assigned treatment, $f(\mathbf{r}_n)$ be the summary score for this response. Without loss of generality we assume $0 \leq f(\cdot) < \infty$, and bigger value of $f(\cdot)$ corresponds to better treatment effect. Let A_i be the event that the trial is under treatment i , $\mu_i = E(f(\mathbf{r}_n)|A_i)$ be the expected performance, or success rate, of the i -th treatment, $\sigma_i^2 = \text{Var}(f(\mathbf{r}_1)|A_i) < \infty$ be its variance ($i = 1, \dots, k$). Denote

$\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_k^2)$. Denote the number of patients assigned to treatment i at time n by $N_i(n)$, let $\mathbf{N}(n) = (N_1(n), \dots, N_k(n))$. The vector of allocation proportions $\mathbf{N}(n)/n$ is the main focus in phase III clinical trial studies, various methods are studied targeting the proportions to achieve, asymptotically, some specified optimality criteria.

For a vector $\mathbf{x} = (x_1, \dots, x_k)$, define $|\mathbf{x}| = x_1 + \dots + x_k$, and for $\mathbf{x}$ and $\mathbf{y}$ of the same length, define $\mathbf{x} - \mathbf{y} = (x_1 - y_1, \dots, x_k - y_k)$. The optimality criterion can be characterized by a specified object functional $G(\cdot, \cdot)$ of the response distributions under the treatments, or of their means $\boldsymbol{\mu}$. The target allocation proportions is the vector $\mathbf{v} = (v_1, \dots, v_k)$ which achieves the given criterion

$$\mathbf{v} = \arg \sup_{\mathbf{v} \in \nabla} G(\boldsymbol{\mu}, \mathbf{v}(\boldsymbol{\mu})),$$

where $\nabla = \{\mathbf{v}(\cdot) = (v_1(\cdot), \dots, v_k(\cdot)) : v_j(\cdot) > 0 (1 \leq j \leq k), |\mathbf{v}| = 1\}$ is the set of all possible vectors of proportions. In the optimal adaptive design, a special randomization rule is constructed such that the design will achieve the asymptotic optimality

$$\frac{\mathbf{N}(n)}{n} \rightarrow \mathbf{v}(a.s.). \quad (1)$$

Some examples of optimality criterion $G(\boldsymbol{\mu}, \mathbf{v}(\boldsymbol{\mu}))$ are given below.

Let I_{ji} be the indicator that the j -th patient is assigned to treatment i . Since $\boldsymbol{\mu}$ is unknown, we plug in its current estimate $\boldsymbol{\mu}_n$, with

$$\mu_{ni} = \sum_{j=1}^n f(\mathbf{r}_j) I_{ji} / \sum_{j=1}^n I_{ji}, \quad (1 \leq i \leq k).$$

More generally, we may consider the conditional distribution $F_i(\cdot)$ of the $f(\mathbf{r}_n)|A_i$'s. Let $\mathbf{F}(\cdot) = (F_1(\cdot), \dots, F_k(\cdot))$. Their empirical versions are

$$F_{n,i}(x_i) = \sum_{j=1}^n \chi(f(\mathbf{r}_j) \leq x_i) I_{ji} / \sum_{j=1}^n I_{ji}, \quad (i = 1, \dots, k)$$

where $\mathbf{x} = (x_1, \dots, x_k)$, and let $\mathbf{F}_n(\mathbf{x}) = (F_{n,1}(x_1), \dots, F_{n,k}(x_k))$, and $\chi(\cdot)$ is the indicator function.

Example 1. The functional $G(\boldsymbol{\mu}, \mathbf{v}(\boldsymbol{\mu})) = \sum_{j=1}^k \mu_j v_j(\boldsymbol{\mu}) = \boldsymbol{\mu}' \mathbf{v}(\boldsymbol{\mu})$ to be optimized is the average performance of all the treatments.

Example 2. Consider the case of integer response, let $\mathbf{q}$ be the allocation proportions and $\mathbf{q}_n$ be its empirical version, so $\boldsymbol{\mu} = \mathbf{q}$ and $\boldsymbol{\mu}_n = \mathbf{q}_n$. Let $c_i > 0$ be the (known) cost for the failure of the i -th treatment, and $\mathbf{I} = \{I_{ij} : 1 \leq i \leq n; 1 \leq j \leq k\}$, $\mathbf{a} = (1, \dots, 1)'$. The average weighted loss after the n -th split is

$$L_n(\mathbf{I}, \mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^k v_j^2 c_j (1 - q_j) I_{ij} / n, \quad \mathbf{v} \in \nabla.$$

Here we use a quadratic loss on $\mathbf{v}$ instead of a linear one, since the latter will assign all the weight to the smallest component. The object functional is the negative asymptotic risk

$$G(\mathbf{q}, \mathbf{v}(\mathbf{q})) = - \lim_n E_P L_n(\mathbf{I}, \mathbf{v}) = - \lim_n \sum_{j=1}^k v_j^2 c_j (1 - q_j) E_P(N_j(n)/n) = - \sum_{j=1}^k v_j^2 c_j (1 - q_j) v_{1,j}.$$

Example 3. Minimax design. Consider the integer response case ($k > 1$) and, for computational convenience, the weighted loss function,

$$L(\mathbf{I}, \mathbf{w}) = \sum_{i=1}^n \sum_{j=1}^k w_j (1 - q_j)^2 I_{ij} / n, \quad \mathbf{w} \in W,$$

where $W = \{\mathbf{w} : 0 < w_i, (1 \leq i \leq k); \sum_{i=1}^k w_i = 1\}$. The risk is asymptotically

$$R(\mathbf{q}, \mathbf{w}) = \sum_{i=1}^k w_i (1 - q_i)^2 v_{1,i}.$$

For fixed $\mathbf{w}$, by Lagrange's multiplier method, the maximum asymptotic risk is attained at $\mathbf{q}^*$ with

$$1 - q_i^* = (k-1) / \left[w_i v_{1,i} \sum_{j=1}^k (w_j v_j)^{-1} \right], \quad (i=1, \dots, k).$$

Plugging in this value we get the maximum asymptotic risk

$$R(\mathbf{q}^*, \mathbf{w}) = (k-1)^2 \left(\sum_{i=1}^k (w_i v_{1,i})^{-1} \right)^{-1}.$$

To minimize this maximum asymptotic risk, use Lagrange method again, we get

$$w_i^* = v_{1,i}^{-1/2} / \sum_{j=1}^k v_{1,j}^{-1/2}, \quad (i=1, \dots, k).$$

Now we choose

$$G(\mathbf{q}, \mathbf{w}(\mathbf{q})) = -R(\mathbf{q}^*, \mathbf{w}^*) = -(k-1)^2 \left(\sum_{i=1}^k v_i^{-1/2} \right)^{-2}.$$

However, in many situations, we are also interested in optimizing the evaluations of the treatments along with the optimization of the assignment proportions. Let $\xi_i = (\xi_{i1}, \dots, \xi_{ik})$ be the current evaluation of the k treatments if the coming patient is assigned treatment i , and $\mathbf{X}_n = (x_{n1}, \dots, x_{nk})$ be the cumulative evaluations of the ξ_i 's for the k treatments up to time n , it is called the urn composition at time n . Denote $|\mathbf{X}_n| = \sum_{j=1}^k x_{nj}$. For example, at time m , we can choose the ξ_i 's just be $\mathbf{v}_{m-1}$, independent of which treatment i is assigned at this time, see the Application section for the construction of the $\mathbf{v}_{m-1}$'s. More generally, we can only require the ξ_i be random with mean $\mathbf{v}$, or to be of some more general nature. A joint optimal assignment for $\mathbf{N}(n)/n$ and $\mathbf{X}_n/n$ to achieve the same criterion $\mathbf{v}$ is a version of the generalized Pólya urn (GPU) design (for example [7, 8]). In this design, to assign the coming n -th patient to one of the treatments, a random variable is drawn from the multinomial distribution with probabilities $\mathbf{X}_n/|\mathbf{X}_n|$. If it is of type i , the patient is assigned to the i -th treatment, a random vector of masses ξ_i is added to the current urn compositions $\mathbf{X}_n$, and the response $\mathbf{r}_n$ is used to update the estimate of $\boldsymbol{\mu}$ in the next step. Let $I\xi = (\xi_{ij})_{i,j=1,\dots,k}$ be the matrix representation of the ξ_{ij} 's, and for each n , and $I\xi_n$ be an i.i.d. version of $I\xi$. To simplify the expressions of the asymptotic variances to be derived later, we assume throughout this article that $I\xi$ is independent of the response observations. The random vector ξ_i is termed the adding rule, and $\mathbf{M} = E(I\xi) = (m_{ij})$ the design matrix with $m_{ij} = E(\xi_{ij})$ (known). Recall that a left eigenvector of a matrix $\mathbf{M}$ corresponding to an eigenvalue λ is the row vector $\mathbf{v}$ such that $\mathbf{v}\mathbf{M} = \lambda\mathbf{v}$ (in contrast a right eigenvector $\mathbf{u}$ is a column vector such that $\mathbf{M}\mathbf{u} = \lambda\mathbf{u}$). The vector is normalized if $|\mathbf{v}| = 1$. The first eigenvalue λ of the design matrix and its normalized first left (row) eigenvector $\mathbf{v}$ plays a key role in the asymptotic properties of the GPU design. Many authors (for instance [21–23]) studied asymptotic properties of $\mathbf{X}_n$ and $\mathbf{N}(n)$, and proved, under suitable conditions, that

$$(\mathbf{X}_n/|\mathbf{X}_n|, \mathbf{N}(n)/n) \rightarrow (\mathbf{v}, \mathbf{v}) \text{ (a.s.)}, \text{ or } (\mathbf{X}_n/n, \mathbf{N}(n)/n) \rightarrow (\lambda\mathbf{v}, \mathbf{v}) \text{ (a.s.)}$$

This give a joint optimality of $(\mathbf{X}_n/n, \mathbf{N}(n)/n)$. For a comprehensive review in this field, see Rosenberger and Lachin [1] and other related recent papers.

However, the existing GPU design can only achieve joint optimality of $(\mathbf{X}_n/n, \mathbf{N}(n)/n)$ for the same criterion $\mathbf{v}$. More generally, we may optimizing $\mathbf{X}_n/n$ and $\mathbf{N}(n)/n$ by different criteria, which will give us more flexibility. The compound adaptive GPU design [10] can achieve this goal, we won't pursue it here.

3 The staggered-start trial and the proposed design

In this case k treatments enter the trial at different times. After treatment 1 under study with n_1 patients, a new treatment 2 is added into the trial,..., after treatment $(k-1)$ under study with n_{k-1} patients, the newest k -th treatment enter the trial, with some initial size $n_{k,0}$. The previous treatments already under way for

long times, and the sample sizes $n_1, \dots, n_{k-1}$ cannot be viewed as initial sample sizes. The questions are can the optimality be achieved in this trial? under what conditions? how to allocate the incoming patients to achieve the overall optimality? It is known that any deterministic assignment is not desirable, so we need a randomized design for this problem. Below we describe the GPU designs for two treatments, three treatments, ..., k treatments accordingly. The general description may seem involved, the simulation example in Section 5 will make it simple and clear with two treatments. For easy of understanding, we first describe the design for two treatments, and then for the case of general k treatments.

3.1 Two treatments

At the 1st stage, only one treatment in the trial with sample size $n_1^{(1)}$. When the second stage begins, the second treatment enters the trial with an initial sample size $n_{2,0}$. The optimality under consideration can be summarized by a 2-dimensional vector $\mathbf{v}^{(2)}$, the urn composition at time n ($\geq n_1^{(1)} + n_{2,0}$) is $\mathbf{X}_{n,2} = (X_{n,1}, X_{n,2})$, the adding matrix is $\mathbf{M}_2$ (2×2 dimensional). Apparently we need to allocate more patients to the second treatment to allow its number of patients to catch up quickly. For this, at time $n = n_1 + n_{2,0}$, define a weight as $\mathbf{W}_{n,2} = (W_{n,1}, W_{n,2})$, with $W_{n,1} = n_{2,0}$, $W_{n,2} = n_1$. Define $\mathbf{W}_{n,2}\mathbf{X}_{n,2} = (W_{n,1}X_{n,1}, W_{n,2}X_{n,2})$, allocate the next $(n+1)$ -th coming patient to one of the two treatments with probability $\mathbf{W}_{n,2}\mathbf{X}_{n,2}/|\mathbf{W}_{n,2}\mathbf{X}_{n,2}|$, and update the weight $\mathbf{W}_{n,2}$ to $\mathbf{W}_{n+1,2}$ as $W_{n+1,1} = n_{2,0} + 1$, $W_{n+1,2} = W_{n,2}$, update $\mathbf{X}_n$ to $\mathbf{X}_{n+m}$ the same way as before, ..., allocate the $(n+m)$ -th patient to one of the two treatments with probability $\mathbf{W}_{n+m,2}\mathbf{X}_{n+m,2}/|\mathbf{W}_{n+m,2}\mathbf{X}_{n+m,2}|$, and update the weight $\mathbf{W}_{n+m-1,2}$ to $\mathbf{W}_{n+m,2}$ as below: if $W_{n+m-1,1} < W_{n+m-1,2}$, increase $W_{n+m-1,1}$ by 1 and $W_{n+m-1,2}$ be kept unchanged, otherwise increase $W_{n+m-1,2}$ by 1 and $W_{n+m-1,1}$ be kept unchanged; when $W_{n+m-1,1} = W_{n+m-1,2}$ reach uniform, we stop updating the weight vector and it stays uniform from then on. Update $\mathbf{X}_{n+m-1,2}$ to $\mathbf{X}_{n+m,2}$ the same way as before, and then allocate the next $(n+m+1)$ -th patient to one of the treatments by the probability $\mathbf{X}_{n+m+1,2}/|\mathbf{X}_{n+m+1,2}|, \dots$, until the end of the trial.

Note that in the above we may define the weights as $\mathbf{W}_{n+1,2} = (\frac{n_{2,0}+1}{n}, \frac{n_1}{n})$. Since we only need the ratio $\mathbf{W}_{n,2}\mathbf{X}_{n,2}/|\mathbf{W}_{n,2}\mathbf{X}_{n,2}|$ for the allocation probability, which result the same as the above defined weights, and also will be easier in our latter computation.

3.2 k Treatments

For the first two treatments in the trial, the procedure is the same as the above case. But to accommodate k treatments at latter stages, notationally we extend the vectors $\mathbf{v}_2$, $\mathbf{X}_{n,2}$ and $\mathbf{M}_2$ to their k -dimensional version as $\mathbf{v}^{(2)}$, $\mathbf{X}_n^{(2)}$ and $\mathbf{M}^{(2)}$, by adding zero components.

In the j -th stage ($j \geq 2$), the j -th treatment enters the trial. At this time the previous $j-1$ treatments already have been allocated to $n_1^{(j)}, \dots, n_{j-1}^{(j)}$ patients. The optimality under consideration can be summarized by a j -dimensional vector $\mathbf{v}_j$, with an urn composition at time n be $\mathbf{X}_{n,j} = (X_{n,1}, 0, \dots, 0, X_{n,j})$ and a $j \times j$ adding matrix $\mathbf{M}_j$. Notationally we extend $\mathbf{v}_j$, $\mathbf{X}_{n,j}$ and $\mathbf{M}_j$ to their k -dimensional version as $\mathbf{v}^{(j)}$, $\mathbf{X}_n^{(j)}$ and $\mathbf{M}^{(j)}$ by adding zero components for $j < k$. The urn composition at time n denoted as a k -vector $\mathbf{X}_n^{(j)}$, which is generated the way described before. Let $n_1^{[j]} \leq \dots \leq n_{j-1}^{[j]}$ be the ordered version of $n_1^{(j)}, \dots, n_{j-1}^{(j)}$. At the beginning of the j -th stage, let $\mathbf{W}_n^{[j]} = (W_{n,1}^{[j]}, \dots, W_{n,j}^{[j]}, 0, \dots, 0) = (n_1^{[j]}, \dots, n_{j-1}^{[j]}, n_1 + \dots + n_{j-1}, 0, \dots, 0)$ and $\mathbf{W}_n^{(j)}$ be the concomitant of $\mathbf{W}_n^{[j]}$ in relationship of $(n_1^{(j)}, \dots, n_{j-1}^{(j)}, n_1 + \dots + n_{j-1}, 0, \dots, 0)$ and $(n_1^{[j]}, \dots, n_{j-1}^{[j]}, n_1 + \dots + n_{j-1}, 0, \dots, 0)$. Allocate the next patient to one of the k treatments by the probability $\mathbf{W}_n^{(j)}\mathbf{X}_n/|\mathbf{W}_n^{(j)}\mathbf{X}_n|$, and increase $n_1^{[j]}$ by 1, and re-define $n_1^{[j]} \leq \dots \leq n_{j-1}^{[j]}$ after this step... Obviously, as n increases, $\mathbf{W}_n$ tends to the uniform weight, and when $\mathbf{W}_n/|\mathbf{W}_n| = (1/k, \dots, 1/k)$, we stop updating it.

At the k -th stage, there are already $(k-1)$ treatments under study with sample sizes $n_1^{(k)}, \dots, n_{k-1}^{(k)}$, and then the k -th treatment enters the trial. The optimality criterion is described by the k -vector $\mathbf{v} = \mathbf{v}(\boldsymbol{\mu})$. Let $n = n_1 + \dots + n_{k-1}$. The urn composition at time n denoted as a k -vector $\mathbf{X}_n$, which is generated the way described before. Let $n_1^{[k]} \leq \dots \leq n_{k-1}^{[k]}$ be the ordered version of $n_1^{(k)}, \dots, n_{k-1}^{(k)}$. At the beginning of the k -th

stage, let $\mathbf{W}_n^{[k]} = (W_{n,1}^{[k]}, \dots, W_{n,k}^{[k]}) = (n_1^{[k]}, \dots, n_{k-1}^{[k]}, n_1 + \dots + n_{k-1})$ and $\mathbf{W}_n^{(k)}$ be the concomitant of $\mathbf{W}_n^{[k]}$ in relationship of $(n_1^{(k)}, \dots, n_{k-1}^{(k)}, n_1 + \dots + n_{k-1})$ and $(n_1^{[k]}, \dots, n_{k-1}^{[k]}, n_1 + \dots + n_{k-1})$. Allocate the next patient to one of the k treatments by the probability $\mathbf{W}_n^{(k)} \mathbf{X}_n / |\mathbf{W}_n^{(k)} \mathbf{X}_n|$, and increase $n_1^{[k]}$ by 1, and re-define $n_1^{[k]} \leq \dots \leq n_{k-1}^{[k]}$ after this step... Obviously, as n increases, $\mathbf{W}_n$ tends to the uniform weight, and when $|\mathbf{W}_n|/|\mathbf{W}_n| = (1/k, \dots, 1/k)$, we stop updating it.

It is seen that by this way, the newly added treatments have a bigger chance to get incoming patients, and with more patients enters the trial, the weights will gradually even out. So, the rule describe above will allow the currently under assigned treatment(s) to quickly catch up while those over assigned treatment(s), so we expect the allocation proportions $\mathbf{N}(n)/n$ will finally converges to the optimality object $\mathbf{v}(\boldsymbol{\mu})$, and essentially the allocation is determined by the adding matrix $\mathbf{M}$, and eventually the optimality will be achieved in the way as the standard trial.

4 Asymptotic properties

Now we study the asymptotic properties of this design for quantitative responses, the case of qualitative responses should be parallel. We distinguish two cases, first the case $n_1, \dots, n_{k-1}$ are large but $n_j/n \rightarrow 0$ as $n \rightarrow \infty$; and secondly the case $\lim n_j/n > 0$ ($j = 1, \dots, k-1$). The asymptotic results are different for the two cases.

Case I. We first consider the case the initial sample sizes n_j 's are not too big relative to the total sample size n , and show that optimality can be achieved as in the corresponding standard clinical trials.

Let $\xrightarrow{D}$ denote convergence in distribution. For vectors $\mathbf{a} = (a_1, \dots, a_k)$ and $\mathbf{b} = (b_1, \dots, b_k)$ with $b_i \neq 0$ ($i = 1, \dots, k$), define $\mathbf{a}/\mathbf{b} = (a_1/b_1, \dots, a_k/b_k)$, and denote $\{\mathbf{a}\}$ the diagonal matrix for $\mathbf{a}$. We impose the following conditions

(B0) $0 < v_{ij} < \infty$, ($1 \leq i, j \leq k$).

(B1) $\mathbf{M}\mathbf{1}' = \lambda_1 \mathbf{1}'$.

(B2) $\varsigma_{ij}(\cdot, \cdot)$ is continuous at $(\boldsymbol{\mu}, \mathbf{y})$ for any $\mathbf{y}$ and there is an $r > 2$ such that $E\|\boldsymbol{\xi}_1\|^r < \infty$ and $\sup_n E(\|\zeta_n\|^r | \mathcal{F}_{n-1}) < \infty$, a.s.

(B3) $\mathbf{M}(\boldsymbol{\mu}) \geq \mathbf{0}$ ($\neq \mathbf{0}$) is differentiable, and there is a $\delta > 0$ such that

$$\mathbf{M}(\mathbf{y}) - \mathbf{M}(\boldsymbol{\mu}) = \sum_{j=1}^k \frac{\partial \mathbf{M}(\mathbf{y})}{\partial y_j} \big|_{\mathbf{y}=\boldsymbol{\mu}} (y_j - \mu_j) + O(\|\mathbf{y} - \boldsymbol{\mu}\|^{1+\delta}), \text{ as } \mathbf{y} \rightarrow \boldsymbol{\mu}.$$

(B4) $\prod_{j=1}^{k-1} n_j = o(n^{1/2})$.

The following results show that under assumption (B4), the staggered-start trial achieves the optimality criterion and with asymptotic behavior of $(\frac{\mathbf{X}_n}{n}, \frac{\mathbf{N}(n)}{n})$ be the same as those for the standard trial.

Theorem 1. Assume (B0)-(B2) (for some $r > 1$) and (B4), then we have

(i) $\boldsymbol{\mu}_n \rightarrow \boldsymbol{\mu}$, a.s.; $\sup_x \|\mathbf{F}_n(\mathbf{x}) - \mathbf{F}(\mathbf{x})\| \rightarrow 0$, a.s.

(ii) $\left(\frac{\mathbf{X}_n(\boldsymbol{\mu}_{n-1})}{n}, \frac{\mathbf{N}(n)}{n}\right) \rightarrow (\lambda_1 \mathbf{v}(\boldsymbol{\mu}), \mathbf{v}(\boldsymbol{\mu}))$ a.s.

(iii) $\frac{\sum_{i=1}^k \sum_{j=1}^n f(\mathbf{x}_j) I_{ji}}{n} \rightarrow \mathbf{v}(\boldsymbol{\mu}) \boldsymbol{\mu}'$, a.s.

Theorem 2. (i) Assume (B0)-(B2) and (B4), we have

$$\sqrt{n}(\boldsymbol{\mu}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_\mu), \sqrt{n}(\mathbf{F}_n(\mathbf{x}) - \mathbf{F}(\mathbf{x})) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Omega}_F), \mathbf{x} \in \mathbf{S}(\mathbf{F}),$$

where $\Omega_\mu = \{\sigma^2/\mathbf{v}\}$, and $\Omega_F = \{\mathbf{F}(\mathbf{x})(1 - \mathbf{F}(\mathbf{x}))/\mathbf{v}\}$.

(ii) Assume (B0)-(B4). If $\lambda_1 > 2\text{Re}(\lambda_2)$, then

$$\sqrt{n} \left(\frac{\mathbf{X}_n(\boldsymbol{\mu}_{n-1})}{n} - \lambda_1 \mathbf{v}(\boldsymbol{\mu}), \frac{\mathbf{N}(n)}{n} - \mathbf{v}(\boldsymbol{\mu}) \right)' \xrightarrow{D} N(\mathbf{0}, \Omega),$$

where Ω is given in the proof.

Remark 1) When $\lambda_1 \leq 2\text{Re}(\lambda_2)$, the convergence rate is generally slower than $n^{1/2}$.

2). The asymptotic covariance matrix Ω can be made very simple by choice of $\mathbf{M}$, as in the Corollary in Yuan and Chai [8], as below. Let $\mathbf{v}_1(\cdot)$ be twice differentiable with $|\mathbf{v}_1(\cdot)| = 1$. Set $\mathbf{M}(\cdot) = \mathbf{1}'\mathbf{v}_1(\cdot)$, Ξ be a constant matrix, then the conditions of Theorem 2(ii) are satisfied with $\lambda_1 = 1$ and $\lambda_2 = 0$, and Ω is simplified as

$$\begin{aligned} \Omega_{11} &= 2\Sigma_v, \quad \Omega_{12} = 3\Sigma_v(\mathbf{I} - \mathbf{1}'\mathbf{v}_1), \\ \Omega_{22} &= \{\mathbf{v}_1\} - \mathbf{v}_1'\mathbf{v}_1 + 6(\mathbf{I} - \mathbf{v}_1'\mathbf{1})\Sigma_v(\mathbf{I} - \mathbf{1}'\mathbf{v}_1), \end{aligned}$$

where $\mathbf{I}$ is the k -dimensional identity matrix, $\Sigma_v = (\partial\mathbf{v}_1(\boldsymbol{\mu})/\partial\boldsymbol{\mu})' \text{diag}(\sigma_1^2/v_{11}, \dots, \sigma_k^2/v_{1k}) \partial\mathbf{v}_1(\boldsymbol{\mu})/\partial\boldsymbol{\mu}$. Especially, when take $\zeta = \text{Const.}$, then $E(\zeta'\{\mathbf{v}_1\}\zeta) - \mathbf{v}_1'\mathbf{v}_1 = \mathbf{0}$ and Ω_{ij} 's can be simplified further.

Case II. In this case, $\lim n_i/n \rightarrow \alpha_i > 0$ ($i = 1, \dots, k-1$), but since the stopping time n is unknown in advance, the proportions $\alpha_1, \dots, \alpha_{k-1}$ are only known when the trial terminates, otherwise the design can be made simpler by using the α_j 's. We will see that in this case, the optimality criterion can only be partially achieved, in that the targeting limits are partially kept unchanged, as seen by comparing results of Theorems 3 and 4 vs those in Theorems 1 and 2.

Theorem 3. Assume $\mathbf{M}_j = \mathbf{1}'\mathbf{v}^{(j)}$, then

- (i) $\boldsymbol{\mu}_n \rightarrow \boldsymbol{\mu}$, a.s.; $\sup_x \|\mathbf{F}_n(x) - \mathbf{F}(x)\| \rightarrow 0$, a.s.
- (ii) with $(\mathbf{y}, \mathbf{u})$ given in the proof,

$$\begin{aligned} & \left(\frac{\mathbf{X}_n(\boldsymbol{\mu}_{n-1})}{n}, \frac{\mathbf{N}(n)}{n} \right) \rightarrow (\mathbf{y}, \mathbf{u}) \quad \text{a.s.}, \\ \text{(iii)} \quad & \frac{\sum_{i=1}^k \sum_{j=1}^n f(\mathbf{x}_j) I_{ji}}{n} \rightarrow \mathbf{v}(\boldsymbol{\mu})\boldsymbol{\mu}', \quad \text{a.s.} \end{aligned}$$

Theorem 4. Assume $\mathbf{M}_j = \mathbf{1}'\mathbf{v}^{(j)}$ and $n_j/n = \alpha_j + o(n^{-1/2})$, then

$$\sqrt{n} \left(\frac{\mathbf{X}_n(\boldsymbol{\mu}_{n-1})}{n} - \mathbf{y}, \frac{\mathbf{N}(n)}{n} - \mathbf{u} \right)' \xrightarrow{D} N(\mathbf{0}, \bar{\Omega}),$$

where $\bar{\Omega}$ is given in the proof.

In Theorem 3 (ii), $(\mathbf{y}, \mathbf{u}) \neq (\mathbf{v}, \mathbf{v})$. Thus, if the sample sizes n_j 's are not small in relation to the total sample size n , full optimality cannot be achieved.

5 Application examples and simulation study

In practice $\mathbf{v}(\cdot)$ should be chosen differentiable around $\boldsymbol{\mu} \neq \mathbf{0}$, and hence satisfies the required conditions. We will continue the three examples given in Section 2 and with two more examples in common applications, with a simulation study for the last example.

Example 1 (continued). The functional $G(\boldsymbol{\mu}, \mathbf{v}(\boldsymbol{\mu})) = \sum_{j=1}^k \mu_j v_j(\boldsymbol{\mu}) = \boldsymbol{\mu}' \mathbf{v}(\boldsymbol{\mu})$. Typically, the components of $\mathbf{v}$ assigns larger values to larger components of $\boldsymbol{\mu}$. By Theorem 3, the “success rate” is asymptotically

$$G(\boldsymbol{\mu}_n, \mathbf{v}(\boldsymbol{\mu}_n)) = \boldsymbol{\mu}_n' \mathbf{v}(\boldsymbol{\mu}_n) \rightarrow \boldsymbol{\mu}' \mathbf{v}(\boldsymbol{\mu}), a.s.$$

To maximize this quantity over $\mathbf{v}(\cdot)$, one may wish to assign all the mass of $\mathbf{v}$ to the largest component (or components, in case of ties) of $\boldsymbol{\mu}$, but this will generally fail the conditions for consistency, which usually requires $v_i(\cdot) > 0$. So we assign $v_i(\cdot) > 0$ by some fixed rule, and let the first left eigenvector to be: $v_{1i}(\boldsymbol{\mu}_{n-1}) = v_i(\boldsymbol{\mu}_{n-1}) (1 \leq i \leq k)$ and normalize it, still denote it by $\mathbf{v}(\cdot)$.

Example 2 (continued). The object functional is the negative asymptotic risk

$$G(\mathbf{q}, \mathbf{v}(\mathbf{q})) = - \sum_{j=1}^k v_j^2 c_j (1 - q_j) v_{1,j}.$$

We are to maximize the object functional over $\mathbf{v} \in \nabla$ and $\mathbf{v}_1 \in \nabla$. Using Lagrange’s multiplier, we get the theoretical optimal $(\mathbf{v}(\mathbf{q}), \mathbf{v}_1(\mathbf{q}))$ with components

$$v_{1,i}(\mathbf{q}) = v_i(\mathbf{q}) = \frac{1}{\sqrt{c_i(1-q_i)}} \left(\sum_{j=1}^k \frac{1}{\sqrt{c_j(1-q_j)}} \right)^{-1} \quad (i = 1, \dots, k)$$

Now, given $\mathbf{q}_{n-1} = (q_{n-1,1}, \dots, q_{n-1,k})'$, we get the optimal design with

$$v_{n-1,1,i} = \frac{1}{\sqrt{c_i(1-q_{n-1,i})}} \left(\sum_{j=1}^k \frac{1}{\sqrt{c_j(1-q_{n-1,j})}} \right)^{-1}, \quad (i = 1, \dots, k).$$

Example 3 (continued). The optimality target functional is

$$G(\mathbf{q}, \mathbf{w}(\mathbf{q})) = -R(\mathbf{q}^*, \mathbf{w}^*) = -(k-1)^2 \left(\sum_{i=1}^k v_i^{-1/2} \right)^{-2}$$

and we are to find $\mathbf{v}_1$ to minimize this quantity. Using the Lagrange method again, we get $\mathbf{v}_1$ for the minimax design $v_{1,i} = 1/k$, $(i = 1, \dots, k)$, which is a fixed design.

Example 4. Consider two qualitative treatments. The Neyman allocation [3] is to allocate $N_1(n)$ patients to treatment I by maximizing the power for testing the difference $\Delta = p_1 - p_2$, where $\mathbf{p} = (p_1, p_2)$ are the success rates of the two treatments. Here $f(\cdot) \equiv 1$ and $\boldsymbol{\mu} = \mathbf{p}$. This strategy leads to the allocation ratio

$$\frac{N_1(n)}{N_2(n)} \rightarrow \sqrt{\frac{p_1(1-p_1)}{p_2(1-p_2)}} =: \frac{v_{11}}{v_{12}},$$

or $\mathbf{v}_1(\mathbf{p}) = (v_{11}(\mathbf{p}), v_{12}(\mathbf{p}))$, with

$$v_{11}(\mathbf{p}) = \frac{\sqrt{p_1(1-p_1)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}, \quad v_{12}(\mathbf{p}) = \frac{\sqrt{p_2(1-p_2)}}{\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)}}.$$

We have $|\mathbf{v}_1| = 1$, and the corresponding optimal $\mathbf{v}_1(\cdot) = (v_{11}(\cdot), v_{12}(\cdot))'$. The generating matrix $M(\cdot)$ can be chosen as in the Corollary, $M(\cdot) = \mathbf{v}_1(\cdot) \mathbf{1}'$. Take $\lambda_1 = 1$, and the $\xi_{n,ij}$ ’s be constants, the corresponding adding rule is $\zeta_{ni} = \mathbf{v}_1(\mathbf{p}_{n-1})$ ($i = 1, 2$). It is easily checked that all the conditions in Theorem 3 and the Corollary are satisfied, so we have

$$\left(\frac{X_{n,1}}{n}, \frac{X_{n,2}}{n} \right) \rightarrow (v_{11}(\mathbf{p}), v_{12}(\mathbf{p})) \leftarrow \left(\frac{N_{n,1}}{n}, \frac{N_{n,2}}{n} \right). \quad a.s.$$

Note $\sigma_1^2 = p_1(1-p_1)$ and $\sigma_2^2 = p_2(1-p_2)$, so by the Corollary we have $\sqrt{n}(\mathbf{X}_n/n, \mathbf{N}(n)/n) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\Omega} = (\Omega_{ij})_{1 \leq i, j \leq 2}$, $\Omega_{12} = \Omega'_{21}$,

$$\mathbf{\Omega}_{11} = 2\Sigma_V = \frac{(1-2p_1)^2\sqrt{p_2(1-p_2)} + (1-2p_2)^2\sqrt{p_1(1-p_1)}}{2(\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)})^3} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

$$\mathbf{\Omega}_{12} = 3\Sigma_V(\mathbf{I} - \mathbf{1}\mathbf{1}') = \frac{3}{4} \frac{(1-2p_1)^2\sqrt{p_2(1-p_2)} + (1-2p_2)^2\sqrt{p_1(1-p_1)}}{(\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)})^3} \begin{pmatrix} 1, & -1 \\ -1, & 1 \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{\Omega}_{22} &= \{\mathbf{v}_1\} - \mathbf{v}_1'\mathbf{v}_1 + 6(\mathbf{I} - \mathbf{v}_1'\mathbf{1})\Sigma_V(\mathbf{I} - \mathbf{1}\mathbf{v}_1') \\ &= \left(\frac{\sqrt{p_1p_2(1-p_1)(1-p_2)}}{(\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)})^2} + \frac{3[(1-2p_1)^2\sqrt{p_2(1-p_2)} + (1-2p_2)^2\sqrt{p_1(1-p_1)}]}{2(\sqrt{p_1(1-p_1)} + \sqrt{p_2(1-p_2)})^3} \right) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \end{aligned}$$

Example 5. The criterion in Rosenberger et al. [4] is to minimize the expected treatment failure and leads to

$$\frac{N_1(n)}{N_2(n)} \rightarrow \sqrt{\frac{p_1}{p_2}} := \frac{v_{11}}{v_{12}} \text{ or } v_{12}(\mathbf{p}) = \left(1 + \sqrt{\frac{p_1}{p_2}}\right)^{-1},$$

so we take the targeting $\mathbf{v}_1(\cdot)$ as

$$v_{11}(\mathbf{p}) = \frac{\sqrt{p_1}}{\sqrt{p_1} + \sqrt{p_2}}, \quad v_{12}(\mathbf{p}) = \frac{\sqrt{p_2}}{\sqrt{p_1} + \sqrt{p_2}}.$$

The $f(\cdot)$, $\boldsymbol{\mu}$, the generating matrix and the adding rule are the same as in Example 1, with $\mathbf{v}_1$ given here. The asymptotic results are similar, with

$$\mathbf{\Omega}_{11} = \frac{(1-p_1)p_2^{3/2} + (1-p_2)p_1^{3/2}}{2\sqrt{p_1p_2}(\sqrt{p_1} + \sqrt{p_2})^3} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

$$\mathbf{\Omega}_{12} = \frac{3}{4} \frac{(1-p_1)p_2^{3/2} + (1-p_2)p_1^{3/2}}{\sqrt{p_1p_2}(\sqrt{p_1} + \sqrt{p_2})^3} \begin{pmatrix} 1, & -1 \\ -1, & 1 \end{pmatrix}$$

and

$$\mathbf{\Omega}_{22} = \frac{(3-p_1)p_2^{3/2} + (3-p_2)p_1^{3/2}}{2\sqrt{p_1p_2}(\sqrt{p_1} + \sqrt{p_2})^3} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

5.1 Simulation study

Below we perform a simulation study for Example 5. At the end of the 1st stage of the trial, treatment 1 has been under study with $n_1^{(1)} = 100$ patients, with responses $\mathbf{r}_{1, n_1^{(1)}} = (r_{11}, \dots, r_{1, n_1^{(1)}})$. In the 2nd stage, treatment 2 enters the trial with initial patient size $n_{2,0} = 15$, with responses $\mathbf{r}_{2, n_{2,0}} = (r_{21}, \dots, r_{2, n_{2,0}})$. The responses for the two treatments are generated from $N(\mu_1, 1)$ and $N(\mu_2, 1)$ respectively, with $\mu_1 = 2.5$ and $\mu_2 = 3$. We define the success of a treatment as a response ≥ 2 . Thus the success rates $\mathbf{p} = (p_1, p_2) = (0.6915, 0.8413)$, and targeting allocation proportion vector is $\mathbf{v}_1 = (v_{11}(\mathbf{p}), v_{12}(\mathbf{p}))$ given in Example 5.

When treatment 2 enters the trial, the total sample size is $n_1^{(1)} + n_{2,0}$. General time $n = n_1 + n_2$, where n_1 is the total number of patients allocated to treatment 1 upto time n (including the $n_1^{(1)}$ patients), and n_2 is the total number of patients allocated to treatment 2 upto time n (including the $n_{2,0}$ patients). We use the adding

matrix with $\lambda_1 = 1$ as in Remark 1 to make it simple to use, then urn composition at time n is $\mathbf{X}_n = (x_{n1}, x_{n2})'$, with $x_{n1} = nv_{11}(\mathbf{p}_n)$, $x_{n2} = nv_{12}(\mathbf{p}_n)$, $\mathbf{p}_n = (p_{1, n_1}, p_{2, n_2})$, and

$$p_{1, n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} I(r_{1i} \geq 2), p_{2, n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} I(r_{2i} \geq 2)$$

are the empirical estimates for p_1 and p_2 , $I(\cdot)$ is the indicator function. For the weight, when treatment 2 first enters the trial, $\mathbf{W}_n = (w_{n1}, w_{n2}) = (n_{2,0}, n_1^{(1)})$. Denote $\mathbf{Q}_n = \mathbf{W}_n \mathbf{X}_n / |\mathbf{W}_n \mathbf{X}_n|$ (recall for a vectors $\mathbf{a} = (a_1, \dots, a_k)$ and $\mathbf{b} = (b_1, \dots, b_k)$, $\mathbf{ab} = (a_1 b_1, \dots, a_k b_k)$, and $|\mathbf{a}| = a_1 + \dots + a_k$), then

$$\mathbf{Q}_n = \frac{(w_{n1} v_{11}(\mathbf{p}_n), w_{n2} v_{12}(\mathbf{p}_n))}{w_{n1} v_{11}(\mathbf{p}_n) + w_{n2} v_{12}(\mathbf{p}_n)}.$$

When the next $(n+1)$ -th patient comes, allocate this patient to treatment 1 or 2 according to the probability $\mathbf{Q}_n$.

After this patient been allocated, record his/her response $r_{2, n_{2,0}+1}$, if he/she is allocated to treatment 2 (or $r_{1, n_1^{(1)}+1}$, if he/she is allocated to treatment 1), then update $\mathbf{p}_n$ to $\mathbf{p}_{n+1}$ accordingly. Update the weight $\mathbf{W}_n$ to $\mathbf{W}_{n+1} = (w_{n+1,1}, w_{n+1,2}) = (n_{2,0} + 1, n_1^{(1)})$, and update $\mathbf{Q}_n$ to $\mathbf{Q}_{n+1}$ accordingly with $(\mathbf{W}_n, \mathbf{p}_n)$ replaced by $(\mathbf{W}_{n+1}, \mathbf{p}_{n+1})$. Then allocate the $(n+2)$ -th coming patient to one of the two treatments by probability $\mathbf{Q}_{n+1}, \dots$, continue this way until $w_{n1} = w_{n2}$, then $\mathbf{W}_n$ is fixed, and then $\mathbf{Q}_n = (v_{11}(\mathbf{p}_n), v_{12}(\mathbf{p}_n))$ after this n . We continue allocate the each coming patient with probability $\mathbf{Q}_n$, and update it (along with the updating of $\mathbf{p}_n$) after each allocation, till the end of the trial.

The allocation proportion at time n is $\mathbf{N}(n)/n = (n_1, n_2)/n$, where n_1 is the total number of all the patients allocated to treatment 1 at time n (including the $n_1^{(1)}$ patients in the first stage), and n_2 is the total number of patients allocated to treatment 2 at time n (including the $n_{2,0}$ initial patients). We want $\mathbf{N}(n)/n \rightarrow \mathbf{v}(\mathbf{p})$ as n increases. The results are presented in Table 1.

Table 1: Allocation proportion with simulated data.

n	n_1/n	n_2/n
115	(0.86957,	0.13043)
150	(0.70667,	0.29333)
200	(0.62500,	0.37500)
300	(0.56667,	0.43333)
500	(0.55400,	0.44600)
700	(0.51429,	0.48571)
900	(0.50889,	0.49111)
1,000	(0.51900,	0.48100)
2,000	(0.50750,	0.49250)
5,000	(0.48040,	0.51960)
8,000	(−.46825,	0.53175)
10,000	(0.47430,	0.52570)
20,000	(−.47085,	0.52915)
50,000	(0.47708,	0.52292)
$\mathbf{v}(\mathbf{p}) =$	(0.47550,	0.5245)

We see that when treatment 2 enters the trial, the allocation proportion is highly skewed to treatment 1. As new patients come in, with the weighted allocation design, the proportion catches up slowly; when the total sample size surpass 300, the difference between two proportions becomes not so significant, when sample size reaches 1,000, the two proportions are close; and when it surpasses 2,000, the proportions begins to shift toward treatment 2, as this treatment has higher success rate. Although in most phase III clinical trial, sample size rarely surpass 5,000, this simulation experiment gives us some sense for how fast the asymptotic results for staggered-start trial can stabilize to its optimal target.

6 Discussion

In this study, we studied adaptive design for staggered-start clinical trial. We used weighted assignment to accelerate the allocation of patients to the newly entered treatment(s). The choice of our weight is intuitive, but different weights can be used. For example, put more heavy weight on the new treatment(s) to let their sample size(s) catch up more quickly. An extreme is to allocate all incoming patients to the new treatment until it has about the same sample size as the early started one(s). As mentioned before, this type of design with a deterministic portion violates the basic principle in clinical trial. Another extreme is allocate the incoming patients using the classical rule, just as all the treatments are started at the same time. As we pointed out, and seen from the simulation example, this design can not achieve, or approximate a given target allocation proportion. So a natural question is: is there an optimal weight for the staggered-start trial? This question seems not simple. First, one should to have a criterion as for optimal in what sense. It should also depend on the number of treatments, how many of them are late started, the sample sizes for the early started treatments, etc. This question can be a good research topic in our future studies.

The choice of the design criteria is an old topic. We listed five commonly used such criteria, and there are many more in application and research literature, a comprehensive review on this can be found in, for example, Rosenberger and Lachin [1], and Jennison and Turnbull [18], Chapter 17).

Another interesting question is: what is a cutoff value on when it is not suggested to enter a new treatment to staggered-start clinical trials? Again this question is not simple. It will depend on the motivation and plan of the trial process. In terms of achieving, or approximating, the targeted allocation proportions, based on our intuition and simulation results, it seems that if the ongoing trial is already under way more than one third of the whole planned process, i.e., the number of patients in the trial already surpassed one third of the planned total number of patients, it is not adequate to start a new treatment to join the trial. However, this is just a referential suggestion, the actual decision should depend on many more technical and ethical factors.

Appendix

Let $N_0 = n_1 + \dots + n_{k-1} + 1$, $\mathbf{I}_n = (I_{n1}, \dots, I_{nk})$ and $\mathcal{F}_n = \sigma(\mathbf{r}_1, \dots, \mathbf{r}_n; \mathbf{I}_1, \dots, \mathbf{I}_n)$ be the sigma field generated by the underlined variables.

Proof of Theorem 1: (i) We show $\mu_{ni} \rightarrow \mu_i$ (a.s.) for each $i = 1, \dots, k$. In fact,

$$\mu_{ni} = \frac{\sum_{j=1}^n f(r_j) I_{ji}}{\sum_{j=1}^n I_{ji}} = \frac{1}{N_i(n)} \sum_{j=1}^{N_i(n)} f(r_j | A_i),$$

thus $\mu_{ni} \rightarrow \mu_i$ (a.s.) as long as $N_i(n) \rightarrow \infty$ (a.s.) which we show below.

We only need to consider $n > N_0$, so the $\mathbf{W}_n$'s and $\mathbf{X}_n$'s are all k -dimensional. By the construction of $\mathbf{W}_n$, as n increases to $n+1$, only the smallest component of $\mathbf{W}_n$ gets one increment, the others stay unchanged. Since $|\mathbf{W}_n| = n$, so for $n > N_1$, $\mathbf{W}_n/n$ will be uniform, i.e., $W_{n,i}/n = 1/k$ ($i = 1, \dots, k$); where $N_1 = \sum_{j=1}^{k-1} \prod_{i=j}^{k-1} n_{(i)}$ and $n_{(1)} \leq \dots \leq n_{(k-1)}$ is the ordered sample sizes of $n_1, \dots, n_{k-1}$.

Note that the sequence of the adding components to $\mathbf{X}_n$ after centering, $\{\zeta_n - E(\zeta_n | \mathcal{F}_{n-1}); n \geq N_0\}$ is a matrix martingale, with

$$\sum_{n=N_0}^{\infty} \frac{E\|\zeta_n - E(\zeta_n | \mathcal{F}_{n-1})\|^r | \mathcal{F}_{n-1}}{(n - N_0)^r} \leq \sum_{n=1}^{\infty} \frac{C}{(n - N_0)^r} < \infty$$

by (B2). So by the law of large numbers for martingales,

$$\frac{1}{n-N_0} \sum_{m=N_0}^n \zeta_m - \frac{1}{n-N_0} \sum_{m=N_0}^n E(\zeta_m | \mathcal{F}_{m-1}) \rightarrow \mathbf{0} \quad (a.s.).$$

Since $E(\zeta_m | \mathcal{F}_{m-1}) = \mathbf{M}(\boldsymbol{\mu}_{m-1})$, we have $\sum_{m=N_0}^n \zeta_m = \sum_{m=N_0}^n \mathbf{M}(\boldsymbol{\mu}_{m-1}) + o(n-N_0)\mathbf{1}'\mathbf{1} = O(n-N_0)\mathbf{1}'\mathbf{1} + o(n-N_0)\mathbf{1}'\mathbf{1}$ (a.s.). Since components of $\mathbf{M}(\boldsymbol{\mu})$ are positive, $\sum_{m=N_0}^n \mathbf{M}(\boldsymbol{\mu}_{m-1}) = O(n-N_0)\mathbf{1}'\mathbf{1}$ (a.s.), so $|\mathbf{X}_n| = O(|\mathbf{X}_{N_0}| + \sum_{m=N_0+1}^n k^{-1}|\zeta_m|) = O(|\mathbf{X}_{N_0}| + \sum_{m=N_0+1}^n k^{-1}|\mathbf{M}(\boldsymbol{\mu}_{m-1})| + o(n-N_0)) = O(n-N_0)$ (a.s.), where $|\mathbf{M}(\cdot)| = \sum_{i,j} m_{ij}(\cdot)$. Hence $\sum_{n=N_0}^\infty |\mathbf{X}_n|^{-1} = \infty$ (a.s.) and so, for some $c > 0$,

$$\begin{aligned} \sum_{n=N_0}^\infty P(I_{ni} = 1 | \mathcal{F}_{n-1}) &= \sum_{n=N_0}^\infty W_{n-1,i} X_{n-1,i} / |\mathbf{W}_{n-1} \mathbf{X}_{n-1}| \\ &\geq c \sum_{n=N_0}^\infty \mathbf{X}_{n-1,i} \left(\frac{1}{k} + O\left(\frac{1}{n-N_0}\right) \right) / \left(|X_{n-1}| \left(\frac{1}{k} + O\left(\frac{1}{n-N_0}\right) \right) \right) \geq c X_{N_0-1,i} \sum_{n=N_0}^\infty |\mathbf{X}_{n-1}|^{-1} = \infty, \quad (as) \end{aligned}$$

this, with the generalized Borel-Cantelli lemma for martingale (Corollary 2.3 [24]), implies that $P(I_{n,i} = 1, i.o.) = 1$, or $N_{n,i} \rightarrow \infty$ (a.s.).

For the second result, since $F_{n,i}(x) = \frac{1}{N_i(n)} \sum_{j=1}^{N_i(n)} \chi(f(\mathbf{r}_j) \leq x | A_i)$, and $N_i(n) \rightarrow \infty$ (a.s.), the result is direct from the uniform almost sure convergence of empirical distribution.

(ii). Define $\mathbf{S}_n = \sum_{m=1}^n \Delta \mathbf{S}_m$ with $\Delta \mathbf{S}_k = \mathbf{I}_k \zeta_k - E(\mathbf{I}_k \zeta_k | \mathcal{F}_{k-1})$. Then $\{\mathbf{S}_n, \mathcal{F}_{n-1} : n \geq N_0\}$ is a martingale. The proof is modified from that in Zhang et al. [7]. Denote $\mathbf{M}_k = \mathbf{M}(\boldsymbol{\mu}_k)$, and note $E(\mathbf{I}_k \zeta_k | \mathcal{F}_{k-1}) = \frac{W_{k-1} X_{k-1}}{|W_{k-1} X_{k-1}|} \mathbf{M}_k$. We have

$$\mathbf{X}_n = \mathbf{X}_{n-1} + \mathbf{I}_n \zeta_n = \mathbf{X}_{N_0} + \sum_{m=N_0+1}^n \mathbf{I}_m \zeta_m = \mathbf{X}_{N_0} + \mathbf{S}_n + \sum_{m=N_0}^{n-1} \frac{\mathbf{W}_m \mathbf{X}_m}{|\mathbf{W}_m \mathbf{X}_m|} \mathbf{M}_{m+1}.$$

Let $\bar{\mathbf{M}} = \mathbf{M} - \lambda_1 \mathbf{1}'\mathbf{v}$. Note $\mathbf{v}\mathbf{1}'\mathbf{v} = \mathbf{v}$, $\mathbf{v}\bar{\mathbf{M}} = \mathbf{0}$, $\mathbf{v}\mathbf{M} = \lambda_1 \mathbf{v}$, $\sum_{m=N_0}^{n-1} \left(\frac{\mathbf{W}_m \mathbf{X}_m}{|\mathbf{W}_m \mathbf{X}_m|} \right) \lambda_1 \mathbf{1}'\mathbf{v} = \lambda_1 \sum_{m=N_0}^{n-1} \left(\frac{|\mathbf{W}_m \mathbf{X}_m| \mathbf{v}}{|\mathbf{W}_m \mathbf{X}_m|} - \mathbf{v} \right) = \mathbf{0}$ and $\frac{\mathbf{W}_m \mathbf{X}_m}{|\mathbf{W}_m \mathbf{X}_m|} = \frac{\mathbf{X}_m}{|\mathbf{X}_m|} (1 + o(m^{-1}))$ (a.s.). We have

$$\begin{aligned} \mathbf{X}_n - (n-N_0)\lambda_1 \mathbf{v} &= \mathbf{S}_n + \sum_{m=N_0}^{n-1} \left(\frac{\mathbf{W}_m \mathbf{X}_m}{|\mathbf{W}_m \mathbf{X}_m|} - \mathbf{v} \right) \mathbf{M}_{m+1} + \sum_{m=1}^n \mathbf{v}(\mathbf{M}_m - \mathbf{M}) + \mathbf{X}_{N_0} \\ &= \mathbf{S}_n + \sum_{m=N_0}^{n-1} \left(\frac{\mathbf{X}_m}{|\mathbf{X}_m|} - \mathbf{v} \right) \mathbf{M}_{m+1} + \sum_{m=N_0+1}^n \mathbf{v}(\mathbf{M}_m - \mathbf{M}) + \mathbf{X}_{N_0} + \sum_{m=N_0}^{n-1} o(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} \mathbf{M}_{m+1} \quad (2) \\ &:= \mathbf{A}_n + \sum_{m=N_0}^{n-1} o(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} \mathbf{M}_{m+1}, \quad (a.s.) \end{aligned}$$

Similarly as in Zhang et al. [7], $\mathbf{A}_n = o(n-N_0)\mathbf{1}$ (a.s.). Apparently, as $\frac{\mathbf{X}_m}{|\mathbf{X}_m|} \mathbf{M}_{m+1}$ is bounded, the last term in eq. (2) is also $o(n-N_0)\mathbf{1}$ (a.s.), and so the claimed result is true.

Now we prove $\mathbf{N}(n)/n \rightarrow \mathbf{v}$ (a.s.). Define $\mathbf{m}_n = \sum_{k=1}^n \Delta \mathbf{m}_k$, with $\Delta \mathbf{m}_k = \mathbf{I}_k - E(\mathbf{I}_k | \mathcal{F}_{k-1})$. Then $\{\mathbf{m}_n, \mathcal{F}_{n-1} : n \geq N_0\}$ is a matrix martingale, and $E(\mathbf{I}_k | \mathcal{F}_{k-1}) = \frac{\mathbf{W}_{k-1} \mathbf{X}_{k-1}}{|\mathbf{W}_{k-1} \mathbf{X}_{k-1}|}$. Since $\mathbf{N}(n) = \sum_{m=N_0}^n \mathbf{I}_m$, we have

$$\begin{aligned} \mathbf{N}(n) - (n-N_0)\mathbf{v} &= \sum_{m=N_0}^n \mathbf{m}_m + \sum_{m=N_0}^n \frac{\mathbf{W}_m \mathbf{X}_m}{|\mathbf{W}_m \mathbf{X}_m|} - (n-N_0)\mathbf{v} \\ &= \sum_{m=N_0}^n \mathbf{m}_m + \sum_{m=N_0}^n \frac{\mathbf{X}_m}{|\mathbf{X}_m|} - (n-N_0)\mathbf{v} + \sum_{m=N_0}^{n-1} o(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} \quad (a.s.) \quad (3) \\ &:= \mathbf{D}_n + \sum_{m=N_0}^{n-1} o(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|}, \quad (a.s.) \end{aligned}$$

and similarly as in Zhang et al. [7] we have $\mathbf{D}_n = o(n-N_0)\mathbf{1}$ (a.s.), and with the fact $\sum_{m=N_0}^{n-1} o(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} = o(n-N_0)\mathbf{1}$ (a.s.), we get the desired result.

(iii) Since

$$\frac{\sum_{i=1}^k \sum_{j=1}^n f(\mathbf{r}_j) I_{ji}}{n} = \sum_{i=1}^k \frac{N_i(n)}{n} \sum_{j=1}^n \frac{f(\mathbf{r}_j) I_{ji}}{N_i(n)}$$

the conclusion is direct from those in (i) and (ii).

Proof of Theorem 2: (i). With the result in Theorem 1 (ii), the proof is the same as that in Theorem 2 (i) in Yuan and Chai [8].

(ii). By eq. (2), we have

$$\sqrt{n} \left(\frac{\mathbf{X}_n}{n} - \lambda_1 \mathbf{v} \right) = n^{-1/2} \mathbf{A}_n + N_0 n^{-1/2} \lambda_1 \mathbf{v} + n^{-1/2} \sum_{m=N_0}^{n-1} o_p(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} \mathbf{M}_{m+1}.$$

Since N_0 is fixed, $N_0 n^{-1/2} \lambda_1 \mathbf{v} \rightarrow 0$ as $n \rightarrow \infty$. Also, with N_1 given in the proof of Theorem 1 (i), for $n < N_1$, $\mathbf{W}_n \mathbf{X}_n / |\mathbf{W}_n \mathbf{X}_n| = \mathbf{X}_n / |\mathbf{X}_n| (1 + o_p(n^{-1}))$; for $n \geq N_1$, $\mathbf{W}_n$ is uniform, so $\mathbf{W}_n \mathbf{X}_n / |\mathbf{W}_n \mathbf{X}_n| = \mathbf{X}_n / |\mathbf{X}_n|$, and so in the third term of the previous expression, the $o(m^{-1})$ is actually 0 for such n 's. Thus $n^{-1/2} \sum_{m=N_0}^{n-1} o_p(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} \mathbf{M}_{m+1} = n^{-1/2} \sum_{m=N_0}^{N_1} o_p(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} \mathbf{M}_{m+1} \xrightarrow{P} 0$, since by assumption (B4), $N_1 n^{-1/2} \rightarrow 0$.

Similarly, by eq. (3),

$$\sqrt{n} \left(\frac{\mathbf{N}(n)}{n} - \mathbf{v} \right) = n^{-1/2} \mathbf{D}_n + N_0 n^{-1/2} \mathbf{v} + n^{-1/2} \sum_{m=N_0}^{n-1} o_p(m^{-1}) \frac{\mathbf{X}_m}{|\mathbf{X}_m|} = n^{-1/2} \mathbf{D}_n + o_p(1).$$

So we get

$$\sqrt{n} \left(\frac{\mathbf{X}_n}{n} - \lambda_1 \mathbf{v}, \frac{\mathbf{N}(n)}{n} - \mathbf{v} \right) = \sqrt{n} (\mathbf{A}_n, \mathbf{D}_n) + o_p(1)$$

Note $\mathbf{A}_n$ is the asymptotic expansion of the urn composition corresponds to $\mathbf{X}_n$ without weighting $\mathbf{W}_n$, and $\mathbf{D}_n$ is that for the allocation accounts $\mathbf{N}(n)$ without weighting $\mathbf{W}_n$, so the result follows from Theorem 4 in Yuan and Chai [8].

To describe the asymptotic covariance matrix $\mathbf{\Omega}$, we need the following notations. Let $\bar{\mathbf{M}} = \mathbf{M} - \mathbf{1}'\mathbf{v}$, $\zeta = \zeta(\boldsymbol{\mu}, \Xi_1)$, $\Sigma_1 = \{\mathbf{v}\} - \mathbf{v}'\mathbf{v}$, $\Sigma_2 = E[(\zeta - \mathbf{M})' \{\mathbf{v}\} (\zeta - \mathbf{M})]$, $\Sigma_3 = \text{diag}(v_1 \sigma_1^2, \dots, v_k \sigma_k^2)$, with $\sigma_j^2 = \text{Var}(\mu_{n,i})$, and $\Sigma_{23} = E[(\zeta - \mathbf{M})' \{\mathbf{v}_1\} (f(\mathbf{r}_1) - \boldsymbol{\mu})]$, where $f(\mathbf{r}_1) = (f(\mathbf{r}_1)|A_1, \dots, f(\mathbf{r}_1)|A_k)$. Define $\mathbf{t}_j = \mathbf{v} \partial \mathbf{M}(\mathbf{x}) / \partial x_j|_{\mathbf{x}=\boldsymbol{\mu}}$ and $\mathbf{T} = (\mathbf{t}_1/v_1, \dots, \mathbf{t}_k/v_k)'$. Recall for a real number a , positive number x and a matrix A , the following definitions will be adopted

$$e^{aA} = \sum_{i=0}^{\infty} \frac{a^i}{i!} A^i, \quad x^A = \sum_{i=0}^{\infty} \frac{\log^i(x)}{i!} A^i.$$

Denote $\mathbf{\Omega} = (\mathbf{\Omega}_{ij})_{1 \leq i, j \leq 2}$

$$\mathbf{\Omega}_{11} = \mathbf{M}' \mathbf{\Lambda}_1^\dagger \mathbf{M} + \mathbf{\Lambda}_2^\dagger + \mathbf{\Lambda}_3^\dagger + \mathbf{\Lambda}_{23}^\dagger + (\mathbf{\Lambda}_{23}^\dagger)', \mathbf{\Omega}_{12} = \mathbf{M}' \mathbf{\Lambda}_1^\dagger + (\mathbf{\Lambda}_2^\diamond + \mathbf{\Lambda}_3^\diamond + \mathbf{\Lambda}_{23}^\diamond + \mathbf{\Lambda}_{32}^\diamond)(\mathbf{I} - \mathbf{1}'\mathbf{v}),$$

and

$$\mathbf{\Omega}_{22} = \mathbf{\Lambda}_1^\dagger + (\mathbf{I} - \mathbf{v}'\mathbf{1})(\mathbf{\Lambda}_2^\sharp + \mathbf{\Lambda}_3^\sharp + \mathbf{\Lambda}_{23}^\sharp + (\mathbf{\Lambda}_{23}^\sharp)')(\mathbf{I} - \mathbf{1}'\mathbf{v}),$$

$$\text{with, } \mathbf{\Lambda}_1^\dagger = \int_0^1 \left(\frac{1}{x} \right)^{\bar{\mathbf{M}}} \Sigma_1 \left(\frac{1}{x} \right)^{\bar{\mathbf{M}}} dx \quad \mathbf{\Lambda}_3^\dagger = \int_0^1 \left[\int_x^1 \frac{1}{y} \left(\frac{1}{y} \right)^{\bar{\mathbf{M}}} dy \right]' \mathbf{T}' \Sigma_1 \mathbf{T} \left[\int_x^1 \frac{1}{y} \left(\frac{1}{y} \right)^{\bar{\mathbf{M}}} dy \right] dx,$$

$$\mathbf{\Lambda}_2^\dagger = \int_0^1 \left(\frac{1}{x} \right)^{\bar{\mathbf{M}}} \Sigma_2 \left(\frac{1}{x} \right)^{\bar{\mathbf{M}}} dx \quad \mathbf{\Lambda}_{23}^\dagger = \int_0^1 \left(\frac{1}{x} \right)^{\bar{\mathbf{M}}} \Sigma_{23} \mathbf{T} \left[\int_x^1 \frac{1}{y} \left(\frac{1}{y} \right)^{\bar{\mathbf{M}}} dy \right] dx,$$

$$\begin{aligned}
\Lambda_2^\# &= \int_0^1 \left[\int_x^1 \frac{1}{y} \left(\frac{y}{x} \right)^{\bar{M}} dy \right]' \Sigma_2 \left[\int_x^1 \frac{1}{y} \left(\frac{y}{x} \right)^{\bar{M}} dy \right] dx & \Lambda_{23}^\# &= \int_0^1 \left[\int_x^1 \frac{1}{y} \left(\frac{y}{x} \right)^{\bar{M}} dy \right]' \mathbf{T}' \Sigma_{23} \mathbf{T} \left[\int_x^1 \frac{1}{y} \left(\frac{y}{x} \right)^{\bar{M}} dy \right] dx, \\
\Lambda_3^\# &= \int_0^1 \left[\int_x^1 \int_x^y \frac{1}{yu} \left(\frac{y}{u} \right)^{\bar{M}} du dy \right]' \mathbf{T}' \Sigma_3 \mathbf{T} \left[\int_x^1 \int_x^y \frac{1}{yu} \left(\frac{y}{u} \right)^{\bar{M}} du dy \right] dx, \\
\Lambda_2^\diamond &= \int_0^1 \left(\frac{1}{x} \right)^{\bar{M}} \Sigma_2 \int_x^1 \frac{1}{y} \left(\frac{y}{x} \right)^{\bar{M}} dy dx, & \Lambda_{23}^\diamond &= \int_0^1 \left(\frac{1}{x} \right)^{\bar{M}} \Sigma_{23} \mathbf{T} \int_x^1 \frac{1}{yu} \int_x^y \left(\frac{y}{u} \right)^{\bar{M}} du dy dx, \\
\Lambda_{32}^\diamond &= \int_0^1 \left[\int_x^1 \frac{1}{y} \left(\frac{1}{y} \right)^{\bar{M}} dy \right]' \mathbf{T}' \Sigma_{23}' \int_x^1 \frac{1}{y} \left(\frac{y}{x} \right)^{\bar{M}} dy dx \text{ and} \\
\Lambda_3^\diamond &= \int_0^1 \int_x^1 \left[\frac{1}{y} \left(\frac{1}{y} \right)^{\bar{M}} dy \right]' \mathbf{T}' \Sigma_3 \mathbf{T} \int_x^1 \int_x^y \frac{1}{yu} \left(\frac{y}{u} \right)^{\bar{M}} du dy dx.
\end{aligned}$$

Proof of Theorem 3: (i). Let $\mathbf{v}^{(1)} = (1, 0, \dots, 0)$, $\mathbf{v}^{(2)} = (v_1^{(2)}, v_2^{(2)}, 0, \dots, 0), \dots, \mathbf{v}^{(k-1)} = (v_1^{(k-1)}, \dots, v_{k-1}^{(k-1)}, 0)$. Let $n_1^{(j)}, \dots, n_j^{(j)}$ be the allocated number of the first j treatments at the start of the j -th stage, and denote the weight at stage j as $\mathbf{W}_n^{(j)}$. The $\mathbf{v}^{(j)}$'s are all k -dimensional vectors, $\mathbf{v}^{(j)}$ has j non-zero components which correspond to the optimality criterion for the first j treatments. Let $\mathbf{M}^{(j)}$ be the adding matrix for the first j treatments, $\lambda^{(j)}$ be its first eigenvalue, and $\alpha_l^{(j)} = \lim n_l^{(j)} / n_j^{(j)}$. Denote $\mathbf{X}_0 = (x_{1,0}, 0, \dots, 0)$. Let $\mathbf{g}_1(\mathbf{v}^{(1)}) = (1, 0, \dots, 0)$, for a k -vector $\mathbf{x} = (x_1, \dots, x_k)$, let $\mathbf{g}_2(\mathbf{x})$ be the k -vector given by

$$\mathbf{g}_2(\mathbf{x}) = \left(\int_0^{\alpha_1^{(2)} \wedge 1} \frac{(w x_1, \alpha_1^{(2)} x_2)}{w x_1 + \alpha_1^{(2)} x_2} dw + \left[1 - (\alpha_1^{(2)} \wedge 1) \right] \left(\frac{(x_1, x_2)}{x_1 + x_2}, 0, \dots, 0 \right) \right).$$

To define $\mathbf{g}_j(\mathbf{x})$, we first re-arrange the first j components of $\mathbf{W}_n^{(j)}$ in increasing order as $\mathbf{W}_n^{[j]} = (n^{[1]}, \dots, n^{[j-1]}, n_1 + \dots + n_{j-1}, 0, \dots, 0)$, with $n^{[1]} \leq \dots \leq n^{[j-1]}$. Define

$$\begin{aligned}
\mathbf{g}^{[3]}(\mathbf{x}) &= \left(\int_{\alpha_1^{(3)}}^{\alpha_1^{(3)} + (1/2) \wedge \alpha_2^{(3)}} \int_{\alpha_2^{(3)}}^{\alpha_2^{(3)} + (1/2) \wedge \alpha_1^{(3)}} \frac{(w_1 x_1, w_2 x_2, (\alpha_1^{(3)} + \alpha_2^{(3)}) x_3)}{w_1 x_1 + w_2 x_2 + (\alpha_1^{(3)} + \alpha_2^{(3)}) x_3} dw_1 dw_2 \right. \\
&\quad \left. + \left[1 - \left(\frac{\alpha_1^{(3)}}{2} \vee \frac{\alpha_2^{(3)}}{2} \right) \wedge 1 \right] \left(\frac{(x_1, x_2, x_3)}{x_1 + x_2 + x_3}, 0, \dots, 0 \right) \right),
\end{aligned}$$

and then define $\mathbf{g}_3(\mathbf{x})$ as $\mathbf{g}^{[3]}(\mathbf{x})$ with components switched back to the original order of $\mathbf{W}_n^{(3)}$, i.e., $\mathbf{g}_3(\mathbf{x})$ is the concomitant of $\mathbf{g}^{[3]}(\mathbf{x})$ in relationship of $\mathbf{W}_n^{(3)}$ to $\mathbf{W}_n^{[3]}$. More generally, define

$$\begin{aligned}
\mathbf{g}^{[j]}(\mathbf{x}) &= \left(\int_{\alpha_1^{(j)}}^{\alpha_1^{(j)} + (j-1)^{-1} \wedge \alpha_{j-1}^{(j)}} \dots \int_{\alpha_{j-1}^{(j)}}^{\alpha_{j-1}^{(j)} + (j-1)^{-1} \wedge \alpha_1^{(j)}} \frac{(w_1 x_1, \dots, w_{j-1} x_{j-1}, (\alpha_1^{(j)} + \dots + \alpha_{j-1}^{(j)}) x_j)}{w_1 x_1 + \dots + w_{j-1} x_{j-1} + (\alpha_1^{(j)} + \dots + \alpha_{j-1}^{(j)}) x_j} dw_1 \dots dw_{j-1} \right. \\
&\quad \left. + \left[1 - \left(\frac{\alpha_1^{(j)}}{j-1} \vee \dots \vee \frac{\alpha_{j-1}^{(j)}}{j-1} \right) \wedge 1 \right] \left(\frac{(x_1, \dots, x_j)}{x_1 + \dots + x_j}, 0, \dots, 0 \right), (j = 2, \dots, k) \right)
\end{aligned}$$

and define $\mathbf{g}_j(\mathbf{x})$ as the concomitant of $\mathbf{g}^{[j]}(\mathbf{x})$.

Lastly set $\mathbf{u} = \sum_{j=1}^k \alpha_j \mathbf{g}_j(\mathbf{v}^{(j)})$, and $\mathbf{y} = \sum_{j=1}^k \alpha_j \mathbf{g}_j(\mathbf{v}^{(j)}) \mathbf{M}^{(j)}$.

(ii). Let $\Delta \mathbf{S}_m$ and $\Delta \mathbf{m}_m$ as in the proof of Theorem 1 (ii). Denote $\mathbf{S}_n^{(1)} = \sum_{m=1}^{n_1} \Delta \mathbf{S}_m := \sum^{(1)} \Delta \mathbf{S}_m$, $\mathbf{S}_n^{(2)} = \sum_{m=n_1+1}^{n_2} \Delta \mathbf{S}_m := \sum^{(2)} \Delta \mathbf{S}_m, \dots, \mathbf{S}_n^{(k-1)} = \sum_{m=n_1+\dots+n_{k-2}+1}^{n_1+\dots+n_{k-1}} \Delta \mathbf{S}_m := \sum^{(k-1)} \Delta \mathbf{S}_m$, $\mathbf{S}_n^{(k)} = \sum_{m=n_1+\dots+n_{k-1}+1}^n \Delta \mathbf{S}_m := \sum^{(k)} \Delta \mathbf{S}_m$.

Similarly as in the proof of Theorem 1 (ii), we have

$$\begin{aligned} \frac{\mathbf{X}_n}{n} - \mathbf{g} &= \sum_{j=1}^k \frac{n_j}{n} \left(\frac{1}{n_j} \mathbf{S}_n^{(j)} + \frac{1}{n_j} \sum_{m=1}^{(j)} \left(\frac{\mathbf{W}_m^{(j)} \mathbf{X}_m}{|\mathbf{W}_m^{(j)} \mathbf{X}_m|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \mathbf{M}_{m+1}^{(j)} \right. \right. \\ &\quad \left. \left. + \frac{1}{n_j} \sum_{m=1}^{(j)} \mathbf{g}_j(\mathbf{v}^{(j)}) (\mathbf{M}_{m+1}^{(j)} - \mathbf{M}^{(j)}) \right) + \sum_{j=1}^k \mathbf{g}_j(\mathbf{v}^{(j)}) \left(\frac{n_j}{n} - \alpha_j \right) + n^{-1} \mathbf{X}_0. \end{aligned}$$

Since $n_j^{-1} \mathbf{S}_n^{(j)} \rightarrow \mathbf{0}$ (a.s.), $\mathbf{M}_{m+1}^{(j)} - \mathbf{M}^{(j)} \rightarrow \mathbf{0}$ (a.s.), $n^{-1} \mathbf{X}_0 \rightarrow \mathbf{0}$ (a.s.), $n_j/n \rightarrow \alpha_j$, and since for m with $n_1 + \dots + n_{j-1} \leq m \leq n_1 + \dots + n_j$, $\mathbf{X}_m/|\mathbf{X}_m| = (\mathbf{X}_m/m)/|\mathbf{X}_m/m| \rightarrow \mathbf{v}^{(j)}/|\mathbf{v}^{(j)}| = \mathbf{v}^{(j)}$ (a.s.). Take $j=2$ for example, when m varies from $n_1^{(2)}$ to $n_2^{(2)}$, the first two components of $\mathbf{W}_m^{(2)}$ varies from $(n_{2,0}, n_1^{(2)})$ to $(n_1^{(2)} \wedge n_2^{(2)}, n_1^{(2)})$, $n_{2,0}$ is fixed, the other components of $\mathbf{W}_m$ are zeros, so for any fixed k -vector $\mathbf{x}$,

$$\frac{1}{n_2} \sum_{m=1}^{(2)} \frac{\mathbf{W}_m^{(2)} \mathbf{x}}{|\mathbf{W}_m^{(2)} \mathbf{x}|} = \frac{n_2^{(2)}}{n_2} \frac{1}{n_2^{(2)}} \sum_{m=1}^{(2)} \frac{\mathbf{W}_m^{(2)} \mathbf{x}}{|\mathbf{W}_m^{(2)} \mathbf{x}|} \rightarrow \mathbf{g}_2(\mathbf{x}), \quad (a.s.).$$

To see how the form of $\mathbf{g}_j(\mathbf{x})$ be identified, we first re-arrange the first j components of $\mathbf{W}_n^{(j)}$ in increasing order as $\mathbf{W}_n^{[j]} = (n^{[1]}, \dots, n^{[j-1]}, n_1 + \dots + n_{j-1}, 0, \dots, 0)$, with $n^{[1]} \leq \dots \leq n^{[j-1]}$. For example, for $\mathbf{W}_n^{[3]}$, at the start of the 3rd stage, when m varies from $n_1 + n_2$ to $n_1 + n_2 + n_3$, the first 3 components of $\mathbf{W}_m^{[3]}$ varies from $(n^{[1]}, n^{[2]}, n_1 + n_2)$ to $(n^{[1]} + [n_3/2] \wedge n^{[2]}, n^{[2]} + [n_3/2] \wedge n^{[1]}, n_1 + n_2)$, and this gives $\frac{1}{n^{[2]}} \sum \frac{\mathbf{W}_m^{[3]} \mathbf{x}}{|\mathbf{W}_m^{[3]} \mathbf{x}|} \rightarrow \mathbf{g}^{[3]}(\mathbf{x})$. Now it is seen that

$$\frac{1}{n_j} \sum_{m=1}^{(j)} \left(\frac{\mathbf{W}_m^{(j)} \mathbf{X}_m}{|\mathbf{W}_m^{(j)} \mathbf{X}_m|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) = \frac{1}{n_j} \sum_{m=1}^{(j)} \frac{\mathbf{W}_m^{(j)} (\mathbf{X}_m/|\mathbf{X}_m|)}{|\mathbf{W}_m^{(j)} (\mathbf{X}_m/|\mathbf{X}_m|)|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \rightarrow \mathbf{0}, \quad (a.s.)$$

these give $\frac{\mathbf{X}_n}{n} \rightarrow \mathbf{u}$ (a.s.).

Similarly,

$$\frac{\mathbf{N}(n)}{n} - \mathbf{u} = \sum_{j=1}^k \frac{n_j}{n} \left(\frac{1}{n_j} \mathbf{m}_n^{(j)} + \frac{1}{n_j} \sum_{m=1}^{(j)} \left(\frac{\mathbf{W}_m^{(j)} \mathbf{X}_m}{|\mathbf{W}_m^{(j)} \mathbf{X}_m|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) + \sum_{j=1}^k \mathbf{g}_j(\mathbf{v}^{(j)}) \left(\frac{n_j}{n} - \alpha_j \right) \right) \rightarrow \mathbf{0} \quad (a.s.).$$

Proof of Theorem 4: (i) For the first part see the proof of Theorem 2.

(ii). Let $\mathbf{h}^{[j]}(\mathbf{w}, \mathbf{x})$ be the integrand of $\mathbf{g}^{[j]}(\mathbf{x})$,

$$\begin{aligned} \mathbf{G}^{[j]}(\mathbf{x}) &= \left(\int_{\alpha_1^{(j)}}^{\alpha_1^{(j)} + (j-1)^{-1} \wedge \alpha_{j-1}^{(j)}} \dots \int_{\alpha_{j-1}^{(j)}}^{\alpha_2^{(j)} + (j-1)^{-1} \wedge \alpha_1^{(j)}} (\mathbf{h}^{[j]}(\mathbf{w}, \mathbf{x}))' \mathbf{h}^{[j]}(\mathbf{W}, \mathbf{X}) d\mathbf{w}_1 \dots d\mathbf{w}_{j-1} \right. \\ &\quad \left. + \left[1 - \left(\frac{\alpha_1^{(j)}}{j-1} \vee \dots \vee \frac{\alpha_{j-1}^{(j)}}{j-1} \right) \wedge 1 \right] \left(\frac{(x_1, \dots, x_j)' (x_1, \dots, x_j)}{(x_1 + \dots + x_j)^2}, 0, \dots, 0 \right), (j=2, \dots, k) \right) \end{aligned}$$

and $\mathbf{G}^{(j)}(\mathbf{x})$ be the concomitant of $\mathbf{G}^{[j]}(\mathbf{x})$.

As in the proof of Theorem 3 (ii),

$$\begin{aligned} \sqrt{n} \left(\frac{\mathbf{X}_n}{n} - \mathbf{y} \right) &= \sum_{j=1}^k \left(\frac{n_j}{n} \right)^{1/2} \left(n_j^{-1/2} \mathbf{S}_n^{(j)} + n_j^{-1/2} \sum_{m=1}^{(j)} \left(\frac{\mathbf{W}_m^{(j)} \mathbf{X}_m}{|\mathbf{X}_m^{(j)} \mathbf{X}_m|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) \mathbf{M}_{m+1}^{(j)} \right. \\ &\quad \left. + n_j^{-1/2} \sum_{m=1}^{(j)} \mathbf{g}_j(\mathbf{v}^{(j)}) (\mathbf{M}_{m+1}^{(j)} - \mathbf{M}^{(j)}) \right) + \sum_{j=1}^k \mathbf{g}_j(\mathbf{v}^{(j)}) \frac{n_j - \alpha_j n}{\sqrt{n}} + n^{-1/2} \mathbf{X}_0 \\ &= \sum_{j=1}^k \left(\frac{n_j}{n} \right)^{1/2} \left(n_j^{-1/2} \mathbf{S}_n^{(j)} + \frac{1}{n_j} \sum_{m=1}^{(j)} \left(\frac{m}{n_j} \right)^{-1/2} \sqrt{m} \left(\frac{\mathbf{W}_m^{(j)} (\mathbf{X}_m/m)}{|\mathbf{W}_m^{(j)} (\mathbf{X}_m/m)|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) \mathbf{M}_{m+1}^{(j)} \right. \\ &\quad \left. + n_j^{-1/2} \sum_{m=1}^{(j)} \mathbf{g}_j(\mathbf{v}^{(j)}) (\mathbf{M}_{m+1}^{(j)} - \mathbf{M}^{(j)}) \right) + O_p(n^{-1/2}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^k \left(\frac{n_j}{n}\right)^{1/2} \left(n_j^{-1/2} S_n^{(j)} + \frac{1}{n_j} \sum^{(j)} \left(\frac{m}{n_j}\right)^{-1/2} \sqrt{m} \left(\mathbf{g}_j(\mathbf{X}_m/m) - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) \mathbf{M}_{m+1}^{(j)} \right. \\
&\quad \left. + n_j^{-1} \sum^{(j)} \left(\frac{m}{n_j}\right)^{-1/2} \mathbf{g}_j(\mathbf{v}^{(j)}) \dot{\mathbf{M}}^{(j)} \sqrt{m} (\boldsymbol{\mu}_{m+1} - \boldsymbol{\mu}) \right) + O_p(n^{-1/2}),
\end{aligned}$$

where $\dot{\mathbf{M}}^{(j)}$ is the derivative of $\mathbf{M}^{(j)}$ with respect to $\boldsymbol{\mu}$ and evaluated at $\boldsymbol{\mu}$, which is a $k \times k \times k$ array.

As in Theorem 2 (ii), $\sqrt{m}(\boldsymbol{\mu}_{m+1} - \boldsymbol{\mu}) \xrightarrow{D} N(\mathbf{0}, \{\sigma^2/\mathbf{v}^{(j)}\})$. With the form of $\mathbf{M}^{(j)}$, by the Corollary in Yuan and Chai [8], for each m in the summation $\sum^{(j)}$, we have $\sqrt{m}(\frac{\mathbf{X}_m}{m} - \mathbf{v}^{(j)}) \xrightarrow{D} N(\mathbf{0}, \Omega^{(j)})$, so by the delta method,

$$\sqrt{m}(\mathbf{g}_j(\mathbf{X}_m/m) - \mathbf{g}_j(\mathbf{v}^{(j)})) \xrightarrow{D} N(\mathbf{0}, \dot{\mathbf{g}}_j \Omega^{(j)} \dot{\mathbf{g}}_j'),$$

where $\dot{\mathbf{g}}_j(\mathbf{v}) = \partial \mathbf{g}_j(\mathbf{v}) / \partial \mathbf{v}'$ and $\dot{\mathbf{g}}_j = \dot{\mathbf{g}}_j(\mathbf{v}^{(j)})$, and so

$$\frac{1}{n_j} \sum^{(j)} \left(\frac{m}{n_j}\right)^{-1/2} \sqrt{m} \left(\mathbf{g}_j(\mathbf{X}_m/m) - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) \mathbf{M}_{m+1}^{(j)} \xrightarrow{D} N(\mathbf{0}, C_j^2 \mathbf{M}^{(j)'} \dot{\mathbf{g}}_j \Omega^{(j)} \dot{\mathbf{g}}_j' \mathbf{M}^{(j)}),$$

with $C_j = \int_{(\alpha_1 + \dots + \alpha_{j-1})/\alpha_j}^{(\alpha_1 + \dots + \alpha_j)/\alpha_j} w^{-1/2} dw$. Similarly, with $\Omega_\mu^{(j)} = \text{diag}\{\sigma_1^2/v_1^{(j)}, \dots, \sigma_j^2/v_j^{(j)}\}$,

$$n_j^{-1} \sum^{(j)} \left(\frac{m}{n_j}\right)^{-1/2} \mathbf{g}_j(\mathbf{v}^{(j)}) \dot{\mathbf{M}}^{(j)} \sqrt{m} (\boldsymbol{\mu}_{m+1} - \boldsymbol{\mu}) \xrightarrow{D} N(\mathbf{0}, C_j^2 (\mathbf{g}_j \dot{\mathbf{M}}^{(j)})' \Omega_\mu^{(j)} \mathbf{g}_j \dot{\mathbf{M}}^{(j)}).$$

Recall $\{S_n, \mathcal{F}_n\}$ is a martingale, and for each m in the summation $\sum^{(j)}$ sign, denote $\mathbf{H}_m = E(\mathbf{M}(\boldsymbol{\mu}_m) | \mathcal{F}_{m-1})$, then

$$\begin{aligned}
\text{Var}(\Delta \mathbf{S}_m | \mathcal{F}_{m-1}) &= E \left(\mathbf{M}_m' \text{diag} \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) \mathbf{M}_m | \mathcal{F}_{m-1} \right) - \mathbf{H}_m' \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right)' \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) \mathbf{H}_m \\
&= E \left((\mathbf{M}_m - \mathbf{H}_m)' \text{diag} \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) (\mathbf{M}_m - \mathbf{H}_m) | \mathcal{F}_{m-1} \right) \\
&\quad + \mathbf{H}_m' \left(\text{diag} \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) - \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right)' \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) \right) \mathbf{H}_m.
\end{aligned}$$

Note in our case, $\lim_m \mathbf{H}_m = \lim_m \mathbf{M}_m = \mathbf{M}^{(j)}$ (a.s.), $\text{diag}\{(\mathbf{W}_{m-1} \mathbf{X}_{m-1})/|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|\}$ is bounded, so the first term above tends to zero. Thus,

$$\begin{aligned}
\text{Var}(n_j^{-1/2} \mathbf{S}_n^{(j)}) &:= n_j^{-1} \sum^{(j)} \text{Var}(\Delta \mathbf{S}_m | \mathcal{F}_{m-1}) \\
&= \frac{1}{n_j} \sum^{(j)} \mathbf{H}_m' \left(\text{diag} \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) - \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right)' \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} \right) \right) \mathbf{H}_m + o_p(1), \\
&\xrightarrow{P} \mathbf{M}^{(j)'} (\text{diag}(\mathbf{g}_j(\mathbf{v}^{(j)})) - \mathbf{G}_j(\mathbf{v}^{(j)})) \mathbf{M}^{(j)} := \mathbf{Q}_j.
\end{aligned}$$

Also, for each m , $\text{Cov}(\Delta \mathbf{S}_m, \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) | \mathcal{F}_{m-1}) = E(\Delta \mathbf{S}_m' | \mathcal{F}_{m-1}) \left(\frac{\mathbf{W}_{m-1} \mathbf{X}_{m-1}}{|\mathbf{W}_{m-1} \mathbf{X}_{m-1}|} - \mathbf{g}_j(\mathbf{v}^{(j)}) \right) = \mathbf{0}$, and

$\text{Cov}(\Delta \mathbf{S}_m, \mathbf{g}_j(\mathbf{v}^{(j)}) \mathbf{M}^{(j)} \sqrt{m} (\boldsymbol{\mu}_{m-1} - \boldsymbol{\mu}) | \mathcal{F}_{m-1}) = \mathbf{0}$, and

$$\begin{aligned}
&\frac{1}{n_j} \sum^{(j)} \text{Cov}((\mathbf{g}_j(\mathbf{X}_m/m) - \dot{\mathbf{g}}_j(\mathbf{v}^{(j)}) \mathbf{M}_{m+1}^{(j)}, \mathbf{g}_j(\mathbf{v}^{(j)}) \mathbf{M}^{(j)} (\boldsymbol{\mu}_{m+1} - \boldsymbol{\mu}) | \mathcal{F}_m) \\
&= \frac{1}{n_j} \sum^{(j)} \mathbf{M}^{(j)'} \dot{\mathbf{g}}_j (\boldsymbol{\mu}_{m+1} - \boldsymbol{\mu})' (\boldsymbol{\mu}_{m+1} - \boldsymbol{\mu}) \mathbf{g}_j \dot{\mathbf{M}}^{(j)} + o_p(1) \xrightarrow{P} \mathbf{0},
\end{aligned}$$

so we get

$$\sqrt{n} \left(\frac{\mathbf{X}_n}{n} - \boldsymbol{\gamma} \right) \xrightarrow{D} \mathbf{N}(\mathbf{0}, \bar{\Omega}_{11}), \bar{\Omega}_{11} = \sum_{j=1}^k \alpha_j \left(\mathbf{Q}_j + C_j^2 (\dot{\mathbf{g}}_j \mathbf{M}^{(j)})' \Omega^{(j)} \dot{\mathbf{g}}_j \mathbf{M}^{(j)} + C_j^2 (\mathbf{g}_j \dot{\mathbf{M}}^{(j)})' \Omega_\mu^{(j)} \mathbf{g}_j \dot{\mathbf{M}}^{(j)} \right).$$

Similarly, we have

$$\sqrt{n}\left(\frac{\mathbf{N}(n)}{n} - \mathbf{u}\right) \xrightarrow{D} N(\mathbf{0}, \bar{\Omega}_{22}), \bar{\Omega}_{22} = \sum_{j=1}^k \alpha_j \left(\text{diag}(\mathbf{g}_j(\mathbf{v}^{(j)})) - \mathbf{G}_j(\mathbf{v}^{(j)}) + \mathbf{C}_j^2(\dot{\mathbf{g}}_j)' \Omega^{(j)} \dot{\mathbf{g}}_j \right),$$

and

$$\sqrt{n}\left(\frac{\mathbf{X}_n}{n} - \mathbf{y}, \frac{\mathbf{N}(n)}{n} - \mathbf{u}\right) \xrightarrow{D} N(\mathbf{0}, \bar{\Omega}), \quad \bar{\Omega} = (\bar{\Omega}_{ij})_{1 \leq i, j \leq 2},$$

where $\bar{\Omega}_{12} = \sum_{j=1}^k \alpha_j \mathbf{M}^{(j)'} \left(\text{diag}(\mathbf{g}_j(\mathbf{v}^{(j)})) - \mathbf{G}_j(\mathbf{v}^{(j)}) + \mathbf{C}_j^2(\dot{\mathbf{g}}_j \mathbf{M}^{(j)})' \Omega^{(j)} \right)$.

Acknowledgment: We thank Dr. Ainong Zhou of MedStar, Inc. for the topic of staggered-start clinical trials. Yuan, Xiong and Tan are partially supported by the National Cancer Institute (NCI) grant R01CA164717 and P30CA 051008. Li is supported in part by the National Science Foundation of China, Grant No. (11371353, 61134013) and the Breakthrough Project of Strategic Priority Program of the Chinese Academy of Sciences, Grant No. XDB13040600.

References

1. Rosenberger WF, Lachin JM. The use of response-adaptive designs in clinical trials. *Controlled Clin Trials* 1993;14:471–84.
2. Eisele JR. The doubly adaptive biased coin design for sequential clinical trials. *J Stat Plann Inference* 1994;38:249–62.
3. Melfi V, Page C. Variability in adaptive designs for estimation of success probabilities. In: Flournoy N, Rosenberger WF, Wong WK, editors. *New developments and applications in experimental design*. Hayward, CA: IMS, 1998:106–114.
4. Rosenberger WF, Stallard N, Ivanova A, Harper C, Ricks M. Optimal adaptive designs for binary response trials. *Biometrics* 2001;57:909–13.
5. Bai ZD, Hu F, Rosenberger WF. Asymptotic properties of adaptive designs for clinical trials with delayed response. *Ann Stat* 2002;30:1–18.
6. Bai ZD, Hu F. Asymptotics in randomized urn models. *Ann Appl Probab* 2005;1B:914–40.
7. Zhang L, Hu F, Cheung S. Asymptotic theorems of sequential estimation-adjusted urn models. *Ann Appl Probab* 2006;16:340–69.
8. Yuan A, Chai GX. Optimal adaptive generalized Pólya urn design for multi-arm clinical trials. *J Multivariate Anal* 2008;99:1–24.
9. Zhu H, Hu F. Sequential monitoring of response-adaptive randomized clinical trials. *Ann Stat* 2010;38:2218–41.
10. Yuan A, Bezandry P, Bonney G. Compound adaptive GPU design for clinical trials. *J Stat Plann Inference* 2010;140:3505–15.
11. Barthel F, Babiker A, Royston P, Parmar M. Evaluation of sample size and power for multi-arm survival trials allowing for non-proportional hazards, loss to follow-up and cross-over. <http://www.fm-sbarthel.de/ISCB04.pdf>, 2004.
12. Bodick N, Forette F, Hadler D, Harvey RJ, Leber P, McKeith IG, et al. Protocols to demonstrate slowing of Alzheimer disease progression. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. The Disease Progression Sub-Group. *Alzheimer Dis Assoc Disord* 1997;11:50–53.
13. Leber P. Observations and suggestions on antidementia drug development. *Alzheimer Dis Assoc Disord* 1996;10:31–5.
14. Kublin J. Advancing HIV vaccines to efficacy testing in the HVTN and linking to other prevention research, HIV Vaccine Trials Network Washington, DC, 30 May 2012. <http://hvtan.org/meeting/ppt/may12/P1/JimKublinMAY30FINAL.pdf>, 2012.
15. Cummings J, Gould H, Zhong K. Advances in designs for Alzheimer's disease clinical trials. *Am J Neurodegener Dis* 2012;1:205–16.
16. Hendrix S, Horton S, Orgogozo JM. Modification of the “randomized withdrawal” and “staggered start” clinical trial designs: toward a practical demonstration of disease modification in Alzheimers disease. <https://www.myriad.com/downloads/EFNS-2007-Disease-Modification-in-AD-Hendrix.pdf>, 2013.
17. Hayre LS. Two-population sequential tests with three hypotheses. *Biometrika* 1979;66:465–74.
18. Jennison C, Turnbull BW. *Group sequential methods with applications to clinical trials*. Boca Raton, FL: Chapman and Hall/CRC Press, 2000.
19. Eisele JR, Woodroffe M. Central limit theorems for doubly adaptive biased coin designs. *Ann Stat* 1995;23:234–54.
20. Hu F, Zhang L, He X. Efficient randomized-adaptive designs. *Ann Stat* 2009;37:2345–560.
21. Athreya KB, Karlin S. Limit theorems for the split times of branching processes. *J Math Mech* 1967;17:257–77.
22. Gouet R. Strong convergence of proportions in a multicolor Pólya urn. *J Appl Probab* 1997;34:426–35.
23. Janson S. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes Appl* 2004;110:177–245.
24. Hall P, Heyde CC. *Martingale limit theory and its application*. New York: Academic Press, 1980.