Published in final edited form as:

Stat Med. 2014 October 15; 33(23): 4029–4042. doi:10.1002/sim.6222.

Outcome-adaptive randomization for a delayed outcome with a short-term predictor: imputation-based designs

Mi-Ok Kim^{a,b,*†}, Chunyan Liu^a, Feifang Hu^c, and J. Jack Lee^d

^aDivision of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, U.S.A

^bDepartment of Pediatrics, University of Cincinnati, Cincinnati, Ohio, U.S.A

^cDepartment of Statistics, George Washington University, Washington, DC, U.S.A

^dDepartment of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, U.S.A

Abstract

Delay in the outcome variable is challenging for outcome-adaptive randomization, as it creates a lag between the number of subjects accrued and the information known at the time of the analysis. Motivated by a real-life pediatric ulcerative colitis trial, we consider a case where a short-term predictor is available for the delayed outcome. When a short-term predictor is not considered, studies have shown that the asymptotic properties of many outcome-adaptive randomization designs are little affected unless the lag is unreasonably large relative to the accrual process. These theoretical results assumed independent identical delays, however, whereas delays in the presence of a short-term predictor may only be conditionally homogeneous. We consider delayed outcomes as missing and propose mitigating the delay effect by imputing them. We apply this approach to the doubly adaptive biased coin design (DBCD) for motivating pediatric ulcerative colitis trial. We provide theoretical results that if the delays, although non-homogeneous, are reasonably short relative to the accrual process similarly as in the iid delay case, the lag is also asymptotically ignorable in the sense that a standard DBCD that utilizes only observed outcomes attains target allocation ratios in the limit. Empirical studies, however, indicate that imputation-based DBCDs performed more reliably in finite samples with smaller root mean square errors. The empirical studies assumed a common clinical setting where a delayed outcome is positively correlated with a short-term predictor similarly between treatment arm groups. We varied the strength of the correlation and considered fast and slow accrual settings.

Keywords

outcome adaptive randomization; delayed outcome; doubly adaptive biased coin design; imputation

1. Introduction

Outcome-adaptive randomization designs are ethical randomization schemes that assign treatments to incoming subjects as a function of the treatment outcomes of subjects accrued thus far, typically skewing treatment allocation ratios to favor better performing treatment arms [1]. Although applied increasingly more, outcome-adaptive randomization designs are not as widely used as they could be. One important reason is incomplete understanding of the designs' properties and performance with delayed outcomes. In theory, complete outcome data are assumed on all previously accrued subjects. In reality, however, outcomes are often observed after some treatment period, and a lag exists between the number of subjects accrued and the information known at the time of the analysis. It is also often infeasible successively suspending the accrual until outcomes on all previously accrued subjects are observed. In this paper, we examine the impact of the lag in a setting where a short-term predictor exists for the delayed outcome and propose designs that mitigate the effect of the lag.

Many trials fit the profile of the concerned case. Survival in a leukemia trial is the ultimate goal of treatment and hence is the primary outcome, whereas it is not immediately measurable. Complete remission (CR), which is achievable and measurable after a few cycles of treatment, on the other hand, is desirable for prolonging survival and may serve as a short-term predictor. In a depressive disorder trial, the primary measure of treatment outcomes was change in the Hamilton Depression Scale from baseline after approximately 8 weeks of treatment [2]. The trial defined greater than 50% reduction in Hamilton Depression Scale in two consecutive visits after at least 3 weeks of therapy as a short-term surrogate and used it to adapt treatment allocation ratios instead. A pediatric ulcerative colitis (UC) study that motivates this paper also fits the profile. Treatments for pediatric UC target to attain steroid-free remission (SFR) as corticosteroids (CS) are known to negatively influence stature and bone turnover regardless of disease activity [3]. A meaningful clinical benchmark of a long-term treatment outcome is attaining SFR within 1 year from diagnosis (year 1 SFR). Preliminary results, on the other hand, reported that the year 1 SFR was observed at a twofold higher rate in those who achieved clinical remission by week 4 after diagnosis. The short-term week 4 clinical remission may serve as a predictor.

In practice, the lag between the number of subjects accrued and the information known at the time of the analysis is often ignored, and allocation ratios are updated on available outcome data. This practice, seemingly motivated for practicality, is in fact theoretically justified: when a short-term predictor is not considered, studies have shown that the asymptotic properties of many outcome-adaptive randomization designs are little affected unless the lag is unreasonably large relative to the accrual process [1, 4–6]. Hu *et al.* [1] specifically showed that a doubly adaptive biased coin design (DBCD) that utilizes only observed outcome data attains target allocation ratios in the limit. With survival outcomes, Zhang and Rosenberger [7] showed that the reasonable delay time assumption of Hu *et al.* [1] is satisfied for their parametric survival time model. Their simulation studies showed that the effect of the delay was minimal in finite samples.

These results are, however, not directly applicable to the case under consideration. They assumed iid delays, whereas the presence of a short-term predictor implies that delays may only be conditionally homogenous. In the motivating pediatric UC trial, a failure as an outcome is defined anytime standard therapies fail to maintain SFR and rescue therapies are prescribed within one year from diagnosis. A success is conversely defined as time to a failed outcome being longer than 1 year. The year 1 SFR was observed higher in those who achieved a clinical remission by week 4, however. This implies that the times to failed outcome or delays may only be homogeneous conditioning on the short-term week 4 clinical remission status. In leukemia trials, survival times may also be homogeneous conditioning on the attainment of CR. Non-homogenous delay times were addressed by Huang *et al.* [8] regarding a leukemia trial. They proposed a Bayesian approach that included both the short-term predictor CR and the survival time by modeling their relationship. Their simulation results showed that the joint model adapted allocation ratios more efficiently. This calls for a formal investigation of the impact of non-homogenous delays and search for outcome-adaptive randomization designs that mitigate the efficiency loss due to the delay.

In this paper, we consider delayed outcomes as missing observations and propose mitigating the delay effect by imputing them. We apply this approach to the DBCD for the motivating pediatric UC trial. We adapt Hu *et al.*'s [1] delay mechanism setup and show that the lag is similarly ignorable asymptotically in the sense that a simple or delay non-mitigating DBCD attains target allocation ratios in the limit provided that delays, even if non-homogeneous, are not unreasonably long. Simulation studies, however, show that imputation-based delay mitigating DBCDs performed more reliably in finite samples with smaller mean squared errors than a delay non-mitigating DBCD.

The rest of the paper is organized as follows: the delay mitigation via imputation approach is described in Section 2. In Section 3, delay mitigating DBCDs will be introduced for the motivating pediatric UC trial. Empirical study results will be presented in Section 4 with a summary and discussion in Section 5.

2. Doubly adaptive biased coin design for a delayed outcome with a short-term predictor

The DBCD is an outcome-adaptive randomization design that is constructed to target pre-specified allocation proportions [1]. With binary outcomes and two treatment arms ($k = 1, 2$), for example, the target allocation proportions (denoted by v_1 and v_2) given by the limiting allocation of the randomize-play-the-winner design are

$$v_1 = \rho_1(\theta_1, \theta_2) = \frac{1 - \theta_2}{(1 - \theta_1) + (1 - \theta_2)}, \quad v_2 = \rho_2(\theta_1, \theta_2) = \frac{1 - \theta_1}{(1 - \theta_1) + (1 - \theta_2)}, \quad (1)$$

where θ_k denote the unknown overall success probability in treatment arm k . The target proportions given by the Neyman allocation are

$$v_1 = \rho_1(\theta_1, \theta_2) = \frac{\sqrt{\theta_1(1-\theta_1)}}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_2(1-\theta_2)}}, \quad v_2 = \rho_2(\theta_1, \theta_2) = \frac{\sqrt{\theta_2(1-\theta_2)}}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_2(1-\theta_2)}}.$$

This target driven nature separates the DBCD from other intuitively motivated designs including urn models. Since first proposed by Eisele [9] and Eisele and Woodroffe [10], the design has gained recognition thanks to its nice properties: the DBCD can be constructed to target various possible desired allocation proportions, and often attains smaller asymptotic variance compared to other designs with the same limiting allocation proportions [11–13]. Smaller asymptotic variance is desirable as the asymptotic variance of sample allocation proportions has been shown directly related to the asymptotic power loss of response-adaptive randomization procedures [11].

The DBCD differs from other biased coin designs in that it uses allocation rules that rely on both the sample allocation proportions and the estimates of the target proportions. Suppose subject $(m+1)$ is ready for treatment assignment. The DBCD employs allocation rules that assess how close the current sample proportions are to the estimated allocation targets based on the outcome data of m previously enrolled subjects and compute allocation ratios for the subject $(m+1)$'s treatment assignment as functions of the discrepancies. With the overall success probabilities θ_1, θ_2 unknown, the allocation targets $\rho_k(\theta_1, \theta_2)$, $k = 1, 2$, are estimated by $\hat{\rho}_k = \rho_k(\hat{\theta}_{m,1}, \hat{\theta}_{m,2})$, where $\hat{\theta}_{m,k}$ denote estimators of θ_k based on the currently enrolled m subjects' outcome data. Studies examine how quickly and reliably the sample proportions converge to the targets for the asymptotic properties of the DBCD. These properties depend on the choices of the target proportions, allocation rules, and the estimators $\hat{\theta}_{m,k}$. Delay in the outcome affects the efficiency of the estimators $\hat{\theta}_{m,k}$, and consequently, the efficiency of the sample proportions $\hat{\rho}_k$ by creating a lag between the number of subjects accrued and the information known at the time of the estimation. In order to investigate the impact of the efficiency loss on $\hat{\rho}_k$, we fix allocation rules and target proportions, and consider two estimators of θ_k , a naïve or delay non-mitigating estimator and a delay mitigating estimator in this section.

2.1. Delay adaptive estimation

We expand Hu *et al.*'s [1] setup for a delayed binary outcome to include a short-term predictor. For simplicity, we consider a binary short-term predictor ($z = 0, 1$), and two treatment arms ($k = 1, 2$) without loss of generality. Suppose subjects arrive sequentially. Let $X_m = (X_{m,1}, X_{m,2})$ with $X_{m,k} = 0, 1$ for $k = 1, 2$ denote the treatment assignment of subject m , and let $Y_m = (Y_{m,1}, Y_{m,2})$ and $Z_m = (Z_{m,1}, Z_{m,2})$ denote the outcome and the short-term predictor, so if the subject is assigned to treatment 1, then $X_m = (1, 0)$ and $Y_{m,1}$ and $Z_{m,1}$ denote the corresponding outcome and short-term predictor. For $m = 1, 2, \dots$ and $k = 1, 2$, we assume $(Y_{m,k}, Z_{m,k})$ are independent random vectors with

$$E(Z_{m,k}=z) = \pi_k \quad \text{and} \quad E(Y_{m,k}|Z_{m,k}=z) = \theta_k^z, \quad (2)$$

where $\pi_1 + \pi_2 = 1$ and $\theta_k^0 \neq \theta_k^1$. Then, the overall mean is given as

$$E(Y_{1,k}) = \theta_k = \pi_k \theta_k^1 + (1 - \pi_k) \theta_k^0, \quad k=1, 2.$$

Suppose subject $(m+1)$ is enrolled and ready for treatment assignment. For the previously enrolled subject $j, j = 1, \dots, m$, we let $\delta_{j,k}(m)$ be indicators denoting whether the outcomes $Y_{j,k}$ are observed and available for the treatment assignment of subject $m+1$. We consider two estimators $\tilde{\theta}_{m,k}$ and $\bar{\theta}_{m,k}$ given as follows:

$$\tilde{\theta}_{m,k} = \frac{\sum_{j=1}^m X_{j,k} \delta_{j,k}(m) Y_{j,k} + \theta_{0,k}}{\sum_{j=1}^m X_{j,k} \delta_{j,k}(m) + 1}, \quad \bar{\theta}_{m,k} = \frac{\sum_{j=1}^m X_{j,k} \hat{Y}_{j,k}(m) + \theta_{0,k}}{\sum_{j=1}^m X_{j,k} + 1}, \quad (3)$$

where $\theta_{0,k}$ is some initial mean estimate for the treatment arm k and

$$\hat{Y}_{j,k}(m) = \begin{cases} Y_{j,k}, & \text{if } \delta_{j,k}(m) = 1 \\ \hat{E}(Y_{j,k} | Z_{j,k} = z), & \text{if } \delta_{j,k}(m) = 0 \end{cases}, \quad j=1, \dots, m.$$

Here, $\hat{E}(Y_{j,k} | Z_{j,k} = z)$ denotes the imputed value for the delayed and hence unobserved outcome. Under the condition (2) and assuming no prior or external knowledge, we may use the sample mean of the currently observed outcomes with the same predictor value in the same treatment arm group as follows:

$$\hat{E} \left[Y_{j,k} | Z_{j,k} = z, \{(X_{i,k}, Y_{i,k}, Z_{i,k}, \delta_{i,k}(m))\}_{1 \leq i \leq m} \right] = \frac{\sum_{j=1}^m X_{j,k} I[Z_{j,k} = z] \delta_{j,k}(m) Y_{j,k}}{\sum_{j=1}^m X_{j,k} I[Z_{j,k} = z] \delta_{j,k}(m)}, \quad (4)$$

where $I[\cdot]$ denotes an indicator function. This is mean imputation conditioning on the short-term predictor. We call $\tilde{\theta}_{m,k}$ a delay adaptive estimator. In contrast, $\bar{\theta}_{m,k}$ is defined by the currently observed outcome data, which is typical with standard outcome-adaptive randomization designs. We call $\bar{\theta}_{m,k}$ a naive or delay non-adaptive estimator.

2.2. Outcome delay mechanism

The outcomes may not be available immediately, while the short-term predictors are, and delays may differ by the short-term predictors. We expand Hu *et al.*'s [1] delay mechanism and define a delay mechanism that is only homogeneous conditioning on the short-term predictor.

Let t_m be the entry time of subject m , so $\{t_m\}_{m=1}^n$ is an increasing sequence of random variables. Let $\tau_{m,k}^z$ be the time from the entry to which the outcome $Y_{m,k}$ is observed when $Z_{m,k} = z$. At the time of subject $(m+1)$'s treatment assignment, we consider time interval $t_{m+1} - t_j$ for some previously enrolled subject $j, 1 \leq j \leq m$. This is the interval between the two

subjects' entry times. The interval can be seen as collection of $m - j + 1$ non-overlapping subintervals progressively defined by entry times of subjects successively enrolled after subject j as follows: $(t_j, t_{j+1}]$, $(t_{j+1}, t_{j+2}]$, ..., $(t_m, t_{m+1}]$ or simply $\{(t_{j+l}, t_{j+l+1})\}_{l=0}^{m-j}$. If $Y_{j,k}$ is observed by the treatment assignment time of subject $m + 1$, then $Y_{j,k}$ must have been observed within one of these subintervals. We consider an indicator function for each of the non-overlapping intervals that takes the value 1 if $Y_{j,k}$ is observed within the particular interval: given $Z_{m,k} = z$,

$$M_k^z(j, l) = I \left\{ t_j + \tau_{j,k}^z \in (t_{j+l}, t_{j+l+1}] \right\}, \quad j=0, \dots, m-j.$$

$M_k^z(j, 0)=1$ indicates $t_j + \tau_{j,k}^z \in (t_j, t_{j+1}]$, that is, $Y_{j,k}$ is observed by the time the next subject $j + 1$ enters the study. This is equivalent to no delay. $M_k^z(j, l)=1$ with $l > 0$ indicates $Y_{j,k}$ is not observed until the treatment assignment of subject $j+l$, but is observed and available for the treatment assignment of subject $j + l + 1$.

We assume $\{M_k^z(j, l), j = 1, 2, \dots\}$ is a sequence of iid random variables across subjects with $\mu_{l,k}^z = E \{M_k^z(j, l)\}$ for each fixed l, k , and z . This condition follows naturally from the usual iid assumption on $\{t_{m+1} - t_m\}$ and $\{\tau_{m,k}^z\}$ for each k and z . Note that given k and z , $\{M_k^z(j, l), l = 0, 1, \dots\}$ defines delay for subject j . As the delay mechanism is not constrained to be same across the short-term predictor, delays are only conditionally homogeneous within each treatment arm. We assume the following condition on the delay times.

Assumption 1—For some $\phi > 0$,

$$\sum_{i=l}^{\infty} \mu_{i,k}^z = P \left(\tau_{m,k}^z > t_{m+l} - t_m \right) = o(l^{-\phi}), \quad k=1, 2. \quad (5)$$

The probability in (5) is the probability that the outcome on a subject assigned to treatment k is observed with the short-term predictor z after treatment assignment has been made on at least next l subjects. Similar to the assumption 1 of Hu *et al* [1], this assumption implies that the delays cannot be unreasonably large relative to the accrual process. We consider a practical setting where the accrual mechanism generates a Poisson process and the delay has an exponential distribution within each treatment conditioning on the short-term predictor, so $\{\tau_{m,k}^z\}$ and $\{t_{m+1} - t_m\}$ are respectively sequences of iid exponential random variables with mean parameters λ_k^z and λ^* . The probability in (5) is simply $\{\lambda^* / (\lambda^* + \lambda_k^z)\}^l$, and the assumption is satisfied.

2.3. Asymptotic results

We denote the total number of subjects assigned to treatment arm k by $N_{n,k}$. The goal of the DBCD is to have the sample allocation probability $N_{n,k}/n$ converges to a certain desired allocation target v_k as $n \rightarrow \infty$. To this end, allocation rules compute subject $(m + 1)$'s

treatment allocation ratios as functions of discrepancies between the sample proportions and the estimated targets. We let $g_k(N_m/m, \hat{\rho}_m)$ denote such allocation rules, where $N_m/m = (N_{m,1}/m, N_{m,2}/m)$ denote the sample allocation proportions and $\hat{\rho}_m = (\hat{\rho}_{m,1}, \hat{\rho}_{m,2})$ with $\hat{\rho}_{m,k} = \rho_k(\hat{\theta}_{m,1}, \hat{\theta}_{m,2})$ denote the estimated targets based on the data up to the currently enrolled m subjects. The strong consistency results in the succeeding text show that both the simple or delay non-adaptive estimator and the delay adaptive estimator provide $\hat{\rho}_m$ that attains the target allocation ratios asymptotically. The proof of Theorem 1 is provided in the Appendix.

Theorem 1—Assume the condition in (2) holds with $0 < \pi_k < 1$ and $E(|Y_{1,k}| | Z_{1,k} = z) < \infty$, $k = 1, 2$. Suppose that Assumption 1 holds and $g_1(v, w)$ is a continuous function that is strictly decreasing in v_1 and strictly increasing in w_1 , and $g_1(v, w) = v_1$, if $v_1 = w_1$. If $\rho_k(\cdot, \cdot)$ is a continuous function, then, $N_{n,k}/n \rightarrow v_k$, $\hat{\theta}_{n,k} \rightarrow \theta_k$, and $\rho_k(\hat{\theta}_{n,1}, \hat{\theta}_{n,2}) \rightarrow v_k$ almost surely, as $n \rightarrow \infty$. Similarly, $\hat{\theta}_{n,k} \rightarrow \theta_k$, and $\rho_k(\hat{\theta}_{n,1}, \hat{\theta}_{n,2}) \rightarrow v_k$ almost surely, as $n \rightarrow \infty$.

As $v_1 + v_2 = w_1 + w_2 = 1$ and $g_1(v, w) + g_2(v, w) = 1$, the strict monotonicity conditions of $g_1(\cdot, \cdot)$ in v_1 and w_1 are required, so the next subject will be assigned to treatment k with a probability that is smaller than $\hat{\rho}_{m,k}$ if the current sample allocation proportion $N_{m,k}/m$ for the treatment k exceeds the estimated target $\hat{\rho}_{m,k}$. With additional regularity conditions assumed, we expect other asymptotic results of Hu *et al.* [1] such as asymptotic normality will similarly hold. This means that the DBCD is little affected if delays, although heterogeneous, are not unreasonably long relative to an accrual process.

3. Doubly adaptive biased coin designs for the motivating pediatric ulcerative colitis study

We propose three different imputation-based delay mitigating DBCDs for the pediatric UC study that motivates this paper. Along with a standard or delay non-mitigating DBCD, the designs differ in the treatment of delayed outcomes, but are identical in other aspects by targeting the same allocation proportions and using the same allocation rules.

3.1. Pediatric ulcerative colitis study

Corticosteroids are known to negatively influence stature and bone turnover regardless of disease activity in pediatric UC [3]. Treatments target to attain SFR. Current treatment regimens, however, are far from optimal with 45% remaining dependent on CS 1 year after diagnosis, and up to 26% requiring colectomy within 5 years [14, 15]. Because of lack of clinical trial data in children, treatment currently relies exclusively upon schemas derived from adults. Aminosalicylates (5-ASA) or CS are used to induce remission, followed by maintenance with 5-ASA and/or an immunomodulator (IM) [14, 16, 17].

Predicting outcome to standardized pediatric colitis therapy (PROTECT) is an ongoing observational study, which aims to develop a risk prediction model based on data available at diagnosis that separates those who will be in a stable remission treated alone with 5-ASA, the least toxic drug to treat UC [18], from those who will not and may potentially benefit from an early introduction of IM. The risk prediction model, if successfully established, will be used in a future randomized controlled trial to create two strata, high risk and low risk, to separately test efficacy of the early IM introduction strategy in each stratum. The early

introduction of IM is hypothesized to only benefit the high risk group that is likely not to attain and maintain SFR if treated alone with 5-ASA.

Figure 1 describes the design of the future randomized controlled trial. Upon enrollment, patients will be placed in low or high risk stratum by the risk prediction model, and subsequently randomly assigned to control or early IM intervention group within each stratum. Randomization rules will be separately applied to allow treatment allocation ratios independently adapted within each stratum in order to test the hypothesized differential efficacy of the early IM introduction strategy between the strata.

The primary outcome will be whether a patient attains SFR within 1 year from diagnosis (year 1 SFR). Preliminary results report that the year 1 SFR was observed at a twofold higher rate in those who achieved clinical remission by week 4 after diagnosis. Clinical remission by week 4 that is observable rather immediately compared with the delayed primary outcome will be considered as a short-term predictor. This implies that delays or time to failed outcomes may only be homogeneous conditioning on the short-term week 4 clinical remission status.

3.2. Naïve doubly adaptive biased coin design

This is a standard DBCD. We consider all outcomes are delayed by 1 year, as successful outcomes are observed at year 1. The indicator function $\delta_{j,k}(m)$ that denotes the availability of the subject j 's outcome for the treatment assignment of subject $m + 1$ is given as follows:

$$\delta_{j,k}(m) = \begin{cases} 1, & \text{if } t_{m+1} - t_j \geq 1 \\ 0, & \text{if } t_{m+1} - t_j < 1 \end{cases} \quad (6)$$

The delay non-adaptive estimator $\tilde{\theta}_{m,k}$ is accordingly defined as in equation (3). We call a DBCD using this naïve estimator *naïve DBCD*.

3.3. Delay mitigation by conditional mean imputation doubly adaptive biased coin design

This design uses the conditional mean imputation defined in equation (3). With the indicator function $\delta_{j,k}(m)$ identically defined as in (6), the imputed conditional means replace delayed outcomes, and the delay adaptive estimator $\tilde{\theta}_{m,k}$ is defined accordingly as in equation (3). We call a DBCD using this conditional mean imputation *conditional mean imputation DBCD*.

3.4. Delay mitigation by continuous weighted conditional mean imputation doubly adaptive biased coin design

We note that a failed outcome becomes available any time when a patient receives rescue therapies within 1 year, while time to a failed outcome can be considered censored for anyone who has not failed but has not survived beyond 1 year without failing yet. This leads us to consider *continuous imputation* using the Kaplan–Meier estimator. We let $T_{m,k}^z$ denote time to a failed outcome of a patient m if the patient is assigned to treatment k and has a short-term outcome z . Then, the delay time, $\tau_{m,k}^z$, is given by $\tau_{m,k}^z = \min(T_{m,k}^z, 1)$. For the

treatment assignment of subject $(m+1)$, we have the following *follow-up time* and the indicator data observed on the previously enrolled subjects:

$$\tilde{\tau}_{j,k}^z(m) = \min(\tau_{j,k}^z, t_{m+1} - t_j) \text{ and } \delta_{j,k}(m) = \begin{cases} 1, & \text{if } t_{m+1} - t_j \geq \tau_{m,k}^z \\ 0, & \text{if } t_{m+1} - t_j < \tau_{m,k}^z \end{cases}.$$

We consider the Kaplan–Meier estimator of times to a failed outcome based on the follow-up time and the indicator data $\{(\tilde{\tau}_{i,k}^z(m), \delta_{i,k}(m))\}_{i=1}^m$ and denote it by $\hat{F}_{m,k}^z$. We modify the imputation in (4) to the continuous nature of the failure time distribution such that

$$\hat{E} \left[Y_{j,k} | Z_{j,k} = z, \left\{ (y_{i,k}, X_{i,k}, Z_{i,k}, \delta_{i,k}(m), \tau_{i,k}^z, t_i) \right\}_{1 \leq i \leq m, i \neq j} \right] = \sum_{i > j} \left(\frac{\Delta \hat{F}_{m,k}^z(\tilde{\tau}_{i,k}^z)}{1 - \hat{F}_{m,k}^z(\tilde{\tau}_{j,k}^z)} \right) I[Z_{i,k} = z] X_{i,k} Y_{i,k}, \quad (7)$$

where $\Delta \hat{F}_{m,k}^z(\tilde{\tau}_{i,k}^z)$ denote the point probability mass assigned to observed failure time $\tilde{\tau}_{i,k}^z$ by the Kaplan–Meier estimator with $\Delta \hat{F}_{m,k}^z(\tilde{\tau}_{i,k}^z) = 0$ for $\delta_{i,k}(m) = 0$, that is, censored or unobserved outcomes. This is conditional weighted mean imputation with the weights defined by the Kaplan–Meier estimator continuously. As $Y_{j,k}^z$ are binary, plugging the continuously imputed values in (7) in the equation (3) leads to a *continuously* delay adaptive estimator:

$$\bar{\theta}_{m,k} = \frac{\sum_{z=0}^1 N_{m,k}^z \left\{ 1 - \hat{F}_{m,k}^z(1^-) \right\} + \theta_{0,k}}{N_{m,k} + 1} \quad (8)$$

where $N_{m,k}^z$ denote the numbers of patients previously assigned to treatment k with observed short-term predictor value z , and $1 - \hat{F}_{m,k}^z(1^-)$ denote the Kaplan–Meier estimated proportion of $N_{m,k}^z$ patients that have not failed by the time subject is up for treatment assignment. The equation (8) follows as

$$\begin{aligned} & \sum_{j=1}^m \left[\delta_{j,k}(m) X_{j,k} Y_{j,k} + \{1 - \delta_{j,k}(m)\} \sum_{z=0}^1 \sum_{i > j} \left(\frac{\Delta \hat{F}_{m,k}^z(\tilde{\tau}_{i,k}^z)}{1 - \hat{F}_{m,k}^z(\tilde{\tau}_{j,k}^z)} \right) I[Z_{i,k} = z] X_{i,k} Y_{i,k} \right] \\ &= \sum_{z=0}^1 \sum_{i=1}^m \delta_{i,k}(m) X_{i,k} Y_{i,k} I[Z_{i,k} = z] \left[1 + \sum_{j < i} \left(\frac{\Delta \hat{F}_{m,k}^z(\tilde{\tau}_{i,k}^z)}{1 - \hat{F}_{m,k}^z(\tilde{\tau}_{j,k}^z)} \right) \right] \\ &= \sum_{z=0}^1 N_{m,k}^z \sum_{i=1}^m X_{i,k} Y_{i,k} I[Z_{i,k} = z] \Delta \hat{F}_{m,k}^z(\tilde{\tau}_{i,k}^z) \end{aligned}$$

by the self-consistency of the Kaplan–Meier estimator and $X_{i,k}$ and $Y_{i,k}$ are binary. We refer to Zhou and Li [19] for the details of the rearrangement in the equation. We call a DBCD using the estimator in (8) *continuous weighted conditional mean imputation DBCD*.

The idea of continuous updating for survival outcomes has been proposed in the literature. Cheung, Inoue, Wathen, and Thall [20] considered survival outcomes as continuous outcomes using Weibull distributions and continuously adapted allocation ratios in their Bayesian outcome-adaptive randomization approach. Rosenberger and Seshaiyer [21] used the log-rank statistics and continuously updated the allocation ratios. The proposed design uses *continuous imputation* for censored outcomes.

3.5. Delay mitigation by continuous weighted mean imputation doubly adaptive biased coin design

One may consider applying the continuous weighted mean imputation using the Kaplan–Meier estimator *unconditionally*, that is, without conditioning on the short-term predictor. This requires pooling the follow-up time data together and creates combined data,

$\left\{ \left(\tilde{\tau}_{i,k}^*(m), \delta_{i,k}(m) \right) \right\}_{i=1}^m$, where $\tilde{\tau}_{i,k}^*(m) = I[Z_{i,k}=1]\tilde{\tau}_{i,k}^1(m) + I[Z_{i,k}=0]\tilde{\tau}_{i,k}^0(m)$. With the Kaplan–Meier estimator of the pooled failure time distribution, this leads to the following estimator:

$$\bar{\theta}_{m,k} = \frac{N_{m,k} \left\{ 1 - \hat{F}_{m,k}(1^-) \right\} + \theta_{0,k}}{N_{m,k} + 1},$$

where $N_{m,k}$ denote the total number of patients previously assigned to treatment k , and $1 - \hat{F}_{m,k}(1^-)$ denote the Kaplan–Meier estimated proportion of $N_{m,k}$ patients that have not failed by the time subject is up for treatment assignment. We call this design *continuous weighted (unconditional) mean imputation DBCD*.

4. Simulation studies

We used the clinical context of the motivating pediatric UC study and compared the four designs proposed in Section 3 in order to assess how different imputation methods (including no imputation) affect finite sample performance of the DBCD. We simulated no delay data and used the performance of a standard DBCD under no delay as reference in the comparison. All designs included in the simulation studies were identical in all other aspects except for the treatment of delayed outcomes. They targeted the same allocation proportions and used the same allocation rules. We used the urn allocation proportions given in the equation (1) for the common target proportions. An allocation rule given in Hu and Zhang [13] was used for the common allocation rule:

$$g_k(r_{m,k}, \hat{\rho}_{m,k}) = \frac{\hat{\rho}_{m,k}(\tilde{\rho}_{m,k}/r_{m,k})^2}{\sum_{j=1}^2 \hat{\rho}_{m,j}(\tilde{\rho}_{m,j}/r_{m,j})^2} \text{ with } r_{m,k} = N_{m,k}/m \text{ and } \hat{\rho}_{m,k} = \rho_k(\tilde{\theta}_{m,1}, \hat{\theta}_{m,2}),$$

where the estimators $\hat{\theta}_{m,1}$, $\hat{\theta}_{m,2}$ are differently defined across the different DBCDs we compared. Reflecting the trial design in Figure 1, this allocation rule was separately applied by risk stratum with $\hat{\theta}_{m,k}$ evaluated within each stratum.

The simulation model based on the motivating pediatric UC study is as follows: let $k = 1$ denote the early IM treatment and $k = 2$ the control. For $i = 1, 2, \dots, n$,

Treatment groups: $X_{i,1} = 1$, if allocated to the early IM group, or $X_{i,2} = 1$.

Entry times: $t_i \sim iid F_t$,

Short-term predictors (week 4 clinical remission): $Z_{i,k}|X_{i,k} = 1 \sim iid Bernoulli(\pi_k)$,

Times to failed outcomes: $\tau_{i,k}^z|X_{i,k}=1, Z_{i,k}=z \sim iid exp(\eta_k^z)$,

Outcomes (year 1 SFR): $Y_{i,k}|X_{i,k}=1, Z_{i,k}=z \begin{cases} =1, & \text{if } \tau_{i,k}^z > 1 \\ =0, & \text{if } \tau_{i,k}^z < 1 \end{cases}$

We selected the value of η_k^z such that $P(\tau_{i,k}^z > 1) = \theta_{i,k}^z$, and $Y_{i,k}|X_{i,k} = 1, Z_{i,k} = z \sim iid Bernoulli(\theta_k^z)$.

Parameter values, π_k and $\theta_{i,k}^z$, and entry time distribution F_t were chosen to emulate the accrual process and anticipated treatment effects in the motivating pediatric UC study (Table I and Figure 2). Thirty-five percent of pediatric UC patients were high risk patients in whom the early IM intervention increased the probability of year 1 SFR from 0.2 to 0.5 compared with the control group. A null effect was assumed in the low risk stratum. The week 4 clinical remission (Z) was predictive of the year 1 SFR outcome with reasonably high sensitivity and specificity in both strata. We fixed the total sample size at $n = 310$. The sample size for the high risk stratum was determined probabilistically by the sampling probability of 35%. This gave an effective sample size of 108.5 on average for the high risk strata, which would have provided 80% power under equal allocation for the anticipated early IM intervention effect in the high risk group specified in Table I. With the parameter values specified in Table I, the target allocation probability given by the urn model in the equation (1) is 0.615 or 61.5% for the early IM intervention group in the high risk stratum. The urn allocation target proportion commanded treating on average 12.5 more patients with the intervention than an equal allocation randomization.

We anticipate that the accrual would be complete within 2 years with the rate constant over time. Accrual setting 1 represents the anticipated setting (Figure 2). Under this setting, on average 48.4% of the total sample was enrolled within the first year, that is, before the first successful outcome would be observed. This corresponds to a case of fast accrual or long delay. We additionally considered a slow accrual or short delay setting (setting 2 in Figure 2) in order to investigate relative performance of the designs in a more outcome-adaptive randomization favorable setting. Under setting 2, on average 12.3% of the total sample was enrolled within the first year.

How strongly the short-term predictor is associated with the outcome may affect the relative performance of the delay mitigating DBCDs. We manipulated the joint probabilities and the marginal probabilities of the short-term predictor (Z) in Table I, while fixing the marginal probabilities of the outcome (Y) at the same values, and considered two extreme cases additionally, a no association ($\rho = 0$) and an extremely strong association ($\rho > 0.95$).

The early introduction of IM is hypothesized to only benefit the high risk group that is likely not to attain SFR if treated alone with the standard 5-ASA treatment. We focus on the high risk stratum results. Figures 3 and 4 summarize the results from 10,000 simulated data per simulation setting. Results were summarized separately by accrual setting and strength of the association between the short-term predictor and the outcome. The different designs were denoted by the imputation methods they employed as *naïve*, *conditional mean*, *continuous weighted mean*, and *continuous weighted conditional mean imputation*, respectively. We use results obtained by a standard DBCD under no delay as reference in each simulation setting and compare the designs. As the designs are identical except for the treatment of delayed outcomes, we compare them for how closely and reliably the sample proportion of the early IM-treated patients approached the target allocation proportion across the 10,000 simulated data.

Figure 3 presents mean % treated patients results. With an effective sample of $n = 108.5$ on average for this high risk stratum, the mean % of treated patients was not much lower than the mean observed for no delay case and was already close to the target allocation 61.5% (denoted by the dimmed solid horizontal line) with all designs. This empirically confirms the results of Theorem 1 that the asymptotic properties of the DBCD are little affected as long as delays are not unreasonably large relative to the accrual process. The delay mitigating designs reported higher means than the delay non-mitigating design across all settings. The improvement was more noticeable under setting 1 with the longer delay (relative to setting 2). Different imputation methods also made noticeable differences under the setting 1. The two continuous weighted mean imputation methods brought more improvement than the conditional mean imputation method. Additionally utilizing the short-term predictor information in the weighted mean imputation via conditioning did not necessarily improve any further. This combined approach, however, reduced the variability by attaining the smaller standard error (SE) for the % treated. This is shown by the root mean squared (RMSE) results of the % early IM treated in Figure 4. The combined approach has uniformly smaller RMSEs than the continuous weighted mean imputation alone approach in all cases: the continuous weighted mean imputation alone reduced the RMSE by 11.3%, 13.5%, and 16.4% compared with the naïve (no imputation) approach for the no association, medium association, and high association cases, respectively. The combined approach, however, further reduced the RMSE additionally by 1.5%, 3.0%, and 3.8%. In no association case where the short-term predictor was not a predictor in fact, the RMSE of the combined imputation approach was not much smaller, but was noticeably smaller in the other two positive association cases. The RMSE results also show that in no association case, the conditional mean imputation may worsen the performance compared with the no imputation. It is because conditioning by the short-term predictor in imputation increased the variability of the estimator, whereas the conditional imputation did not add any information as the predictor was not a predictor in the no association case. The improvement brought by incorporating the short-term predictor via conditional mean imputation was comparable with the improvement by the continuous weighted imputation only in the high yet unrealistic association case.

The improvement brought by the imputation-based designs can be better appreciated when the reduced SE results are translated to estimated sample size increase needed for the naïve (no imputation) design to attain the same precision (the same SE) (Table II). Here, the SE measures precision of the designs as to how reliably the sample proportion of the early IM-treated patients approached the target allocation proportion across the 10,000 simulated data with different designs. Using the relationship between the SE and the sample size, we find that under the accrual setting 1 where the accrual is fast relative to the delay as anticipated in the real-life UC trial, the naïve design needs a larger sample by 5.1%~16.7% in order to attain the same precision as the continuous weighted mean imputation-based design. The estimated sample size increase is much larger, 8.8%~28.4%, that is needed for the naïve design to attain the same precision as the combined approach (continuous weighted conditional mean imputation). Under the slow accrual setting, the estimated sample size increase is smaller yet appreciable (5.1%~11.5%).

These simulation results are quite general albeit based on one clinical trial setting. The motivating real-life UC trial epitomizes clinical trials for a delayed outcome with a short-term predictor. Delayed outcomes are most commonly positively correlated with short-term predictors similarly between treatment arm groups. We simulated this clinical context under two accrual settings (fast/slow) with the delayed outcome differently correlated with the short-term predictors. Hence, the simulated settings cover representative cases of zero, medium, and high correlations each paired with fast and slow accruals (relative to delay in the outcome). Unique to the real life UC trial in the simulation setting are the values of the marginal probabilities, which have been shown in the literature not affecting comparison of trial designs except the magnitude of differences.

5. Conclusion

We examined the impact of delay in the outcome variable on the performance of outcome-adaptive randomization designs for a case where a short-term predictor is available for the delayed outcome. When a short-term predictor is not considered and delays are iid, the asymptotic properties of many outcome-adaptive randomization designs have been shown little affected [1,4–7]. On the contrary, Huang *et al.* [8] reported that a Bayesian joint modeling that incorporates the short-term predictor information adapted allocation ratios more efficiently. These seemingly conflicting results motivated this study. We found that delays, even if non-homogeneous, little affect the performance of the DBCD in the sense that a standard DBCD utilizing only completely observed outcomes attains target allocation ratios in the limit unless they are unreasonably large relative to the accrual process.

Empirical studies, however, show that finite sample performance of the DBCD was improved by considering delayed outcomes as missing and imputing them. Three delay mitigating DBCDs that differ by the imputation method were considered for the motivating pediatric UC trial. The conditional mean imputation DBCD imputed delayed outcomes conditionally on by the observed value of the short-term predictor. The continuous weighted mean imputation designs were motivated by the fact that delay times are times to failures. The RMSE results show that imputation, particularly, continuous imputation, improved the performance in finite samples. In comparison to the effect of continuous imputation,

incorporating the short-term predictor only via conditional mean imputation improved the performance moderately. We conjecture that the relatively smaller gain by the conditional mean imputation itself may be due to the fact that both the short-term and the outcome are binary. If both the predictor and the outcome had been continuous variables, the conditional imputation might have brought greater improvement, as the conditioning tends to reduce the SE of the estimator much more significantly with continuous outcomes.

Acknowledgments

This research was supported in part by a National Cancer Institute (NCI) award CA16672, an Institutional Clinical and Translational Science (CTSA) Award 8UL1TR000077-04, and a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) award U01 DK095745-01.

References

1. Hu F, Zhang LX, Cheung SH, Chan WS. Doubly adaptive biased coin designs with delayed responses. *Canadian Journal of Statistics-Revue Canadienne De Statistique*. 2008; 36:541–559.
2. Tamura RN, Faries DE, Andersen JS, Heiligenstein JH. A case-study of an adaptive clinical-trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association*. 1994; 89:768–776.
3. Hyams JS, Carey DE. Corticosteroids and growth. *Journal of Pediatrics*. 1988; 113:249–254. [PubMed: 3294360]
4. Bai ZD, Hu F, Rosenberger WF. Asymptotic properties of adaptive designs for clinical trials with delayed response. *Annals of Statistics*. 2002; 30:122–139.
5. Hu F, Zhang LX. Asymptotic normality of urn models for clinical trials with delayed response. *Bernoulli*. 2004; 10:447–463.
6. Sun RB, Cheung SH, Zhang LX. A generalized drop-the-loser rule for multi-treatment clinical trials. *Journal of Statistical Planning and Inference*. 2007; 137:2011–2023.
7. Zhang LJ, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *Journal of the Royal Statistical Society Series C-Applied Statistics*. 2007; 56:153–165.
8. Huang XL, Ning J, Li YS, Estey E, Issa JP, Berry DA. Using short-term response information to facilitate adaptive randomization for survival clinical trials. *Statistics in Medicine*. 2009; 28:1680–1689. [PubMed: 19326367]
9. Eisele JR. The doubly adaptive biased coin design for sequential clinical-trials. *Journal of Statistical Planning and Inference*. 1994; 38:249–261.
10. Eisele JR, Woodroffe MB. Central limit theorems for doubly adaptive biased coin designs. *The Annals of Statistics*. 1995; 23:234–254.
11. Hu, F.; Rosenberger, WF. *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley & Sons; Hoboken, New Jersey: 2006.
12. Rosenberger WF, Hu F. Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials*. 2004; 1:141–147. [PubMed: 16281886]
13. Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics*. 2004; 32:268–301.
14. Hyams JS, Markowitz J, Lerer T, Griffiths A, Mack D, Bousvaros A, Otley A, Evans J, Pfefferkorn M, Rosh J, Rothbaum R, Kugathasan S, Mezooff A, Wyllie R, Tolia V, Delrosario JF, Moyer MS, Oliva-Hemker M, Leleiko N. The natural history of corticosteroid therapy for ulcerative colitis in children. *Clinical Gastroenterology and Hepatology*. 2006; 4:1118–1123. [PubMed: 16820327]
15. Hyams JS, Davis P, Grancher K, Lerer T, Justinich CJ, Markowitz J. Clinical outcome of ulcerative colitis in children. *Journal of Pediatrics*. 1996; 129:81–88. [PubMed: 8757566]
16. Hyams JS, Lerer T, Griffiths A, Pfefferkorn M, Stephens M, Evans J, Otley A, Carvalho R, Mack D, Bousvaros A, Rosh J, Grossman A, Tomer G, Kay M, Crandall W, Oliva-Hemker M, Keljo D, LeLeiko N, Markowitz J, Coll PIBD. Outcome following infliximab therapy in children with

- ulcerative colitis. *American Journal of Gastroenterology*. 2010; 105:1430–1436. [PubMed: 20104217]
17. Hyams JS, Lerer T, Mack D, Bousvaros A, Griffiths A, Rosh J, Otley A, Evans J, Stephens M, Kay M, Keljo D, Pfefferkorn M, Saeed S, Crandall W, Michail S, Kappelman MD, Grossman A, Samson C, Sudel B, Oliva-Hemker M, LeLeiko N, Markowitz J, Coll PIBD. Outcome following thiopurine use in children with ulcerative colitis: a prospective multicenter registry study. *American Journal of Gastroenterology*. 2011; 106:981–987. [PubMed: 21224840]
 18. Rosh JR, Gross T, Mamulo P, Griffiths A, Hyams JS. Hepatosplenic T-cell lymphoma in adolescents and young adults with Crohn's disease: a cautionary tale? *Inflammatory Bowel Diseases*. 2007; 13:1024–1030. [PubMed: 17480018]
 19. Zhou M, Li G. Empirical likelihood analysis of the Buckley-James estimator. *Journal of Multivariate Analysis*. 2008; 99:649–664. [PubMed: 19018294]
 20. Cheung YK, Inoue LYT, Wathen JK, Thall PF. Continuous Bayesian adaptive randomization based on event times with covariates. *Statistics in Medicine*. 2006; 25:55–70. [PubMed: 16025549]
 21. Rosenberger WF, Seshaiyer P. Adaptive survival trials. *Journal of Biopharmaceutical Statistics*. 1997; 7:617–624. [PubMed: 9358333]

Appendix A: Proof of Theorem 1

Let $T_{n,k}^z$ denote the number of the outcomes with the short-term predictor values z that are observed up to the time when subject $(n + 1)$ is ready for treatment assignment. We let $S_{n,k}^z$ denote the corresponding sum of the completely observed outcomes. Then, note that

$$\begin{aligned} S_{n,k}^z &= \sum_{m=1}^n \sum_{l=m}^n M_k^z(m, l-m) X_{m,k} I[Z_{m,k}=z] Y_{m,k} = \sum_{m=1}^n X_{m,k} I[Z_{m,k}=z] Y_{m,k} \sum_{l=0}^{n-m} M_k^z(m, l), \\ T_{n,k}^z &= \sum_{m=1}^n X_{m,k} I[Z_{m,k}=z] \sum_{l=0}^{n-m} M_k^z(m, l). \end{aligned} \quad (\text{A.1})$$

As $Y_{j,k}$ is observed only once, there exists only one l for every triad of j, k , and z , so that $M_k^z(j, l) = 1$ and $M_k^z(j, l^*) = 0$ for all $l \neq l^*$. As

$$\delta_{j,k}(n) = Z_{j,k} \sum_{l=0}^{n-j} M_k^1(j, l) + (1 - Z_{j,k}) \sum_{l=0}^{n-j} M_k^0(j, l), \text{ we have}$$

$$\hat{E} \left[Y_{j,k} | Z_{j,k}=z, \{(X_{i,k}, Y_{i,k}, Z_{i,k}, \delta_{i,k}(n))\}_{1 \leq i \leq n} \right] = S_{n,k}^z / T_{n,k}^z.$$

After some algebraic operations, we have

$$\begin{aligned} \tilde{\theta}_{n,k} &= \frac{S_{n,k}^1 + S_{n,k}^0 + \theta_{0,k}}{T_{n,k} + 1} = \frac{S_{n,k}^1}{N_{n,k}^1} \frac{N_{n,k}^1}{N_{n,k}} \frac{N_{n,k}}{T_{n,k}} + \frac{S_{n,k}^0}{N_{n,k}^0} \frac{N_{n,k}^0}{N_{n,k}} \frac{N_{n,k}}{T_{n,k}} + O(1/T_{n,k}), \\ \tilde{\theta}_{n,k} &= \frac{(N_{n,k}^1/T_{n,k}^1) S_{n,k}^1 + (N_{n,k}^0/T_{n,k}^0) S_{n,k}^0 + \theta_{0,k}}{N_{n,k} + 1} = \frac{S_{n,k}^1}{N_{n,k}^1} \frac{N_{n,k}^1}{T_{n,k}^1} \frac{N_{n,k}^1}{N_{n,k}} + \frac{S_{n,k}^0}{N_{n,k}^0} \frac{N_{n,k}^0}{T_{n,k}^0} \frac{N_{n,k}^0}{N_{n,k}} + O(1/N_{n,k}). \end{aligned} \quad (\text{A.2})$$

By Lemma 1 of Hu *et al* [1], for each $k = 1, 2$ and $z = 0, 1$, in the event that $\{N_{n,k}^z \rightarrow \infty\}$,

$$\frac{T_{n,k}^z}{N_{n,k}^z} = \frac{\sum_{m=1}^n \sum_{l=0}^{n-m} X_{m,k} I[Z_{m,k}=z] M_k^z(m, l)}{N_{n,k}^z} \rightarrow 1 \text{ a.s.} \quad (\text{A.3})$$

Also by Lemma 1 of Hu *et al.* [1] and (A.2), for each $k = 1, 2$ and $z = 0, 1$, in the event that $\{N_{n,k}^z \rightarrow \infty\}$,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{\sum_{m=1}^n \sum_{l=n-m+1}^{\infty} X_{m,k} I[Z_{m,k}=z] M_k^z(m, l) |Y_{n,k}|}{N_{n,k}^z} \\ & \leq 0 + E[I[Z_{m,k}=z] | Y_{n,k} | I\{|Y_{n,k}| \geq A\}] \rightarrow 0 \text{ a.s., as } A \rightarrow \infty. \end{aligned} \quad (\text{A.4})$$

By (A.3) and (A.4), we have that for each $k = 1, 2$ and $z = 0, 1$, in the event that $\{N_{n,k}^z \rightarrow \infty\}$,

$$\frac{S_{n,k}^z}{N_{n,k}^z} = \frac{\sum_{m=1}^n X_{m,k} I[Z_{m,k}=z] Y_{n,k}}{N_{n,k}^z} - \frac{\sum_{m=1}^n \sum_{l=n-m+1}^{\infty} X_{m,k} I[Z_{m,k}=z] M_k^z(m, l) Y_{n,k}}{N_{n,k}^z} \rightarrow \theta_k^z \text{ a.s.} \quad (\text{A.5})$$

Similarly, by using the arguments of Lemma 1 of Hu *et al* [1], under condition in (2), for each $k = 1, 2$, in the event that $\{N_{n,k} \rightarrow \infty\}$,

$$\frac{N_{n,k}^1}{N_{n,k}} = \frac{\sum_{m=1}^n X_{m,k} I[Z_{m,k}=1]}{\sum_{m=1}^n X_{m,k}} \rightarrow \pi_k \text{ a.s.} \quad (\text{A.6})$$

It follows from (A.2), (A.3), (A.5), and (A.6) that for each $k = 1, 2$, in the event that $\{N_{n,k}^z \rightarrow \infty\}$,

$$\tilde{\theta}_{n,k} \rightarrow \theta_k \text{ and } \bar{\theta}_{n,k} \rightarrow \theta_k \text{ a.s.}$$

By the continuity of $\rho_k(\cdot, \cdot)$ and the arguments of the proof of Theorem 1 of Hu *et al.* [1], as $n \rightarrow \infty$, $\rho_k(\tilde{\theta}_{n,1}, \tilde{\theta}_{n,2}) \rightarrow \rho_k(\theta_1, \theta_2) = v_k$ and $\rho_k(\bar{\theta}_{n,1}, \bar{\theta}_{n,2}) \rightarrow \rho_k(\theta_1, \theta_2) = v_k$ a.s.

Figure 1.
Flow chart of the future randomized controlled pediatric ulcerative colitis trial.

Figure 2.
Accrual process.

Figure 3.
Mean % early immunomodulator-treated patient results.

Figure 4.
Root mean squared error (RMSE) results of the % early immunomodulator-treated patients.

Table 1

Simulation setting.

	Low risk (65%)				High risk (35%)			
	control		Early immunomodulator intervention		control		Early immunomodulator intervention	
	Y = 1	Y = 0	Y = 1	Y = 0	Y = 1	Y = 0	Y = 1	Y = 0
Z = 1	0.5	0.2	0.7	0.144	0.2	0.344	0.36	0.185
Z = 0	0.1	0.2	0.3	0.056	0.6	0.656	0.14	0.315
	0.6	0.4	1	0.2	0.8	1	0.5	0.5
Cor* (Y,Z)		0.356		0.396				0.351

* Correlation between the delayed outcome (Y) and the short-term predictor (Z)

Table II

Estimated % increase in the sample size for the naïve design

Accrual setting	Association with the short-term predictor	Design	Standard error ratio	Estimated % increase*
1 (Fast accrual)	Independent	Naïve (no imputation)	1	—
		Conditional mean imputation	1.047	−8.8%
		Continuous weighted mean imputation	0.976	5.1%
		Continuous conditional weighted mean imputation	0.958	8.8%
		No delay	0.829	45.3%
	Medium	Naïve (no imputation)	1	—
		Conditional mean imputation	0.993	1.3%
		Continuous weighted mean imputation	0.968	6.8%
		Continuous conditional weighted mean imputation	0.931	15.5%
		No delay	0.828	45.8%
2 (Slow accrual)	High	Naïve (no imputation)	1	—
		Conditional mean imputation	0.883	28.4%
		Continuous weighted mean imputation	0.926	16.7%
		Continuous conditional weighted mean imputation	0.883	28.4%
		No delay	0.815	50.4%
		Naïve (no imputation)	1	—
	Independent	Conditional mean imputation	1.003	−0.6%
		Continuous weighted mean imputation	0.976	5.1%
		Continuous conditional weighted mean imputation	0.969	6.4%
		No delay	0.931	15.3%
		Naïve (no imputation)	1	—
		Conditional mean imputation	0.981	4.0%
	Medium	Continuous weighted mean imputation	0.983	3.4%
		Continuous conditional weighted mean imputation	0.965	7.4%
		No delay	0.949	11.0%
		Naïve (no imputation)	1	—
		Conditional mean imputation	0.965	7.5%
		Continuous weighted mean imputation	0.963	7.8%

Accrual setting	Association with the short-term predictor	Design	Standard error ratio	Estimated % increase*
		Continuous conditional weighted mean imputation	0.947	11.5%
		No delay	0.920	18.2%

* Estimated % of sample size increase needed for the naive design to attain the same precision (standard error) as the imputation based or no delay designs.