Published in final edited form as:

J Biopharm Stat. 2012 ; 22(4): 700–718. doi:10.1080/10543406.2012.676534.

ESTIMATION OF TREATMENT EFFECT FOLLOWING A CLINICAL TRIAL WITH ADAPTIVE DESIGN

Xiaolong Luo^{1,2}, Mingyu Li¹, Weichung Joe Shih², and Peter Ouyang¹

¹Celgene Corporation, Summit, New Jersey, USA

²University of Medicine and Dentistry of New Jersey, New Jersey, USA

Abstract

Parameter estimation following an adaptive design or group sequential design has been extremely challenging due to potential random high from its face value estimate. In this paper, we introduce a new framework to model clinical trial data flow based on a marked point process (MPP). The MPP model allows us to use methods of stochastic calculus for analyses of any adaptive clinical trial. As an example, we apply this method to a two stage treatment selection design and derive a procedure to estimate the treatment effect. Numerical examples will be used to evaluate the performance of the proposed procedure.

Keywords

Adaptive design; Clinical trial; Parameter estimation; Treatment selection

1. INTRODUCTION

Clinical trials are often used to compare experimental treatments and standard of care in a patient population. They serve as a basis for regulatory approval for marketing a new drug. A good clinical trial design should minimize resources required to gain sufficient information regarding the safety and efficacy of the study treatments. However, designing an efficient trial can be a challenge task due to uncertainty about the ideal study drug dosage and the natural variation of the study endpoint. For this reason, there has been much interest in considering adaptive clinical trial design in which some design features can be optimized as more information becomes available during the trial conduct itself. This concept was considered by Box and Jenkins (1962) and studied in earlier literature report, including those on group sequential design (Amitage, 1958; Jennison and Turnbull, 1999; Stein, 1945) and adaptive randomization (Pocock and Simon, 1975; Zelen, 1969). Recently, there has been a great deal of renewed interest in this area, and numerous papers were published in sample size reestimation (Bauer and Köhne, 1994; Cui et al., 1999; Chow and Chang, 2007; Fisher,

Address correspondence to Xiaolong Luo, Biostatistics, Clinical R&D, Celgene Corporation, 86 Morris Avenue, Summit, NJ 07901, USA; xluo@celgene.com.

Copyright of Journal of Biopharmaceutical Statistics is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

1998; Liu and Chi, 2001; Proschan and Hunsberger, 1995), adaptive treatment selection (Li et al., 2009; Luo et al., 2010; Stallard and Friede, 2008), population enrichment (Wang et al., 2007) and covariate adjustment and response adaptive designs (Hu and Rosenberger, 2003; Zhang et al., 2007). A comprehensive publications can be found in the draft guidance on adaptive design recently published by the U.S. Food and Drug Administration (FDA).

There are two fundamental questions after a clinical trial is completed. One concerns the magnitude of the experimental treatment efficacy as compared with the control, which is emphasized in the European Medicines Agency (EMA) guidance for adaptive trial design. The other concerns whether such magnitude is significantly different from zero (i.e., no treatment effect), which is important in making a regulatory decision. In practical application, one would like to know the actual outcome improvement from a new treatment. Therefore, in some sense, an appropriate estimation of the potential treatment effect is fundamentally important.

With adaptive clinical trial design, the estimation of treatment effect could be controversial. In the recent US. FDA draft guidance on adaptive design, there are concerns about a *random high* in the observed treatment effect as a result of interim analyses and adapted design. In practice, there is a potential *random low* associated with futility analysis, which would undesirable for clinical development. The biased nature of a naive estimate following adaptive design is well known. Emerson and Fleming (1990) proposed a method for parameter estimation adjusting for interim analyses for group sequential design. However, their estimator requires a specific sampling plan that cannot always be followed in practice (see Liu and Anderson, 2008, suppl.) Liu and Chi (2001) provided median unbiased estimates for a two-stage adaptive design. For a general two stage adaptive design, Liu et al. (2002) proposed unbiased weighted estimates. The median unbiased estimates for two-stage adaptive designs are generalized to sequential median unbiased estimates for (adaptive) group sequential designs by Liu and Anderson (2008) following the repeated confidence interval approach of Jennison and Turnbull (1999). With sample size adjustments, these median unbiased estimates do not use sufficient statistics, which, along with combination tests of Cui et al. (1999), is at the center of controversy with many existing adaptive design procedures today. In a single-arm setting with maximum selection design, Shen (2001) identified bias from naive estimate and Luo et al. (2010) provided estimation based on conditional moments. In a comparative setting, Sampson and Sill (2005) proposed confidence interval estimation via reversing hypothesis testing. Most of those methods are unique to specific underlined design and/or outcome distribution.

The conduct of any clinical trial follows a series of events of subject enrollment, treatment assignment, and outcome evaluation, which fit well with a marked point process (MPP) framework. Many clinical trial design features can be naturally seen as part of the MPP specification. In particular, the MPP model provides a flexible framework for implementing different types of adaptive designs, such as sample size reestimation and treatment selection. There has been substantial research on the statistical inference for an MPP, including Cox and Lewis (1972), Arjas and Haara (1984), Arjas (1989), Feng (1999), and Martinussen and Scheike (2006). In this paper, we formulate a MPP framework for clinical trials and derive a general approach for treatment effect estimation following an adaptive clinical trial design

based on the martingale structure of the MPP. We illustrate the method via a two stage adaptive selection design that complements the work by Sampson and Sill (2005, 2009), Stallard and Friede (2008), and Li et al. (2009). It can be viewed as an extension of Luo et al. (2010), which includes both clinical trial and numerical examples but based on binomial distribution only. The outline of the paper is as follows: In section 2, we describe a clinical trial design with adaptive treatment selection and an estimating procedure for the treatment effect following the design. In section 3, we use simulation analyses to evaluate the performance of the proposed procedure under different outcome distributions. In section 4, we provide some discussions. For ease of reading, we leave most of technical details in the appendix.

2. CLINICAL TRIALS WITH ADAPTIVE TREATMENT SELECTION

In contrast with the traditional random variable-based setup for clinical trials, we adopt a stochastic process framework for three reasons. First, the conduct of any clinical trial is sequential and dynamic in nature. A stochastic process framework can accurately and easily carry out the design of a clinical trial protocol, including its amendments that often occur. Second, using a stochastic framework is for subsequent technical convenience, since it will allow more analytical tools such as stochastic calculus to be utilized easily. Third, the traditional random variable-based setup can be easily seen as a special case, while the opposite is not true in many adaptive design situations. However, it would be helpful to compare the corresponding forms we use here with conventional terms such as the sum of individual outcome variables or the sample mean.

Suppose we have a two-stage clinical trial to compare K experimental treatment groups $l = 1, 2, \dots, K$ to a control group $l = 0$. The trial will be conducted in two stages. Subjects are sequentially enrolled and their enrollment times are denoted as $0 < u_1 < u_2 < u_3 < \dots$ in calendar time. For simplicity we assume no two subjects are enrolled at the same time. We assume that the treatment group is randomly assigned according to the assignment probability $r_s(l)$, where s is the subject's arrival time and l is the assigned treatment group. We use X_s to denote the randomized treatment for the subject arrived at s . Thus, $P(X_s = l) = r_s(l)$ for any $l = 0, 1, 2, \dots, K$. As indicated later, the probability $r_s(l)$ may not be fixed but randomly decided based on data up to time s . Denote Y_s as the treatment outcome of the subject enrolled at time s . We assume that Y_s follows a distribution with an unknown density function, $f(l, y)$, conditional on $X_s = l$, and they are independent from each other among different s . For any treatment group l , $0 \leq l \leq K$, denote m_l and v_l as the mean and variance of Y_s given $X_s = l$, respectively, that is,

$$m_l = \int y f(l, y) dy$$

$$v_l = \int (y - m_l)^2 f(l, y) dy \quad (1)$$

Let the process

$$R_t = \sum_{i \geq 1} 1(u_i \leq t), \quad t > 0 \quad (2)$$

count the number of subjects enrolled up to t . Suppose the first stage of the trial is based on the data from the first n_1 subjects and the second stage is planned after the total N subjects are enrolled and evaluated. We can write τ_1 and τ_2 as the calendar times at which the planned interim and final analysis will be performed so that

$$\tau_1 = \inf\{t > 0: R_t \geq n_1\};$$

$$\tau_2 = \inf\{t > 0: R_t \geq N\}.$$

In general, we can fix n_1 prior to the trial start. As explained in section 2.3, the total sample size N may not be fixed but adaptively determined based on information available at the time τ_1 , including the observed treatment effect. It can be noted that while the basic model framework and its relevant results can be applied to all $K \geq 1$, we would consider the case $K > 1$ due to the nature of treatment selection in this paper.

2.1. Interim Analysis and Adaptation

At the time τ_1 , an interim analysis is performed. We calculate the number of subjects enrolled in the treatment group l by

$$n(t, l) = \int_0^t 1(X_s = l) dR_s \quad (3)$$

and the sum of the Y from the treatment group l as

$$Y(t, l) = \int_0^t 1(X_s = l) Y_s dR_s \quad (4)$$

for $t = \tau_1$. Thus,

$$D_l \equiv \frac{Y(\tau_1, l)}{n(\tau_1, l)}, \quad l = 0, 1, 2, \dots, K \quad (5)$$

is the sample mean of all Y_s , $s < \tau_1$ for the treatment group l . Then, for all treatments $l = 1, 2, \dots, K$, we rank the means and denote their order statistics as $D_{(1)} \leq D_{(2)} \leq \dots \leq D_{(K)}$ with

$G_{(l)}$ as the corresponding anti-rank, that is, $D_{G_{(l)}} = D_{(l)}$. Thus, $G_{(1)}$ and $G_{(2)}$ are the most and second most promising treatment group, respectively. When there are ties, we assign the treatment group with a lower index to the higher rank. Thus, $G_{(1)}$ represents the most promising group and $D_{(1)}$ is the corresponding mean. Suppose ρ is a prespecified futility threshold; that is, the trial will be terminated if no treatment is better than the control group by the margin ρ , that is, $D_{(1)} - D_{(0)} \leq \rho$. Otherwise, the subject enrollment would continue if $D_{(1)} - D_{(0)} > \rho$ and the randomization will be between the control group and the treatment group $G_{(1)}$. Thus, we have the intensity for the enrollment process

$$\lambda_s = \begin{cases} \lambda_1(s) & 0 < s < \tau_1 \\ \lambda_2(s) & \tau_1 \leq s < e, D_{(1)} - D_{(0)} > \rho \\ 0 & \tau_1 \leq s < e, D_{(1)} - D_{(0)} \leq \rho \end{cases}$$

where $\lambda_1(s)$ is a deterministic function representing enrollment rate prior to the interim analysis and $\lambda_2(s)$ is the enrollment rate after the interim analysis that may be random but determined by information up to τ_1 , e represents the end of enrollment time, and the treatment assignment probability is

$$r_s(l) = \begin{cases} \frac{1}{K+1} & \text{for any } l, 0 < s < \tau_1 \\ \frac{1}{2} & l = 0, G_{(1)}, \tau_1 \leq s < e \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

2.2. Approximation of the Treatment Effect $m_{G_{(1)}} - m_0$

At the final analysis time τ_2 , we need to estimate the treatment effect. In contrast with traditional framework, the treatment effect $m_{G_{(1)}} - m_0$ is not a fixed parameter since the selected group $G_{(1)}$ is random in nature. In the the current adaptive design framework, the key is to approximate $m_{G_{(1)}}$ and we avoid the usual term of parameter estimation and consider treatment effect approximation.

Without considering how the data is generated, it is tempting to approximate $m_{G_{(1)}}$ as the naive mean from the post hoc selected treatment group. This naive estimator and its variance can be written as

$$\hat{m}_{\text{naive}} = \frac{\int_0^{\tau_2} 1(X_s = G_{(1)}) Y_s dR_s}{N_1}$$

$$\hat{v}_{\text{naive}} = \frac{\int_0^{\tau_2} 1(X_s = G_{(1)})(Y_s - \hat{m}_{\text{naive}})^2 dR_s}{(N_1 - 1)N_1} \quad (7)$$

where

$$N_1 = \frac{n_1}{K+1} + \frac{N-n_1}{2} 1(D_{(1)} - D_0 > \rho) \quad (8)$$

In this paper, we propose to approximate $m_{G_{(1)}}$ based on a general estimating equation from the first conditional moment

$$E\left(Y(\tau_2, G_{(1)}) \middle| \mathcal{F}_{\tau_1}, G_{(1)}, D_{G_{(2)}}\right) = Y(\tau_2, G_{(1)})$$

where $\mathcal{F}_{\tau_1}$ represents the information available at the time τ_1 , and $Y(\tau_2, G_{(1)})$, from equation (4), is the final sum of all Y_s from the selected group (see Luo et al., 2010 for motivation). The utility of this general estimating equation is easy to see once we decompose the final sum from the selected group, $Y(\tau_2, G_{(1)})$, into components that are prior and post the interim analysis at τ_1 , that is,

$$Y(\tau_2, G_{(1)}) = Y(\tau_1, G_{(1)}) + \int_{\tau_1}^{\tau_2} 1(X_s = G_{(1)}) Y_s dR_s \quad (9)$$

The first component involves with a maximization process, namely;

$$Y(\tau_1, G_{(1)}) = n(\tau_1, G_{(1)}) D_{G_{(1)}} = n(\tau_1, G_{(1)}) \max \{D_l, 1 \leq l \leq K\}$$

and the second component may depend on the data from the first component. From the appendix, the first component can be approximated by

$$E\left(Y(\tau_1, G_{(1)}) \middle| \mathcal{F}_{\tau_1}, G_{(1)}, D_{G_{(2)}}\right) \approx \frac{n_1}{K+1} \left[m_{G_{(1)}} + \sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}} K \left(\frac{D_{G_{(2)}} - m_{G_{(1)}}}{\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}} \right) \right] \quad (10)$$

where

$$\kappa(u) = \frac{\phi(u)}{1 - \Phi(u)} \quad (11)$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution function of the standard normal distribution. The second component is

$$E\left(\int_{\tau_1}^{\tau_2} 1(X_s = G_{(1)})Y_s dR_s \middle| \mathcal{F}_{\tau_1-}\right) = m_{G_{(1)}}[N - n_1]1(D_{(1)} - D_0 > \rho)/2 \quad (12)$$

Thus, the general estimating equation becomes

$$\begin{aligned} & \frac{n_1}{K+1} \left[m_{G_{(1)}} + \sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}} \kappa\left(\frac{D_{G_{(2)}} - m_{G_{(1)}}}{\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}}\right) \right] + m_{G_{(1)}}[N - n_1]1(D_{(1)} - D_0 > \rho)/2 \quad (13) \\ & = Y(\tau_2, G_{(1)}) \end{aligned}$$

Letting $\nu = \frac{D_{(2)} - m}{\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}}$ or $m = D_{(2)} - \sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}\nu$, it is equivalent to solve the following equation of ν :

$$\frac{n_1}{K+1} \sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}} \kappa(\nu) - \sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}} N_1 \nu = Y(\tau_2, G_{(1)}) - N_1 D_{(2)}$$

Now, dividing $\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}$ on both sides leads to the specific estimating equation

$$g(\nu) = \tilde{D} \quad (14)$$

where

$$g(\nu) = \frac{n_1}{K+1} \kappa(\nu) - N_1 \nu$$

$$\tilde{D} = \frac{Y(\tau_2, G_{(1)}) - N_1 D_{(2)}}{\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}}$$

Let $\hat{v}$ be solution of the equation (14), which exists as explained in the appendix. We approximate $m_{G_{(1)}}$ by

$$\hat{m} = D_{(2)} - \sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}} \hat{v} \quad (15)$$

with its asymptotic variance approximated by

$$\hat{v} = \frac{v_{G_{(1)}}}{\hat{N}_1} \quad (16)$$

where

$$\hat{N}_1 = \frac{n_1}{K+1} \left[1 - \delta \left(\frac{D_{(2)} - \hat{m}}{\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}} \right) \right] + \frac{N - n_1}{2} 1(D_{(1)} - D_0 > \rho)$$

$$\delta(u) = \kappa(u)(\kappa(u) - u) \quad (17)$$

We can see $\hat{N}_1$ is close to N_1 but with the first stage sample size per group $\left(\frac{n_1}{K+1}\right)$ deflated

by a factor $1 - \delta \left(\frac{D_{(2)} - \hat{m}}{\sqrt{\frac{(K+1)v_{G_{(1)}}}{n_1}}} \right)$, which reflects the variation from the treatment group selection

process. In practice, the variances $v_{G_{(1)}} \quad I \quad K$, may be unknown. The preceding procedure can be applied by having a prior estimate for the variance. One simple approach may be to replace $v_{G_{(1)}}$ by a pooled naive sample estimate,

$$\hat{v}_{G(1)} = \frac{\int_0^{\tau_2} 1(X_s = 0, G_{(1)}) (Y_s - \bar{m})^2 dR_s}{2N_1 - 1}$$

$$\bar{m} = \frac{\int_0^{\tau_2} 1(X_s = 0, G_{(1)}) Y_s dR_s}{2N_1} \quad (18)$$

In addition, the control group is not subject to selection and its mean and variance of the mean can be approximated by

$$\hat{m}_0 = \frac{\int_0^{\tau_2} 1(X_s = 0) Y_s dR_s}{N_1}$$

$$\hat{v}_0 = \frac{\int_0^{\tau_2} 1(X_s = 0) (Y_s - \hat{m}_0)^2 dR_s}{(N_1 - 1)N_1} \quad (19)$$

Finally, the bias and variance of the bias by the proposed method and by the naive method can be written as

$$\hat{m}_b = (\hat{m} - \hat{m}_0) - \left(m_{G(1)} - m_0 \right)$$

$$\hat{v}_b = \frac{\hat{v}_{G(1)}}{\hat{N}_1} + \hat{v}_0$$

$$\hat{m}_{b,naive} = (\hat{m}_{naive} - \hat{m}_0) - \left(m_{G(1)} - m_0 \right)$$

$$\hat{v}_{b,naive} = \hat{v}_{naive} + \hat{v}_0 \quad (20)$$

2.3. Sample Size Reestimation

Suppose the study continues following the planned interim analysis, that is, $D_{(1)} - D_0 > \rho$. However, the observed difference $D_{(1)} - D_0$ may not look large enough and there is a chance that the true difference may be buried in the sample variance. In this case, it is possible to increase the planned sample size in the second stage to increase the accuracy of the estimation. From equations (16) and (20), it may be natural to choose N so that

$$\frac{D_{(1)} - D_0}{\sqrt{2 \frac{\hat{v}_{G(1)}}{\hat{N}_1}}} > c$$

for some constant c . For example, we can choose $c = 1.96$. Thus, from equation (17), we can choose

$$N = n_1 + \frac{4c^2 \hat{v}_{G(1)}}{(D_{(1)} - D_0)^2} - 2 \frac{n_1}{K+1} \left[1 - \delta \left(\frac{D_{(2)} - D_{(1)}}{\sqrt{\frac{(K+1) \hat{v}_{G(1)}}{n_1}}} \right) \right]$$

However, we do not discuss this further due to the scope of current paper.

3. SIMULATIONS

In this section, we mimic some typical clinical trial setup and use simulation analyses to evaluate the performance of the procedure proposed in section 2 and compare it with the naive procedure. Suppose we have a two-stage clinical trial to compare two experimental treatment groups $I = 1, 2$ to a control group $I = 0$. We assume that the outcome variable has probability density functions $f_I(\cdot)$ for the treatment group $I = 0, 1, 2$. These distributions are generally known but will be specified later for the simulation purpose. However, in contrast with previous work (e.g., Luo et al., 2010; Sampson and Sill, 2005, 2009), the distribution will not be limited to binomial or normal distributions. Let $n_1 > 0$ be a constant representing the sample size in the first stage. Let $\rho > 0$ be the threshold for futility. In the first stage, we randomly generate response variables $Y_{1I}^j, j = 1, 2, \dots, n_{1I}$ from the corresponding distribution $f_I(\cdot), I = 0, 1, 2$ for $n_{1I} = n_1/3$. Then we calculate the sample mean D_I for each group and determine the order statistics $D_{(1)}, D_{(2)}$ and anti-ranks $G_{(1)}, G_{(2)}$ as done earlier. If $D_{(1)} - D_0 > \rho$, we determine N and randomly generate responses $Y_{2I}^j, j = 1, 2, \dots, n_{2I}$ for $I = 0, G_{(1)}$ from the corresponding distribution, where $n_{2I} = (N - n_1)/2$. Otherwise, if $D_{(1)} - D_0 \leq \rho$, the data generation process stops.

For the selected treatment group, we construct two estimates $\hat{m}_{\text{naive}}$ and $\hat{m}$ from equations (7) and (15), respectively. In addition, bias from both estimators is calculated from equation (20).

This procedure is repeated for $B = 10,000$ times. Then we use two measures to evaluate the bias in $\hat{m}_{\text{naive}} - \hat{m}_0$ and $\hat{m} - \hat{m}_0$ for the true difference $m_{G(1)} - m_0$. One is the overall bias calculated as the mean of those 10,000 simulated differences, that is,

$$\text{Bias}(\hat{m} - \hat{m}_0) = \frac{1}{B} \sum_{k=1}^B \left[(\hat{m}^{(k)} - \hat{m}_0) - (m_{G(1)}^{(k)} - m_0) \right]$$

which approximates $E \left[(\hat{m} - \hat{m}_0) - (m_{G(1)} - m_0) \right]$, where k is used to index each simulation.

The other is the root mean squared error (RMSE) of those 10,000 simulated differences, that is,

$$\text{RMSE}(\hat{m} - \hat{m}_0) = \sqrt{\frac{1}{B} \sum_{k=1}^B \left[(\hat{m}^{(k)} - \hat{m}_0) - (m_{G(1)}^{(k)} - m_0) \right]^2}$$

which approximates $\sqrt{E \left[(\hat{m} - \hat{m}_0) - (m_{G(1)} - m_0) \right]^2}$ that incorporates both bias and variance.

In practice, $m_{G(1)}$ and m_0 are unknown since $m_{j, 0} \quad I = K$, are unknown. Thus, we need to approximate the mean square error $E \left[(\hat{m} - \hat{m}_0) - (m_{G(1)} - m_0) \right]^2$ through some estimator of the variance. Denote $MSE(\hat{m} - \hat{m}_0) = (\text{RMSE}(\hat{m} - \hat{m}_0))^2$, which will be estimated from equations (16), (19), and (20). We report the bias and RMSE of $\hat{v} + \hat{v}_0$ via the following:

$$\text{Bias}(\hat{v} + \hat{v}_0) = \sum_{k=1}^B (\hat{v}^{(k)} + \hat{v}_0) / B - MSE(\hat{m} - \hat{m}_0)$$

$$\text{RMSE}(\hat{v} + \hat{v}_0) = \sqrt{\sum_{k=1}^B [(\hat{v}^{(k)} + \hat{v}_0) - MSE(\hat{m} - \hat{m}_0)]^2 / B} \quad (21)$$

3.1. Normal Distribution

In this section, we consider that $f(\cdot)$ follows a normal distribution with $f(x; m_l, \sigma_l^2) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left\{-\frac{(x - m_l)^2}{2\sigma_l^2}\right\}$, where m_l and σ_l^2 are its mean and variance. For the simulations, we let $\sigma_l^2 = 1$, $l = 0, 1, 2$, $m_0 = 1$, and m_1 and m_2 vary as indicated in Table 1. On the design part, we let $\rho = 0.1$ and $n_1 = 90$ and $N = 230$.

Table 1 shows the results of the simulations (see the appendix for the approximation of $g(\nu)$). The last three rows represent cases when both experimental groups differ from the

control group. The results suggest the new procedure has smaller bias compared with the naive estimator. The first row represents when there is no mean difference in all three groups. The results suggest that the new procedure may not be as accurate as the naive estimator. This appears consistent with the MSE estimation where the naive estimator results smaller MSE, likely due to ignoring the variation from the selection process. It should be pointed out that, in practice, we would only have concerns about any estimate when there appears to be a meaningful difference between the selected experimental group and the control group. When treatment groups are not considered different, there is little concern about whether an estimation is exaggerated. Thus, the new procedure here would help to reduce concerns about any random high in the estimation with a selection process.

3.2. Log-Normal Distribution

In this section, we consider that $f(\cdot)$ follows a log-normal distribution with

$$f(x; m_l, \sigma_l) = \frac{1}{\sqrt{2\pi x} \tilde{\sigma}_l} \exp\left\{-\left(\ln x - \tilde{m}_l\right)^2 / 2\tilde{\sigma}_l^2\right\}, \quad x > 0, \text{ where}$$

$$\tilde{m}_l = \ln \frac{m_l}{\sqrt{1 + \left(\frac{\sigma_l}{m_l}\right)^2}}$$

$$\tilde{\sigma}_l = \sqrt{\ln\left(1 + \left(\frac{\sigma_l}{m_l}\right)^2\right)}$$

and m_l and σ_l^2 are the corresponding mean and variance of the log-normal distribution. For the simulations, we let $\sigma_l^2 = 1$, $l = 0, 1, 2$, $m_0 = 1$, and m_1 , and m_2 vary as indicated in Table 2. The $\tilde{m}_l$ and $\tilde{\sigma}_l$ are then calculated for log-normal specification. On the design part, we let $\rho = 0.1$ and $n_1 = 90$ and $N = 230$.

Table 2 shows that the results of the simulations and discussions for the normal case would generally apply for the log-normal case as well.

3.3. Binomial Distribution

In this section, we consider that $f(\cdot)$ follows a binomial distribution with

$$f(x; p_l) = x^{p_l} (1-x)^{1-p_l}, \quad x = 0, 1. \text{ Here, the mean and variance are } p_l \text{ and } p_l(1-p_l). \text{ For the simulations, we let } p_0 = 0.1 \text{ and } p_1 \text{ and } p_2 \text{ vary as indicated in Table 3. On the design part, we let } \rho = 0.05 \text{ and } n_1 = 150 \text{ } N = 250.$$

In case of a binomial distribution, it is possible to have $D_{(1)} = D_{(2)}$ and $D_{(1)} - D_0 < \rho$ so that there is no solution for equation (14). In these situations, we use a tie-breaking rule by

adding a small number to $D_{(1)}$. Here, we use $D_{(1)}^*$ such that $D_{(1)}^* = D_{(1)} + \frac{0.0001}{n_1}$ when $D_{(1)} = D_{(2)}$ and $D_{(1)}^* = D_{(1)}$ when $D_{(1)} > D_{(2)}$.

Table 3 shows the results of the simulations. The last four rows represent cases when both experimental groups differ from the control group. The results suggest that the new procedure has smaller bias compared with the naive estimator. The first row represents when there is no difference in all three groups. The results suggest that the new procedure may not be as accurate as the naive estimator. This appears consistent with the MSE estimation where the naive estimator results in smaller MSE, likely due to ignoring the variation from the selection process. As mentioned for the normal case, in practice, we would only have concerns about any estimate when there appears to be a meaningful difference between the selected experimental group and the control group. When treatment groups are not considered different, there is little concern about whether an estimation is exaggerated. Thus, the new procedure here would help to reduce concerns about a random high in estimation with a selection process.

4. DISCUSSIONS

Most people are familiar with the example of the potato growth experiment that Fisher used to explain the statistics theory he developed (Fisher, 1925). In that experiment, all subjects (potato seeds) entered the experiment at the same time and independently from any prior knowledge. Most of our statistical methods for analyzing clinical trials today assume the same *static* subject entry mechanism, which is mathematically simple but unfortunately problematic for clinical trials, given the sequential nature of patient enrollment and especially the adaptive design when applied. The adaptive treatment selection design discussed in this paper illustrates how the mechanism of subject entry may be changed following the interim analysis, and this process variation clearly changes the distribution of any naive statistics that may be used for statistical inference. Part of the objectives of this paper is to introduce a dynamic data framework, that is, via a marked point process, to analyze clinical trial data, which are often generated dynamically.

The theory of the marked point process has been developed for a long time but its application in clinical trials is not common, except in its simple case of a counting process. In the case of adaptive clinical trial design, any mechanism of adaptation would be a predictable process since it is based on information available up to the time of its interim analysis. Thus, we can say that any adaptive design would fit into a MPP and thus most of the tools available for MPP can be used for analyzing data from the adaptive design. This paper provides an example of the kind of applications of MPP framework in which we would dynamically select promising treatments for confirmation in order to optimize the sampling resource. Here, we focus on treatment effect approximation. However, the issue of related asymptotic distribution may be nonstandard in case there are more than one treatment group ($K > 1$), as described by Luo et al. (2010) for some special case. We will explore this further in the future and also examine the hypothesis testing problem in case of $K = 1$. In addition, this framework can be similarly applied to dynamic randomization (e.g., Pocock and Simon, 1975), response adaptive design (Zelen, 1969) and more general

covariate adjustment and response adaptive designs (see Hu and Rosenberger, 2003; Zhang et al., 2007). We will explore those expansions in the future.

Acknowledgments

We thank an editor for valuable comments and suggestions that greatly improved the presentation.

APPENDIX

Justification of Estimating Equation

In this section, we provide the rationale of the proposed estimation procedure described in section 2. In order to analyze the terms of equation (9), we first introduce a general martingale framework for the clinical trial data flow as described in section 2.

A Marked Point Process (MPP) Framework of Clinical Trial

Continuing the notations in section 2 and using the terminology of marked point process, we say we would mark the event of a subject enrollment at time s by the subject's randomized treatment group X_s and the outcome Y_s . Thus, the mark space can be written as $\mathcal{X} = \{0, 1, 2, \dots, K\} \times R$ and we associate it with a natural σ -algebra $\mathcal{A}$. Suppose $t > 0$ is fixed. For any $l = 0, 1, 2, \dots, K$, $C \subset R$, let

$$p_t(\{l\} \times C) = \int_0^t 1(X_s = l, Y_s \in C) dR_s$$

Count the number of subjects enrolled by time t and marked by treatment l with treatment outcome in C . Thus, for any given t , p_t defines a random measure on the mark space $\mathcal{X} = \{0, 1, 2, \dots, K\} \times R$. When fixing $\{l\}$ and $C \subset R$, $p_t(\{l\} \times C)$ is a nondecreasing function in t and thus defines a Lebesgue Stieltjes measure on $[0, \mathcal{J}]$. Combining these two measures together, we can define a random measure, $p(\cdot)$, on the time and mark space $[0, \mathcal{J}] \times \mathcal{X}$ and it can be described as

$$p(ds dx dy) = 1(X_s = x, Y_s \in dy) dR_s \quad (22)$$

for any s, ds, x, dx, y, dy , where ds and dy are usual Lebesgue measures and dx is the counting measure.

For any calendar time $t \geq 0$, all study information of enrollments, treatments, and outcomes up to the time t can be described by the so-called internal filtration

$$\mathcal{F}_t = \sigma(R_u, p(A \times \{l\} \times C): \forall u < t, A \times \{l\} \times C \subset [0, t] \times \{0, 1, 2, \dots, K\} \times \mathcal{R})$$

It can be noted that the enrollment process $\{R_t, t \geq 0\}$ is $\mathcal{F}_t$ predictable, that is, R_t is $\mathcal{F}_{t-}$ measurable. From Chapter II Theorem 1.8 of Jacod and Shiryaev (2003), the random

measure $p(\cdot)$ has a compensator $q(\cdot)$ such that $q(\cdot)$ is $\mathcal{F}_t$ predictable and, for any optional and integrable function g ,

$$\int_0^t \int_{\mathcal{X}} g(s, x, y, \omega) p(ds dx dy) - \int_0^t \int_{\mathcal{X}} g(s, x, y, \omega) q(ds dx dy)$$

is an $\mathcal{F}_t$ martingale. In particular, we have from equation (22) the compensator

$$q(ds dx dy) = E(p(ds dx dy) | \mathcal{F}_{s-}) = E(1(X_s = x, Y_s \in dy) dR_s | \mathcal{F}_{s-}) = r_s(x) f(x, y) dx dy dR_s = q_s(dx dy) dR_s$$

where

$$q_s(dx dy) \equiv r_s(x) f(x, y) dx dy \quad (23)$$

For any fixed measurable $E \subset \mathcal{X}$, define

$$M_t(E) = p_t(E) - \int_0^t \int_E r_s(x) f(x, y) dx dy dR_s$$

Then, $M_t(E)$ is an $\mathcal{F}_t$ martingale. Thus, $M_t(\cdot)$ is a martingale measure for the mark space $\mathcal{X}$.

Following similar development for the stochastic integral with respect to martingales (see chapter 3 of Kallianpur and Xiong, 1995, or Chapter II of Jacod and Shiryaev, 2003), we can define a square integrable martingale

$$M_t^g = \int_0^t \int_{\mathcal{X}} g(s, x, y, \omega) dM_s = \int_0^t g(s, X_s, Y_s, \omega) dR_s - \int_0^t \left[\int_{\mathcal{X}} g(s, x, y, \omega) r_s(x) f(x, y) dx dy \right] dR_s$$

for any $\mathcal{F}_s$ predictable $g(s, x, y, \omega)$ with

$$E \int_0^T \int_{\mathcal{X}} |g(s, x, y, \omega)| p(ds dx dy) < \infty$$

$$E \int_0^T \int_{\mathcal{X}} g(s, x, y, \omega)^2 p(ds dx dy) < \infty$$

and M_t^g has the predictable variation process

$$\langle M^g \rangle_t = \int_0^t \left[\int_{\mathcal{X}} g(s, x, y, \omega)^2 r_s(x) f(x, y) dx dy \right] dR_s$$

Application of the MPP Framework

First, we can note that the explicit form of the compensator for the model in section 2 can be derived from equations (23) and (6), that is,

$$q(ds dx dy) = \left[\frac{1(s < \tau_1)}{K+1} + \frac{1(s \geq \tau_1, x \in \{0, G_{(1)}\})}{2} \right] f(x, y) dx dy dR_s \quad (24)$$

and the analysis time points τ_1 and τ_2 are $\mathcal{F}_s$ predictable stopping times.

Now, we can look at equations (10) and (12). Indeed, note that

$$\int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)})(y - m_{G_{(1)}}) q(ds dx dy) = \int_{\tau_1}^{\tau_2} \sum_x 1(x = G_{(1)}) \int (y - m_{G_{(1)}}) f(G_{(1)}, y) dy dR_s = 0$$

We can write

$$\begin{aligned} \int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)}) y p(ds dx dy) &= \int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)})(y - m_{G_{(1)}}) dM_s \\ &+ \int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)}) m_{G_{(1)}} p(ds dx dy) \end{aligned}$$

Thus,

$$\begin{aligned} E \left(\int_{\tau_1}^{\tau_2} 1(X_s = G_{(1)}) Y_s dR_s \middle| \mathcal{F}_{\tau_1} \right) &= E \left(\int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)}) y p(ds dx dy) \middle| \mathcal{F}_{\tau_1} \right) \\ &= m_{G_{(1)}} [n(\tau_2, G_{(1)}) - n(\tau_1, G_{(1)})] = m_{G_{(1)}} [N - n_1] 1(D_{(1)} - D_0 > \rho) \end{aligned}$$

which is equation (12).

For equation (10), denote

$$W_l = \frac{\int_0^{\tau_1} \int_{\mathcal{X}} 1(x = l)(y - m_l) dM_s}{\sqrt{\frac{n_1 v_1}{K+1}}}$$

It is easy to see that W_l is asymptotically $n(0, 1)$ for large n_1 from the martingale central limit theorem. We can write

$$\begin{aligned}
D_l &= \frac{Y(\tau_1-, l)}{n(\tau_1-, l)} = \frac{\int_0^{\tau_1-} \int_{\mathcal{X}} 1(x=l)(y-m_l)p(ds dx dy) + m_l n(\tau_1-, l)}{n(\tau_1-, l)} \\
&= \frac{\int_0^{\tau_1-} \int_{\mathcal{X}} 1(x=l)(y-m_l)dM_s + m_l n(\tau_1-, l)}{n(\tau_1-, l)} = \frac{W_l + m_l n(\tau_1-, l) \sqrt{\frac{n_1 v_l}{K+1}}}{n(\tau_1-, l) \sqrt{\frac{n_1 v_l}{K+1}}} \approx \frac{W_l + m_l \sqrt{\frac{n_1}{K+1}} \sqrt{v_l}}{\sqrt{\frac{n_1}{K+1}} \sqrt{v_l}} \\
&= \frac{W_l \sqrt{v_l} + m_l \sqrt{\frac{n_1}{K+1}}}{\sqrt{\frac{n_1}{K+1}}}
\end{aligned}$$

Thus, we can view D_l as asymptotically normal with mean m_l and variance $\frac{(K+1)v_l}{n_1}$, that is,

$$n\left(m_l, \frac{(K+1)v_l}{n_1}\right).$$

For any j_1 and j_2 of $1, \dots, K$, denote

$$Q \equiv Q_{j_1, j_2} = \{G_{(1)} = j_1, G_{(2)} = j_2\}$$

and $\mathcal{E}_{\tau_1} = \sigma(\tau_1, R_u, u < \tau_1)$ is the collection of all enrollment information up to τ_1 . Note that

$$\begin{aligned}
P(D_{j_1} \in dx, D_{j_2} \in dz, Q | \mathcal{E}_{\tau_1}) &= P(D_{j_1} \in dx, D_{j_2} \in dz; D_{j_2} > D_{j_1}, j \neq j_1, j_2; x > z | \mathcal{E}_{\tau_1}) \quad (25) \\
&= P(D_{j_1} \in dx | \mathcal{E}_{\tau_1}) 1(x > z) P(D_{j_2} \in dz; D_{j_2} > D_{j_1}, j \neq j_1, j_2 | \mathcal{E}_{\tau_1})
\end{aligned}$$

Integrating over x , we have

$$P(D_{j_2} \in dz; Q | \mathcal{E}_{\tau_1}) = P(D_{j_1} > z | \mathcal{E}_{\tau_1}) P(D_{j_2} \in dz; D_{j_2} > D_{j_1}, j \neq j_1, j_2 | \mathcal{E}_{\tau_1}) \quad (26)$$

Dividing equation (25) by equation (26), we have

$$P(D_{j_1} \in dx | D_{j_2} = z, Q, \mathcal{E}_{\tau_1}) = \frac{P(D_{j_1} \in dx | \mathcal{E}_{\tau_1})}{P(D_{j_1} > z | \mathcal{E}_{\tau_1})} 1(x > z)$$

Since D_{j_1} is asymptotically normal with mean m_{j_1} and variance $\frac{(K+1)v_{j_1}}{n_1}$, the conditional distribution of D_{j_1} given $D_{j_2} = z, Q, \mathcal{E}_{\tau_1}$ is asymptotically a truncated normal distribution.

Thus,

$$E\left(D_{j_1} \mid D_{j_2} = z, Q, \mathcal{E}_{\tau_1}\right) \approx m_{j_1} + \sqrt{\frac{(K+1)v_{j_1}}{n_1}} \kappa\left(\frac{z - m_{j_1}}{\sqrt{\frac{(K+1)v_{j_1}}{n_1}}}\right)$$

$$\text{Var}\left(D_{j_1} \mid D_{j_2} = z, Q, \mathcal{E}_{\tau_1}\right) \approx \frac{(K+1)v_{j_1}}{n_1} \left(1 - \delta\left(\frac{z - m_{j_1}}{\sqrt{\frac{(K+1)v_{j_1}}{n_1}}}\right)\right) \quad (27)$$

(see Johnson and Kotz, 1970) where $\kappa(\cdot)$ and $\delta(\cdot)$ are defined in equations (11) and (17), which leads to equation (10).

It should be pointed out that since

$$g'(\nu) = -\frac{n_1}{K+1}(1 - \delta(\nu)) - (N - n_1)1(D_{(1)} - D_0 > \rho)/2 < 0$$

$g(\cdot)$ is decreasing. Since $\lim_{\nu \rightarrow \infty} g(\nu) = -\infty$ and $\lim_{\nu \rightarrow -\infty} g(\nu) = +\infty$, equation (14) has a unique solution $\hat{\nu}$, which leads to the estimate $\hat{m} = D_{(2)} - \sqrt{\frac{(K+1)v_{G(1)}}{n_1}} \hat{\nu}$.

For a standard error estimate of $\hat{m}$, we can use the Delta method based on equation (13).

From equation (24),

$$E\left[\left(\int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)}) (y - m_{G_{(1)}}) dM_s \right)^2 \mid \mathcal{F}_{\tau_1} -\right] = E\left(\int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)}) (y - m_{G_{(1)}})^2 q(ds dx dy) \mid \mathcal{F}_{\tau_1} -\right)$$

$$= E\left(\frac{1}{2} E \int_{\tau_1}^{\tau_2} \int_{\mathcal{X}} 1(x = G_{(1)}) (y - m_{G_{(1)}})^2 f(x, y) dx dy v dR_s \mid \mathcal{F}_{\tau_1} -\right) = \frac{1}{2} v_{G_{(1)}} E\left([R_{\tau_2} - R_{\tau_1}] \mid \mathcal{F}_{\tau_1} -\right)$$

$$= \frac{1}{2} v_{G_{(1)}} (N - n_1) 1(D_{(1)} - D_0 > \rho)$$

From equation (27), we have

$$EY(\tau_{2,G(1)})^2 \approx \left(\frac{n}{K+1}\right)^2 \left[\frac{(K+1)^{v_{G(1)}}}{n_1} \left(1 - \delta \left(\frac{D_{(2)} - \hat{m}}{\sqrt{\frac{(K+1)^{v_{G(1)}}}{n_1}}} \right) \right) \right] + \frac{1}{2} v_{G(1)} (N - n_1) 1(D_{(1)} - D_0 > \rho)$$

$$= v_{G(1)} \hat{N}_1$$

Since the derivative of the left-hand side of equation (13) evaluated at $\hat{m}$ is $\hat{N}_1$, the asymptotic variance of $\hat{m}$ is

$$\hat{v} = \frac{v_{G(1)}}{\hat{N}}$$

which is equation (16).

Approximation of $g(\nu)$

We use an approximating form for $g(\nu)$ of equation (14) in the simulations of section 3. The numerical calculation is based on R language, where the built in functions have technical difficulty in evaluating the function $\kappa(\cdot)$ of equation (14) when the variable becomes large. In general, a better approximation would lead to better consistency for the new procedure. The approximation for the function $g(\nu)$ implemented in this paper is described next.

Denote $n_{1g} = \frac{n_1}{K+1}$ and $n_{2g} = (N - n_1) 1(D_{(1)} - D_0 > \rho)/2$. Then, $g(\nu) = n_{1g}[\kappa(\nu) - \nu] - n_{2g}\nu$. From equation (14), we have

$$-n_{1g}\nu - n_{2g}\nu \leq \tilde{D} \leq n_{1g}[\kappa(\nu) - \nu]$$

Thus, we have a lower bound $\nu \geq -\tilde{D}/(n_{1g} + n_{2g})$.

It is easy to see that

$$\lim_{\nu \rightarrow \infty} \nu[\kappa(\nu) - \nu] = \lim_{\nu \rightarrow \infty} \frac{\nu\phi(\nu) - \nu^2 + \nu^2\Phi(\nu)}{1 - \Phi(\nu)} = \lim_{\nu \rightarrow \infty} \frac{3 - \nu^2}{1 - \nu^2} = 1$$

Thus, there is a $\nu_M > 0$ such that $0.5 < \nu[\kappa(\nu) - \nu] < 1.5$ for all $\nu > \nu_M$ or

$$\frac{0.5}{\nu} < \kappa(\nu) - \nu < \frac{1.5}{\nu}$$

If $\tilde{D} \leq 0$, then $\tilde{D} < \frac{0.5}{\nu} < \kappa(\nu) - \nu$ for $\nu > \nu_M$. If $\tilde{D} \leq 0$, we have $\nu \leq \nu_M \vee 1.5n_{1g}/\tilde{D}$.

For any given $\theta > 0$, we can choose $K > 1.5$ so that $\kappa(\nu) - \nu \leq \frac{K}{\nu}$ for $\theta \leq \nu \leq \nu_M$. Thus,

$$-\tilde{D}/(n_{1g} + n_{2g}) \leq \nu \leq \max \left\{ \nu_M, K n_{1g}/|\tilde{D}| \right\}$$

Therefore, for numerical approximation of $g(\nu)$, we can use

$$g_{\theta}(\nu) = \begin{cases} g(\nu) & \nu < \theta; \\ (1 + ae^{-\nu^2})/\nu - \frac{1}{2}(N - n_1)1(D_{(1)} - D_0 > \rho)\nu & \nu \geq \theta \end{cases}$$

where θ is a constant (e.g., $\theta = 7$ as in our simulations) and a is chosen so that

$$\frac{n_1}{K+1}(\kappa(\theta) - \theta) = (1 + ae^{-\theta^2})/\theta, \text{ that is, } a = \frac{\left[\theta \frac{n_1}{K+1}(\kappa(\theta) - \theta) - 1 \right]}{e^{-\theta^2}}.$$

References

- Arjas E, Haara P. A marked point process approach to censored failure data with complicated covariates. *Scandinavian Journal of Statistics*. 1984; 11:193–209.
- Arjas E. Survival models and martingale dynamics (with discussion). *Scandinavian Journal of Statistics*. 1989; 16:177–225.
- Amitage P. Sequential methods in clinical trials. *American Journal of Public Health*. 1958; 48:1395–1402. [PubMed: 13583284]
- Bauer P, Köhne K. Evaluations of experiments with adaptive interim analyses. *Biometrics*. 1994; 50:1029–1041. [PubMed: 7786985]
- Box, GEP., Jenkins, GM. Some statistical aspects of adaptive optimization and control. In: Kotz, S., Johnson, NL., editors. *Breakthroughs in Statistics*. Vol. II. New York: Inc. Springer-Verlag; 1962. p. 370
- CHMP/EWP/2459/02. Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design. 2007 Oct 18.
- Chow, SC., Chang, M. *Adaptive Design Methods in Clinical Trials*. Boca Raton, FL: Chapman & Hall; 2007.
- Cox DR, Lewis PAW. Multivariate point processes. *Proc 6th Berkeley Symp Math Statist Prob*. 1972; 3:401–448.
- Cui L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999; 55:321–324. [PubMed: 11318176]
- Emerson SS, Fleming TR. Parameter estimation following group sequential hypothesis testing. *Biometrika*. 1990; 77(4):875–892.
- Feng J. Estimations in survival analysis – a stochastic filtering perspective. *Applied Probability Trust*. 1999 Oct 28.
- Fisher, RA. Statistical methods for research workers. In: Kotz, S., Johnson, NL., editors. *Breakthroughs in Statistics*. Vol. II. New York: Springer-Verlag; 1925. p. 66–81.
- Fisher LD. Self-designing clinical trials. *Statistics in Medicine*. 1998; 17:1551–1562. [PubMed: 9699229]
- Hu F, Rosenberger WF. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association*. 2003; 98(463):671–678.

- Jennison, C., Turnbull, BW. Turnbull Group Sequential Methods with Applications to Clinical Trials. Boca Raton, FL: Chapman & Hall/CRC; 1999.
- Johnson, NL., Kotz, S. Continuous Univariate Distributions-1. New York: John Wiley & Sons; 1970. chap. 13
- Li G, Zhu J, Ouyang SP, Xie J, Deng L, Law G. Adaptive design for interim dose selection. *Statistics in Biopharmaceutical Research*. 2009; 1(4):366–376.
- Liu, Q., Anderson, KM. Theory of inference for adaptively extended group sequential designs with applications in clinical trials. *Journal of the American Statistical Association, Supplemental Archive*. 2008. http://www.amstat.org/publications/jasa/supplemental_materials
- Liu Q, Chi GYH. On sample size and inference for two-stage adaptive designs. *Biometrics*. 2001; 57:172–177. [PubMed: 11252594]
- Liu Q, Proschan MA, Pledger GW. A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association*. 2002; 97:1034–1041.
- Luo X, Wu S, Xiong J. Parameter estimation following an adaptive treatment selection trial design. *Biometrical Journal*. 2010; doi: 10.1002/bimj.200900134
- Martinussen, T., Scheike, TH. *Dynamic Regression Models for Survival Data*. New York, NY: Springer Science + Business Media Inc.; 2006.
- Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*. 1975; 31:103–115. [PubMed: 1100130]
- Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics*. 1995; 51:1315–1324. [PubMed: 8589224]
- Sampson AR, Sill MW. Drop-the-losers design: normal case. *Biometrical Journal*. 2005; 47:257–268. [PubMed: 16053251]
- Sampson AR, Sill MW. Drop-the-losers design: binomial case. *Computational Statistics and Data Analysis*. 2009; 53:586–595. [PubMed: 20047003]
- Shen L. An improved method of evaluating drug effect in a multiple dose clinical trial. *Statistics in Medicine*. 2001; 20:1913–1929. [PubMed: 11427949]
- Stallard N, Friede T. A group-sequential design for clinical trials with treatment selection. *Statistics in Medicine*. 2008; 27:6209–6227. [PubMed: 18792085]
- Stein C. A two sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics*. 1945; 16:243–258.
- U.S. Food and Drug Administration. Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics. 2010 Feb. Available at: www.fda.gov/DrugsGuidance-ComplianceRegulatoryInformation/Guidances/UCM201790.pdf
- Wang SJ, O'Neill RT, Hung HMJ. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharmaceutical Statistics*. 2007; 6:227–244.
- Zelen M. Play the winner rule and the controlled clinical trial. *Journal of the American Statistics Association*. 1969; 64:131–146.
- Zhang L, Hu F, Cheung S, Chan W. Asymptotic properties of covariate-adjusted response-adaptive designs. *Annals of Statistics*. 2007; 35(3):1166–1182.

Table 1

Bias and root mean squared error (RMSE) of estimators, mean difference, and its MSE from simulations

m_1 m_2		Bias of mean difference (RMSE)*		Bias of MSE (RMSE)	
		Naive	New	Naive	New
1	1	0.001 (0.159)	-0.192 (0.625)	0.018 (0.031)	-0.267 (0.356)
1.5	1	-0.014 (0.18)	-0.031 (0.231)	-0.01 (0.015)	-0.024 (0.064)
2.0	1	-0.001 (0.145)	-0.001 (0.145)	-0.001 (0.002)	0.002 (0.003)
1.5	1.5	-0.026 (0.153)	-0.017 (0.203)	-0.003 (0.007)	-0.014 (0.061)
1.5	2.0	0.004 (0.144)	-0.003 (0.146)	-0.001 (0.002)	0.002 (0.007)
2.0	2.0	0.03 (0.141)	-0.008 (0.147)	0 (0.002)	0.003 (0.004)

* Normal distribution. $\mu = 0$, $\sigma = 1$, $\rho = 0.1$, $N = 90$, and $N = 230$. Bias and RMSE of estimators are defined in section 3. All values were obtained based on $B = 10,000$ simulations.

Table 2

Bias and root mean squared error (RMSE) of estimators, mean difference, and its MSE from simulations

m_1 m_2		Bias of mean difference (RMSE)*		Bias of MSE (RMSE)	
		Naive	New	Naive	New
1	1	-0.002 (0.168)	-0.214 (0.584)	0.02 (0.05)	-0.190 (0.414)
1.5	1	-0.016 (0.185)	-0.036 (0.250)	-0.009 (0.034)	-0.027 (0.118)
2.0	1	-0.002 (0.144)	-0.002 (0.145)	-0.0005 (0.015)	0.002 (0.022)
1.5	1.5	0.024 (0.159)	-0.03 (0.281)	-0.002 (0.029)	-0.037 (0.252)
1.5	2.0	0.005(0.144)	-0.003 (0.161)	-0.0003 (0.031)	0.001 (0.271)
2.0	2.0	0.030 (0.145)	-0.010 (0.187)	0 (0.034)	-0.005 (0.261)

* Log-normal distribution. $m_0 = 1$, $\rho = 0.1$, $m = 90$, and $N = 230$. Bias and RMSE of estimators are defined in section 3. All values were obtained based on $B = 10,000$ simulations.

Table 3

Bias and root mean squared error (RMSE) of estimators, mean difference, and its MSE from simulations

p_1	p_2	Bias of mean difference (RMSE)*		Bias of MSE (RMSE)	
		Naive	New	Naive	New
0.1	0.1	0.01(0.045)	-0.027(0.071)	0.001(0.002)	0.003(0.007)
0.25	0.1	-0.004(0.061)	-0.008(0.069)	-0.001(0.001)	-0.002(0.004)
0.25	0.2	0.01(0.056)	-0.007(0.071)	0(0.001)	-0.001(0.006)
0.25	0.25	0.016(0.055)	-0.003(0.067)	0(0.001)	-0.001(0.005)
0.3	0.25	0.014(0.056)	-0.002(0.064)	0(0)	-0.001(0.003)
0.3	0.3	0.018(0.056)	0(0.062)	0(0)	-0.001(0.002)

* Binomial distribution. $p_0 = 0.1$, $\rho = 0.05$, $n_1 = 150$, and $N = 250$. Bias and RMSE of estimators are defined in section 3. All values were obtained based on $B = 10,000$ simulations.