Flexible Stopping Boundaries When Changing Primary Endpoints after Unblinded Interim Analyses

Liddy M. Chen[1], **Joseph G. Ibrahim**[2], and **Haitao Chu**[3]

[1]Global Research Operation, Biostatistics, PAREXEL International, 2520 Meridian Parkway, Durham, North Carolina 27713, U.S.A

[2]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A

[3]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A

Abstract

It has been widely recognized that interim analyses of accumulating data in a clinical trial can inflate type I error. Different methods, from group sequential boundaries to flexible alpha spending functions, have been developed to control the overall type I error at pre-specified level. These methods mainly apply to testing the same endpoint in multiple interim analyses. In this paper, we consider a group sequential design with pre-planned endpoint switching after unblinded interim analyses. We extend the alpha spending function method to group sequential stopping boundaries when the parameters can be different between interim, or between interim and final analyses.

Keywords

1 Introduction

It is fundamental to have clinical trials that are properly designed to answer specific scientific questions, such as whether a drug improves overall survival. Every trial design is striving to answer this question with as much robustness and accuracy as possible while involving the fewest number of patients, reasonable costs and the shortest duration of time. Methodology for group sequential clinical trials has been developed largely during the past few decades so that a trial can be stopped early if there is strong evidence of efficacy during any planned interim analysis. Pocock (1977) first proposed that the crossing boundary be constant for all equally spaced analyses. O'Brien and Fleming (1979) suggested that the crossing boundaries for the kth analysis, $z_c(k)$, be changed over the total number of analyses K such that $z_c(k) = z_{OBF} \sqrt{K/k}$. In both procedures, the number of interim analyses and the timing of the interim analyses need to be pre-determined. The O'Brien-Fleming boundaries

Correspondence to: Liddy M. Chen, liddy.chen@parexel.com.

have been used more frequently because they preserve a nominal significance level at the final analysis that is close to that of a single test procedure.

The alpha spending function developed by Lan and DeMets (1983) is a more flexible group sequential procedure that does not require the total number nor the exact time of the interim analyses to be specified in advance. Earlier development of the alpha spending function was based on the assumption that the information accumulated between each interim analysis is statistically independent. However, this assumption does not apply to longitudinal studies in a sequential test of slopes for which the total information is unknown. Sequential analysis using the linear random-effects model suggested by Laird and Ware (1982) has been considered by Lee and DeMets (1991), and Wu and Lan (1992). There have been debates on whether the alpha spending function can still be used since the independent increments structure does not hold and the information fraction is unknown (Wei et al 1990, Su & Lachin 1992). It was argued by DeMets and Lan (1994) that the alpha spending function can still be used with a more complex correlation between successive test statistics. The key to using the alpha spending function is being able to define the information fraction. Tsiatis et al. (1995) derived the joint distribution of sequentially computed score tests and maximum likelihood estimates for general parametric survival models subject to censoring and staggered entry, and show that the joint distribution of the sequentially computed score test for any single parameter has an asymptotically independent increments structure. Scharfstein et al. (1997) further demonstrated that sequential Wald statistics behave like a standardized partial sum of independent normal variables and thus have independent increments.

A high degree of flexibility has been well developed with respect to the timing of the analyses and how much type I error (a) to spend at each analysis. However, these methods mainly apply to testing the same endpoint or the same hypothesis in multiple testing. Changing study endpoints is not uncommon in clinical trials. Chan et. al. (2004) compared published articles from 102 randomized clinical trials and reported that 62% of the trials had at least one primary endpoint that was changed. For example the PROactive study concluded that pioglitazone, a peroxisome proliferator-activated receptor agonist, can significantly reduce macrovascular events in diabetes based on a composite endpoint of multiple macrovascular events and death (Dormandy et al 2006). However, the primary endpoint is time to occurrence of a new macrovascular event or death by design (Charbonnel et al 2004).

Changing the primary endpoint after an unblinded interim analysis has been a challenging statistical issue. The EMEA reflection paper finalized in 2007 (EMEA 2007) does not recommend any endpoint change after an unblinded interim analysis although the language has been more flexible in the final version compared to the draft versions. The recently published draft FDA guideline on adaptive designs (FDA 2010) recognizes the need for endpoint change in some cases. However, statistical procedures to control the type I error rate are not well developed in this setting. As the design of the clinical trials is becoming increasingly complex, being able to test different hypotheses while preserving the pre-specified significance level is an emergent need. In this paper, we consider combining a group sequential design with pre-planned endpoint switching after unblinded interim analyses. The major motivation for this is 1) to maximize the information at the interim

analysis so that the interim analysis will have a higher chance of success, or 2) to maximize the probability of success at the final analysis by switching endpoints.

Chen et al (2003) considered a switching endpoint design based on the log-rank statistic where mortality was used as the primary endpoint at the interim analysis while a composite endpoint was used as the primary endpoint at the final analysis. The rationale for the design is that a composite endpoint that includes mortality would have a higher event rate and thus have a higher power should the mortality outcome fail to show significance at the interim analysis. However, using a composite endpoint at the interim analysis may not be feasible as regulatory agencies may be reluctant to terminate a trial early to allow the efficacy claim unless there are convincing results based on a more conventional outcome, such as mortality. Since how much type I error (α) to spend at an interim analysis and the timing of the interim analysis are flexible, one can allocate most of the α to the interim analysis so that there is a sufficiently high probability of terminating the trial early should there be a significant difference in mortality. If the mortality endpoint fails, switching to a composite endpoint can still be successful. The PROactive study described above(Dormandy et al 2006) was successful in a composite endpoint. As another example, the MERIT-HF study (MERIT-HF Study Group 1999) was designed to switch to a composite endpoint after an interim analysis based solely on mortality. The study was terminated early due to a convincing mortality outcome at the interim analysis.

The original motivation of this paper came from a design of a phase III trial in patients with glioblastoma multiforme (GBM). The investigator wished to design the study using progression-free survival (PFS) as the primary endpoint in the interim analysis, while using overall survival as the primary endpoint to be tested at the final analysis. A major consideration for the design is the low event rate for overall survival at time of the interim analysis whereas the PFS data is much more mature and will have a much higher probability to cross the stopping boundary. Goldman et al (2008), from the Southwest Oncology Group Statistical Center, recommend the use of an intermediate endpoint, such as PFS, for interim futility testing of Phase III trials. Other authors have also explored the possibility of using an intermediate endpoint to shorten the time lines for drug approval (Scher et al 2009, Olmos et al 2009, Hallstrom 2009). Although regulatory agencies have not been very open to the idea of terminating a trial early based on intermediate endpoints, the development of targeted therapy and a better understanding of the relationship between biomarkers and disease progression may change the landscape of drug approval. In the future, there may be more studies that would be allowed to use an intermediate endpoint in an early interim analysis.

No paper has formally considered endpoint switching and the implied stopping boundaries in general. In this paper, we extend the alpha spending function methodology to derive stopping boundaries when our interest focuses on switching endpoints or parameters at different analysis times. Statistically, this is equivalent to testing different hypotheses at different interim analyses. The derivation is based on the joint distribution of the test statistics and the alpha spending function so that the overall type I error α will be strictly preserved. After a brief review of the alpha spending function in Section 2, in Section 3, the newly derived stopping boundaries are compared to the boundaries without changing the parameters, using the Pocock and O'Brien-Fleming like spending functions proposed by Lan

and DeMets (1983). Applications to a biviarate survival model and a joint model of longitudinal and time-to-event data are discussed in Sections 4 and 5. We close the article with a discussion in Section 6.

2 Preliminaries: The Alpha Spending Function and Stopping Boundaries

Let T denote the scheduled end of the trial, and t^* denote the fraction of information that has been observed at calendar time t ($t \in [0, T]$). Also let i_k, $k = 1, 2,\dots, K$ denote the information available at the kth interim analysis at calendar time t_k, so $t_k^* = i_k/I$, where I is the total information. Lan and DeMets specified an alpha spending function such that $\alpha(0) = 0$ and $\alpha(1) = \alpha$. Boundary values $z_c(k)$ can be determined successively so that

$$P_0\{|Z(1)| \geq z_c(1), \text{or} |Z(2)| \geq z_c(2), \text{or} \dots, \text{or} |Z(k)| \geq z_c(k)\} = \alpha(t_k^*), \quad (2.1)$$

where $\{Z(1),\dots, Z(k)\}$ are the test statistics from the interim analyses $1,\dots, k$ and P_0 denotes the probability.

Alpha spending functions that approximate O'Brien-Fleming or Pocock Boundaries are as follows:

$$\text{O'Brien−Fleming:} \alpha_1(t^*) = 2 - 2\Phi(Z_{\alpha/2}/\sqrt{t^*}),$$
$$\text{Pocock:} \alpha_2(t^*) = \alpha\ln(1 + (e-1)t^*),$$

where Φ denotes the standard normal cumulative distribution function. To solve for the boundary values $z_c(k)$, we need to obtain the joint distribution of $Z\{(1), Z(2),\dots, Z(k)\}$. In most cases, this distribution is asymptotically multivariate normal, and the covariance matrix Σ is simple when the test statistics involve the same parameter at each interim analysis. In particular,

$$\sigma_{lk} = \text{cov}\{Z(l), Z(k)\}$$
$$= \sqrt{t_l^*/t_k^*}$$
$$= \sqrt{i_l/i_k}, l \leq k,$$

where σ_{lk} is teh lk^{th} element of Σ.

If the information increments have an independent distributional structure, which is usually the case, $\sqrt{i_l/i_k} = \sqrt{n_l/n_k}$, where n_l and n_k are the number of subjects included in the lth and kth interim analyses. The derivation of $z_c(k)$ based on $\alpha(t^*)$ is relatively straightforward with this covariance structure using the methods of Armitage et al (1969).

3 Stopping Boundaries for Testing Different Parameters at the Interim and Final Analysis

3.1 Independent Information Increments

Classical group sequential methods assume that each patient has only one response, such as a 1-year response rate, thus test statistics at each interim analysis will have independent increments. In this Section, we expand upon the classical group sequential methods to derive the stopping boundaries for testing different parameters.

Similar to testing the same hypothesis, boundary values $z_c(k)$ can still be determined successively by equation (2.1) when testing different hypotheses. Because the alpha spending functions $a_1(t^*) = a_2(t^*) = a$ when $t^* = 1$, the overall type I error will be a with the set of crossing boundaries, $z_c(k)$, determined from equation (2.1). The joint distribution of the test statistics, $\{Z(1),\ldots, Z(k)\}$ still asymptotically follows a multivariate normal distribution. In general, if different interim analyses involve different parameters, the covariance structure is unknown, and therefore we cannot obtain the asymptotic joint distribution of $\{Z(1), Z(2),\ldots, Z(k)\}$. Thus, deriving $z_c(k)$ will be problematic. When the parameters we are testing at the interim analysis and the final analysis are from the same likelihood function, however, the covariance is known and can be computed from the Fisher information matrix.

To make our ideas clear, let β_1 denote the parameter to be tested at the lth interim analysis, and let β_2 be the parameter to be tested at the kth interim analysis. The null hypotheses are $H_0: \beta_1 = \beta_{01}$ for testing β_1, $H_0: \beta_2 = \beta_{02}$ for testing β_2. Let l_k denote the log-likelihood at the kth analysis from n_k independent samples, $l_k = \log(L(\beta_1, \beta_2|\mathbf{y_{n_k}}))$. Similarly, let $l_l = \log(L(\beta_1, \beta_2|\mathbf{y_{n_1}}))$ and $l_{k-l} = \log(L(\beta_1, \beta_2|\mathbf{y_{n_{k-1}}}))$. Further assume that $Z(l)$ and $Z(k)$ are the score statistics at the lth and kth interim analysis, and the information accumulated between each interim analysis is independent. Define

$$S_l = \frac{\partial l_l}{\partial \beta_1}\Big|_{\beta_1=\beta_{01}}, \quad S_k^* = \frac{\partial l_k}{\partial \beta_2}\Big|_{\beta_2=\beta_{02}},$$
$$S_l^* = \frac{\partial l_l}{\partial \beta_2}\Big|_{\beta_2=\beta_{02}}, \quad S_{k-l}^* = \frac{\partial l_{k-l}}{\partial \beta_2}\Big|_{\beta_2=\beta_{02}},$$
$$I_l = -\mathrm{E}\left[\frac{\partial^2 l_l}{\partial \beta_1^2}\right]\Big|_{\beta_1=\beta_{01}}, \quad I_k^* = -\mathrm{E}\left[\frac{\partial^2 l_k}{\partial \beta_2^2}\right]\Big|_{\beta_2=\beta_{02}}.$$

It can be shown that

$$
\begin{aligned}
\mathrm{Cov}\{Z(l), Z(k)\} &= \mathrm{E}(Z(l)Z(k)) = \mathrm{E}\left(\frac{S_l S_k^*}{\sqrt{I_l I_k^*}}\right) \\
&= \frac{\mathrm{E}\{S_l(S_l^* + S_{k-l}^*)\}}{\sqrt{I_l I_k^*}} = \frac{\mathrm{E}\{S_l S_l^*\} + \mathrm{E}\{S_l S_{k-l}^*\}}{\sqrt{I_l I_k^*}} \\
&= \frac{\mathrm{E}\left(\frac{\partial l_l}{\partial \beta_1}\frac{\partial l_l}{\partial \beta_2}\right)\big|_{\beta_1=\beta_{01},\beta_2=\beta_{02}}}{\sqrt{I_l I_k^*}} \\
&= \sqrt{\frac{n_l}{n_k}}\frac{I_{12}(\beta_{01},\beta_{02})}{\sqrt{I_{11}(\beta_{01})I_{22}(\beta_{02})}},
\end{aligned}
\tag{3.1}
$$

where $I_{12}(\beta_{01}, \beta_{02})$ is the off diagonal element, and $I_{11}(\beta_{01})$, $I_{22}(\beta_{02})$ are the diagonal elements of the Fisher information matrix. Note that $E\{S_l S_{k-l}^*\}=0$ under the independent information increment assumption. Therefore, when we test different hypotheses at different interim analyses, the stopping boundaries will not only depend on the information fraction, they will also depend on the information matrix of the two parameters under H_0. Thus, there will not be one set of stopping boundaries that can be used for all likelihood functions or all parameters. The investigators in this case must derive their own stopping boundaries for different study designs.

Letting $w = \dfrac{I_{12}(\beta_{01}, \beta_{02})}{\sqrt{I_{11}(\beta_{01}) I_{22}(\beta_{02})}}$, it can be shown that

$$w = \frac{E\left(\frac{\partial l}{\partial \beta_1} \frac{\partial l}{\partial \beta_2}\right)}{\sqrt{E\left(\left(\frac{\partial l}{\partial \beta_1}\right)^2\right) E\left(\left(\frac{\partial l}{\partial \beta_2}\right)^2\right)}}\Bigg|_{\beta_1 = \beta_{01}, \beta_2 = \beta_{02}}$$

$$= \frac{\text{Cov}\left(\frac{\partial l}{\partial \beta_1}, \frac{\partial l}{\partial \beta_2}\right)}{\sqrt{\text{Var}\left(\frac{\partial l}{\partial \beta_1}\right) \text{Var}\left(\frac{\partial l}{\partial \beta_2}\right)}}\Bigg|_{\beta_1 = \beta_{01}, \beta_2 = \beta_{02}},$$

where l denotes the log-likelihood based on a sample of size 1. Thus, w is the correlation coefficient (Corr) of the score function, and $|w| \leq 1$. Since the covariance matrix, Σ of the test statistics $(Z(1), Z(2),\ldots, Z(k))$ is positive definite, the value of w is also bounded by a number that is > -1. When we test the same endpoint between the lth and the kth interim analysis, $w = 1$. It should be noted that the covariance in (3.1) is derived under the global hypothesis of $\beta_1 = \beta_{10}$ and $\beta_2 = \beta_{20}$. However, the score statistic $Z(l)$ at the lth analysis only tests the hypothesis of $\beta_1 = \beta_{10}$, and $Z(k)$ at the kth analysis only tests the hypothesis of $\beta_2 = \beta_{20}$. The other parameter will be replaced by its maximum likelihood estimate when calculating the score statistics. If we reject the null hypothesis at the lth interim analysis, the covariance of $Z(l)$ and $Z(k)$ will be irrelevant, as $Z(k)$ will not be calculated. If we cannot reject the null hypothesis at the lth analysis, it is reasonable to derive the covariance under the global hypothesis of β_1 and β_2. Therefore, the false positive rate will be under the global hypothesis.

We next calculate different stopping boundaries by assuming different values of w. In Table 1, we compare the boundaries computed from $a_1(t^*)$ and $a_2(t^*)$, the O'Brien-Fleming-like, and the Pocock-like alpha spending functions proposed by Lan & DeMets (1983). The comparison is made for a one-sided test with significance level $a = 0.025$, $K = 5$, equally spaced, and the test parameter is β_1 for $j = 1, 2$, β_2 for $j = 3, 4, 5$ ($j = 1,\ldots, 5$), and $t_j^* = j/5$.

Note that between ($j = 1, 2$) and ($j = 3, 4, 5$), the value of w is still 1. The covariance matrix for $(Z(1),\ldots, Z(5))$ is

$$\sum = \begin{pmatrix} 1 & \sqrt{1/2} & \sqrt{1/3}w & \sqrt{1/4}w & \sqrt{1/5}w \\ \sqrt{1/2} & 1 & \sqrt{2/3}w & \sqrt{2/4}w & \sqrt{2/5}w \\ \sqrt{1/3}w & \sqrt{2/3}w & 1 & \sqrt{3/4} & \sqrt{3/5} \\ \sqrt{1/4}w & \sqrt{2/4}w & \sqrt{3/4} & 1 & \sqrt{4/5} \\ \sqrt{1/5}w & \sqrt{2/5}w & \sqrt{3/5} & \sqrt{4/5} & 1 \end{pmatrix}, \quad (3.2)$$

where $w \geq 1$. We can see that the covariance matrix can be partitioned into four sub-matrices

$\begin{pmatrix} \sum_{\beta_1} & \sum_{\beta_1,\beta_2} \\ \sum'_{\beta_1,\beta_2} & \sum_{\beta_2} \end{pmatrix}$. Solving for $(z_c(1),\ldots, z_c(K))$ in equation (2.1) requires numerical integration. The quadrature method by Armitage et al (1969) cannot be applied here with this covariance structure since the statistics are not the same in the sequential procedure. Here, we used the adaptive integration method by Genz (1992) to evaluate $z_c(k)$. Compared to a Monte Carlo algorithm and the subregion adaptive algorithm, the adaptive integration method of Genz (1992) reliably computes multivariate normal probabilities with as many as ten variables in a few seconds. When we compare our boundaries to a group sequential procedure that does not change parameters at different interim analyses, the boundaries are very close when the alpha spending function is $a_1(t^*)$. However, the boundaries are substantially different when the alpha spending function is $a_2(t^*)$.

We next consider a scenario where we change the parameter at the final (5*th*) analysis. Our stopping boundary for the first 4 analyses will be the same as the ones in Lan and DeMets (1983). The 5*th* boundary was calculated for different values of w (Table 2). The boundary is substantially different than the Lan-Demets boundary. This shows when a is minimally spent prior to the endpoint change, as in early interim analyses using $a_1(t^*)$, the impact on the stopping boundaries is small (smaller penalty). The more the a is spent prior to the endpoint change, the more significant the impact is on the boundaries. In Table 3, we compare the boundaries computed from $a_1(t^*)$ and $a_2(t^*)$ for a one-sided $a = 0.025$ test with $K = 2$ and $t_j^* = j/2$.

When the test parameter is changed more than once, the covariance matrix for $(Z(1),\ldots, Z(k))$ will have more elements where $w \geq 1$. For example, when $K = 5$, if we test β_1 for $j = 1$, 2, test β_2 for $j = 3$, 4, and test a third parameter β_3 for $j = 5$, the covariance matrix in (3.2) will become

$$\sum = \begin{pmatrix} 1 & \sqrt{1/2} & \sqrt{1/3}w_1 & \sqrt{1/4}w_1 & \sqrt{1/5}w_2 \\ \sqrt{1/2} & 1 & \sqrt{2/3}w_1 & \sqrt{2/4}w_1 & \sqrt{2/5}w_2 \\ \sqrt{1/3}w_1 & \sqrt{2/3}w_1 & 1 & \sqrt{3/4} & \sqrt{3/5}w_3 \\ \sqrt{1/4}w_1 & \sqrt{2/4}w_1 & \sqrt{3/4} & 1 & \sqrt{4/5}w_3 \\ \sqrt{1/5}w_2 & \sqrt{2/5}w_2 & \sqrt{3/5}w_3 & \sqrt{4/5}w_3 & 1 \end{pmatrix},$$

where $w_1 = \dfrac{I_{12}(\beta_{01}, \beta_{02})}{\sqrt{I_{11}(\beta_{01})I_{22}(\beta_{02})}}$, $w_2 = \dfrac{I_{13}(\beta_{01}, \beta_{03})}{\sqrt{I_{11}(\beta_{01})I_{33}(\beta_{03})}}$, and $w_3 = \dfrac{I_{23}(\beta_{02}, \beta_{03})}{\sqrt{I_{22}(\beta_{02})I_{33}(\beta_{03})}}$. As a result, the critical values will be larger.

The value of w is determined by the Fisher information matrix. However, when the likelihood function is complex, the Fisher information matrix can be dificult to obtain. In this case, researchers can use more conservative boundaries by assuming a smaller w so that the overall type I error rate will be a. The observed information matrix can also be used for the calculation. Since $w = 1$ when the test parameter is the same, the critical values for the interim analyses prior to endpoint switch can be calculated without the information matrix. The observed information matrix based on data collected prior to endpoint switch can then be used to calculate the critical values when or after the endpoint is switched. The number of interim analyses and timing of the switch should depend on the duration and the size of the study, as well as the probability of success at the interim or final analyses. To avoid extra penalties, we recommend that the switch not be done in a late stage interim analysis.

Although w can be less than 0, there seems to be no further impact on the stopping boundaries in the scenarios we presented above. For a given alpha spending function, solving equation (2.1) sequentially involves finding the smallest critical values in a sequential order such that the tail probability is no larger than the value defined on the right hand side of the equation. For a fixed set of critical values in the region that we are interested in, the tail probability increases as w decreases, with negligible increases beyond $w = 0$. Thus, solving equation (2.1) will result in smaller critical values (smaller penalty) when w is larger. Recall that w is the correlation coefficient of two efficient scores $\frac{\partial l}{\partial \beta_1}$ and $\frac{\partial l}{\partial \beta_2}$, and the sign of w is determined by the off diagonal element of the Fisher information matrix, which is determined by the covariance of β_1 and β_2. When $w < 0$, the covariance of β_1 and β_2 is positive, the penalty to switch endpoints is also higher.

In this Section, we have provided stopping boundaries for different values of w. For convenience and easy comparison to classical stopping boundaries, we have assumed that the information fractions are equally spaced. When the information fraction is not equally spaced, for example, we just need to replace $\sqrt{\frac{1}{2}}$, $\sqrt{\frac{2}{3}}$ etc. in the covariance matrix (3.2) with $\sqrt{\frac{n_1}{n_2}}$, $\sqrt{\frac{n_2}{n_3}}$ etc., where n_1, n_2 and n_3 are the sample size at the 1st, 2nd, and fird analysis. The numerical method we discussed above can still be used to derive the stopping boundaries.

3.2 Non-independent Information Increments

There are scenarios where complete independent increments of information cannot be assumed. Classical group sequential designs have similar issues and there has been a large literature devoted to this topic. Many argued that the sequentially computed statistics has an asymptotically independent increment structure under usual conditions (Tsiatis et al. 1995, Scharfstein at al. 1997). In Section 3.1, we showed that

$$\text{Cov}\{Z(l), Z(k)\} = \frac{\text{E}\{S_l S_l^*\}}{\sqrt{I_l I_k^*}}. \quad (3.3)$$

If the sequentially computed score statistics have an asymptotically independent increments structure, the information fraction between the lth analysis and the kth analysis will be

asymptotically $\frac{n_l}{n_k}$. The covariance structure we discussed in Section 3.1 can be applied in time-to-event analysis. Assuming independent increments is more controversial in other settings such as sequential testing of the slope in longitudinal studies (Wei et al 1990, Su & Lachin 1992, DeMets & Lan 1994) where the total information is unknown. When the sequentially computed score statistics are not independent or asymptotically independent, (3.3) can still be expressed as

$$\mathrm{Cov}\{Z(l), Z(k)\} = \sqrt{\frac{n_l}{n_k}}\mathrm{Corr}(S_l, S_l^*),$$

and the absolute value of the correlation coefficient (Corr) of S_l and S_l^* is still bounded by 1. However, $\mathrm{Corr}(S_l, S_l^*)$ reflects both the additional contribution of the n_l subjects to the kth analysis and the correlation coefficient of the score function. We re-write $\sqrt{\frac{n_l}{n_k}}\mathrm{Corr}(S_l, S_l^*)$ as

$$\sqrt{\text{Information fraction between the analyses}} \times w,$$

where w is the correlation coefficient of the score function, as defined in Section 3.1. Information fraction between the analyses includes both incremental of information between subject and within subject. It follows that the total information estimation methods suggested in classical group sequential designs (Lan & Zucker 1993, Lan et al 1994) can then be applied in this setting. As discussed in Section 3.1, a larger w will result in smaller stopping boundaries, and there is no further impact on the boundaries when $w < 0$ compared to $w = 0$. Therefore, when $w \quad 0$, an accurate estimate of the information fraction is not needed.

4 Application to a Bivariate Survival Model

In this section, we present an application for a bivariate survival model where we are interested in testing whether time-to-event A or the first recurrence time is associated with treatment (H_0: $\beta_1 = 0$) in an earlier interim analysis, but change to testing whether time to event B or second recurrence time is associated with treatment (H_0: $\beta_2 = 0$) in a later interim, or final analysis.

We assume that the event time for the ith subject and the jth event type ($i = 1,\ldots, N, j = 1, 2$) is follows a Weibull distribution with shape parameter ν and frailty ψ_i. Thus, the hazard function of the event time of the ith subject and the jth event type, T_{ij}, is

$$\lambda_{ij}(t_{ij}) = \psi_i \nu t_{ij}^{\nu-1} \exp(\beta_0 + \beta_j x_{ij}),$$

where x_{ij} denotes the explanatory variable for subject i and the jth event type, and β_0 and β_j are the intercept and the coefficient of the explanatory variable, respectively. We consider a

model with gamma frailty ψ_i, and thus $f(\psi_i) = \frac{\theta^\theta}{\nu(\theta)} \psi_i^{\theta-1} \exp(-\theta\psi_i)$, with mean 1 and variance $\frac{1}{\theta}$. Conditional on ψ_i, the survival times are assumed to be independent. Thus, the observed-data likelihood function is given by

$$L(\theta, \beta) = \prod_{i=1}^n \int_0^\infty \prod_{j=1}^2 \left[\psi_i \nu t_{ij}^{\nu-1} \exp(\beta_0 + \beta_j x_{ij}) \exp(-\psi_i t_{ij}^\nu \exp(\beta_0 + \beta_j x_{ij})) \right]^{C_{ij}}$$
$$\times \left[\exp(-\psi_i t_{ij}^\nu \exp(\beta_0 + \beta_j x_{ij})) \right]^{1-C_{ij}} \frac{\theta^\theta}{\nu(\theta)} \psi_i^{\theta-1} \exp(-\theta\psi_i) d\psi_i, \tag{4.1}$$

where C_{ij} is the right censoring indicator (equals 0 for right censoring, 1 otherwise), and t_{ij} denotes the event time for subject i and the jth event type. After ψ_i is integrated out in (4.1), the observed-data likelihood is given by

$$L(\theta, \beta) = \prod_{i=1}^n \frac{\Gamma(\theta + D_i)}{\Gamma(\theta)} \left(\frac{\theta}{\theta + t_i^\nu \cdot (\beta)} \right)^\theta \left(\frac{\nu}{\theta + t_i^\nu \cdot (\beta)} \right)^{D_i} \times \exp\left(\sum_{j=1}^2 C_{ij}(\beta_0 + \beta_j x_{ij}) \right) \prod_{j=1}^2 t_{ij}^{(\nu-1)C_{ij}}, \tag{4.2}$$

where $D_i = \sum_{j=1}^2 C_{ij}$ and $t_i^\nu \cdot (\beta) = \sum_{j=1}^2 t_{ij}^\nu \exp(\beta_0 + \beta_j x_{ij})$.

For ease of exposition, let treatment be the only explanatory variable and therefore $x_i = x_{i1} = x_{i2}$ in this particular setting. In most confirmatory clinical trials, there are only two treatment arms. Based on the likelihood function (4.2) and using the reference cell coding for convenience ($x_i = \{0, 1\}$),

$$w = \frac{I_{12}(\beta_{01}, \beta_{02})}{\sqrt{I_{11}(\beta_{01}) I_{22}(\beta_{02})}}$$
$$= \frac{-E\left(\frac{\partial^2 l(\beta)}{\partial \beta_1 \partial \beta_2} \right)}{\sqrt{\left(-E(\frac{\partial^2 l(\beta)}{\partial \beta_1^2}) \right) \left(-E(\frac{\partial^2 l(\beta)}{\partial \beta_2^2}) \right)}} \bigg|_{\beta_1=0, \beta_2=0}$$
$$= \frac{-E[b(t_1)b(t_2)]}{\sqrt{E[(b(t_1) - b^2(t_1))] E[(b(t_2) - b^2(t_2))]}},$$

where $l(\beta) = \ln L(\theta, \beta)$ and $b(t_k) = \frac{t_k^\nu \exp(\beta_0)}{\theta + \sum_{j=1}^2 t_j^\nu \exp(\beta_0)}$. The quantities $E[b(t_1)b(t_2)]$ and $E[(b(t_1) - b^2(t_1))]$ can be solved numerically, since the joint density of the survival times, (t_1, t_2), is given by

$$f(t_1, t_2) = \frac{\Gamma(\theta+2)}{\Gamma(\theta)} \frac{\theta^\theta \exp\left(\sum_{j=1}^2 \beta_0 + \beta_j x_i \right) \nu^2 t_1^{\nu-1} t_2^{\nu-1}}{\left[\theta + \sum_{j=1}^2 \left(t_j^\nu e^{\beta_0 + \beta_j x_i} \right) \right]^{\theta+2}}.$$

However, regardless of the choice of θ, ν, or β_0, the value of w will be negative. As discussed in Section 3, the boundaries will be the same as the boundaries for the case when $w = 0$. Further calculation of w in this case will not be necessary.

5 Application to Joint Modeling of Longitudinal and Time-to-Event Data

Most time-to-event studies also collect repeated measurements of potential biomarkers. A powerful method to take into account the dependency of time-to-event data and repeated measurements of biomarkers is joint modeling of these two data types (Wulfsohn & Tsiatis 1997, Henderson et al 2000, Tsiatis & Davidian 2004). Applications of joint models in studying surrogate endpoints was particularly discussed in Taylor and Wang (2002). It has been demonstrated through simulation studies that use of joint modeling leads to correction of biases and improvement of efficiency (Hsieh et al 2006, Chen et al 2011, Ibrahim et al 2010). Since joint models contain multiple parameters that may be related to the treatment effect in a joint likelihood, this modeling situation presents a unique opportunity and advantage of testing different parameters at different interim analyses. Although joint models have not been widely accepted as the primary analysis models, we believe that there is great potential in this type of modeling and hence provide another demonstration of the advantages of the proposed methodology in this Section.

5.1 Notation

For subject i, ($i = 1,\ldots, N$), let T_i and C_i denote the event and censoring times, respectively; $S_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$. Let Z_i be a treatment indicator, and let $X_i(u)$ be the longitudinal process (also referred to as the trajectory) at time $u \ge 0$. In a more general sense, Z_i can be a q-dimensional vector of baseline covariates including treatment. To simplify the notation, Z_i denotes the binary treatment indicator in this section. Values of $X_i(u)$ are measured intermittently at times $u \le S_i$, $j = 1,\ldots, m_i$, for subject i. Let $Y(t_{ij})$ denote the observed value of $X_i(t_{ij})$ at time t_{ij}, which may be prone to measurement error.

The joint modeling approach links two sub-models, one for the longitudinal process $X_i(u)$ and one for the event time T_i, by including the trajectory in the hazard function of T_i (Taylor & Wang 2002, Chen et al 2011). Thus,

$$\lambda_i(t) = \lambda_0(t)\exp\{\beta X_i(t) + \xi Z_i\}, \quad (5.1)$$

where β measures the degree of association between the longitudinal marker and time-to-event, and ξ is the direct treatment effect on survival. A general polynomial linear model is frequently used to model the trajectory function of the longitudinal data (Wulfsohn & Tsiatis 1997, Ibrahim et al 2004), given by.

$$X_i(u) = \theta_{0i} + \theta_{1i}u + \theta_{2i}u^2 + \cdots + \theta_{pi}u^p + \gamma Z_i, \quad (5.2)$$

where $\theta_i = \{\theta_{0i}, \theta_{1i}, \ldots, \theta_{pi}\}^T$ is distributed as a multivariate normal with mean μ_θ and variance-covariance matrix Σ_θ. The parameter γ is a fixed treatment effect for the longitudinal marker, which may also have an indirect effect on survival. The observed

longitudinal measures are modeled as $Y_i(t_{ij}) = X_i(t_{ij}) + e_{ij}$, where $e_{ij} \sim N(0, \sigma_e^2)$, the $\theta_{\mathbf{i}}$'s are independent and $\mathrm{Cov}(e_{ij}, e_{ij'}) = 0$, for $j \neq j' \neq 0$. The observed-data likelihood function for subject i is given by

$$\int_{-\infty}^{\infty} \left[\prod_{j=1}^{m_j} f(Y_{ij}|\theta_{\mathbf{i}}, \gamma, \sigma_e^2) \right] f(\theta_i|\mu_\theta, \sum\nolimits_\theta) f(S_i, \Delta_i|\theta_{\mathbf{i}}, \beta, \gamma, \xi) d\theta_{\mathbf{i}} \quad (5.3)$$

(Wulfsohn & Tsiatis 1997).

In expression (5.3), the density function for the time-to-event, $f(S_i, \Delta_i|\theta_{\mathbf{i}}, \beta, \gamma, \xi)$, can be based on any model, such as the Weibull model, exponential model, or piecewise exponential model.

5.2 Motivation for Testing Different Parameters at Different Interim Analyses in Joint Models

Based on the hazard function (5.1) and the trajectory model (5.2), we can see that the overall treatment effect is $\beta\gamma + \xi$, where γ is the treatment effect on the longitudinal marker, ξ is the direct treatment effect on time-to-event, and β is the association between the longitudinal marker and time-to-event. The parameter β can also be viewed as measuring the degree of "surrogacy" between the longitudinal marker and the time-to-event. It was suggested by Taylor and Wang (2002) that the quantity $\frac{\beta\gamma}{\beta\gamma + \xi}$ represents

$$\frac{\text{treatment effect on survival through marker}}{\text{overall treatment effect on survival}},$$

which is a measure of surrogacy suggested by Freedman et al (1992). If Y_{ij} is a good surrogate, the values of β and γ will be relatively large compared to the value of ξ.

In the case of a real surrogate, directly testing the treatment effect may require substantially more subjects and may take longer to observe enough events. A natural question is whether we can test β and γ jointly. If $\beta \neq 0$ and $\gamma \neq 0$, then $\beta\gamma \neq 0$, and if $\beta\gamma$ and ξ have the same sign, which is typically the case, the overall treatment effect $\beta\gamma + \xi \neq 0$. Simulations were carried out to examine the power of testing β, γ and $\beta\gamma + \xi$ from the joint model (5.3). We simulated 1000 datasets and each dataset had 200 subjects (100 subjects per treatment group). Power was determined as the % of datasets with a p-value ≤ 0.05 from the score test for testing

$$\begin{cases} H_0\text{:} & \beta = 0 \text{ or } \gamma = 0 \\ H_1\text{:} & \beta \neq 0 \text{ and } \gamma \neq 0 \end{cases} \quad (5.4)$$

versus

$$\begin{cases} H_0: & \beta\gamma+\xi=0 \\ H_1: & \beta\gamma+\xi \neq 0 \end{cases}. \quad (5.5)$$

Rejecting H_0 in (5.4) implies rejecting H_0 in (5.5) unless the direct treatment effect on the time-to-event ξ has a completely opposite effect compared to $\beta\gamma$. Table 4 shows substantial power advantages for testing β and γ jointly instead of testing $\beta\gamma + \xi$ alone, especially when the size of ξ is relatively small.

By rejecting the null hypothesis in (5.4), we showed that there is a treatment effect on survival through the biomarker. This is different from the idea of switching endpoints because the endpoint is still survival. The number of events required to reject (5.4) is smaller than the number of events required to reject (5.5), and therefore provides a higher probability to succeed in the interim analysis. This is more evident when the marker is a good surrogate of survival.

5.3 Stopping Boundaries in a Hypothetical Design

Let $\varphi = \beta\gamma+\xi$. The likelihood function (5.3) can be reparameterized in terms of β, γ and φ by replacing ξ with $\varphi - \beta\gamma$. Let $Z_\beta(l)$, $Z_\gamma(l)$, $Z_\varphi(k)$ denote the score test statistics of β and γ at the lth analysis, and of φ at the kth analysis. Then

$$\mathrm{Cov}\,(Z_\beta(l), Z_\varphi(k)) = \sqrt{F_{lk}}\frac{I(\beta_0,\varphi_0)}{\sqrt{I(\beta_0)I(\varphi_0)}}, \text{ and } \mathrm{Cov}\,(Z_\gamma(l), Z_\varphi(k)) = \sqrt{F_{lk}}\frac{I(\gamma_0,\varphi_0)}{\sqrt{I(\gamma_0)I(\varphi_0)}},$$

where F_{lk} is the information fraction between the lth and the kth analysis, $I(\beta_0)$, $I(\gamma_0)$, $I(\varphi_0)$, $I(\beta_0, \varphi_0)$ and $I(\gamma_0, \varphi_0)$ are elements of the Fisher information matrix under the null hypothesis.

Suppose that in a study with two planned analyses, we are interested in testing hypothesis (5.4) in the interim analysis, and testing hypothesis (5.5) in the final analysis. To ensure the type I error will not exceed the planned level of 0.05 in two-sided tests, the boundary values $z_c(1)$ and $z_c(2)$ can be determined successively so that

$$P_0\left\{|Z_\beta(1)| \geq z_{c1}(1) \cup |Z_\varphi(2)| \geq z_{c1}(2)\right\} = \alpha_\beta(t_k^*) \quad (5.6)$$

$$P_0\left\{|Z_\gamma(1)| \geq z_{c2}(1) \cup |Z_\varphi(2)| \geq z_{c2}(2)\right\} = \alpha_\gamma(t_k^*). \quad (5.7)$$

Note that both (5.6) and (5.7) need to be satisfied as the parameter space under H_0 for the first interim analysis is the set $\{\beta: \beta = 0\}\cup\{\gamma: \gamma = 0\}$. The two sets in the null hypothesis parameter space can be completely disjoint.

The likelihood function of (5.3) does not have a closed form, thus a direct estimate of the Fisher information matrix will be difficult. One possible solution is to approximate the likelihood function by a Laplace approximation and obtain the approximate Fisher information matrix. Based on our simulated data, we obtained negative value of

$$\frac{I_{(n)}(\beta_0, \varphi_0)}{\sqrt{I_{(n)}(\beta_0)I_{(n)}(\varphi_0)}} \text{ and } \frac{I_{(n)}(\gamma_0, \varphi_0)}{\sqrt{I_{(n)}(\gamma_0)I_{(n)}(\varphi_0)}},$$ where $I_{(n)}$ stands for the observed information. This is expected since the correlation between φ and β (or γ) is usually positive, resulting in $w < 0$. Therefore it is fairly safe to derive a set of boundaries by assuming $w = 0$ between β (or γ) and φ. Even if $w > 0$, assuming $w = 0$ will only result in more conservative crossing boundaries and the overall type I error will be less than α.

It is possible that a different alpha spending function can be used in (5.6) and (5.7), and therefore result in different crossing boundaries for β and γ. However the second stopping boundary should be based on the maximum of $z_{c1}(2)$ and $z_{c2}(2)$ so that the overall type I error will not exceed α.

6 Discussion

In this paper, we have extended the concept of the alpha spending function to testing different endpoints between interim or between interim and final analyses. Correlations between successive test statistics not only depend on the information fraction between interim analyses, but they also depend on the Fisher information matrix. The correlation between the two parameter is inversely related to the off diagonal elements of the Fisher information matrix. Thus, when the correlation coefficient of the score function (w) is positive, the two parameters have negative association, and the additional penalty to pay for switching endpoints is smaller. When w is negative, the two parameters are positively associated, and the additional penalty for switching endpoints is larger.

The general covariance structure of the sequential score test statistics in this paper was developed based on the independent increments of information assumption. The proposed methodology can be applied to any parameter test statistics where the assumption is satisfied asymptotically, which applies to most test statistics (Tsiatis et al 1995, Scharfstein et al. 1997). When the assumption is violated, the covariance structure of the sequential test statistics will be more complex, although the alpha spending function can still be applied as long as the covariance is known or can be numerically calculated based on a known likelihood function. Therefore, the assumption is not necessary for changing hypotheses in general. Furthermore, assuming independence of the test statistics ($w = 0$) returns the most conservative stopping boundaries. When $w = 0$, the independent increments of information assumption is no longer important.

In the case where $w = 0$ and there are only two analyses, the critical values based on the method discussed in this paper will be equivalent to the ones based on a simple Bonferroni adjustment. When there are more than 2 analyses, the method will return smaller critical values even when $w = 0$, as it takes into account the correlation between the test statistics when the parameter is the same. The simple Bonferroni method assumes complete independence of $(Z(1),\ldots, Z(k))$, that is, $w = 0$ for all off diagonal elements in the covariance matrix. This will return the largest critical values.

Sophisticated parametric models have not been widely used in late stage clinical trials. It is recognized that despite the many promising innovations in bio-markers to measure drug

effect and disease progression, little has changed in the style of drug development over the past 30 years. The Food and Drug Administration (FDA) Pharmacometrics Group released two advisory meeting proceedings in 2006 and 2008 concerning disease-drug-trial models, with the objective of improving the success rate of late stage drug development (FDA 2006, FDA 2008). We believe that sophisticated parametric models that give better power to detect treatment differences will become more popular in late phase clinical trials and will be more acceptable to FDA reviewers.

The application to joint modeling of longitudinal and time-to-event data is promising in the case that the longitudinal marker may be a good surrogate for the time-to-event. In reality, time-to-event data, such as overall survival, may be lengthy to obtain. With recent advances in genetic research and other biomarker research, many potential surrogates are being identified. For example, Circulating tumor cells (CTC's) have been found to be associated with progression-free survival and overall survival in patients with metastatic breast cancer (Dawood et al 2008, Liu et al 2009). However, there is also great uncertainty in terms of how good these "surrogates" are. By showing both $\beta \quad 0$ and $\gamma \quad 0$, this not only demonstrates that the longitudinal marker is a strong potential surrogate, it also demonstrates that there is a treatment effect on survival through the longitudinal marker. This may be considered sufficient efficacy evidence to terminate the trial early. If the longitudinal marker is a weak surrogate, this will allow the investigator to proceed to the next analysis to test the hypothesis that the overall treatment effect, $\beta\gamma + \xi$, is not 0. Note that if we know whether the longitudinal marker is or is not a good surrogate, testing the same parameter would be more powerful as it will not result in any extra penalty.

As discussed by Fleming and DeMets (1993), early termination of a clinical trial is a complex process and cannot be simply reached by pre-specified stopping rules. For example, even when the efficacy stopping boundary is crossed, a trial may need to be continued to collect sufficient safety information. In this paper, we simply provide a tool to facilitate the decision process and ensure that the type I error will be strictly under control to its pre-specified level when testing different endpoints at different interim analyses. Furthermore, we are not advocating that when a particular endpoint is not performing well in terms of significance at a particular interim analysis, we can just switch to another endpoint at a subsequent interim analysis. The hypothesis tests should be pre-specified before the trial begins. If the analyses are based on different likelihood functions, or the value of w cannot be reliably estimated, w should be set to 0, assuming no correlation between the test statistics.

References

Armitage P, McPherson CK, Rowe BC. Repeated significance tests on accumulating data. Journal of the Royal Statistical Society. 1969; 132:232–244.

Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials. Journal of the American Medical Association. 2004; 291(20):2457–2465. [PubMed: 15161896]

Charbonnel B, Dormandy J, Erdmann E, Massi-Benedetti M, Skene A. The prospective pioglitazone clinical trial in macrovascular events (PROactive): Can pioglitazone reduce cardiovascular events in diabetes? Study design and baseline characteristics of 5238 patients. Diabetes Care. 2004; 27:16471653.

Chen LM, Ibrahim JG, Chu H. Sample Size and Power Determination in Joint Modeling of Longitudinal and Survival Data. Statistics in Medicine. 2011; 30:2295–2309. [PubMed: 21590793]

Chen YHJ, DeMets DL, Lan KKG. Monitoring mortality at interim analyses while testing a composite endpoint at the final analysis. Controlled Clinical Trials. 2003; 24:16–27. [PubMed: 12559639]

Dawood S, Broglio K, Valero V, et al. Circulating tumor cells in metastatic breast cancer: from prognostic stratification to modification of the staging system? Cancer. 2008; 113(9):2422–2430. [PubMed: 18785255]

DeMets DL, Lan KKG. Interim analysis: the alpha spending function approach. Statistics in Medicine. 1994; 13:1341–1352. [PubMed: 7973215]

Dormandy J, Charbonnel B, Eckland D, Erdmann E, Massi-Benedetti M, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive study (PROspective pioglitAzone Clinical Trial In macroVascular Events): A randomized controlled trial. Lancet. 2006; 366:12791289.

European Medicines Agency. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. 2007. http://www.ema.europa.eu/ema/pages/includes/document/open_document.jsp?webContentId=WC500003616

Fleming TR, DeMets DL. Monitoring of clinical trials: issues and recommendations. Control Clinical Trials. 1993; 14:183197.

Food and Drug Administration. Proceedings of the Clinical Pharmacology Sub-Committee Advisory Committee Meeting. 2006. http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm180485.htm

Food and Drug Administration. Proceedings of the Clinical Pharmacology Sub-Committee Advisory Committee Meeting. 2008. http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm180485.htm

Food and Drug Administration. Guidance for Industry - Adaptive design clinical trials for drugs and biologics. 2010. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm201790.pdf

Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic disease. Statistics in Medicine. 1992; 11:167178.

Genz A. Numerical Computation of Multivariate Normal Probabilities. J Comp Graph Stat. 1992; 1:141–149.

Goldman B, LeBlanc M, Crowley J. Interim futility analysis with intermediate endpoints. Clinical Trials. 2008; 5:14–22. [PubMed: 18283075]

Hallstrom, A. Is survival the only or even the right outcome for evaluating treatments for out-of-hospital cardiac arrest? A proposed test based on both an intermediate and ultimate outcome. UW Biostatistics Working Paper Series. 2009. Working Paper 352. http://www.bepress.com/uwbiostat/paper352

Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics. 2000; 1:465–480. [PubMed: 12933568]

Hsieh F, Tseng YK, Wand JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. Biometrics. 2006; 62:1037–1043. [PubMed: 17156277]

Ibrahim JG, Chen MH, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. Statistica Sinica. 2004; 14:863–883.

Ibrahim JG, Chu H, Chen LM. Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. Journal of Clinical Oncology. 2010; 28(16):2796–2808. [PubMed: 20439643]

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983; 70:659–663.

Lan KKG, Zucker D. Sequential monitoring of clinical trials: The role of information in Brownian motion. Statistics in Medicine. 1993; 12:753–765. [PubMed: 8516592]

Lan KKG, Reboussin DM, DeMets DL. Information and information fractions for design and sequential monitoring of clinical trials. Communications in Statistics-Theory and Methods. 1994; 23(2):403–420.

Lee JW, DeMets DL. Sequential comparison of change with repeated measurement data. Journal of the American Statistical Association. 1991; 86:757–762.

Liu MC, Shields PG, Warren RD, et al. Circulating tumor cells: a useful predictor of treatment efficacy in metastatic breast cancer. Journal of Clinical Oncology. 2009; 27(31):5153–5159. [PubMed: 19752342]

MERIT-HF Study Group. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL randomized intervention trial in congestive heart failure (MERIT-HF). Lancet. 1999; 353:20012007.

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979; 35:549–556. [PubMed: 497341]

Olmos D, Arkenau HT, Ang JE, Ledaki I, Attard G, Carden CP, Reid AHM, A'Hern R, Fong PC, et al. Circulating tumour cell (CTC) counts as intermediate end points in castration-resistant prostate cancer (CRPC): a single-centre experience. Annals of Oncology. 2009; 20:27–33. [PubMed: 18695026]

Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977; 64:191–199.

Scharfstein DO, Tsiatis AA, Robins JM. Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. Journal of the American Statistical Association. 1997; 92:1342–1350.

Scher HI, Jia X, Bono JS, Fleisher M, Pienta KJ, Raghavan D, Heller G. Circulating tumour cells as prognostic markers in progressive, castration-resistant prostate cancer: a reanalysis of IMMC38 trial data. The Lancet Oncology. 2009; 10:233– 239. [PubMed: 19213602]

Su JQ, Lachin JU. Group sequential distribution-free methods for the analysis of multivariate observations. Biometrics. 1992; 48:1033–1042. [PubMed: 1290797]

Taylor JMG, Wang Y. Surrogate markers and joint models for longitudinal and survival data. Controlled Clinical Trials. 2002; 23:626–634. [PubMed: 12505241]

Tsiatis AA, Boucher H, Kim K. Sequential methods for parametric survival model. Biometrika. 1995; 82:165–173.

Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: and overview. Statistica Sinica. 2004; 14:809–834.

Wei LJ, Su JQ, Lachin JM. Interim analyses with repeated measurements in a sequential clinical trial. Biometrika. 1990; 77:359–364.

Wu MC, Lan KKG. Sequential monitoring for comparison of changes in a response variable in clinical trials. Biometrics. 1992; 48:765–779. [PubMed: 1420840]

Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997; 53:330–339. [PubMed: 9147598]

Table 1

One-sided Boundaries for Different Values of w with $\alpha = 0.025$ and $K = 5$ (The test parameter is assumed to be β_1 for $j = 1, 2, \beta_2$ for $j = 3, 4, 5$, where $j = 1, \ldots, 5, t_j^* = j/5$.)

w	O'Brien-Fleming Like Alpha Spending Function $\alpha_1(t^*)$					Pocock Like Alpha Spending Function $\alpha_2(t^*)$				
	$z_c(1)$	$z_c(2)$	$z_c(3)$	$z_c(4)$	$z_c(5)$	$z_c(1)$	$z_c(2)$	$z_c(3)$	$z_c(4)$	$z_c(5)$
1	4.88	3.36	2.68	2.29	2.03	2.44	2.42	2.41	2.40	2.39
0.8	4.88	3.36	2.69	2.29	2.03	2.44	2.42	2.50	2.43	2.42
0.5	4.88	3.36	2.70	2.30	2.03	2.44	2.42	2.57	2.46	2.44
0	4.88	3.36	2.70	2.30	2.03	2.44	2.42	2.60	2.50	2.45
−0.5	4.88	3.36	2.70	2.30	2.03	2.44	2.42	2.60	2.50	2.45
−0.7	4.88	3.36	2.70	2.30	2.03	2.44	2.42	2.60	2.50	2.45

Table 2

One-sided Boundaries for the 5th Analysis $z_c(5)$ When $\alpha = 0.025$ and $K = 5$ (The test parameter is assumed to be β_1 for $j = 1 - 4$, β_2 for $j = 5$, where $j = 1$, ..., 5, $t_j^* = j/5$.)

Alpha Spending Function	w					
	1	**0.8**	**0.5**	**0**	**−0.5**	**−0.7**
O'Brien-Fleming Like Function, $a_1(t^*)$	2.03	2.13	2.19	2.23	2.23	2.23
Pocock Like Function, $a_2(t^*)$	2.39	2.54	2.64	2.70	2.70	2.70

Table 3

One-sided Boundaries for Different Values of w with $\alpha = 0.025$, $K = 2$ and $t_j^* = j/2$.)

	O'Brien-Fleming Like Alpha Spending Function $\alpha_1(t^*)$		Pocock Like Alpha Spending Function $\alpha_2(t^*)$	
w	$z_c(1)$	$z_c(2)$	$z_c(1)$	$z_c(2)$
1	2.96	1.97	2.16	2.20
0.8	2.96	1.98	2.16	2.25
0.5	2.96	1.98	2.16	2.30
0	2.96	1.99	2.16	2.34
−0.5	2.96	1.99	2.16	2.34
−0.8	2.96	1.99	2.16	2.34
−1	2.96	1.99	2.16	2.34

Table 4

Comparison of Power for Testing $\{\beta = 0 \text{ or } \gamma = 0\}$, and $\beta\gamma + \xi = 0$ from the Joint Model

β	γ	ξ	$H_0: \beta = 0 \text{ or } \gamma = 0$					$H_0: \beta\gamma + \xi = 0$	
			β Estimate (SE)	γ Estimate (SE)	Power			Estimate (SE)	Power
0.2	0.25	0.05	0.210 (0.056)	0.250 (0.130)	46.6%			0.095 (0.170)	9.7%
0.2	0.15	0.15	0.210 (0.056)	0.149 (0.130)	25.6%			0.176 (0.168)	18.8%

a The event time was simulated from an exponential distribution with $\lambda_i(t) = \lambda_0 \exp\{\beta X_i(t) + \xi Z_i\}$, where $X_i(t) = \theta_{0i} + \theta_{1i} t + \gamma Z_i\}$ and $\lambda_0 = 0.85$. To ensure a minimum follow-up time of 0.75 years (9 months) and maximum follow-up time of 2 years, right censoring was generated from a uniform [0.75, 2] distribution. The $(\theta_{0i}, \theta_{1i})$ were assumed to follow a bivariate normal distribution with

$$\mu_\theta = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \text{ and } \Sigma_\theta = \begin{pmatrix} 1.2 & 0 \\ 0 & 0.7 \end{pmatrix}.$$