Published in final edited form as:

Stat Biopharm Res. 2019 ; 11(4): 387–397. doi:10.1080/19466315.2019.1629996.

A Bayesian Sequential Design for Clinical Trials with Time-to-Event Outcomes

Lin Zhu, Qingzhao Yu*, Donald E. Mercante

School of Public Health, Louisiana State University Health Sciences Center

Abstract

There is increasing interest in Bayesian group sequential design because of its potential to improve efficiency in clinical trials, to shorten drug development time, and to enhance statistical inference precision without undermining the clinical trial's integrity or validity. We propose a Bayesian sequential design for clinical trials with time-to-event outcomes and use alpha spending functions to control the overall type I error rate. Bayes factor is adapted for decision-making at interim analyses. Algorithms are presented to make decision rules and to calculate power of the proposed tests. Sensitivity analysis is implemented to evaluate the impact of different choices of prior parameters on choosing critical values. The power of tests, the expected event size of the proposed design, and the quality of estimators are studied through simulations, and compared with the frequentist group sequential design. Simulations show that at fixed total number of events, the proposed design can achieve greater power and require smaller expected event size when appropriate priors are chosen, compared with the frequentist group sequential design. The feasibility of the proposed design is further illustrated on a real data set.

Keywords

Alpha spending function; Multiple stage analysis; Survival data; Type I error rate

1 Introduction

Sequential designs, in which patients are recruited in pairs and the data can be analyzed as the results from each pair become available, were first proposed by Armitage (1975). Sequential designs were extended by Pocock (1977) and O'Brien and Fleming (1979) to group sequential designs, in which patients are enrolled in successive groups instead of pairs, have been utilized for decades by the statistical and clinical trial communities (U.S. Food and Drug Administration, 2016). These designs allow for multiple interim looks at the accumulated data as the clinical trial proceeds, with the possibility of stopping the trial early due to efficacy, futility, or safety reasons. Early termination for efficacy may occur when the superiority of a treatment under study is clearly established, for futility when the demonstration of a relevant treatment difference is not likely, or for safety concerns when

*Corresponding author, qyu@lsuhsc.edu.

Supplementary material

The supplementary material is available online and contains the following tables and R codes.

unacceptable adverse events become manifest (U.S. Food and Drug Administration, 1998). Group sequential design has the potential to improve the efficiency of a clinical trial by shortening its duration without lowering scientific and regulatory standards. These designs have been widely and successfully used in large, long-term trials, such as phase III trials of drug development for fatal diseases, particularly where the endpoints were progression-free survival time and/or overall survival time. The relatively long recruiting period and follow-up time for assessing the primary outcome allow for the implementation of multiple interim analyses and opportunities for increasing efficiency. Moreover, for trials seeking accelerated approval, group sequential designs may offer an advantage over conventional designs by shortening the time to market.

Group sequential designs can be implemented under both frequentist and Bayesian frameworks. Bayesian inference encourages learning from evidence as it accumulates (U.S. Food and Drug Administration, 2010b). It provides a natural way to incorporate prior information into an ongoing trial. With the accumulation of historical data from previous studies and the advance of computational techniques, Bayesian designs have gained increasing popularity. Bayesian designs have been used more frequently in early-phase trials to assess pharmacokinetics/pharmacodynamics or to find dose ranges (Lee and Chu, 2012; Biswas et al., 2009), however, their use has been rapidly expanding to other applications. Furthermore, the potential gain in efficiency using Bayesian sequential designs may offer advantages over traditional designs for studies involving rare diseases, where there are limited available numbers of patients (U.S. Food and Drug Administration, 2014).

A major statistical challenge in group sequential design is control of the study-wide overall type I error rate (U.S. Food and Drug Administration, 2010a). The type I error rate is defined as the probability of rejecting the null hypothesis when it is true (Casella and Berger, 2002). The overall type I error rate is inflated because of the multiple testing performed at each interim analysis of the study data. Statistical adjustment is needed to control the overall type I error rate. Key early work to control the overall type I error includes the methods developed by Pocock (1977) and O'Brien and Fleming (1979). Both methods require that a fixed number of interim analyses at fixed times with a prescribed fraction of type I error spent at each analysis be specified in advance. Alpha spending methods, proposed by Lan and DeMets (1983) and extended by Kim and DeMets (1987), offer a more flexible approach to controlling the overall type I error rate and have been widely used in frequentist group sequential designs (Bang et al., 2010, 2012; Pitt et al., 2001). Alpha spending allows for systematic allocation of type I error at various time points in the trial.

Controlling the type I error rate in Bayesian sequential design is more complicated since the choices of prior distributions may also influence final decisions (U.S. Food and Drug Administration, 2010b). Simulations are usually needed to help decision making and to assess the operating characteristics of the study design under different scenarios (U.S. Food and Drug Administration, 2010b, 2016). Zhu and Yu (2017) proposed a Bayesian sequential design using alpha spending functions to control the overall type I error in phase III clinical trials (BSDASF). The basic idea is to strictly control the overall type I error rate using alpha spending functions that allocate the type I error rate proportional to the amount of data used for interim analysis. Using the idea to control the overall type I error rate, Zhu et al. (2017)

considered BSDASF for binary outcome with the implementation of a new stopping rule for futility. Further, Yu et al. (2017) proposed a Bayesian sequential design that allows for adaptively changing randomization rates. The rates of allocating patients to different treatment arms can be changed at each interim analysis to reduce the required sample size for a given level of power in 2-sided hypothesis tests.

In this paper, we propose a Bayesian sequential design for time-to-event outcomes. Alpha spending functions are used to control the study-wide overall type I error rate. We put a prior distribution on the parameter of interest, log hazard ratio, and Bayes factor is adapted for decision-making. Note that a Bayes factor represents the change in odds of an alternative model over a null model going from prior to posterior distribution. Therefore, compared with posterior probability or posterior odds, the Bayes factor may be more robust to the choice of prior distribution (U.S. Food and Drug Administration, 2010b).

The remainder of this paper is organized as follows: In Section 2, we introduce the proposed Bayesian sequential design for survival outcomes, and present the algorithms to calculate critical values and powers of hypothesis tests. In Section 3, we discuss the effects of the choices of prior parameters on the decision-making process of the proposed design through sensitivity analyses. In Section 4, a series of simulations are conducted to compare (1) power, maximum sample size, and expected sample size between the proposed design and the frequentist group sequential design using the same alpha spending function; and (2) power and expected sample size of the clinical trials when different alpha spending functions are used with the proposed design. In Section 5, the proposed design is applied to a real data set. Conclusions and future research are given in Section 6.

2 A Bayesian sequential design for time-to-event outcomes

2.1 Settings

For a specific subject under study, we assume that there is a survival time T to a pre-specified event and a right censoring time C . The event time T and right censoring time C are independent. Assume the total number of patients enrolled in the study is n with n_0 in the control (established treatment) group and n_1 in the treatment (novel treatment) group. The data collected for the i th patient is represented by (X_{li}, δ_{li}) , where $X_{li} = \min(T_{li}, C_{li})$, with $l = 0$ denoting the control group and $l = 1$ for the treatment group, and $i = 1, 2, \dots, n$. δ_{li} is an indicator for censoring status, where $\delta_{li} = 1$ if $X_{li} = T_{li}$ and $\delta_{li} = 0$ if $X_{li} = C_{li}$. The primary objective is to test whether the novel treatment results in a better survival outcome compared to the established treatment. We further assume that the hazard rates of the treatment group and the control group are proportional over time. Denote θ as the log hazard ratio of the novel treatment relative to the established treatment. Under the proportional hazard assumption, θ is a constant independent of time. The one-sided hypotheses to be tested are,

$$H_0: \theta = 0 \text{ vs. } H_a: \theta < 0.$$

Suppose that events occur at D ordered times $\tilde{t}_1 \leq \tilde{t}_2 \leq \dots \leq \tilde{t}_D$. Ties between event times might be observed in data because of the way times are recorded. Without loss of generality,

we assume there is no tie between the event times. At time $\tilde{t}_k$, an event happens either in the treatment group or in the control group. We assume there are d_{1k} events observed in the treatment group, out of Y_{1k} subjects at risk in the treatment group at time $\tilde{t}_k$. By assuming no event happens at the same time, d_{1k} could take the values of 0 or 1. Correspondingly, Y_{0k} represents the number of subjects at risk in the control group. Note that $Y_k = Y_{1k} + Y_{0k}$ is the number of subjects such that $X_{li} \geq \tilde{t}_k$.

Assume that

$$d_{1k} \mid \theta, y_{1k}, y_{0k} \sim \text{Bernoulli}(p_{1k}), \quad k = 1, 2, \dots, D,$$

where p_{1k} is the probability of observing the event in the treatment group at t_k . Under the pre-stated assumptions, p_{1k} can be estimated by

$$\hat{p}_{1k} = \frac{\exp(\theta)y_{1k}}{y_{0k} + \exp(\theta)y_{1k}}.$$

We set up a prior distribution for θ , $p(\theta)$, to be a normal distribution with mean θ_0 and variance σ^2 / τ . σ^2 is assumed to be known, θ_0 is chosen as an estimate of the log hazard ratio based on historical data and prior knowledge, and the parameter τ is selected based on how much we trust the prior distribution. A large τ indicates smaller uncertainty of the choice.

Denote n_e as the total targeted number of events for the trial. J interim analyses are planned when a series of pre-specified number of events are observed in both groups. Let the time of the j th interim analysis be t_j and denote $n_e(t_j)$ as the observed number of events in both groups at t_j , $j = 1, 2, \dots, J$. Accordingly, denote the total number of patients recruited for the treatment and the control groups at the j th interim analysis as $n(t_j)$, where $n(t_j) = n_1(t_j) + n_0(t_j)$, and $n_1(t_j)$ and $n_0(t_j)$ are numbers of patients in the treatment and control group respectively. Denote $(\vec{x}_{1j}, \vec{\delta}_{1j})$ as the set of $n_1(t_j)$ observations from the treatment group, and $(\vec{x}_{0j}, \vec{\delta}_{0j})$ the set of $n_0(t_j)$ observations from the control group at the j th interim analysis.

Conditional on the interim data at t_j , the posterior distribution of θ is,

$$\begin{aligned} p(\theta \mid (\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j})) &\propto p(\theta \mid d_{1n_e(t_j)}, y_{1n_e(t_j)}, y_{0n_e(t_j)}, \dots, d_{11}, y_{11}, y_{01}) \\ &\propto f(d_{1n_e(t_j)}, y_{1n_e(t_j)}, y_{0n_e(t_j)}, \dots, d_{11}, y_{11}, y_{01} \mid \theta) p(\theta) \\ &\propto \prod_{s=1}^{n_e(t_j)} \left(\frac{\exp(\theta)y_{1s}}{y_{0s} + \exp(\theta)y_{1s}} \right)^{d_{1s}} \left(\frac{y_{0s}}{y_{0s} + \exp(\theta)y_{1s}} \right)^{1-d_{1s}} \\ &\quad \exp\left(-\frac{\tau(\theta - \theta_0)^2}{2\sigma^2}\right). \end{aligned} \quad (1)$$

Take second derivative of $\log p(\theta \mid (\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}))$ with regards to θ , we have that

$$\frac{\partial^2 \log p(\theta \mid (\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}))}{\partial \theta^2} \propto - \sum_{s=1}^{n_e(t_j)} \frac{\exp(\theta) y_{1s} y_{0s}}{(y_{0s} + \exp(\theta) y_{1s})^2} - \frac{\tau}{\sigma^2}. \quad (2)$$

The right-hand-side of (2) is negative. Therefore, $p(\theta \mid (\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}))$ is log-concave in θ . Thus, we use adaptive rejection sampling to draw samples of θ from its posterior distribution.

The Bayes factor in favor of H_0 is calculated as

$$\begin{aligned} \text{BF}_{01} &= \frac{m((\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}) \mid \theta = 0)}{m((\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}) \mid \theta < 0)} \\ &= \frac{f((\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}) \mid \theta = 0)}{\int_{\theta < 0} f((\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}) \mid \theta) p(\theta) d\theta (\int_{\theta < 0} p(\theta) d\theta)^{-1}}, \end{aligned} \quad (3)$$

where $m(\cdot)$ is the marginal likelihood of observed data by integrating out θ with respect to the prior distribution. In this paper, we calculate the marginal likelihood by generating a sample $\{\theta^{(h)}\}_{h=1}^H$ from the posterior distribution and computing the harmonic mean of the posterior sample conditional likelihoods by (Newton and Raftery, 1994)

$$\begin{aligned} \hat{m}((\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}) \mid \theta < 0) \\ = \left[\frac{1}{H} \sum_{h=1}^H \frac{1}{f((\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}) \mid \theta^{(h)})} \right]^{-1}. \end{aligned} \quad (4)$$

2.2 A Bayesian sequential design for time-to-event outcomes (BSD4TEO)

We propose a Bayesian sequential design for time-to-event outcomes (BSD4TEO). To control the study wide overall type I error α , alpha spending functions (Lan and DeMets, 1983; Zhu and Yu, 2017; Zhu et al., 2017; Yu et al., 2017) are used. Let w_j^* be the information fraction at the j th interim analysis, which can be defined as $w_j^* = n_e(t_j) / n_e$, the ratio of observed events to the total targeted number of events in both groups for survival settings (DeMets and Lan, 1995; Bang et al., 2010, 2012; Pitt et al., 2001). The alpha spending function $g(w^*)$ is a non-decreasing function of the information fraction, satisfying $g(0) = 0$ and $g(1) = \alpha$, the preset type I error rate. The type I error rate allocated up to the j th interim analysis is $g(w_j^*)$. Similar to the Bayesian sequential design using alpha spending function (BSDASF) proposed by Zhu and Yu (2017) and Zhu et al. (2017), we use the following three types of alpha spending functions: 1) O'Brien–Fleming alpha spending function ($g_1(w^*) = 2 - 2\Phi(z_\alpha / 2 / \sqrt{w^*})$, where Φ is the cumulative distribution function of the standard normal distribution); 2) Pocock alpha spending function ($g_2(w^*) = \alpha \log\{1 + (e-1)w^*\}$); 3) Uniform alpha spending function ($g_3(w^*) = w^* \alpha$). For simplicity, we skip the step of stopping for futility.

A Bayesian sequential design for time-to-event outcomes (BSD4TEO)—At the j th interim analysis, which is planned when $n_e(t_j)$ events out of total targeted n_e events are observed at time $t_j, j = 1, \dots, J (w_J^* = 1)$.

1. Posterior Distributions: Update the posterior distribution of Θ , $p(\theta \mid (\vec{x}_{1j}, \vec{\delta}_{1j}), (\vec{x}_{0j}, \vec{\delta}_{0j}))$, using the current data.
2. Testing statistics: Calculate the Bayes factor BF_{01} in favor of H_0 .
3. Stopping for Efficacy: If $BF_{01} < \lambda_j$, stop the trial for efficacy. Otherwise go to Step 4. The choice of the critical value λ_j , depending on the observed number of events at the j th interim and the choice of α , is discussed in Section 2.3.
4. Continue the Trial: If the trial is not stopped and the maximum number of events has not been observed, continue the trial until the next interim analysis.

2.3 Critical values and power of the BSD4TEO

To generate time-to-event data, we assume that the event times T_0 for control group and T_1 for treatment group are independently following distributions f_{T0} and f_{T1} , respectively. Accordingly, the true value of the log hazard ratio of the treatment group over control group is Θ . For example, if f_{T0} and f_{T1} are exponential distributions with median survival time of five months and ten months, respectively, the corresponding hazard ratio is 0.5. We further assume that the censoring times C_0 for control group and C_1 for treatment group follow the same distribution f_C . The entry time S_I into the trial for subject I is defined as the time from starting the trial until the I th subject is recruited, where $I = 0$ for the control group and $I = 1$ for the treatment group. We assume S_0 and S_1 follow the same distribution f_S .

Algorithm 1 is used to calculate the critical value λ_j at the j th interim analysis, where time t_j is the time when the preset required number of events for j th interim analysis are observed. Information fraction w_j^* is set as $n_e(t_j)/n_e, j = 1, \dots, J$, and J is the total number of planned interim analyses.

Algorithm 1. Calculate the critical value λ_j at time $t_j, j = 1, \dots, J$ with alpha spending function $g(w^*)$ at desired overall type I error rate α . Assume the log hazard ratio $\Theta = 0$ and let n_1 be the maximum allowed sample size for the treatment group and n_0 for the control group. Preset the planned number of events at the j th interim analysis as $n_e(t_j)$.

1. For $m = 1, 2, \dots, M$, where M is the number of simulated trials:

(a) Generated $\vec{t}_0^m = (t_{01}^m, \dots, t_{0n_0}^m)^T$ from f_{T0} as event times,
 $\vec{c}_0^m = (c_{01}^m, \dots, c_{0n_0}^m)^T$, from f_C as censoring times, and $\vec{s}_0^m = (s_{01}^m, \dots, s_{0n_0}^m)^T$
 from f_S as entry times for control group. Calculate $(\vec{x}_0^m, \vec{\delta}_0^m)$ from $\vec{t}_0^m$ and
 $\vec{c}_0^m$, using $x_{0i}^m = \min(t_{0i}^m, c_{0i}^m), \delta_{01}^m = 1$ if $x_{0i}^m = t_{0i}^m$, and $\delta_{01}^m = 0$ if
 $x_{0i}^m = c_{0i}^m, i = 1, 2, \dots, n_0$;

(b) Under the null hypothesis, generate $\vec{\mathbf{t}}_1^m = (t_{11}^m, \dots, t_{1n_1}^m)^T$ from f_{T0} as event times, $\vec{\mathbf{c}}_1^m = (c_{11}^m, \dots, c_{1n_1}^m)^T$ from f_C as censoring times, and $\vec{\mathbf{s}}_1^m = (s_{11}^m, \dots, s_{1n_1}^m)^T$ from f_S as entry time for treatment group. Similarly as in 1a, calculate $(\vec{\mathbf{x}}_1^m, \vec{\delta}_1^m)$ from $\vec{\mathbf{t}}_1^m$ and $\vec{\mathbf{c}}_1^m$;

(c) For $j = 1, \dots, J$, where J is the total number of planned interim analyses:

- i. Update the currently collected data such that those subjects who have not left the trial and do not have an event at time t_j are considered as censored at time t_j . Update $x_{li}^m = t_j$, $\delta_{li}^m = 0$ for $(li) \in \{(l, i) \mid s_{li}^m + x_{li}^m > t_j\}$;
- ii. Update the posterior distribution $p(\theta \mid (\vec{\mathbf{x}}_{1j}, \vec{\delta}_{1j}), (\vec{\mathbf{x}}_{0j}, \vec{\delta}_{0j}))$ using the data in 1(c)i;
- iii. Obtain a sample of Θ from the posterior distribution $p(\theta \mid (\vec{\mathbf{x}}_{1j}, \vec{\delta}_{1j}), (\vec{\mathbf{x}}_{0j}, \vec{\delta}_{0j}))$;
- iv. Estimate the marginal likelihood $\hat{m}((\vec{\mathbf{x}}_{1j}, \vec{\delta}_{1j}), (\vec{\mathbf{x}}_{0j}, \vec{\delta}_{0j}) \mid \theta < 0)$ based on the sample from 1(c)iii;
- v. Calculate the Bayes factor in favor of H_0 , $\text{BF}_{01} = \lambda_j^m$;

2. Let $\vec{\lambda}^m = (\lambda_1^m, \lambda_2^m, \dots, \lambda_J^m)^T$;

2. Denote $\Lambda_1 = (\vec{\lambda}^1, \dots, \vec{\lambda}^M)^T$, a $M \times J$ matrix;
3. The critical value at the 1st interim analysis λ_1 is set as the $g(w_1^*)$ th quantile of the 1st column of matrix Λ_1 ;
4. For $j = 2, \dots, J$:
 - (a) Let Λ_j be the matrix composed of the rows of Λ_{j-1} where the $(j-1)$ th element of the row is larger than or equal to λ_{j-1} ;
 - (b) λ_j is set as the $\Delta g(w_j^*)$ th quantile of the j th column of matrix Λ_j , where $\Delta g(w_j^*) = g(w_j^*) - g(w_{j-1}^*)$

Given the calculated critical values λ_j , Algorithm 2 is proposed to calculate the power of the BSD4TEO at fixed total sample size n and total targeted number of observed events n_e .

Algorithm 2. Power calculation for the BSD4TEO with alpha spending function $g(w^*)$ at desired overall type I error rate α . Let $n = n_1 + n_0$ be the maximum allowed sample size, n_1 for the treatment group and n_0 for the control group. Preset the planned number of events to be observed at the j th interim analysis as $n_e(t_j)$.

1. For $m = 1, 2, \dots, M$, where M is the number of simulated trials:

(a) Generate $\vec{\mathbf{t}}_0^m = (t_{01}^m, \dots, t_{0n_0}^m)^T$ from f_{T0} as event times, $\vec{\mathbf{c}}_0^m = (c_{01}^m, \dots, c_{0n_0}^m)^T$ from f_C as censoring times, and $\vec{\mathbf{s}}_0^m = (s_{01}^m, \dots, s_{0n_0}^m)^T$ from f_S as entry times for control group. Calculate $(\vec{\mathbf{x}}_0^m, \vec{\delta}_0^m)$ from $\vec{\mathbf{t}}_0^m$ and $\vec{\mathbf{c}}_0^m$, using $x_{0i}^m = \min(t_{0i}^m, c_{0i}^m)$, $\delta_{01}^m = 1$, if $x_{0i}^m = t_{0i}^m$, and $\delta_{01}^m = 0$ if $x_{0i}^m = c_{0i}^m$, $i = 1, 2, \dots, n_0$, $i = 1, 2, \dots, n_0$;

(b) Under the alternative hypothesis, generate $\vec{\mathbf{t}}_1^m = (t_{11}^m, \dots, t_{1n_1}^m)^T$ from f_{T1} as event times, $\vec{\mathbf{c}}_1^m = (c_{11}^m, \dots, c_{1n_1}^m)^T$ from f_C as censoring times, and $\vec{\mathbf{s}}_1^m = (s_{11}^m, \dots, s_{1n_1}^m)^T$ from f_S as entry time for treatment group. Similarly as in 1a, calculate $(\vec{\mathbf{x}}_1^m, \vec{\delta}_1^m)$ from $\vec{\mathbf{t}}_1^m$ and $\vec{\mathbf{c}}_1^m$;

(c) For $j = 1, \dots, J$, where J is the total number of planned interim analyses:

- i. Update the currently collected data such that those subjects who have not left the trial and do not have an event at time t_j are considered as censored at time t_j . Update $x_{li}^m = t_j$, $\delta_{li}^m = 0$ for $(li) \in \{(l, i) \mid s_{li}^m + x_{li}^m > t_j\}$;
- ii. Update the posterior distribution $p(\theta \mid (\vec{\mathbf{x}}_{1j}, \vec{\delta}_{1j}), (\vec{\mathbf{x}}_{0j}, \vec{\delta}_{0j}))$ using the data in 1(c)i;
- iii. Obtain a sample of θ from the posterior distribution $p(\theta \mid (\vec{\mathbf{x}}_{1j}, \vec{\delta}_{1j}), (\vec{\mathbf{x}}_{0j}, \vec{\delta}_{0j}))$;
- iv. Estimate the marginal likelihood $\hat{m}((\vec{\mathbf{x}}_{1j}, \vec{\delta}_{1j}), (\vec{\mathbf{x}}_{0j}, \vec{\delta}_{0j}) \mid \theta < 0)$ based on the sample from 1(c)iii;
- v. Calculate the Bayes factor in favor of H_0 , $\text{BF}_{01} = \lambda_j^m$;

2. Let $\vec{\lambda}^m = (\lambda_1^m, \lambda_2^m, \dots, \lambda_J^m)^T$;

2. Denote $\Lambda_1 = (\vec{\lambda}^1, \dots, \vec{\lambda}^M)^T$, a $M \times J$ matrix;
3. For $j = 2, \dots, J+1$, let Λ_j be the matrix composed of the rows of Λ_{j-1} , where the $(j-1)$ th element of the row is larger than or equal to λ_{j-1} ;
4. $\beta = (\text{the number of rows of } \Lambda_{J+1}) / M$. Power = $1 - \beta$.

The R code to implement these algorithms are included in the supplemental material.

3 Sensitivity analysis

The study-wide overall type I error rates are well controlled in the BSD4TEO by using alpha spending functions, adjusting for choices of parameters of the prior distribution. However,

the choice of the prior mean, θ_0 , and the precision parameter, τ , may have an impact on the critical values through the calculation process by simulated trials. A desirable critical value should be consistent through different simulations.

Without loss of generality, we assume an exponential distribution for T_0 with a median survival time of seven months. We further assume that the censoring time C for both groups follows an exponential distribution with a median censoring time of nine months, and the entry time S are uniformly distributed over 6 months. The maximum sample sizes for the treatment group and the control group are both set at $n = 248$. The total targeted number of events is 248. Five interim analyses are planned at 50, 100, 150, 200, and 248 events. We set $\sigma^2 = 0.09^2$. The critical values are calculated for significance level 0.05. For simplicity, we use the O'Brien–Fleming alpha spending function for sensitivity analysis.

To examine how the choice of θ_0 influences the distribution of Bayes factors at each interim analysis, we fix $\tau=1$ and change θ_0 from $\log(0.5)$ to $\log(1)$. Figure 1 displays the changes of critical values for the BSD4TEO at different θ_0 . For better illustration, the critical values for $\theta_0 = \log(0.5)$ are shown in the left panel and those for $\theta_0 = \log(0.6)$ to $\log(1)$ are shown in the right using different scales for the vertical axis. We find that in general, with the O'Brien–Fleming alpha spending function, the critical values increase as the observed number of events increases. Figure 1 also indicates that smaller θ_0 (prior information in favor of the alternative hypothesis) leads to smaller critical values at early stages (more difficult to reject the null hypothesis) but is associated with larger increases in critical values when more events are observed so that the overall type I error is maintained at 0.05.

Figure 2 displays the histograms of BF_{01} at each interim analysis with $\theta_0 = \log(0.6)$, $\log(0.8)$, and $\log(1)$. (The horizontal axis is truncated at 10 in Figure 2 for better illustration.) We find that a θ_0 closer to $\log(1)$ is associated with a smaller mean of BF_{01} and a smaller variance, which partially explains the changes of critical values when θ_0 varies as shown in Figure 1. Figure 2 also indicates that the density of BF_{01} moves towards the right as more data are collected. This explains the overall increasing trend of critical values.

The spread of the distribution of BF_{01} at the interim analysis has an impact on how many simulations are needed to determine the critical values. If the variance of BF_{01} is large, we need more simulations to achieve stable critical values. To further investigate its impact, we calculate the critical values based on different number of simulations and obtain the coefficients of variation from ten repeated calculations. The critical value is considered stable when the corresponding coefficient of variation is smaller than 0.1. In Table 1, we list the coefficients of variation at different numbers of simulations when various priors for θ_0 are chosen. In general, the coefficient of variation decreases as θ_0 gets closer towards the null ($\exp(\theta_0) = 1$). That is, critical values tend to be more stable with θ_0 closer to the null since the corresponding distribution of BF_{01} is less dispersed as demonstrated by Figure 2. It takes fewer simulations for critical values to converge when θ_0 is closer to $\log(1)$. Table 1 can be used as a reference to determine how many simulations are needed to ensure convergence of critical values with different θ_0 . For example, when $\exp(\theta_0) = 0.7$, at least 1800 simulations should be used to calculate the critical values.

To investigate how the choice of τ can affect the critical values, we fix θ_0 and change τ from 0.01 to 100. The change of the critical values for the BSD4TEO at different τ are shown in Figure 3 with θ_0 fixed at $\log(1)$, $\log(0.85)$, and $\log(0.7)$, respectively. We find that using the O'Brien–Fleming alpha spending function, the critical values increase more rapidly with smaller τ as more events are observed. As θ_0 moves away from null, the change of critical values becomes insignificant for larger τ . When we rely less on the prior information, that is, $\tau < 1$, the critical values are similar across different θ_0 . Therefore, we suggest to choose small τ unless researchers are very confident about the prior information.

4 Study design comparison by simulation

In this section, a series of simulations are generated to compare the BSD4TEO with the frequentist group sequential design when the same alpha spending function is used to control the overall type I error rate. We also compare the usage of different alpha spending functions with BSD4TEO in terms of the power of tests, expected event sizes, and the mean squared errors of $\hat{\theta}$. The expected event size is defined as the expected number of observed events when the simulated trial stops with a correct conclusion.

In the simulations, assume the time of event, T_0 , for the control group has an exponential distribution with a median survival time of seven months. To compare power for testing the one-sided hypothesis between the BSD4TEO and the frequentist group sequential design, we set the true $\exp(\theta)$ to a sequence of values ranging from 0.5 to 0.9 by 0.1, with a common variance of $\sigma^2 = 0.09^2$. The censoring times, C , for both groups are assumed to follow an exponential distribution with a median censoring time of nine months. The entry times S are set to be uniformly distributed over six months. The total sample sizes for the treatment group and the control group are both fixed at $n = 248$ and the trial is planned to stop when a maximum of 248 events are observed in both groups. Five interim analyses are planned when 50, 100, 150, 200, and 248 events are observed. The expected time of observing 248 events is two years when $\theta = \log(0.7)$. The desired significance level is set as 0.025. Let $\theta_0 = \theta$, where correct prior information is used and τ is set to 1.

The simulation results for the frequentist group sequential design are presented in Table 2 for further comparison with the BSD4TEO. In the table, size represents the expected event sizes. For frequentist group sequential design, the log-rank statistic is used for interim analyses and θ is estimated from a Cox proportional hazards model. The corresponding variance of the estimator for $\theta(\text{Var}(\hat{\theta}))$, the squared bias of the estimator (bias^2), and the mean squared error (MSE) are also shown in Table 2.

The simulation results based on 5000 replicates for the BSD4TEO are shown in Table 3 for comparison with results for the frequentist group sequential design in Table 2. The empirical type I error rates from the BSD4TEO, $\tilde{\alpha}$, are also reported. For the BSD4TEO, the posterior mean is used as an estimator for θ , and the corresponding variance ($\text{Var}(\hat{\theta})$), the squared bias (bias^2), and the mean squared error of the estimator are also presented. We find that in general, the BSD4TEO results in greater power and smaller expected event size than the frequentist group sequential design with any prespecified alpha spending function. When θ gets closer to 0, the advantages from using Bayesian method become less significant. We

also find that in this simulation, the estimates from the BSD4TEO have uniformly smaller variance and squared bias than the Cox proportional hazards estimates obtained by the frequentist group sequential design. Therefore, the BSD4TEO gives more precise estimates of θ compared to the Cox proportional hazards estimate from the frequentist group sequential design as measured by the mean squared errors, when the prior information is reliable.

In addition, we conclude from Table 3 that power of the BSD4TEO is uniformly the greatest when using the O'Brien–Fleming alpha spending function, followed by the Uniform and Pocock alpha spending functions. On the other side, the O'Brien–Fleming alpha spending function is also associated with the largest expected event size while the Pocock alpha spending function requires the smallest expected event size. This is because the O'Brien–Fleming alpha spending function tends to allocate smaller type I error rate at early stages and therefore the trial is less likely to stop early compared to other alpha spending functions. The reported empirical type I error rates in Table 3 verify that the BSD4TEO with the use of alpha spending functions appropriately maintains the type I error rate at 0.025. As we have seen in Figure 1 and Figure 3, the critical values, calculated through simulations with enough replicates, are adjusted for the choice of prior parameters.

Table 4 shows the simulation results for the BSD4TEO when $\theta_0 = \log(0.9)$ while all other assumptions are kept as before. Comparing with results for the frequentist group sequential design in Table 2, we reach the same conclusion in terms of power and the expected event size. However, we find that as the difference between θ_0 and true θ becomes larger, the squared bias of $\hat{\theta}$ from the BSD4TEO increases. In general, when compared with the frequentist group sequential design, the variances of $\hat{\theta}$ from the BSD4TEO is smaller, while the squared bias of $\hat{\theta}$ is larger. As a result, the mean squared errors of $\hat{\theta}$ obtained from the BSD4TEO is comparable to those provided by the frequentist group sequential design when θ_0 is not too far away from the true θ . Bias is brought in when inaccurate prior information is used. Note that a very off choice of the prior may result in considerably big bias, and therefore a big mean squared error. However, the variances of estimates are generally reduced by using the Bayesian method. For the null case of $\theta_0 = \log(1)$ and the case of $\theta_0 = \log(0.5)$ that strongly favors the treatment group, readers are referred to Table 7 and Table 8 in the Supplementary material for the simulation results of the BSD4TEO, respectively.

We further apply a non-informative prior to the BSD4TEO and compare the results with those from the frequentist group sequential design. We use a vague prior to approximate non-informative prior by setting $\tau = 0.001$. θ_0 is set to be $\log(1)$. Table 5 presents simulation results. The results are compared with the frequentist results in Table 2. We find that the power and expected event size are quite similar for the BSD4TEO and the frequentist group sequential design. We conclude that the advantage of the BSD4TEO is significant when correct informative prior is used. If a vague prior is used, the BSD4TEO may not be better over the frequentist group sequential design in terms of power and expected event size.

5 Application on a real data set

We apply the BSD4TEO on a Chronic Granulomatous Disease (CGD) Dataset. It was described in a report by the International Chronic Granulomatous Disease Cooperative Study Group in 1991 (Gallin et al., 1991). A randomized, double-blind, placebo-controlled study was designed to determine whether interferon gamma decreases the frequency and severity of serious episodes of infection and improves chronic infectious and inflammatory conditions in patients with all genetic types of chronic granulomatous disease. Patients received interferon gamma ($50 \mu\text{g}$ per square meter of body-surface area) or placebo subcutaneously, three times a week for up to a year. The primary endpoint of the study was the time to the first serious infection, which was calculated as the number of days elapsed between the date of random assignment to treatment and the date of diagnosis of the infection.

During a seven-month period (from 1988-08-28 to 1989-03-21), a total of 128 eligible patients were enrolled. Thirty out of 65 patients in the placebo group and 14 out of 63 in the interferon gamma group had at least one serious infection. The last patient left the trial on 1990-01-17, approximately one year and five months after the first patient was randomized. The p-value of a two-sided log-rank test comparing interferon gamma to placebo was 0.0006. The estimated log hazard ratio of the interferon gamma over placebo was -1.0940 (standard error 0.3348). The conclusion from the final analysis was that interferon gamma had significant benefit with respect to the time to serious infection, as compared with placebo.

To apply the BSD4TEO to this data set, we assume that three interim analyses are planned at the time that 15, 30, and 44 events are observed, respectively. The information fractions are $t_1^* = 15 / 44$, $t_2^* = 30 / 44$, $t_3^* = 1$. The prior parameters are set as $\theta_0 = -1.0940$ ($\exp(\theta_0) = 0.3349$). τ is set to 1 since we rely heavily on the prior information. We generate entry times from the uniform distribution, $U(0, 205)$. The event times for control group are assumed to follow an exponential distribution with a mean survival time of 457 days, and censoring times are assumed to be exponentially distributed with a mean censoring time of 391 days based on the maximum likelihood estimates from the original data under the independent censoring assumption. Furthermore, we would like to check how the choice of prior parameters can influence the interim analysis results. For this purpose, $\theta_0 = \log(0.5)$ and $\log(1)$ are also used in the application.

Table 6 presents the analysis results when $\tau=1$, in which n represents the actual total sample size used and d represents the total number of events observed when a trial stops for efficacy. n_1 and n_0 are the sample sizes used by treatment group and control group, while d_1 and d_0 are the number of events observed in the treatment group and the control group, respectively. t_{stop} represents the time in days from the start to the conclusion of the trial. BF_{01} is the Bayes factor in favor of H_0 at the time of stopping the trial. $\hat{\theta}$ is the posterior mean of log hazard ratio θ when the trial stops. We find that using the BSD4TEO with different alpha spending functions will significantly reduce the length of the clinical trial. O'Brien–Fleming alpha spending function is related with the longest trial time, which still reduces the duration

of the trial by 34%. The choice of different prior parameters does not affect the stopping time of the trial.

The prior and estimated posterior densities of θ for the BSD4TEO with O'Brien–Fleming alpha spending function when trial stops at the second interim analysis are shown in Figure 4. We find that the choice of θ_0 in the prior distribution does not influence the final conclusion significantly, although the posterior mean of θ gets closer to the true value more quickly when the prior information is less accurate. We also apply weaker priors with smaller τ , 0.1, in the BSD4TEO for its application on the CGD data set, and the results are presented in Table 9 of Supplementary material.

6 Discussion and future research

In this paper, we propose a Bayesian sequential design for time-to-event outcomes and use alpha spending functions to control the study-wide type I error rate. The Bayes factor is adapted for making decisions at interim and final analyses. We find that BSD4TEO can achieve larger power at a fixed total sample size and event size, compared with the frequentist group sequential design when using the same alpha spending function. We also find that when appropriate priors are chosen, the posterior mean from the BSD4TEO provides a better estimator of log hazard ratio compared to the one based on the Cox proportional hazards model in the frequentist group sequential design, as assessed by the mean squared errors. In our simulations and real data analysis, the BSD4TEO has the potential to stop a clinical trial early with the right conclusion and to give an accurate estimate of treatment difference.

The choices of prior parameters θ_0 and τ can affect the critical values and thereby the final decisions. We perform a sensitivity analysis and find that the number of simulations for critical values to converge increases rapidly as θ_0 moves away from the null. We provide a table as a reference on how many simulations are required for the convergence of critical values when different θ_0 are chosen. As for the choice of τ , we suggest setting τ at small values unless there is enough confidence that the prior information is of high quality and can be well validated in similar settings (Zhu and Yu, 2017; Yu et al., 2017). Simulation studies show that the BSD4TEO with O'Brien–Fleming alpha spending function has the greatest power but also is associated with the largest expected event size. We recommend the use of Pocock or Uniform alpha spending function only under special circumstances, such as limited number of patients, limited access to resources, critical need for early stopping, etc. Otherwise, O'Brien–Fleming alpha spending function is recommended since it is more conservative at early stages and related with greater power.

An advantage of the Bayesian group sequential design is that the prior information can be included in the analysis and together with what is learned during the trial, are available to use for prospectively specified adaptations. The BSD4TEO we proposed provides a flexible framework for time-to-event outcomes to facilitate the introduction of further design modifications. Various hierarchical testing approaches have been proposed along with different methods for decision making in group sequential designs. Those methods could potentially be applied with the BSD4TEO with necessary adjustments. Limitations exist that

the type I error rate may not be well preserved by the BSD4TEO when interim analyses cannot occur as planned. Note that the BSD4TEO is efficient and obtain the desirable operating characteristics when trials are carefully designed, and the plans are followed. In addition, there are alternative bias-adjusted estimators for group sequential designs that are not included in the simulation studies. Our conclusions regarding estimators are specific to the Cox proportional hazards estimator for the frequentist design. Other bias-adjusted estimators should also be able to be used with the BSD4TEO. As future research, we will work on a Bayesian sequential design with adaptive randomization rates for multi-arm, multi-stage clinical trials, where the rates of allocating patients to different treatment arms can be changed adaptively at each interim analysis. We will also extend the design to allow for sample size adjustment and more flexible schedule of interim analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the two reviewers, the associate editor and editor for their constructive comments that improve the paper. This work was supported in part by U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Armitage P (1975). Sequential medical trials. Oxford: Blackwell Scientific Publications.
- Bang Y-J, Cutsem EV, Feyereislova A, Chung HC, Shen L, Sawaki A, Lordick F, Ohtsu A, Omuro Y, Satoh T, Aprile G, Kulikov E, Hill J, Lehle M, Rschoff J, and Kang Y-K (2010). Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of her2-positive advanced gastric or gastro-oesophageal junction cancer (toga): a phase 3, open-label, randomised controlled trial. *The Lancet* 376 (9742), 687–697.
- Bang Y-J, Kim Y-W, Yang H-K, Chung HC, Park Y-K, Lee KH, Lee K-W, Kim YH, Noh S-I, Cho JY, Mok YJ, Kim YH, Ji J, Yeh T-S, Button P, Sirzn F, and Noh SH (2012). Adjuvant capecitabine and oxaliplatin for gastric cancer after d2 gastrectomy (classic): a phase 3 open-label, randomised controlled trial. *The Lancet* 379 (9813), 315–321.
- Biswas S, Liu DD, Lee JJ, and Berry DA (2009). Bayesian clinical trials at the university of texas m. d. anderson cancer center. *Clinical Trials* 6 (3), 205–216. [PubMed: 19528130]
- Casella G and Berger RL (2002). Statistical inference (Second ed.). Pacific Grove, CA: Duxbury.
- DeMets DL and Lan G (1995). The alpha spending function approach to interim data analyses, pp. 1–27. Boston, MA: Springer.
- Gallin J, Malech H, Curnutte W, Quie P, Jaffe H, and Ezkowitz R (1991). A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. *The New England journal of medicine* 324 (8), 509–516. [PubMed: 1846940]
- Kim K and DeMets DL (1987). Design and analysis of group sequential tests based on the type i error spending rate function. *Biometrika* 74 (1), 149–154.
- Lan KKG and DeMets DL (1983). Discrete sequential boundaries for clinical trials. *Biometrika* 70(3), 659–663.
- Lee JJ and Chu CT (2012). Bayesian clinical trials in action. *Statistics in Medicine* 31 (25), 2955–2972. [PubMed: 22711340]
- Newton MA and Raftery AE (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)* 56 (1), 3–48.

- O'Brien PC and Fleming TR (1979). A multiple testing procedure for clinical trials. *Biometrics* 35 (3), 549–556. [PubMed: 497341]
- Pitt B, O'Neill B, Feldman R, Ferrari R, Schwartz L, Mudra H, Bass T, Pepine C, Texter M, Haber H, Uprichard A, Cashin-Hemphill L, and Lees RS (2001). The quinapril ischemic event trial (quiet): evaluation of chronic ace inhibitor therapy in patients with ischemic heart disease and preserved left ventricular function. *The American Journal of Cardiology* 87 (9), 1058–1063. [PubMed: 11348602]
- Pocock SJ (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64(2), 191–199.
- U.S. Food and Drug Administration (1998). Guidance for industry: E9 statistical principles for clinical trials. <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf>.
- U.S. Food and Drug Administration (2010a). Guidance for industry: Adaptive design clinical trials for drugs and biologics. <https://www.fda.gov/downloads/Drugs/Guidances/ucm201790.pdf>.
- U.S. Food and Drug Administration (2010b). Guidance for the use of bayesian statistics in medical device clinical trials. <https://www.fda.gov/RegulatoryInformation/Guidances/ucm071072.htm>.
- U.S. Food and Drug Administration (2014). Complex issues in developing drugs and biological products for rare diseases and accelerating the development of therapies for pediatric rare diseases including strategic plan: Accelerating the development of therapies for pediatric rare diseases. <https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAact/FDASIA/UCM404104.pdf>.
- U.S. Food and Drug Administration (2016). Adaptive designs for medical device clinical studies: Guidance for industry and food and drug administration staff. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf>.
- Yu Q, Zhu L, and Zhu H (2017). A bayesian sequential design with adaptive randomization for 2-sided hypothesis test. *Pharmaceutical Statistics* 16 (6), 451–465. [PubMed: 28980435]
- Zhu H and Yu Q (2017). A bayesian sequential design using alpha spending function to control type i error. *Statistical Methods in Medical Research* 26 (5), 2184–2196. [PubMed: 26187736]
- Zhu H, Yu Q, and Mercante DE (2017). A bayesian sequential design with binary outcome. *Pharmaceutical Statistics* 16(3), 192–200. [PubMed: 28251815]

Fig. 1.

Comparison of critical values λ_j with O'Brien-Fleming alpha spending function for the BSD4TEO at different θ_0 , calculated from 10000 simulated trials.

Fig. 2.

The histograms for BF_{01} at each interim analysis in 5000 simulated trials, when $\theta_0 = 0.6$, 0.8, 1, along with the vertical dashed lines at the corresponding 0.05 quantiles.

Fig. 3. Comparison of critical values λ_j with O'Brien-Fleming alpha spending function for the BSD4TEO at different τ , calculated from 10000 simulated trials.

Fig. 4.

The prior density (blue) and estimated posterior density of θ with histogram (black), and the corresponding posterior mean (red) when trial stops at the second interim analysis, for the BSD4TEO application on CGD data set with O'Brien–Fleming alpha spending function.

Table 1

Compare the coefficients of variation of the critical values at the final analysis with different θ_0 , calculated using O'Brien–Fleming alpha spending function from ten replicates.

	number of simulations					
exp(θ_0)	300	500	1500	1800	3200	6900
1	0.084	0.054	0.047	0.044	0.023	0.011
0.9	0.140	0.092	0.048	0.050	0.029	0.021
0.8	0.243	0.174	0.097	0.090	0.060	0.020
0.7	0.492	0.273	0.144	0.097	0.051	0.041
0.6	0.455	0.239	0.123	0.161	0.097	0.066
0.5	0.400	0.460	0.212	0.222	0.015	0.096

Table 2

Power, expected event size, variance, squared bias and mean squared error of $\hat{\theta}$ from the frequentist group sequential design for different true θ , when using different alpha spending functions.

$\exp(\theta)$		Frequentist				
		Power	Size	$\text{Var}(\hat{\theta})$	bias^2	MSE
0.5	OF	0.999	120	0.0410	0.0020	0.0430
0.6	OF	0.979	151	0.0311	0.0022	0.0333
0.7	OF	0.819	182	0.0278	0.0017	0.0296
0.8	OF	0.436	199	0.0266	0.0005	0.0271
0.9	OF	0.148	205	0.0224	<.0001	0.0224
0.5	Pocock	0.996	86	0.0530	0.0084	0.0613
0.6	Pocock	0.924	121	0.0476	0.0098	0.0574
0.7	Pocock	0.729	147	0.0454	0.0071	0.0525
0.8	Pocock	0.345	157	0.0431	0.0020	0.0451
0.9	Pocock	0.128	148	0.0355	0.0004	0.0359
0.5	Uniform	0.998	90	0.0541	0.0084	0.0625
0.6	Uniform	0.937	126	0.0479	0.0090	0.0569
0.7	Uniform	0.748	153	0.0442	0.0061	0.0504
0.8	Uniform	0.365	167	0.0402	0.0017	0.0419
0.9	Uniform	0.126	165	0.0327	0.0002	0.0329
OF: O'Brien–Fleming.						

Table 3

Power, expected event size, variance, squared bias, and mean squared error of $\hat{\theta}$ from the BSD4TEO, when correct prior information is used. Empirical type I error rates for the BSD4TEO are reported as $\tilde{\alpha}$ in the table.

$\exp(\theta)$		BSD4TEO					
		Power	Size	$\text{Var}(\hat{\theta})$	bias^2	MSE	$\tilde{\alpha}$
0.5	OF	1.000	110	0.0011	<.0001	0.0011	0.0252
0.6	OF	0.990	140	0.0013	<.0001	0.0013	0.0246
0.7	OF	0.883	173	0.0014	<.0001	0.0014	0.0253
0.8	OF	0.525	189	0.0018	<.0001	0.0018	0.0248
0.9	OF	0.187	196	0.0019	<.0001	0.0019	0.0242
0.5	Pocock	1.000	76	0.0008	<.0001	0.0008	0.0253
0.6	Pocock	0.976	104	0.0010	<.0001	0.0010	0.0249
0.7	Pocock	0.827	136	0.0012	<.0001	0.0012	0.0255
0.8	Pocock	0.466	152	0.0017	<.0001	0.0017	0.0250
0.9	Pocock	0.169	139	0.0019	<.0001	0.0019	0.0252
0.5	Uniform	1.000	79	0.0008	<.0001	0.0008	0.0250
0.6	Uniform	0.983	111	0.0010	<.0001	0.0010	0.0243
0.7	Uniform	0.851	143	0.0013	<.0001	0.0013	0.0250
0.8	Uniform	0.485	161	0.0017	<.0001	0.0017	0.0246
0.9	Uniform	0.175	153	0.0019	<.0001	0.0019	0.0252

OF: O'Brien–Fleming.

Table 4

Power, expected event size, variance, squared bias, and mean squared error of $\hat{\theta}$ from the BSD4TEO, when θ_0 is set to $\log(0.9)$.

$\exp(\theta)$		BSD4TEO				
		Power	Size	$\text{Var}(\hat{\theta})$	bias^2	MSE
0.5	OF	1.000	110	0.0004	0.2376	0.2381
0.6	OF	0.990	139	0.0004	0.1020	0.1024
0.7	OF	0.849	169	0.0008	0.0350	0.0358
0.8	OF	0.512	187	0.0014	0.0069	0.0083
0.9	OF	0.187	196	0.0019	<.0001	0.0019
0.5	Pocock	1.000	75	0.0004	0.2667	0.2670
0.6	Pocock	0.979	104	0.0003	0.1137	0.1140
0.7	Pocock	0.793	133	0.0005	0.0379	0.0384
0.8	Pocock	0.450	142	0.0012	0.0072	0.0084
0.9	Pocock	0.169	139	0.0019	<.0001	0.0019
0.5	Uniform	1.000	80	0.0004	0.2629	0.2633
0.6	Uniform	0.983	110	0.0003	0.1119	0.1122
0.7	Uniform	0.812	141	0.0006	0.0372	0.0378
0.8	Uniform	0.467	153	0.0013	0.0071	0.0084
0.9	Uniform	0.175	153	0.0019	<.0001	0.0019

OF: O'Brien–Fleming.

Table 5

Power, expected event size, variance, squared bias, and mean squared error of $\hat{\theta}$ from the BSD4TEO, when θ_0 is set to $\log(1)$ and $\tau = 0.001$.

$\exp(\theta)$		BSD4TEO				
		Power	Size	$\text{Var}(\hat{\theta})$	bias^2	MSE
0.5	OF	1.000	115	0.0557	0.0061	0.0618
0.6	OF	0.970	153	0.0429	0.0036	0.0465
0.7	OF	0.766	185	0.0369	0.0013	0.0382
0.8	OF	0.399	204	0.0275	0.0002	0.0277
0.9	OF	0.131	204	0.0227	<.0001	0.0227
0.5	Pocock	0.999	85	0.0536	0.0072	0.0608
0.6	Pocock	0.956	120	0.0467	0.0075	0.0541
0.7	Pocock	0.717	148	0.0477	0.0052	0.0529
0.8	Pocock	0.346	161	0.0406	0.0015	0.0421
0.9	Pocock	0.127	153	0.0311	0.0003	0.0314
0.5	Uniform	0.999	90	0.0537	0.0076	0.0614
0.6	Uniform	0.958	126	0.0472	0.0071	0.0542
0.7	Uniform	0.732	157	0.0467	0.0043	0.0510
0.8	Uniform	0.363	173	0.0376	0.0010	0.0386
0.9	Uniform	0.122	167	0.0291	0.0002	0.0293

OF: O'Brien–Fleming.

Table 6

The required event size, stopping time, Bayes factor, and estimated log hazard ratio for the BSD4TEO application on CGD data set, using different θ_0 and alpha spending functions, when τ is set to 1.

$\exp(\theta_0)$		n	d	n_1	d_1	n_0	d_0	t_{stop}	BF_{01}	$\hat{\theta}$
0.3349	O'Brien-Fleming	128	30	63	7	65	23	333	0.0021	-1.237
	Pocock	128	15	63	3	65	12	211	0.0352	-1.208
	Uniform	128	15	63	3	65	12	211	0.0352	-1.208
0.5	O'Brien-Fleming	128	30	63	7	65	23	333	0.0047	-0.9968
	Pocock	128	15	63	3	65	12	211	0.0615	-0.9076
	Uniform	128	15	63	3	65	12	211	0.0615	-0.9076
1	O'Brien-Fleming	128	30	63	7	65	23	333	0.0374	-0.6015
	Pocock	128	15	63	3	65	12	211	0.2245	-0.4096
	Uniform	128	15	63	3	65	12	211	0.2245	-0.4096