Switching endpoint in clinical trials with simple and composite time-to-event outcomes

Joshua Y.H. Chen

David L. DeMets, Ph.D.
Switching Endpoint in Clinical Trials with Simple and Composite Time-to-event Outcomes

Y.H. Josh Chen  
Department of Biostatistics and Medical Informatics  
University of Wisconsin-Madison  
ychen@biostat.wisc.edu

David L. DeMets  
Department of Biostatistics and Medical Informatics  
University of Wisconsin-Madison

K.K. Gordon Lan  
Mathematical and Statistical Science  
Pfizer Central Research
1 Introduction

1.1 Multiple Endpoints

In some clinical trials, there are many key outcome measures which need to be taken into consideration. No single outcome can be identified as the primary outcome. O’Brien(1984), Pocock, Geller and Tsiatis(1987) formulated this as a global testing problem in which all the endpoints are combined to form a new single global test statistic. Monitoring this univariate global statistic is straightforward. A global testing procedure assumes that the active treatment affects all the outcomes in the same direction which may not be true. It may also be difficult to interpret the results from a combined global test statistic. A multivariate method, which is based on performing several multiplicity-adjusted univariate analyses, is often more attractive as it allows one to address several individual hypotheses separately. Interim analyses in this setting are discussed in several papers as, for example, in Cook(1996), Jennison and Turnbull(1993), Williams(1996). Jennison and Turnbull(1993) and Cook(1996) studied the interim analyses of bivariate endpoints (safety and efficacy outcomes). Williams(1996) considered several survival endpoints in a group sequential setting.

1.2 Simple versus Composite Endpoints

The above multivariate methods are appropriate if all of these endpoints are equally important. However, in some clinical trials, the endpoints of interest may have a natural ordering. For example, mortality is a clinical relevant endpoint for acute diseases and often would take precedence over any other outcome. Due to some practical reasons, a composite outcome such as death plus disease occurrence (or recurrence) or death plus hospitalization may also be considered. In some clinical trials, the composite outcome is even considered as the primary outcome.

1.2.1 Scenario 1

Since it may take a long time or a very large sample size to obtain enough information or events for a mortality outcome (simple outcome), a composite outcome is often used as the primary outcome. Investigators believe that the treatment effect for each of the several components of the composite outcome is in the same direction and the trial would be more powerful if the composite endpoint is used as the primary endpoint.

For such a clinical trial, investigators usually do not expect the trial to be powerful enough to claim a benefit in mortality reduction based on the preliminary results from some early studies. This is the major reason why a composite endpoint is used as the primary endpoint. Should we spend alpha on the mortality endpoint which is not expected to show significance? In the COREG study(Packer et al(1996)), the mortality was not a primary or
a secondary endpoint and the treatment effect in mortality reduction was not expected to be statistically significant. When the mortality benefit was shown to be significant during the study, it raised a debate about the appropriateness of claiming treatment benefit in mortality reduction while the mortality analysis was not prospectively defined in the protocol (Fisher(1999), Moye(1999)). In addition, the mortality outcome can be viewed as a safety outcome and should almost always be included in the interim analyses for ethical purposes.

Regulatory agencies as well as investigators are often reluctant to terminate a trial early for a composite endpoint, which includes softer or less well defined components. Regulatory agencies may recommend or even "require" that a trial not be stopped early unless a hard outcome such as mortality becomes convincing. In practice, a composite event also needs to be adjudicated by an independent Endpoint Committee so that the confirmation of the composite outcome often lags behind the ascertainment of the simple outcome such as death. Thus, in practice the Data and Safety Monitoring Board (DSMB) has to place more emphasis on the simple outcome, even though the trial is properly powered for the composite outcome. The trial would be stopped early if this simple mortality outcome shows a significant mortality benefit or severe harm at an interim analysis, while the primary or composite outcome will be tested at the final analysis if the trial continues to the end.

This switching-endpoint scenario is more prevalent in recent years. A recent example is the WIZARD trial in which the treatment efficacy for mortality reduction was repeatedly tested at the interim analyses while the composite endpoint was to be tested at the final analysis. Since the trial was not expected to show a statistically significant mortality benefit, only a little alpha (0.001 out of 0.05) was spent on the mortality endpoint.

1.2.2 Scenario 2

In some other clinical trials, the mortality outcome is used as the primary endpoint and a composite outcome is the secondary endpoint. If the mortality endpoint fails to show significance, the secondary composite outcome may consequently become the most relevant issue. It might be ethically difficult to ignore a compelling result about the secondary composite endpoint even the primary mortality endpoint is not statistically significant. However, if all of the nominal alpha has already been spent on the primary mortality outcome but the primary endpoint fails to show significance, claiming a treatment benefit due to the secondary composite outcome may raise a concern about the inflation of the false positive rate. A proposed design is to spend most of the nominal alpha on the primary mortality endpoint and save the rest for the secondary composite outcome, in case the mortality endpoint fails to show significance. A recent cardiovascular study, MERIT-HF 1999, employed this design. As mush alpha as 0.04 out of 0.05 was spent on the primary outcome, all-cause death, and the rest 0.01 was saved for the composite endpoint, all-cause death plus all-cause hospitalization. The primary mortality outcome would be tested at each analysis. This trial was stopped early due to mortality benefit. As in the protocol, if the mortality endpoint had
failed to show significance through the final analysis, the composite endpoint would be tested at the level 0.01.

In this article, we will introduce two so-called "switching-endpoint" procedures in which Lan-DeMets alpha spending function group sequential methods are applied to monitor the simple outcome (mortality) and the composite endpoint is tested at the final analysis. In section 2, Lin's(1991) results about the asymptotic normality of a class of general linear rank statistics and their consistent covariance estimators are extended to a weaker hypothesis which assumes the marginal survival distributions of the simple and composite events are identical for the control and treatment arms. Two switching-endpoint procedures are proposed in section 3 and the estimation of information fractions is discussed in section 4. We use simulated trials to investigate the two switching-endpoint procedures in section 5 and a real clinical trial PRAISE-I is used to illustrate the methods in section 6. Some discussion and recommendations are in section 7.

2 Distributions Under Null Hypotheses

2.1 A Generalization of Lin(1991)'s Result

When multiple time-to-event endpoints are considered in non-informative censoring setting, Wei and Lachin(1984) derive the asymptotic multivariate normal distribution for these linear rank statistics under the null hypothesis that the joint distributions of these endpoints are identical for the control and the treatment arms. This work was extended to the group sequential setting by Lin(1991). In this part, we will derive the asymptotic distribution for these linear rank statistics under the null hypothesis that these marginal survival curves are identical for the control and the treatment arms. This hypothesis is weaker than that in Wei and Lachin(1984) and Lin(1991), and it is more natural in our switching endpoint scenario.

We follow the notation of Lin(1991) and Tsiatis(1982). For \( i = 1, \ldots, n \), let \( Z_i \) be the indicator of treatment group: \( Z_i = 0 \) if patient \( i \) comes from the control group and \( Z_i = 1 \) if he/she comes from the treatment group. Let \( Y_i \) be the entry time for the \( i \)th patient. There are \( L \) time-to-event endpoints to be considered for each patient. For \( l = 1, \ldots, L \), let \( V_{li} \) be the time from entry to the \( l \)th type of event for patient \( i \), and let \( C_{li} \) denote the time from entry to censoring with respect to the \( l \)th type of event for patient \( i \). It is assumed that the entry time \( Y_i \), the failure time vector \( V_i = (V_{i1}, \ldots, V_{iL}) \) and the censoring time vector \( C_i = (C_{i1}, \ldots, C_{iL}) \) are independent conditional on \( Z_i \). Further, \( (Z_i, Y_i, V_i, C_i) \) \( (i = 1, \ldots, n) \) are i.i.d vectors. There are \( L \) marginal hypotheses to be considered and we let \( H_l \) be the hypothesis that the marginal distributions of the \( l \)th failure time are identical for the control and treatment arms, i.e. \( H_l : F_i^{(0)}(x) = F_i^{(1)}(x) \), where \( F_i^{(r)}(x) \) is the cumulate distribution
function for the $l$th time-to-event endpoint in the treatment group $Z = r$ and it is assumed to be absolutely continuous. The overall hypothesis is $H = \{H_1, ..., H_L\}$ which assumes that the control and treatment arms have identical marginal survival distributions for the $L$ outcomes.

Suppose the data are reviewed at a calendar time $T$. From patient $i$, the observation is $(Z_i, Y_i, X_i(T), \Delta_i(T))$, where $X_i(T) = (X_{i1}(T), ..., X_{il}(T))$ is the vector of observed or censored times and $\Delta_i(T) = (\Delta_{i1}(T), ..., \Delta_{il}(T))$ is the vector of censoring indicator, i.e.

$$X_{il}(T) = \max(\min(T - Y_i, V_{il}, C_{il}), 0)$$

$$\Delta_{il}(T) = 1 \text{ if } X_{il}(T) = V_{il}; \ 0 \text{ otherwise.}$$

The general linear rank statistic for testing $H_l$ at calendar time $T$ is defined as

$$U_l(T) = \sum_{i=1}^{n} \Delta_{il}(T)Q_l(T, X_{il}(T))\{Z_i - \frac{Y_l^{(1)}(T, X_{il}(T))}{Y_l^{(0)}(T, X_{il}(T)) + Y_l^{(1)}(T, X_{il}(T))}\}, \hspace{1cm} (1)$$

where, $Q_l(T, x)$ is a uniformly bounded non-negative predictable process with respect to $x$ and converges uniformly on any closed interval in probability to a bounded function $q_l(T, x)$. At different calendar time $T$ and for different endpoint $l$, this weight function could be different. If $Q_l(T, x) = 1$, the linear rank statistic $U_l(T)$ is the logrank statistic. Other choices of the weight function may lead to different test statistics such as Wilcoxon statistic and Gehan statistic. For outcome $l$ and at time $T$, $Y_l^{(j)}(T, x)$ is defined as the number of patients in group $j$ who are at risk at time $x$, i.e. for $j = 0, 1$,

$$Y_l^{(j)}(T, x) = \sum_{i=1}^{n} I(X_{il}(T) \geq x, \text{ and } Z_i = j),$$

where, $I(\cdot)$ is the indicator function.

**Result 2.1.** Suppose the data are reviewed at the calendar times $(T_1, ..., T_K)$. Under the overall hypothesis $H = \{H_1, ..., H_L\}$, the random vector

$$n^{-1/2}(U_1(T_1), ..., U_L(T_1), U_1(T_2), ..., U_L(T_2), ..., U_1(T_K), ..., U_L(T_K))$$

is asymptotically normally distributed with mean vector zero and covariance matrix $\Sigma$ with entry $\sigma_{lm}(k, j)$ as the covariance between the $l$th outcome which is reviewed at time $T_k$ and the $m$th outcome which is reviewed at time $T_j$. 

4
This result is similar to Lin(1991) except that the asymptotic normality is derived under the marginal hypotheses $H = \{H_1, ..., H_L\}$. The proof is the same as in Lin(1991) and Wei and Lachin(1984) since only the marginal hypotheses are needed in their proof. The proof is given in Appendix.

**Result 2.2.** Under $H = \{H_1, ..., H_L\}$, the covariance $\sigma_{lm}(k, j)$ can be consistently estimated by,

\[
\hat{\sigma}_{lm}(k, j) = \frac{1}{n} \sum_{i=1}^{n} \hat{W}_{li}(T_k) \hat{W}_{mj}(T_j),
\]

where,

\[
\hat{W}_{li}(T_k) = \Delta_{li}(T_k) Q_l(T_k, X_{li}(T_k)) \left\{ Z_i - \frac{Y_{l(1)}(T_k, X_{li}(T_k))}{Y_{l(0)}(T_k, X_{li}(T_k)) + Y_{l(1)}(T_k, X_{li}(T_k))} \right\}
\]

\[
- \sum_{j=1}^{n} \frac{\Delta_{lj}(T_k) Q_l(T_k, X_{lj}(T_k)) I(X_{lj}(T_k) \leq X_{li}(T_k))}{Y_{l(0)}(T_k, X_{lj}(T_k)) + Y_{l(1)}(T_k, X_{lj}(T_k))} \left\{ Z_i - \frac{Y_{l(1)}(T_k, X_{lj}(T_k))}{Y_{l(0)}(T_k, X_{lj}(T_k)) + Y_{l(1)}(T_k, X_{lj}(T_k))} \right\}
\]

This estimator is the same as that in Lin(1991). The cumulate hazard function $\Lambda_l(x)$ for the $l$th endpoint is estimated from all the patients in both the control and treatment groups. In Wei and Lachin(1984), $\Lambda_l(x)$ is estimated separately for the control and treatment groups. The consistency of this estimator can be established by the same argument as in Wei and Lachin(1984).

### 2.2 Application to Simple and Composite Endpoints

For our purposes, we consider the case where mortality is the simple outcome and death plus hospitalization is the composite outcome. Our results can be naturally extended to situations where the composite outcome has many other or alternative components. The hypotheses of interests are:

- **H$_1$**: control and treatment have identical distributions for the simple outcome
- **H$_2$**: control and treatment have identical distributions for the composite outcome

The corresponding alternative hypotheses are that the distributions for the composite and simple outcomes are stochastically ordered. Suppose there are $K$ analyses which are
performed at calendar times $T_1, \ldots, T_K$ respectively and the simple endpoint is tested at each of the $K$ analyses while the composite endpoint will be tested only at the final analysis. In survival analysis, logrank statistic is commonly used for two-group comparisons of survival curves. The general linear rank statistic $U_1(T)$ defined in section 2.1 is the logrank statistic if the weight function $Q_1(T, x) = 1$. Suppose this is a randomized, balanced and double blinded clinical trial. From Tsiatis(1982), the variance of a logrank statistic can be consistently estimated by \( \{\text{number of events}\}/4 \). Let $R_{1,k}$ be the normalized logrank statistic for the simple endpoint at the $k$th analysis ($k = 1, \ldots, K$) and $R_{2,K}$ denote the normalized logrank statistic for the composite endpoint at the final analysis. The normalized logrank statistics are defined as follows:

$$
R_{1,k} = \frac{2}{\sqrt{d_k}} \sum_{i=1}^{n} \Delta_{1,i}(T_k)\{Z_i - 1/2\}, \quad k = 1, \ldots, K,
$$

$$
R_{2,K} = \frac{2}{\sqrt{s}} \sum_{i=1}^{n} \Delta_{2,i}(T_K)\{Z_i - 1/2\},
$$

where, $d_k$ is the observed number of deaths at the $k$th analysis and $s$ is the number of composite events at the end of the study. The censoring indicators $\Delta_{1,i}(T_k)$ for the simple endpoint and $\Delta_{2,i}(T_K)$ for the composite outcome are defined in section 2.1. The treatment indicator $Z_i$ is 0 if the $i$th patient is the control group and 1 if he/she is in the active treatment group. For Scenario 1, the corresponding test statistics are $(R_{1,1}, \ldots, R_{1,K-1}, R_{2,K})$ since the mortality endpoint will not be tested at the final analysis; for Scenario 2, the statistics are $(R_{1,1}, \ldots, R_{1,K}, R_{2,K})$. We will consider the latter case since it is more general.

**Corollary 2.1.** Under the overall hypothesis $\{H_1, H_2\}$ that the control and treatment have identical marginal distributions for the simple and the composite outcomes, the normalized logrank statistics $(R_{1,1}, \ldots, R_{1,K}, R_{2,K})$ asymptotically have a multivariate normal distribution with mean vector $\theta$ and correlation matrix $\Sigma$. A consistent estimate of the correlation matrix $\Sigma$ is,

$$
\hat{\Sigma} = \begin{pmatrix}
1 & \sqrt{\frac{d_1}{d_2}} & \sqrt{\frac{d_1}{d_3}} & \cdots & \sqrt{\frac{d_1}{d_{K-1}}} & \sqrt{\frac{d_1}{d_K}} & \nu_1 \\
1 & \sqrt{\frac{d_2}{d_3}} & \sqrt{\frac{d_2}{d_4}} & \cdots & \sqrt{\frac{d_2}{d_{K-1}}} & \sqrt{\frac{d_2}{d_K}} & \nu_2 \\
1 & \sqrt{\frac{d_3}{d_4}} & \sqrt{\frac{d_3}{d_5}} & \cdots & \sqrt{\frac{d_3}{d_{K-1}}} & \sqrt{\frac{d_3}{d_K}} & \nu_3 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\
1 & \sqrt{\frac{d_{K-1}}{d_K}} & \sqrt{\frac{d_{K-1}}{d_K}} & \cdots & \sqrt{\frac{d_{K-1}}{d_K}} & \sqrt{\frac{d_{K-1}}{d_K}} & \nu_{K-1} \\
1 & 1 & 1 & \cdots & 1 & 1 & \nu_K
\end{pmatrix},
$$

where, $\nu = (\nu_1, \ldots, \nu_K)'$ is the estimated correlation vector between $(R_{1,1}, \ldots, R_{1,K})$ and $R_{2,K}$. For $k = 1, \ldots, K$,
\[ \nu_k = \left( 1 / \sqrt{d_k} \cdot s \right) \left\{ \sum_{i=1}^{n} \Delta_{1i}(T_k) \Delta_{2i}(T_K) \right. \\
- \int_{0}^{\infty} \left\{ \sum_{i=1}^{n} I(\Delta_{1i}(T_k) = 1, X_{2i}(T_K) \geq x) \right\} d\hat{\Lambda}_2(x) \\
- \int_{0}^{\infty} \left\{ \sum_{i=1}^{n} I(\Delta_{2i}(T_K) = 1, X_{1i}(T_k) \geq x) \right\} d\hat{\Lambda}_1(x) \\
+ \int_{0}^{\infty} \int_{0}^{\infty} \left\{ \sum_{i=1}^{n} I(X_{1i}(T_k) \geq x, X_{2i}(T_K) \geq y) \right\} d\hat{\Lambda}_1(x) d\hat{\Lambda}_2(y) \right\}, \]

where, \( \hat{\Lambda}_1(x) \) and \( \hat{\Lambda}_2(x) \) are respectively the empirical estimates of the cumulative hazard functions for the simple and composite endpoints. The observed times \( X_{1i}(T_k) \) for the simple outcome and \( X_{2i}(T_K) \) for the composite outcome are defined in section 2.1.

The asymptotic normality is obtained from Result 2.1 and the consistency of the correlation estimator is concluded from Result 2.2 and Tsiatis(1982) Theorem 2. The correlation estimate \( \nu_k \) indicates that the correlation matrix does not have an independent increment structure.

### 3 Switching-endedpoint Design

Assume that two-sided tests are used for both outcomes. We want to spend a part of the nominal \( \alpha \) in monitoring the mortality outcome. If a significant benefit or harm in survival has been established, we may terminate the trial early; if not, we will test the composite outcome at the end of the trial as scheduled. Let \( \alpha \) be the prespecified nominal type I error level and \( \alpha_1 < \alpha \) be the maximum alpha we want to spend on the simple mortality endpoint. In Scenario 1, to preserve the power of detecting the difference in the primary composite endpoint, \( \alpha_1 \) is usually much smaller than the nominal \( \alpha \). For example, in WIZARD trial, \( \alpha_1 = 0.001 \) which is much smaller than the nominal 0.05. In Scenario 2 where the mortality is the primary endpoint, \( \alpha_1 \) is typically very close to the nominal \( \alpha \). In MERIT-HF study, \( \alpha_1 = 0.04 \) which is close to the nominal level 0.05.

#### 3.1 Monitoring the Simple Endpoint

An alpha spending function \( \alpha^*_1(t) \) with \( \alpha^*_1(1) = \alpha_1 \) is applied to monitor the simple outcome as described by Lan and DeMets(1983). Suppose the information fraction \( t_k \) at the \( k \)th analysis can be estimated and is independent of the test statistics. The estimation of the information fraction will be discussed in section 4. Then as much alpha as \( \alpha^*_1(t_k) - \alpha^*_1(t_{k-1}) \) will be spent at the \( k \)th interim analysis.
3.1.1 Scenario 1

Since the mortality outcome will not be considered at the final analysis, the sequential boundaries \((b_1, \ldots, b_{K-1})\) for the normalized logrank statistics \((R_{1,1}, \ldots, R_{1,K})\) are calculated recursively by:

\[
\Pr(|R_{1,l}| \geq b_l | H_1) = \alpha_1^*(t_1) \\
\Pr(|R_{1,k}| < b_l, \text{ for all } l = 1, \ldots, k - 1) \cap \{ |R_{1,k} | \geq b_k | H_1 \} = \alpha_1^*(t_k) - \alpha_1^*(t_{k-1}) \\
k = 2, \ldots, K - 1
\]

From Corollary 2.1, \((R_{1,1}, \ldots, R_{1,K-1})\) have a multivariate normal distribution with mean 0 and covariance matrix with an independent increment structure. The covariance between \(R_{1,k}\) and \(R_{1,l}\) can be estimated by \(\sqrt{d_k/d_l} (k < l)\), where \(d_k\) is the number of observed deaths at the \(k\)th analysis. The mortality outcome, which is sometimes called the hard outcome, can be easily verified and \(d_k\) can be obtained quickly. The boundaries \((b_1, \ldots, b_{K-1})\) calculated by the above method are dependent on information fractions \((t_1, \ldots, t_{K-1})\) and the observed number of deaths \((d_1, \ldots, d_{K-1})\). The information times specify the fraction of \(\alpha_1\) to be spent at each interim analysis; the observed numbers of deaths \((d_1, \ldots, d_{K-1})\) are used to construct the covariance matrix. The estimation of the information time will be discussed in section 4. After we have obtained the boundaries \((b_1, \ldots, b_{K-1})\), we are able to define the acceptance region of \(H_1\) before the final analysis for Scenario 1 as

\[
A_{1,\alpha_1} = \{|R_{1,k}| < b_k, \text{ for all } k = 1, \ldots, K - 1\}
\]

Let \(A_{1,\alpha_1}^c\) be the completion of \(A_{1,\alpha_1}\). \(A_{1,\alpha_1}^c\) is then the rejection region of \(H_1\) and it is clear that \(\Pr(A_{1,\alpha_1}^c | H_1) = \alpha_1^*(t_{K-1})\) which is the false positive rate for the simple outcome. Since the information fraction \(t_{K-1} \leq 1\), the alpha which is actually spent on the simple outcome is \(\alpha_1^*(t_{K-1})\), which is no more than \(\alpha_1\). If we want to spend all \(\alpha_1\) on the simple outcome, we may need to standardize the estimated information fractions such that \(t_{K-1} = 1\) and then \(\alpha_1^*(t_{K-1}) = \alpha_1\). Otherwise, the actual alpha spent on the simple endpoint is strictly less than \(\alpha_1\).

3.1.2 Scenario 2

The calculations of the efficacy boundaries \((b_1, \ldots, b_{K-1}, b_K)\) for the normalized logrank statistics \((R_{1,1}, \ldots, R_{1,K-1}, R_{1,K})\) are the same except that the simple endpoint will be tested at the final analysis. The acceptance region of \(H_1\) for Scenario 2 is,

\[
A_{2,\alpha_1} = \{|R_{1,k}| < b_k, \text{ for all } k = 1, \ldots, K\}
\]

Let \(A_{2,\alpha_1}^c\) be the completion of \(A_{2,\alpha_1}\) and \(\Pr(A_{2,\alpha_1}^c | H_1) = \alpha_1\). The false positive rate for \(H_1\) is exactly \(\alpha_1\).
3.2 Switching to the Composite Endpoint

If the trial continues to the final analysis as in Scenario 1, or the test for mortality benefit fails at the final analysis as in Scenario 2, we will test the composite outcome. Two methods are proposed to calculate the efficacy boundary for the composite test.

3.2.1 Proposal 1

Let $\hat{\alpha}_1$ be the actual alpha spent on the simple endpoint. For Scenario 1, $\hat{\alpha}_1 = \alpha_1^*(t_{K-1})$ which may be strictly less than the nominal $\alpha_1$; for Scenario 2, $\hat{\alpha}_1 = \alpha_1$. The Bonferroni method to calculate the efficacy boundary for the composite logrank statistic $R_{2,K}$ at the final analysis is,

$$\Pr(|R_{2,K}| \geq c^{(1)}|H_2) = \alpha - \hat{\alpha}_1$$

where, $c^{(1)}$ is the Bonferroni efficacy boundary for the composite logrank statistic $R_{2,K}$. The following result shows that the familywise type I error rate is controlled by the nominal alpha.

Result 3.1: The Bonferroni switching-endpoint procedure Proposal 1 considers a sequence of hypotheses $(H_2, H_1)$. It protects the familywise type I error rate at the nominal level $\alpha$, i.e.

1. $\Pr(\text{reject } H_1|H_1) \leq \alpha$

2. $\Pr(\text{reject } H_2|H_2) \leq \alpha$

3. $\Pr(\text{reject } H_1 \text{ or } H_2|H_1 \text{ and } H_2) \leq \alpha$

The proof of Result 3.1 is given in Appendix. Proposal 1 is computationally simple since we do not need to estimate the correlation vector $\nu$ between the simple test statistics and the composite statistic. Two questions are asked before the trial begins: is the treatment effective in reducing the mortality rate ($H_1$)? is it effective in reducing the composite event rate ($H_2$)? By using Bonferroni method, the two null hypotheses $H_1$ and $H_2$ are marginally tested and the two questions are directly answered.

3.2.2 Proposal 2

Since the correlation vector $\nu$ can be consistently estimated, the following method is proposed to calculate the boundary $c^{(2)}$ for $R_{2,K}$ at the final analysis which makes use of the full information about the joint distribution under the null hypothesis $\{H_1, H_2\}$. Let $A_{\alpha_1}$ be the acceptance region for the simple endpoint. As defined in section 3.1, $A_{\alpha_1} = A_{1,\alpha_1}$ for Scenario 1 and $A_{\alpha_1} = A_{2,\alpha_1}$ for Scenario 2.
\[
\Pr(A_{\alpha_1} \cap \{|R_{2,K}| \geq c^{(2)}\}|H_1, H_2) = \alpha - \hat{\alpha}_1
\]

The calculation of \(c^{(2)}\) requires the asymptotic distribution of \((R_{1,1}, ..., R_{1,K-1}, R_{1,K}, R_{2,K})\) which can be derived under \(\{H_1, H_2\}\) as in Corollary 2.1. A consistent estimator of the correlation vector can also be obtained under \(\{H_1, H_2\}\) as in Corollary 2.1.

**Result 3.2:** The switching-endpoint procedure Proposal 2 considers a sequence of hypotheses \((H_2, \{H_1, H_2\})\). The following are correct:

1. \(\Pr(\text{reject } H_1|H_1) \leq \alpha\)
2. \(\Pr(\text{reject } H_1 \text{ or } H_2|H_1, H_2) \leq \alpha\)

The proof of Result 3.2 is given in Appendix. The assumption \(\{H_1, H_2\}\) in Proposal 2 is stronger than \(H_2\) in Proposal 1. The two hypotheses \(H_1\) and \(\{H_1, H_2\}\) in Proposal 2 have a hierarchical structure since \(\{H_1, H_2\}\) always indicates \(H_1\). If we fail to reject \(H_1\), we then "switch" to test a stronger hypothesis using a test statistic which is based on the composite endpoint. In general, the second question, whether the treatment is effective in reducing the composite event rate, may not be answered directly since the hypothesis is \(\{H_1, H_2\}\) instead of \(H_2\). However, if we believe that the treatment effect for each component of the composite event is in the same direction, \(H_2\) implies \(H_1\) and then \(\{H_1, H_2\} = H_2\). In such situation, the test result can be naturally interpreted. For example, if the composite event is defined as the first event of either death or stroke, and we believe the treatment effects for mortality reduction and stroke prevention are in the same direction, then the hypothesis \(\{H_1, H_2\}\) is equivalent to \(H_2\).

**4 Information Fractions**

Calculations of boundaries \((b_1, ..., b_{K-1})\) for the simple outcome require estimates of information fractions \(t_k\)'s for \(k = 1, ..., K - 1\). Because the covariance matrix of \((Z_2(1), ..., Z_2(K-1))\) has an independent increment structure, a natural estimator of information fraction \(t_k\) is then \(t_k = d_k/d_K\), where \(d_K\) is the total observed number of deaths at the end of the trial. For Scenario 2, since the mortality is the primary endpoint and we often target a specific number of events - so called "event driven" trials for the primary outcome, \(d_K\) is usually known and the information fractions can be estimated by \(t_k = d_k/d_K\). For Scenario 1, since the composite outcome is the primary endpoint, the power calculation is typically based on the composite endpoint. The maximum number of composite events \(s\) is usually projected before the trial starts. If we have knowledge about how many of these composite events are to be simple events, we are able to estimate \(d_K\). For example, if a power calculation based on the composite endpoint gives \(s = 500\) and 40% of these composite events are expected to be
deaths, an estimate of the maximum number of simple events is then \( d_K = 500(40\%) = 200 \). If we do not have such information, a simple way to estimate the information fractions is using the calendar times. For example, if four interim analyses are scheduled at the 6th, 14th, 23rd, 28th month and the final analysis is at the 36th month, the corresponding information fractions can be estimated by \( (6/36, 14/36, 23/36, 28/36, 1) \). If we want to standardize the information fractions such that \( t_4 = 1 \), the estimated information times for the four interim analyses are \( (6/28, 14/28, 23/28, 1) \). For reference, see Lan and DeMets(1989) for detailed discussion about the calendar times and information fractions for group sequential procedures.

5 Simulation Study

In this section, we use hypothetical clinical trials to investigate the switching-endpoint procedures. Our simulation study will focus on Scenario 1, where only a little alpha will be spent on the simple endpoint. We also investigate the effect on the power by increasing \( \alpha_{\text{1}} \), the amount of alpha spent on the simple endpoint. Each simulated trial is a realization of a two-armed, double-blinded, randomized clinical trial. In each simulated trial, 1200 patients are randomized to the control and treatment groups with 600 patients in each group. Patients are uniformly recruited in the first 8 months and the study duration is 36 months. The primary endpoint is the elapsed time from randomization to the first occurrence of either of the following two events: death or hospitalization. For Scenario 1, the simple mortality outcome is used in the interim analyses which are scheduled at the 12th and 24th month. For Scenario 2, the simple mortality outcome will be tested at the 12th, 24th and 36th month. If the tests fail to show significance for the mortality outcome, a test for the composite endpoint will be performed at the 36th month. We assume that there is no loss to follow up.

5.1 Simulation Parameters

We use the bivariate exponential distribution described in Gumbel(1960) to generate the data. Let \( Z \) be the indicator of the treatment group \( (Z = 0 \text{ for control and } Z = 1 \text{ for treatment}) \), \( V_1 \) denote the time-to-death and \( V_2 \) be the hospitalization time. The joint density function of \( (V_1, V_2) \) in group \( Z = j \) \( (j = 0, 1) \) is,

\[
    f^{(j)}(x, y) = (\lambda_j e^{-x\lambda_j})(\theta_j e^{-y\theta_j})[1 + r_j(1 - 2e^{-x\lambda_j})(1 - 2e^{-y\theta_j})]
\]

In treatment group \( Z = j \) \( (j = 0, 1) \), the marginal distributions of the mortality endpoint \( V_1 \) and the hospitalization endpoint \( V_2 \) are both exponential,

\[
    V_1 \sim \lambda_j \text{exp}\{-x\lambda_j\} \quad \text{and} \quad V_2 \sim \theta_j \text{exp}\{-y\theta_j\},
\]

\[
    \lambda_j = \lambda_0 \cdot \left(1 + \alpha_{\text{1}}\right) \quad \text{and} \quad \theta_j = \theta_0 \cdot \left(1 + \alpha_{\text{1}}\right)
\]


11
where, \( \lambda_j \) is the marginal hazard rate for mortality and \( \theta_j \) is the marginal hazard rate for hospitalization in group \( Z = j \). The parameter \( r_j \) is assumed identical for the control and treatment groups, i.e. \( r_0 = r_1 = 1 \).

We assume that the treatment may reduce either the mortality event rate or the hospitalization event rate or both. Three cases about the baseline parameters in the control group will be investigated in our simulation. The percentage of the 3-year composite events which are deaths is increased from 20.5\% in Case 1 to 46.8\% in Case 2, and 75.2\% in Case 3. More specifically,

- **Case 1.** The one-year mortality rate is 5\% and the one-year hospitalization probability is 15.8\%. The one-year composite event rate is then 19.4\% by the distribution of the composite outcome. As a result, the three-year composite event rate is 45.9\% and 20.5\% of the composite events are deaths.

- **Case 2.** The one-year mortality rate is 10\% and the one-year hospitalization probability is 11.1\%. The one-year composite event rate is then 19.1\%. As a result, the three-year composite event rate is 44.7\% and 46.8\% of the composite events are deaths.

- **Case 3.** The one-year mortality rate is 15\% and the one-year hospitalization probability is 5.9\%. The one-year composite event rate is then 19.3\%. As a result, the three-year composite event rate is 45.5\% and 75.2\% of the composite events are deaths.

Suppose the overall type I error rate \( \alpha = 0.05 \). An O'Brien-Fleming type alpha spending function is applied to monitor the simple endpoint; the information times for the interim analyses are simply estimated by calendar times (1/3, 2/3). We will consider two situations. In the first situation, the alpha spent on the simple endpoint is very small: \( \hat{\alpha}_1 = 0.0044 \) (8.8\% of the total alpha). This number is calculated from the O'Brien-Fleming type alpha spending function and the information fraction 2/3 by \( 2[1 - \Phi(Z_{0.02}/\sqrt{2/3})] = 0.004383 \). In the second situation, the total alpha spent on the simple endpoint (two analyses) is \( \hat{\alpha}_1 = 2[1 - \Phi(Z_{0.05}/\sqrt{2/3})] = 0.0164 \) (32.8\% of the total alpha).

The simulation size is nsim=5,000 if the improvement for either the mortality or the hospitalization is 0; if the improvements for both the mortality and the hospitalization are positive, the simulation size is nsim=2,500. We may expect the simulated rejection probabilities to be accurate within 0.002 if either the mortality or the hospitalization is not improved, and within 0.01 if both the mortality and the hospitalization rates are improved.

### 5.2 Results

#### 5.2.1 Correlation

Table 1 gives the correlation coefficients between the normalized logrank statistics. For the bivariate exponential model with \( r = 1 \), the true correlations between the logrank statistics
for the simple endpoint and the logrank statistic for the composite endpoint can be calculated by using the true distribution to calculate $\nu_1$ and $\nu_2$ in Corollary 2.1. As shown in Table 1, if the mortality outcome and the hospitalization outcome are correlated with $\tau = 1$, the correlations between the standardized logrank statistics $(R_{1,1}, R_{1,2})$ and $R_{2,3}$ could be as high as 0.725 in Case 3 when about 75% of the composite events are deaths.

In our simulation studies, the correlation estimates $\nu_1$ and $\nu_2$ are calculated by the formula in Corollary 2.1. The covariance matrix estimate $\tilde{\Sigma}$ in Corollary 2.1 is used in the boundary calculation. Our simulation shows that these nonparametric estimates are close to the true values in Table 1.

Table 1: Correlations

<table>
<thead>
<tr>
<th></th>
<th>corr$(R_{1,1}, R_{1,2})$</th>
<th>corr$(R_{1,1}, R_{2,3})$</th>
<th>corr$(R_{1,2}, R_{2,3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>0.640</td>
<td>0.267</td>
<td>0.392</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.648</td>
<td>0.391</td>
<td>0.580</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.657</td>
<td>0.485</td>
<td>0.725</td>
</tr>
</tbody>
</table>

†Gumbel’s Model with $\tau = 1$. $R_{1,1}$ and $R_{1,2}$ are the normalized logrank statistics for the simple endpoint at the 1st and 2nd analyses, respectively. $R_{2,3}$ is the normalized logrank statistics for the composite endpoint at the 3rd (final) analysis. The percentages of death/composite are 20.5% in Case 1, 46.8% in Case 2, and 75.2% in Case 3.

5.2.2 Rejection Probabilities

The rejection probabilities of Proposal 1 and Proposal 2 are compared with those of the fixed composite-only test which only tests the composite endpoint at the end of trial. For Proposal 1 and 2, we may stop at certain interim analysis or go to the final analysis of the composite endpoint. For the fixed method, the probability of rejecting $H_2$ is given. For Proposal 1 and Proposal 2, the probability of stopping the trial early due to mortality benefit is given; if the mortality fails to show significance, the probability to reject $H_2$ (for Proposal 1) or the probability to reject $\{H_1, H_2\}$ (for Proposal 2) is also given for each case. Table 2 summarizes the rejection probabilities for Case 1 - 3 when the amount of alpha spent on the mortality outcome is small, i.e. $\alpha_1 = 0.0044$ out of 0.05. We assume that the treatment effect is the 3-year event reduction in either mortality or the hospitalization rates or both. The switching-endpoint procedures enable us to stop the trial early due to a convincing benefit in mortality. This is important for ethical reasons and for our medical research interest. In Case 1 where approximately 20% of the three-year composite events are deaths in the control group and the three-year mortality event rate for the treatment arm is reduced by 30%, switching-endpoint procedures Proposal 1 and Proposal 2 give a 12.3% chance to terminate the trial due to the benefit in mortality almost regardless of the reduction in the composite event. This probability goes up to 39.3% when approximately
Table 2: When $\alpha_1$ Is Small †

<table>
<thead>
<tr>
<th>Reduction</th>
<th>Mortality</th>
<th>Methods</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rej. $H_1$</td>
<td>Reduction in hospitalization</td>
<td>Rej. $H_1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0% 10% 20% 30%</td>
<td>0% 10% 20% 30%</td>
<td>0% 10% 20% 30%</td>
</tr>
<tr>
<td>0%</td>
<td>1</td>
<td>.004</td>
<td>.040 .195 .596 .916</td>
<td>.005 .043 .096 .246</td>
<td>.003 .042 .066 .110</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.004</td>
<td>.042 .199 .598 .918</td>
<td>.005 .043 .099 .253</td>
<td>.005 .044 .069 .116</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.005</td>
<td>.045 .210 .611 .926</td>
<td>.005 .048 .106 .264</td>
<td>.005 .049 .071 .118</td>
</tr>
<tr>
<td>10%</td>
<td>1</td>
<td>.010</td>
<td>.048 .25 .68 .94</td>
<td>.025 .073 .21 .45</td>
<td>.045 .136 .21 .30</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.010</td>
<td>.049 .25 .68 .94</td>
<td>.025 .074 .21 .45</td>
<td>.045 .144 .22 .31</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.005</td>
<td>.054 .28 .70 .96</td>
<td>.008 .24 .48 .73</td>
<td>.008 .178 .27 .36</td>
</tr>
<tr>
<td>20%</td>
<td>1</td>
<td>.038</td>
<td>.061 .32 .74 .94</td>
<td>.125 .131 .33 .58</td>
<td>.261 .288 .39 .50</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.038</td>
<td>.061 .32 .74 .94</td>
<td>.125 .134 .33 .58</td>
<td>.261 .297 .40 .51</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.028</td>
<td>.072 .36 .79 .98</td>
<td>.208 .45 .71 .91</td>
<td>.208 .535 .65 .77</td>
</tr>
<tr>
<td>30%</td>
<td>1</td>
<td>.123</td>
<td>.069 .35 .73 .86</td>
<td>.393 .134 .33 .51</td>
<td>.687 .237 .25 .28</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.123</td>
<td>.069 .35 .73 .86</td>
<td>.393 .138 .33 .51</td>
<td>.687 .209 .25 .29</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.101</td>
<td>.101 .46 .86 .99</td>
<td>.388 .67 .90 .98</td>
<td>.872 .972 .97 .99</td>
</tr>
<tr>
<td>50%</td>
<td>1</td>
<td>.570</td>
<td>.039 .20 .38 .42</td>
<td>.953 .017 .03 .05</td>
<td>.999 .001 .00 .00</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.570</td>
<td>.040 .20 .38 .42</td>
<td>.953 .017 .03 .05</td>
<td>.999 .001 .00 .00</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.183</td>
<td>.183 .62 .95 1.00</td>
<td>.777 .95 .99 1.00</td>
<td>1.00 .1.00 .1.00</td>
</tr>
</tbody>
</table>

†: The amount of alpha spent on the mortality outcome is small, i.e. $\alpha_1 = 0.0044$ out of 0.05. Three methods are compared: Proposal 1, Proposal 2 and the fixed (non-sequential) method which tests the composite endpoint only. For the fixed method, the probability of rejecting $H_2$ is given. For Proposal 1 and Proposal 2, a probability of stopping the trial early due to mortality benefit is given in bold (Rej. $H_1$). If the mortality fails to show significance, the probability to reject $H_2$ (for Proposal 1) or the probability to reject $\{H_1, H_2\}$ (for Proposal 2) is given for each case.
Table 3: When $\alpha_1$ Is Moderate $\dagger$

<table>
<thead>
<tr>
<th>Reduction Mortality</th>
<th>Methods</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rej. $H_1$</td>
<td>Reduction in hospitalization</td>
<td>Rej. $H_1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>0%</td>
<td>1</td>
<td>.014</td>
<td>.027</td>
<td>.159</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.014</td>
<td>.029</td>
<td>.165</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.045</td>
<td>.210</td>
<td>.611</td>
</tr>
<tr>
<td>10%</td>
<td>1</td>
<td>.030</td>
<td>.032</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.030</td>
<td>.034</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.054</td>
<td>.23</td>
<td>.70</td>
</tr>
<tr>
<td>20%</td>
<td>1</td>
<td>.098</td>
<td>.038</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.098</td>
<td>.040</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.072</td>
<td>.36</td>
<td>.79</td>
</tr>
<tr>
<td>30%</td>
<td>1</td>
<td>.240</td>
<td>.041</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.240</td>
<td>.043</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.101</td>
<td>.46</td>
<td>.86</td>
</tr>
<tr>
<td>50%</td>
<td>1</td>
<td>.737</td>
<td>.017</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.737</td>
<td>.017</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>fixed</td>
<td>.183</td>
<td>.62</td>
<td>.95</td>
</tr>
</tbody>
</table>

$\dagger$: The amount of alpha spent on the mortality outcome is moderate, i.e. $\alpha_1 = 0.0164$ out of 0.05. Three methods are compared: Proposal 1, Proposal 2 and the fixed (non-sequential) method which tests the composite endpoint only. For the fixed method, the probability of rejecting $H_2$ is given. For Proposal 1 and Proposal 2, a probability of stopping the trial early due to mortality benefit is given in bold (Rej. $H_1$). If the mortality fails to show significance, the probability to reject $H_2$ (for Proposal 1) or the probability to reject $\{H_1, H_2\}$ (for Proposal 2) is given for each case.
half of the composite events are deaths (Case 2) and 68.7% when most of the composite events are deaths (Case 3). Note here the alpha actually spent on the simple endpoint is only .0044 out of .05. This shows that the interim monitoring of the simple outcome could be very beneficial, especially for a realistic case such as Case 2.

One concern about the switching-endpoint procedures is that they may impair the power of detecting the difference in the composite outcome if they fail to detect the difference in the mortality. Since the amount of alpha spent on the simple endpoint is small (.0044 out of .05), the power loss (if any) for the two switching-endpoint procedures is not substantial. For example, in the extreme situation when there is no treatment benefit due to mortality (reduction in mortality is 0%) and the reduction in hospitalization is 30%, the power difference between the switching-endpoint procedures and the fixed method is no greater than 2%. When the reduction in mortality is larger than the reduction in the composite, the two proposals may have larger rejection probabilities (the probability that at least one hypothesis will be rejected) than the composite-only method does.

Proposal 1, which applies a Bonferroni method to adjust for the multiplicity of outcomes, turns out to be essentially as powerful as Propose 2 which makes use of the correlation structure between the simple and composite endpoints. Here the term “power” may have different meanings for Proposal 1 and Proposal 2 since they have somewhat different hypotheses. As in Table 2 when most of the composite events are deaths (Case 3) and the correlations between the mortality test statistics \( R_{1,1}, R_{1,2} \) and the composite test statistic \( R_{2,3} \) are moderately high (correlations (.485, .725) as in Table 1), the maximum difference of the rejection probabilities for Proposal 1 and Proposal 2 is about 0.01.

When the amount of alpha spent on the mortality outcome is moderate, \( \hat{\alpha}_1 = 0.0164 \) out of 0.05, Table 3 gives the rejection probabilities for Case 1 - 3. If more alpha is spent on the simple endpoint, we have a greater chance to detect the difference in mortality. In the extreme situation when there is no treatment benefit due to mortality, the two switching-endpoint procedures are less powerful than the fixed trial but the maximum difference is only about 5%. In some situations, the two switching-endpoint procedures are more powerful if the benefit in mortality is significant enough. The power difference between Proposal 1 and Proposal 2 is not substantial. In fact, the difference is no more than 3% in all of the three cases.

Note in previous discussion, the alpha spending function used to monitor the simple endpoint is the O'Brien-Fleming type spending function. Table 4 gives the results when a Pocock-type alpha spending function is used to monitor the mortality outcome. The amount of alpha spent on the mortality outcome, \( \hat{\alpha}_1 \), is .03817 out of .05, which is close to 0.04 as in MERIT-HF study. This number is calculate by \( (0.05)\log\{1 + (e - 1)(2/3)\} = .03817 \). Only Case 3 is considered. As shown from the results, this type of design is not appropriate for Scenario 1 as introduced in section 1. If a lot of alpha is spent on mortality which is not the primary endpoint, the two switching procedures do lose substantial power when the reduction in hospitalization is greater than the reduction in mortality. For example, if reduction in
Table 4: Pocock Type Alpha-spending Function †

<table>
<thead>
<tr>
<th>Reduction in mortality</th>
<th>Methods</th>
<th>Rej. $H_1$</th>
<th>Reduction in hospitalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>1</td>
<td>.040</td>
<td>0% 0.009 0.016 0.031 0.094</td>
</tr>
<tr>
<td></td>
<td>2 fixed</td>
<td>.040</td>
<td>0.011 0.014 0.025 0.045 0.126</td>
</tr>
<tr>
<td>10%</td>
<td>1</td>
<td>.144</td>
<td>0.025 0.058 0.10 0.16 0.34</td>
</tr>
<tr>
<td></td>
<td>2 fixed</td>
<td>.144</td>
<td>0.033 0.048 0.08 0.13 0.21</td>
</tr>
<tr>
<td>20%</td>
<td>1</td>
<td>.483</td>
<td>0.059 0.11 0.17 0.26 0.40</td>
</tr>
<tr>
<td></td>
<td>2 fixed</td>
<td>.483</td>
<td>0.078 0.14 0.21 0.29 0.42</td>
</tr>
<tr>
<td>30%</td>
<td>1</td>
<td>.865</td>
<td>0.09 0.06 0.09 0.11 0.10</td>
</tr>
<tr>
<td></td>
<td>2 fixed</td>
<td>.865</td>
<td>0.047 0.07 0.09 0.11 0.13</td>
</tr>
<tr>
<td>50%</td>
<td>1</td>
<td>1.000</td>
<td>0.00 0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td></td>
<td>2 fixed</td>
<td>1.000</td>
<td>1.00 1.00 1.00 1.00 1.00</td>
</tr>
</tbody>
</table>

†A Pocock type alpha spending function is applied to monitor the simple endpoint. $\alpha_1 = 0.038$ out of 0.05.

mortality is 10% and the reduction in hospitalization is 50%, the two procedures give rejection probabilities 0.49 and 0.55 respectively, compared to 69% power of the composite-only test. It is recommended to spend most of the nominal alpha on the primary endpoint, which is the composite outcome in Scenario 1 and the mortality endpoint in Scenario 2. The power difference between Proposal 1 and Proposal 2 could be as much as 6%. In Scenario 2 when the mortality endpoint will be tested at the final analysis, the correlation between $R_{1,3}$ for the mortality outcome and $R_{2,3}$ for the composite endpoint should be higher. If an O'Brien-Fleming type spending function is used to monitor the mortality endpoint, the power difference between Proposal 1 and Proposal 2 is expected to be larger. This indicates that Proposal 1 may be conservative in Scenario 2.

6 An Example: the PRAISE-I Trial

PRAISE-I (Prospective Randomized Amlodipine Survival Evaluation; see Packer et al (1996)) is a double-blinded, randomized, placebo-controlled trial conducted from March 9, 1992 to December 31, 1994. The objective of this study is to investigate the effect of amlodipine in patients with severe chronic heart failure. 1153 patients were recruited during March 1992 and June 1994 and randomized to placebo and amlodipine groups. The randomization was stratified on the basis of whether patients had ischemic or nonischemic causes of heart failure. The primary endpoint is the death plus hospitalization. For our purpose, we only consider the 421 patients with nonischemic heart disease. Among these patients, 212 were in
the placebo group and 209 were in the amlodipine group. At the end of the trial, there were 136 primary events with 78 in the placebo group and 58 in the amlodipine group. Among the 136 primary events, there were 103 deaths (66 in placebo group and 37 in amlodipine group). Most of the primary events are deaths (103/136=75.7%). There were another 16 patients who died after hospitalization. So, at the end of the study, there were a total of 119 deaths. The Kaplan-Meier estimates for primary and death events are in Figure 1. The logrank test statistic for primary event is 2.119 which crosses the nominal fixed design boundary $c_{fz}=1.960$ (two-sided and $\alpha=.05$). The logrank test statistic for mortality is 3.307.

Since the composite endpoint is the primary outcome, we consider the switching-endpoint designs for Scenario 1, which spend a small amount of alpha on the mortality endpoint. Suppose there are two interim analyses one year apart, say on March 9, 1993 and March 9, 1994. The mortality outcome is used in the interim analyses since the data collection and adjudication process delayed the currentness of the primary outcome. If the benefit for mortality is not convincing, the composite endpoint will be tested at the scheduled termination on December 31, 1994. We specify $\alpha_1=.02$ and an O'Brien-Fleming (OBF) type alpha-spending function is applied to monitor the mortality outcome. We use information
Table 5: PRAISE-I: Switching-Endpoints

<table>
<thead>
<tr>
<th>endpoint:</th>
<th>analysis 1</th>
<th>analysis 2</th>
<th>final analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>death(1)</td>
<td>death(2)</td>
<td>composite</td>
</tr>
<tr>
<td>correlation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>death(1)</td>
<td>.536</td>
<td>.376</td>
<td></td>
</tr>
<tr>
<td>death(2)</td>
<td></td>
<td>.723</td>
<td></td>
</tr>
<tr>
<td>logrank statistics:</td>
<td>.896</td>
<td>3.218</td>
<td>2.119</td>
</tr>
<tr>
<td>boundaries:</td>
<td>Fixed</td>
<td></td>
<td>1.960</td>
</tr>
<tr>
<td>Proposal 1</td>
<td>4.029</td>
<td>2.852</td>
<td>1.999</td>
</tr>
<tr>
<td>Proposal 2</td>
<td>4.029</td>
<td>2.852</td>
<td>1.973</td>
</tr>
</tbody>
</table>

fractions \((1/3, 2/3)\) and the actual alpha spent on the mortality outcome is \(\hat{\alpha}_1 = 2[1 - \Phi(Z_{0.02/2}/\sqrt{2/3})]\) = 0.0044, the same as in our simulation.

The results are summarized in Table 5. At the first interim analysis, there are 23 deaths (10 in amlodipine group and 13 in placebo group). The logrank test statistic is .896 which does not cross the OBF boundary of 4.029. At the second analysis, there are 80 deaths (28 in the amlodipine group and 52 in the placebo group). The logrank test statistic is now 3.218 which crosses the OBF boundary of 2.852. The significant benefit of amlodipine has been established!

To compare Proposal 1 and Proposal 2, we may continue the trial to the final analysis. As we calculated before, the logrank test statistic for the composite event at the end of the trial is 2.119. Proposal 1 gives a boundary \(c^{(1)}=1.999\) and we can reject the hypothesis for the composite endpoint. The estimated correlation vector between the two mortality statistics and the final composite statistic is \((.376, .723)\). Proposal 2 gives a boundary \(c^{(2)}=1.973\) which is also crossed. The three boundaries, \(c^{fix}=1.960, c^{(1)}=1.999\) and \(c^{(2)}=1.973\), are very close. Even if we fail to reject the hypothesis for the mortality outcome, we may still have enough power to reject the second hypothesis. Proposal 1 gives a boundary which is very close to the boundary by Proposal 2. In this example, the mortality outcome is strongly correlated with the composite outcome (recall that 75.7% primary events are deaths), but Proposal 1 is not substantially more conservative than Proposal 2. This is because that the alpha we spent on the mortality outcome is quite small (.004383) and this is usually true for Scenario 1 where the composite outcome is the primary.

7 Discussion

We should be careful about the interpretation of the results from the switching-endpoint designs Proposal 1 and 2. They have different hypotheses to be tested. Usually switching the marginal hypothesis \(H_1\) to \(H_2\) as in Proposal 1 seems more natural since the two questions
about the simple and composite endpoints are directly answered. Usually investigators do believe the treatment is beneficial for both mortality and other clinically relevant outcomes, and by using a composite outcome they can make the clinical trial powerful enough to detect the benefit. In such situation, the hypothesis $H_2$ implies $H_1$ and then $\{H_1, H_2\} = H_2$. The test result from Proposal 2 can be naturally interpreted. For example, consider the bivariate exponential model with $r = 0$. In the treatment arm $Z = j$ ($j = 0, 1$ for control and treatment arms), the hazard rate for mortality is $\lambda_j$ and the hazard rate for hospitalization is $\theta_j$. Since $r = 0$, the hazard rate for the composite event is $\lambda_j + \theta_j$. If we believe that either $\lambda_0 \geq \lambda_1$ and $\theta_0 \geq \theta_1$ (treatment reduces both mortality and hospitalization event rates), or $\lambda_0 \leq \lambda_1$ and $\theta_0 \leq \theta_1$ (treatment increases both mortality and hospitalization event rates), the hypothesis about the composite endpoint $H_2$, which assumes that $\lambda_0 + \theta_0 = \lambda_1 + \theta_1$, implies $\lambda_0 = \lambda_1$ which is $H_1$.

How to choose $\alpha_1$ needs to be explored. It is recommended to spend most of alpha on the primary endpoint, which is the composite outcome in Scenario 1 and the mortality endpoint in Scenario 2. For Scenario 1 when the amount of alpha spent on the mortality endpoint is small, Proposal 1 is essentially as powerful as Proposal 2. Since Proposal 1, the Bonferroni method, is simple and easy to interpret, it is recommended in such situation. In Scenario 2 where we may switch from the primary mortality endpoint to the secondary composite endpoint, most of the nominal alpha should be spent on the mortality endpoint. Proposal 1 could be conservative in this situation. If we believe that the treatment effect for each component of the composite event is in the same direction, Proposal 2, which is more powerful than the Bonferroni method, is recommended.

Although we primarily consider logrank test statistics in this article, the switching-endpoint procedures can be easily extended to a general setting where the general class of weighted linear rank statistics defined in section 1 is used. If the random weight function converges to a bounded function which does not depend on the time when the interim analysis is performed, the asymptotic covariance matrix of the repeated test statistics for the simple endpoint has an independent increment structure and the alpha-spending function can be naturally applied.

8 Appendix

8.1 Proof of Result 2.1

Our proof is similar to Wei and Lachin(1984), and Lin(1991). When the data are reviewed at time $T$, the linear rank statistic $U_l(T)$ for the $l$th failure time can be written as follows,
\begin{equation}
U_l(T) = \sum_{i=1}^{n} \int_{0}^{\infty} Q_l(T, x) \{ Z_i - \frac{Y_i^{(1)}(T, x)}{Y_i^{(0)}(T, x) + Y_i^{(1)}(T, x)} \} dM_{li}(T, x) 
+ \int_{0}^{\infty} Q_l(T, x) \frac{Y_i^{(0)}(T, x)Y_i^{(1)}(T, x)}{Y_i^{(0)}(T, x) + Y_i^{(1)}(T, x)} [d\Lambda_i^{(1)}(x) - d\Lambda_i^{(0)}(x)],
\end{equation}

where, for \( r = 0, 1 \), \( \Lambda_i^{(r)}(x) \) is the cumulative hazard function for the \( l \)th failure time in the treatment group \( r \). Under \( H_l \), \( \Lambda_i^{(0)}(x) = \Lambda_i^{(1)}(x) = \Lambda_l(x) \) and the second term is zero. \( M_{li}(T, x) \) is defined as follows,

\begin{equation}
M_{li}(T, x) = N_{li}(T, x) - \int_{0}^{x} I(X_{li}(T) \geq u) d\Lambda_l(u),
\end{equation}

where, \( N_{li}(T, x) = Z_{li}(T)I(X_{li}(T) \leq x) \) is a counting process. By Gill(1980) Lemma 2.3, under \( H_l \) and the non-informative censoring condition, \( M_{li}(T, x) \) is a square integrable martingale with respect to the filtration \( \{ \mathcal{F}_{li}(T, x); 0 \leq x \leq T \} \), where \( \mathcal{F}_{li}(T, x) \) is the information about the \( l \)th event obtained from the \( l \)th patient after he/she has been in the study for a time period \( x \). If this patient has not entered the study yet \( (Y_i > T) \), no information except that about the entry time can be obtained for any \( 0 \leq x \leq T \). Note here \( T \) is the calendar time when the data are reviewed and \( x \) has been corrected for the entry time. A rigorous definition of \( \mathcal{F}_{li}(T, x) \) is that it is the \( \sigma \)-field generated by

\[ \{ I(Y_i \leq T), I(Y_i \leq T)(Z_i, Y_i, X_{li}(T)I(X_{li}(T) \leq x), \Delta_{li}(T)I(X_{li}(T) \leq x)) \} \]

Note \( \mathcal{F}_{li}(T, x) \) is increasing as the review time \( T \) increases since more information can be available at a later time. Define:

\begin{equation}
W_{li}(T) = \int_{0}^{\infty} q_l(T, x) \{ Z_i - \mu_i^{(1)}(T, x) \} dM_{li}(T, x),
\end{equation}

where, \( \mu_i^{(1)}(T, x) \) is defined by (3). Let \( \bar{U}_l(T) = \sum_{i=1}^{n} W_{li}(T) \). Under \( H_l \) and some regularity conditions, we have \( n^{-1/2} \{ U_l(T) - \bar{U}_l(T) \} \to 0 \) in probability. See Wei and Lachin(1984)
for a similar proof. The basic method is using the Gill’s(1980) martingale central limit theorem (Theorem 4.2.1) to show that $n^{-1/2}\{U_i(T) - \bar{U}_i(T)\}$ converges in distribution to a degenerate normal variable with variance 0. The proof is based on the condition that $Q_i(T, x)$ converges to $q_i(T, x)$ in sup norm in probability for every fixed $T$ and the uniform convergence of empirical estimator. Only the marginal hypotheses $H_i$’s for $i = 1, \ldots, L$ are needed in this proof. The random vector

$$n^{-1/2}(U_1(T_1), \ldots, U_L(T_1), U_1(T_2), \ldots, U_L(T_2), \ldots, U_1(T_K), \ldots, U_L(T_K))^\prime$$

is then asymptotically equivalent to

$$n^{-1/2}(<\bar{U}_1(T_1), \ldots, \bar{U}_L(T_1), \bar{U}_1(T_2), \ldots, \bar{U}_L(T_2), \ldots, \bar{U}_1(T_K), \ldots, \bar{U}_L(T_K))^\prime$$

which can be written as $n^{-1/2}\sum_{i=1}^n W_i$, where,

$$W_i = (W_{i1}(T_1), \ldots, W_{Li}(T_1), \ldots, W_{i1}(T_K), \ldots, W_{Li}(T_K))^\prime$$

For $i = 1, \ldots, n$, $W_i$ are i.i.d. random vectors from the assumption that $(Z_i, Y_i, V_i, C_i)$ are i.i.d. Under the null hypothesis $H$, $E\{W_i\} = 0$ since each component is a martingale from the martingale transformation theorem. Each entry of the covariance matrix $\Sigma = E\{W_iW_i^\prime\}$ exists and is finite. By the multivariate central limit theorem, $n^{-1/2}\sum_{i=1}^n W_i$ asymptotically has a multivariate normal distribution with mean vector zero and covariance matrix $\Sigma$ under $H$. This finishes the proof of Result 2.1.

### 8.2 Proof of Result 2.2

The proof is the same as that in Wei and Lachin(1984). In fact, the proof in Wei and Lachin(1984) does not require the assumption that the joint distributions of the multiple time-to-event endpoints are identical for the control and treatment groups. Similar proof has also been used in Tsiatis(1981).

### 8.3 Proof of Result 3.1

Since at most $\alpha_1$ will be spent on the simple endpoint, we have in fact

$$\Pr(\text{reject } H_1|H_1) = \hat{\alpha}_1 \leq \alpha$$

Since we switch the endpoints by a Bonferroni method, the alpha spent on the composite endpoint is exactly $\alpha - \hat{\alpha}_1$ which is smaller than $\alpha$.

$$\Pr(\text{reject } H_2|H_2) = \alpha - \hat{\alpha}_1 < \alpha$$
Let $A^c_{a_1}$ be the completion of $A_{a_1}$ which is the acceptance region for the simple outcome.

$$\Pr(\text{reject } H_1 \text{ or } H_2|H_1 \text{ and } H_2)$$

$$= \Pr(A^c_{a_1}|H_1 \text{ and } H_2) + \Pr(A_{a_1} \cap \{|U_2(t_K)/\sqrt{s}| \geq c^{(1)}\}|H_1 \text{ and } H_2)$$

$$\leq \hat{\alpha}_1 + \Pr(|U_2(t_K)/\sqrt{s}| \geq c^{(1)}|H_2)$$

$$= \hat{\alpha}_1 + [\alpha - \hat{\alpha}_1]$$

$$= \alpha$$

8.4 Proof of Result 3.2

We only need to prove (2) since (1) has been proved. Since the hypothesis $\{H_1, H_2\}$ indicates $H_1$, we have

$$\Pr(\text{reject } H_1 \text{ or } H_2|H_1, H_2)$$

$$= \Pr(A^c_{a_1}|H_1, H_2) + \Pr(A_{a_1} \cap \{|U_2(t_K)/\sqrt{s}| \geq c^{(2)}\}|H_1, H_2)$$

$$= \hat{\alpha}_1 + [\alpha - \hat{\alpha}_1]$$

$$= \alpha$$

References


