HHS Public Access

Author manuscript

Lifetime Data Anal. Author manuscript; available in PMC 2021 April 01.

Published in final edited form as:

Lifetime Data Anal. 2020 April ; 26(2): 266–291. doi:10.1007/s10985-019-09470-4.

Group-sequential logrank methods for trial designs using bivariate non-competing event-time outcomes

Tomoyuki Sugimoto,

Department of Mathematics and Computer Science, Graduate School of Science and Technology, Kagoshima University, 1-21-35 Korimoto, Kagoshima 890-8580, Japan

Toshimitsu Hamasaki,

Department of Data Science, National Cerebral and Cardiovascular Center, 5-7-1 Fujishiro-dai, Suita, Osaka 565-8565, Japan

Scott R. Evans,

Epidemiology and Biostatistics and the Center for Biostatistics, George Washington University, 6110 Executive Boulevard Suite 750 Rockville 20852-3943, MD, USA

Susan Halabi

Department of Biostatistics and Bioinformatics,Duke University School of Medicine, Durham, 27705, NC, USA

Abstract

We discuss the multivariate ($2L$-variate) correlation structure and the asymptotic distribution for the group-sequential weighted logrank statistics formulated when monitoring two correlated event-time outcomes in clinical trials. The asymptotic distribution and the variance-covariance for the $2L$-variate weighted logrank statistic are derived as available in various group-sequential trial designs. These methods are used to determine a group-sequential testing procedure based on calendar times or information fractions. We apply the theoretical results to a group-sequential method for monitoring a clinical trial with early stopping for efficacy when the trial is designed to evaluate the joint effect on two correlated event-time outcomes. We illustrate the method with application to a clinical trial and describe how to calculate the required sample sizes and numbers of events.

Keywords

Bivariate dependence; Error-spending method; Independent censoring; Logrank statistic; Non-fatal events; Normal approximation

sugimoto@sci.kagoshima-u.ac.jp.

1 Introduction

Event-time outcomes are commonly used for evaluating the effect of a test intervention compared with a control. In some disease areas, e.g. HIV, oncology or cardiovascular disease, several event-time outcomes are used as the primary endpoints to more completely characterize the effect of an intervention on participants. Clinical trials with more than one primary endpoint can be designed to evaluate effects for *all* of the outcomes (i.e. co-primary endpoints) or to evaluate effects for *at least one* outcome (i.e. multiple primary endpoints). However, clinical trials with multiple event-time outcomes can be expensive and resource intensive as they often require large numbers of participants, collection of massive amounts of data, and long duration of follow-up. The use of group-sequential designs has the potential to improve efficiency, i.e. offering potentially fewer required trial participants, shortening the duration of clinical trials, and thus reducing the costs. Several authors have discussed group-sequential designs for multiple continuous or binary endpoints (e.g., Tang et al. 1989; Cook and Farewell 1994; Jennison and Turnbull 2000; Kosorok et al. 2004; Hung et al. 2007; Glimm et al. 2009; Tamhane et al. 2010, 2012; Asakura et al. 2014). Group-sequential theory and methods for single event-time outcomes have been studied (e.g., Tsiatis 1982; Slud and Wei 1982; Gordon and Lachin 1990; Gu and Lai 1991; Tsiatis et al. 1995; Lin et al. 1996; Lai and Shih 2004; Gombay 2008; Wu and Xiong 2017), and extended for multiple event-time outcomes (e.g., Wei and Lachin 1984; Pocock et al. 1987; Wei et al. 1990; Lin 1991; Cook and Farewell 1994) and for paired event-time data (e.g., Murray 2000; Andrei and Murray 2005; Jung 2008). Despite the extensive literature in group-sequential methods, there is a lack in the theory regarding the asymptotic structure of the weighted logrank statistics when group-sequentially comparing multiple event-time outcomes. Absence of this theory slows implementation and applying the group-sequential methodologies and creates challenges in calculating the power and the required sample size for multiple event-time outcomes.

We discuss a fundamental theory and methodology for group-sequential designs based on the weighted logrank statistic when monitoring several correlated event-time outcomes in clinical trials. We focus on bivariate event-time data rather than multivariate event-times, and consider a scenario where both events are non-fatal, as an extension of the existing method (Sugimoto et al. 2013). When considering the asymptotic distribution of the group-sequential logrank statistics, and the two martingale components with event-time outcomes are correlated on the different time axes, it is difficult to directly apply standard martingale theory for survival analysis, such as Rebolledo's central limit theorem. We overcome this challenge by combining a martingale approach and Ito's formula, and provide an asymptotic formula for group-sequential bivariate logrank statistic. We then apply the asymptotic result to group-sequential designs to evaluate a joint effect on both outcomes. We illustrate the design methodology with a clinical trial example.

This paper is organized as follows: in Sect. 2 we describe how the group-sequential weighted logrank statistic is applied to bivariate event-time data in a clinical trial. In Sect. 3, we discuss the asymptotic distribution with an explicit variance-covariance form for the bivariate version of group-sequential weighted logrank statistic, fundamental for determining the information fraction for each outcome and evaluating the probability of

rejecting the null hypotheses. In Sect. 4, we apply the asymptotic result to a group-sequential clinical trial evaluating the joint effect on the co-primary endpoints. We outline how both or one of the outcomes are monitored and evaluated. In Sect. 5, we summarize the findings and discuss their implications.

2 Group-sequential bivariate event-time data and the logrank statistic

Consider designing a randomized group-sequential clinical trial comparing two interventions evaluating bivariate event-time outcomes. Suppose that up to the planned maximum number of participants n_L will be recruited during an entry period and followed to observe the bivariate survival outcomes. Further, suppose interim analyses are planned with the pre-specified maximum number of analyses L. Let n_ℓ and τ_ℓ be the cumulative total number of participants and the analysis time at the ℓth interim analysis, respectively, with $n_1 \leq \cdots \leq n_L$ and $\tau_1 < \cdots < \tau_L$, and let $[0, \tau_A]$ be a period on which the trial recruiting is performed or which is planned in advance. The group index of intervention is denoted by $j = 2$ if the ith participant belongs to the test group and $j = 1$ otherwise. Let $n_{1\ell}$ and $n_{2\ell}$ denote the numbers of participants assigned to the control and test interventions at the ℓth analysis, respectively ($n_\ell = n_{1\ell} + n_{2\ell}$), where the fractions $n_{j1}/n_1, \ldots, n_{jL}/n_L$ may be often assumed to be approximately equal in each intervention. For $i = 1, \ldots, n_L$ and $k = 1, 2$, let O_i be the ith participant's entry time into the trial, let T^*_{ik} be the ith participant's underlying continuous event time for the kth outcomes, and let C_i be the ith participant's underlying censoring time common for the two outcomes, where O_i is the origin time of T^*_{ik} and C_i and is usually generated from the uniform distribution on the entry period $[0, \tau_A]$, the bivariate time (T^*_{i1}, T^*_{i2}) follows the joint survival distribution denoted by

$$S_j(t, s) = P(t < T^*_{i1}, s < T^*_{i2} \mid g_i = j),$$

g_i is the ith group index of intervention, and all of the C_i's follow the identical survival distribution $C(t) = P(t < C_i)$ independently of (T^*_{i1}, T^*_{i2}). Thus, the ith right-censoring time occurring at the ℓth analysis is $C_i^{(\ell)}$, where

$$C_i^{(\ell)} = \min(C_i, \max(\tau_\ell - O_i, 0)).$$

We will assume *no dropouts* where we observe $C_i = \tau_L - O_i$ because of well-controlled trial. Suppose that T^*_{i1} and T^*_{i2} are non-competing event-times, that is neither event-time is censored by the occurrence of the other event, which is typical in the case of non-fatal events (Sugimoto et al. 2013). For simplicity on notation, we write $O_1 \leq \cdots \leq O_{n_\ell} \leq \tau_\ell$ although we assume that O_i and $O_{i'}$ for $i \neq i'$ are mutually independent. Hence, we have a series of cumulative data set denoted by $\{(T_{i1}^{(\ell)}, T_{i2}^{(\ell)}, \Delta_{i1}^{(\ell)}, \Delta_{i2}^{(\ell)}, g_i)\}_{i=1}^{n_\ell}$, $\ell = 1, \ldots, L$, where $T_{ik}^{(\ell)} = \min(T^*_{ik}, C_i^{(\ell)})$ and $\Delta_{ik}^{(\ell)} = \mathbb{1}\{T^*_{ik} < C_i^{(\ell)}\}$ are the ith observable time and censoring indicator for the kth outcome at the ℓth analysis, respectively, and $\mathbb{1}\{\cdot\}$ is the indicator

function. The information of $(T_{ik}^{(\ell)}, \Delta_{ik}^{(\ell)})$ is also represented by the counting process $N_{ik}^{(\ell)}(t) = \mathbb{1}\{T_{ik}^{(\ell)} \le t, \Delta_{ik}^{(\ell)} = 1\}$ and the at-risk process $Y_{ik}^{(\ell)}(t) = \mathbb{1}\{T_{ik}^{(\ell)} \ge t\}$. Denote their sums on the group j and the kth outcome by

$$\overline{N}_{jk}^{(\ell)}(t) = \sum_{i=1}^{n_\ell} \mathbb{1}\{g_i = j\} N_{ik}^{(\ell)}(t), \quad \overline{Y}_{jk}^{(\ell)}(t) = \sum_{i=1}^{n_\ell} \mathbb{1}\{g_i = j\} Y_{ik}^{(\ell)}(t),$$

$$\overline{N}_{\bullet k}^{(\ell)}\left(t\right) = \overline{N}_{1k}^{(\ell)}\left(t\right) + \overline{N}_{2k}^{(\ell)}\left(t\right) \text{ and } \overline{Y}_{\bullet k}^{(\ell)}\left(t\right) = \overline{Y}_{1k}^{(\ell)}\left(t\right) + \overline{Y}_{2k}^{(\ell)}\left(t\right).$$

Also, let $\lambda_{jk}(t)$ and $\Lambda_{jk}(t)$ be the marginal hazard function and its cumulative function for the kth event time T_{ik}^* in the group j, respectively. Denote the marginal hazard ratio for the kth outcome between the two groups by $\psi_k(t) = \lambda_{2k}(t)/\lambda_{1k}(t)$ and let $\boldsymbol{\psi}(t) = (\psi_1(t), \psi_2(t))^{\mathrm{T}}$.

We are interested in testing sequentially either hypothesis $\mathrm{H}_0^{\mathrm{cp}} = \mathrm{H}_{01} \cup \mathrm{H}_{02}$ (for joint effect) or $\mathrm{H}_0^{\mathrm{mp}} = \mathrm{H}_{01} \cap \mathrm{H}_{02}$ (for at least one effect) using the weighted log-rank statistics, where H_{0k} is the single null hypothesis for the kth outcome, "$\psi_k(t) = 1$ for all t". For the bivariate event-time outcome with L maximum analyses, we have a set of $2L$ group-sequential weighted logrank statistics,

$$\widehat{\boldsymbol{Z}} = (\widehat{Z}_1(\tau_1), \ldots, \widehat{Z}_1(\tau_L), \widehat{Z}_2(\tau_1), \ldots, \widehat{Z}_2(\tau_L))^{\mathrm{T}}$$

composed of

$$\widehat{Z}_k(\tau_\ell) = \sqrt{n_\ell} U_k^{(\ell)}(\tau_\ell) \,/\, \sqrt{\widehat{V}_{kk}^{0(\ell)}(\tau_\ell)}, \; k = 1, 2, \ell = 1, \ldots, L$$

where $\sqrt{n_\ell} U_k^{(\ell)}(t)$ is the weighted logrank process accompanied with the analysis time τ_ℓ

$$U_k^{(\ell)}(t) = \int_0^t \widehat{H}_k^{(\ell)}(s)\{\mathrm{d}\widehat{\Lambda}_{1k}^{(\ell)}(s) - \mathrm{d}\widehat{\Lambda}_{2k}^{(\ell)}(s)\},$$

$\widehat{V}_{kk}^{0(\ell)}(t)$ is the conditional variance of $\sqrt{n_\ell} U_k^{(\ell)}(t)$ under the null hypothesis H_{0k},

$$\widehat{V}_{kk}^{0(\ell)}(t) = \int_0^t \widehat{H}_k^{(\ell)}(s)^2 \left\{1 - \frac{\mathrm{d}\overline{N}_{\bullet k}^{(\ell)}(s) - 1}{\overline{Y}_{\bullet k}^{(\ell)}(s) - 1}\right\} \left\{\frac{\mathrm{d}\widehat{\Lambda}_{1k}^{(\ell)}(s)}{n_\ell^{-1} \overline{Y}_{2k}^{(\ell)}(s)} + \frac{\mathrm{d}\widehat{\Lambda}_{2k}^{(\ell)}(s)}{n_\ell^{-1} \overline{Y}_{1k}^{(\ell)}(s)}\right\}.$$

Also, $\widehat{\Lambda}_{jk}^{(\ell)}(t) = \int_0^t \mathrm{d}\overline{N}_{jk}^{(\ell)}(s) \,/\, \overline{Y}_{jk}^{(\ell)}(s)$ is the Nelson-Aalen estimator at the ℓth analysis for the kth outcome in the group j, $\widehat{H}_k^{(\ell)}(s)$ is the following function including the weight $\widehat{W}_k^{(\ell)}$ of the class $\mathscr{K}$ (Fleming and Harrington 1991)

$$\widehat{H}_k^{(\ell)}(s) = n_\ell^{-1}\widehat{W}_k^{(\ell)}(s)\bar{Y}_{1k}^{(\ell)}(s)\bar{Y}_{2k}^{(\ell)}(s) / \bar{Y}_{\bullet k}^{(\ell)}(s),$$

$\widehat{W}_k^{(\ell)}(s) = f(\hat{S}_{\bullet k}^{(\ell)}(s))$ or $\widehat{W}_k^{(\ell)}(s) = f(n_\ell^{-1}\bar{Y}_{\bullet k}^{(\ell)}(s))$, $f(\cdot)$ is a nonnegative bounded continuous function with bounded variation on [0, 1], and $\hat{S}_{\bullet k}^{(\ell)}(s)$ is the Kaplan-Meier estimator for the kth outcome in the pooled sample at the ℓth analysis time τ_ℓ. A well-known fact is that the logrank and Prentice-Wilcoxon statistics use $\widehat{W}_k^{(\ell)}(s) = 1$ and $\widehat{W}_k^{(\ell)}(s) = \hat{S}_{\bullet k}^{(\ell)}(s_-)$, respectively, where s_- is a time just prior to s. The weight $\widehat{W}_k^{(\ell)}$ should be selected effectively to detect a clinically significant difference. If there is no prior assumption on a specific difference in the clinical significance, the logrank statistic may be adopted, which can be interpreted as detecting the difference in the mean hazard rate. Also, one can consider an optimality for testing using a special weight into the design, if pilot data or registry database are available.

3 Asymptotic structure of the group-sequential bivariate logrank statistic

Asymptotic results regarding the univariate statistic $\widehat{Z}_k(\tau_\ell)$ and its group-sequential version $(\widehat{Z}_k(\tau_1), \ldots, \widehat{Z}_k(\tau_L))^{\mathrm{T}}$ have been developed well (e.g., Andersen et al. 1993, X.2). For example, Lin (1991) shows that $\widehat{\boldsymbol{Z}}$ converges to a multivariate normal distribution with zero means and discuss the estimated variance-covariance matrix, although an explicit form for the asymptotic covariance of $\widehat{\boldsymbol{Z}}$ is not provided. Andrei and Murray (2005) provide a more detailed expression for the asymptotic covariance among weighted logrank statistics, but it is in the context of paired event-time data on the same time axes. To the best of our knowledge, a computable explicit form for the asymptotic variance-covariance of $\widehat{\boldsymbol{Z}}$ is not available in the literature. Extending the result for $\widehat{\boldsymbol{Z}}$ when $L = 1$, i.e. for fixed-sample design (Sugimoto et al. 2013), we provide the result of the asymptotic distribution of $\widehat{\boldsymbol{Z}}$ with an explicit variance-covariance structure for group-sequential design (Theorem 1).

We next provide details for expressing an asymptotic distribution of $\widehat{\boldsymbol{Z}}$. The limit forms of $\widehat{H}_k^{(\ell)}(t)$ and $n_\ell^{-1}\bar{Y}_{jk}^{(\ell)}(s)$ are different among the analysis time points as the censoring distributions vary with each analysis-time τ_ℓ. Let $H_k^{(\ell)}(t)$, $y_{jk}^{(\ell)}(t)$ and $y_{\bullet k}^{(\ell)}(t)$ denote the limit forms of $\widehat{H}_k^{(\ell)}(t)$, $n_{j\ell}^{-1}\bar{Y}_{jk}^{(\ell)}(t)$ and $n_\ell^{-1}\bar{Y}_{\bullet k}^{(\ell)}(t)$, respectively. Denote $\hat{a}_{j\ell} = n_{j\ell} / n_\ell$ for the sample rate of participants assigned to the group j at the ℓth analysis and $\hat{\gamma}_\ell = n_\ell / n_L$ for the sample size ratio between the ℓth and final analyses. Let $\xrightarrow{P}$ denote the convergence in probability. We assume the following regularity conditions.

Condition 1. For each j, ℓ $0 < a_{j\ell} < 1$ is satisfied, where $a_{j\ell}$ is a constant such that $\hat{a}_{j\ell} \xrightarrow{P} a_{j\ell}$ as $n_\ell \to \infty$.

Condition 2. For each ℓ $0 < \gamma_\ell \leq 1$ is satisfied with $\gamma_1 \leq \cdots \leq \gamma_L$, where γ_ℓ is a constant such that $\hat{\gamma}_\ell \xrightarrow{P} \gamma_\ell$ as $n_L \to \infty$.

Condition 3. For each j, k, ℓ, $y_{jk}^{(\ell)}(t) > 0$ on $[0, \tau_\ell]$ is satisfied with $\tau_\ell = \sup\{t : y_{jk}^{(\ell)}(t) > 0\}$, where $y_{jk}^{(\ell)}$ is a deterministic function such that, as $n_{j\ell} \to \infty$,

$$\sup_{t \in [0, \tau_\ell]} \left| n_{j\ell}^{-1} \bar{Y}_{jk}^{(\ell)}(t) - y_{jk}^{(\ell)}(t) \right| \to_P 0 .$$

Under our setting, the convergences provided in Conditions 1-2 and Condition 3 are derived by the law of large numbers and Glivenko-Cantelli theorem, respectively. Hence, we have $\gamma_\ell = E(\hat{\gamma}_\ell)$, $a_{j\ell} = E(\hat{a}_{j\ell})$ and $y_{\bullet k}^{(\ell)}(t) = a_{1\ell} y_{1k}^{(\ell)}(t) + a_{2\ell} y_{2k}^{(\ell)}(t)$. Note that $a_{j\ell}$ permits changing on the analysis time τ_ℓ but each $a_{j\ell}$ should be fixed at the design stage to control Type I error rate. The type of convergence in Condition 1 is usually replaced with the non-probabilistic version based on an allocation proceduce. Condition 3 provides $\lim_{t \to \tau_\ell + 0} y_{jk}^{(\ell)}(t) = 0$, which means that all at-risk individuals are once censored at the ℓth analysis time τ_ℓ.

Let $C_\ell(t)$ be the survival function of censoring times $C_i^{(\ell)}$ when the analysis time is τ_ℓ. Under the independent censoring assumption, we can easily show that

$$y_{jk}^{(\ell)}(t) = C_\ell(t_-) S_{jk}(t_-) \text{ and } y_{\bullet k}^{(\ell)}(t) = C_\ell(t_-) S_{\bullet k}^{(\ell)}(t_-), \tag{1}$$

where $S_{jk}(t) = P(t < T_{ik}^* \mid g_i = j)$ is the marginal survival function of T_{ik}^* assigned to the group j, and $S_{\bullet k}^{(\ell)}(t) = a_{1\ell} S_{1k}(t) + a_{2\ell} S_{2k}(t)$. Hence, given the condition that bivariate event-time outcomes are non-fatal, for $t \le \tau_\ell$ we have

$$H_k^{(\ell)}(t) = W_k^{(\ell)}(t) C_\ell(t_-) \frac{a_{1\ell} S_{1k}(t_-) a_{2\ell} S_{2k}(t_-)}{S_{\bullet k}^{(\ell)}(t_-)}, \tag{2}$$

where $W_k^{(\ell)}(t)$ is either $f(S_{\bullet k}^{(\ell)}(t_-))$ or $f(y_{\bullet k}^{(\ell)}(t))$ corresponding to the selection of $\widehat{W}_k^{(\ell)}$ in the class $\mathscr{K}$, so that $H_k^{(\ell)}(t)$ is a deterministic continuous function of bounded variation. In particular, when considering a typical group sequential trial, we will assume that participants are recruited uniformly on $[0, \tau_A]$, followed up with no dropouts and then will be analyzed at the times $t = \tau_1, \ldots, \tau_L$. Then we can specify the censoring survival distribution as

$$C_\ell(t) = \begin{cases} 1, & 0 \le t \le \tau_\ell - \min(\tau_\ell, \tau_A) \\ (\tau_\ell - t) \,/\, \min(\tau_\ell, \tau_A), & \tau_\ell - \min(\tau_\ell, \tau_A) < t \le \tau_\ell, \\ 0, & \tau_\ell < t \end{cases} \tag{3}$$

(recall τ_A is the length of the entry period planned in advance). Hence, we have

$$\gamma_\ell = \min(\tau_\ell, \tau_A) \,/\, \tau_A \tag{4}$$

under the censoring assumption (3), because it is the averaged ratio of the number of participants recruited until the analysis time τ_ℓ.

Suppose that $\boldsymbol{Z}^* = (Z_1^*(t_1), \ldots, Z_1^*(t_L), Z_2^*(t_1), \ldots, Z_2^*(t_L))^{\mathrm{T}}$ follows $2L$-variate normal distribution $N(D_{\boldsymbol{n}}\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector

$$D_{\boldsymbol{n}}\mu = D_{\boldsymbol{n}}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} = (\sqrt{n_1}\mu_{11}, \ldots, \sqrt{n_L}\mu_{1L}, \sqrt{n_1}\mu_{21}, \ldots, \sqrt{n_L}\mu_{2L})^{\mathrm{T}}$$

and variance-covariance matrix

$$\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} = \begin{pmatrix}\sigma_{1111} & \cdots & \sigma_{111L} & \sigma_{1211} & \cdots & \sigma_{121L}\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ \sigma_{11L1} & \cdots & \sigma_{11LL} & \sigma_{12L1} & \cdots & \sigma_{12LL}\\ \sigma_{2111} & \cdots & \sigma_{211L} & \sigma_{2211} & \cdots & \sigma_{221L}\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ \sigma_{21L1} & \cdots & \sigma_{21LL} & \sigma_{22L1} & \cdots & \sigma_{22LL}\end{pmatrix},$$

where $D_n = \mathrm{diag}(\sqrt{n_1}, \sqrt{n_2}, \ldots, \sqrt{n_L}, \sqrt{n_1}, \sqrt{n_2}, \ldots, \sqrt{n_L})$, $\boldsymbol{\mu}_k = (\mu_{k1}, \ldots, \mu_{kL})^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{kk'} = (\sigma_{kk'\ell\ell'})_{\ell\ell'}$. That is, for $k, k' = 1, 2$ and $\ell, \ell' = 1, \ldots, L$, the elements of means and covariances for $Z_k^*(t_\ell)$ and $Z_{k'}^*(t_{\ell'})$ are written as

$$E(Z_k^*(\tau_\ell)) = \sqrt{n_\ell}\mu_{k\ell} = \sqrt{n_\ell}\frac{m_k^{(\ell)}(\tau_\ell)}{\sqrt{V_{kk}^{0(\ell)}(\tau_\ell)}},$$

$$\mathrm{Cov}(Z_k^*(\tau_\ell), Z_{k'}^*(\tau_{\ell'})) = \sigma_{kk'\ell\ell'} = \frac{V_{kk'}(\tau_\ell, \tau_{\ell'} \mid \tau_\ell, \tau_{\ell'})}{\sqrt{V_{kk}^{0(\ell)}(\tau_\ell)V_{k'k'}^{0(\ell')}(\tau_{\ell'})}},$$

where we assume that the elements $m_k^{(\ell)}$, $V_{kk}^{0(\ell)}$ and $V_{kk'}$ are defined by

$$m_k^{(\ell)}(t) = \int_0^t H_k^{(\ell)}(x)\{\mathrm{d}\Lambda_{2k}(x) - \mathrm{d}\Lambda_{1k}(x)\},$$

$$V_{kk}^{0(\ell)}(t) = \int_0^t H_k^{(\ell)}(x)^2\left\{\frac{\mathrm{d}\Lambda_{1k}(x)}{a_{2\ell}y_{2k}^{(\ell)}(x)} + \frac{\mathrm{d}\Lambda_{2k}(x)}{a_{1\ell}y_{1k}^{(\ell)}(x)}\right\},$$

$$V_{kk}(t, s \mid \tau_\ell, \tau_{\ell'}) = \sqrt{\frac{\gamma_{\ell\wedge\ell'}}{\gamma_{\ell\vee\ell'}}}\int_0^{t\wedge s} H_k^{(\ell)}(x)H_k^{(\ell')}(x)\left\{\frac{\mathrm{d}\Lambda_{1k}(x)}{a_{1\ell\vee\ell'}y_{1k}^{(\ell\vee\ell')}(x)} + \frac{\mathrm{d}\Lambda_{2k}(x)}{a_{2\ell\vee\ell'}y_{2k}^{(\ell\vee\ell')}(x)}\right\},$$

$$V_{12}(t, s \mid \tau_\ell, \tau_{\ell'}) = \sqrt{\frac{\gamma_{\ell\wedge\ell'}}{\gamma_{\ell\vee\ell'}}}\int_0^t\int_0^s H_1^{(\ell)}(x)H_2^{(\ell')}(y)C_{\ell\wedge\ell'}(x\vee y) \times\left\{\frac{A_1(\mathrm{d}x, \mathrm{d}y)}{a_{1\ell\vee\ell'}y_{11}^{(\ell)}(x)y_{12}^{(\ell')}(y)} + \frac{A_2(\mathrm{d}x, \mathrm{d}y)}{a_{2\ell\vee\ell'}y_{21}^{(\ell)}(x)y_{22}^{(\ell')}(y)}\right\},$$

$$A_j(\mathrm{d}x, \mathrm{d}y) = S_j(\mathrm{d}x, \mathrm{d}y) + S_j(x, \mathrm{d}y)\mathrm{d}\Lambda_{j1}(x) + S_j(\mathrm{d}x, y)\mathrm{d}\Lambda_{j2}(y) + S_j(x, y)\mathrm{d}\Lambda_{j1}(x)\mathrm{d}\Lambda_{j2}(y),$$

$$S_j(\mathrm{d}x, \mathrm{d}y) = S_j(x, y) - S_j(x_-, y) - S_j(x, y_-) + S_j(x_-, y_-),$$

$S_j(\mathrm{d}x, y) = S_j(x, y) - S_j(x_-, y)$, $S_j(x, \mathrm{d}y) = S_j(x, y) - S_j(x, y_-)$, $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$. The forms provided in (4), (1) and (2) are applied into these elements $m_k^{(\ell)}$, $V_{kk}^{0(\ell)}$

and $V_{kk'}$. Under Conditions 1 and 3, it is well-known that the univariate weighted logrank statistic can be normally approximated (e.g., Fleming and Harrington 1991, Theorem 7.2.1). We have the following asymptotic result for the group-sequential weighted logrank statistic $\widehat{\boldsymbol{Z}}$ with correlated two outcomes.

Theorem 1 Suppose that Conditions 1-3 are satisfied ($a_{j\ell} \in (0, 1)$, $r_\ell \in (0, 1]$, $\tau_\ell = \sup\{t : y_{1k}^{(\ell)}(t) y_{2k}^{(\ell)}(t) > 0\}$), and that $S_j(t, s)$, $j = 1, 2$ are continuous on $(0, \tau_L] \times (0, \tau_L]$. Suppose that $f(\cdot)$ is a nonnegative bounded continuous function with bounded variation on [0, 1]. For sufficiently large n_ℓ's ($n_1 \le \cdots \le n_L$), the distribution of the $2L$-variate weighted logrank statistic $\widehat{\boldsymbol{Z}}$ can then be approximated by $N(\boldsymbol{D_n\mu}, \boldsymbol{\Sigma})$. That is, as $n_L \ge \cdots \ge n_1 \to \infty$, $\widehat{\boldsymbol{Z}} - \boldsymbol{D_n}\widehat{\boldsymbol{\mu}}$ converges in distribution to $\boldsymbol{Z}^* - \boldsymbol{D_n\mu}$ distributed as $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\widehat{\boldsymbol{\mu}}$ converges in probability to $\boldsymbol{\mu}$, $\widehat{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\mu}}_1^{\mathrm{T}}, \widehat{\boldsymbol{\mu}}_2^{\mathrm{T}})^{\mathrm{T}}$, $\widehat{\boldsymbol{\mu}}_k = (\hat{\mu}_{k1}, \ldots, \hat{\mu}_{kL})^{\mathrm{T}}$, $\hat{\mu}_{k\ell} = \widehat{m}_k^{(\ell)}(\tau_\ell) / \sqrt{V_{kk}^{0(\ell)}(\tau_\ell)}$,

$$\widehat{m}_k^{(\ell)}(t) = \int_0^t \widehat{H}_k^{(\ell)}(x)\{\mathrm{d}\Lambda_{2k}(x) - \mathrm{d}\Lambda_{1k}(x)\},$$

and $\boldsymbol{0}$ is the $2L$-dimensional zero vector.

This proof is provided in Appendix A. By conducting simulation stduies to evalaute the finite sample behavior for Theorem 1, we found that the asymptotic distribution works well in most practical situations if the event rate or sample size is not so small.

Several authors (e.g., Wei and Lachin 1984; Lin 1991) have indicated that the proof can be completed by the multivariate central limit theorem and the Cramér-Wald device, leading to asymptotic normality, but the asymptotic form of the variance-covariance was not clearly defined. The asymptotic form of variance-covariance as described in Theorem 1 has not been provided in the context of comparing independent groups with respect to several possibly correlated co-primary endpoints. In fact, when two martingale components with event-time outcomes are correlated on the different time axes as in this context, it is difficult to directly apply standard martingale theory, such as Rebolledo's central limit theorem, for survival analysis (Fleming and Harrington 1991) considering how the covariance of martingale components converges. As a reference to overcome the problem, we provide our solution based on a martingale approach through the proof of Theorem 1 in Appendix A.

Based on the result of Theorem 1 that the distribution of the weighted logrank statistics, $\widehat{\boldsymbol{Z}}$, can be approximated by $N(\boldsymbol{D_n\mu}, \boldsymbol{\Sigma})$, we can consider a group-sequential design and the asymptotic power for the testing procedure. In our setting, the distribution parameters of the mean vector $\boldsymbol{\mu}$ and the diagonal block matrix $\boldsymbol{\Sigma}_{kk}$ of $\boldsymbol{\Sigma}$ are determined by the setting of the marginal survival distributions $S_{jk}(t)$, the censoring survival distributions $C_k(t)$, and the sample rates $a_{j\ell}$ ($k = 1, 2$, $j = 1, 2$, $\ell = 1, \ldots, L$). In fact, the proportions $\gamma_1, \ldots, \gamma_L$ of sample sizes are determined by $\tau_1, \ldots, \tau_L$ and τ_A under the censoring assumption (3). On the other hand, in determining the non-diagonal block matrix $\boldsymbol{\Sigma}_{12}$ (and $\boldsymbol{\Sigma}_{21}$) of $\boldsymbol{\Sigma}$, the assumption of the joint survival distributions $S_j(t, s)$, $j = 1, 2$ are required. At the design stage of a trial, one convenient setting is to model $S_j(t, s)$ by

$$S_j(t, s) = \mathscr{C}(S_{j1}(t), S_{j2}(s); \theta) \tag{5}$$

where $\mathscr{C}(\cdot, \cdot)$ is a copula function (such as Clayton, Gumbel and Frank models), and the association parameter θ characterizes the level of dependence between $S_{j1}(t)$ and $S_{j2}(t)$ and is a one-to-one function of a dependence measure (Hsu and Prentice 1996)

$$\rho_j = \mathrm{Corr}\,[\Lambda_{j1}(T^*_{i1}), \Lambda_{j2}(T^*_{i2})] = \int_0^\infty \int_0^\infty S_j(t, s)\mathrm{d}\Lambda_{j1}(t)\mathrm{d}\Lambda_{j2}(s) - 1\,.$$

The mean vector $\boldsymbol{\mu}$ and the diagonal block matrix $\boldsymbol{\Sigma}_{kk}$ depend on the assumptions of the censoring distribution and the hazard ratios $\psi_1(t)$ and $\psi_2(t)$. The weighted logrank statistic is nonparametric, so that it is reasonable to assume the exponential distribution for marginals $S_{j1}(t)$ and $S_{j2}(t)$ in one group. Given the hazard ratios independent of times, such as the proportional hazard hypothesis $\{\psi_k(t) \equiv \psi_k$ for $t \in (0, \tau_L]$, $k = 1, 2\}$ ($\boldsymbol{\mu} = \mathbf{0}$ if $\psi_1 = \psi_2 = 1$), the marginals $S_{j1}(t)$ and $S_{j2}(t)$ in one group may model those of another group. Hence, a typical design calculation may be based on four exponential marginals $S_{jk}(t)$, $j = 1, 2$, $k = 1, 2$ and the setting of the analysis times ($\tau_1, \ldots, \tau_L$), the entry period (τ_A), the dependence measures (ρ_1, ρ_2) and the selection of some copula function. Numerically, calculations included in $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ρ_j can be sufficiently precisely by using the numerical integration method, such as the Trapezoidal rule or Simpson's rule (e.g., Sugimoto et al. 2013).

In group-sequential designs, the concept of information fraction is important in determining the critical boundary to preserve overall Type I error rate. This can be generalized to a bivariate event-time setting by analogy to a single event-time outcome. The information at τ_ℓ for each outcome can be characterized using an asymptotic form of the Fisher information, i.e., $\mathscr{I}_{k\ell} = n_\ell V^{0(\ell)}_{kk}(\tau_\ell)$, $k = 1, 2$, which corresponds to the information under the null hypothesis for the log of the hazard ratios (e.g., Jennison and Turnbull 2000; Yin 2012). As information is accumulated from τ_ℓ to τ_L, the standardized internal time $R\mathscr{I}_{k\ell}$ for each outcome is defined by the fraction of the maximum information of $\mathscr{I}_{kL}$, i.e., $R\mathscr{I}_{k\ell} = \mathscr{I}_{k\ell} / \mathscr{I}_{kL}$. Theorem 1 provides that the components of $\boldsymbol{\Sigma}$, under the null hypothesis of $\psi_1 = \psi_2 = 1$, are obtained as

$$V^{0(\ell)}_{kk}(t) = a_{1\ell}a_{2\ell}\int_0^t W^{(\ell)}_k(x)^2 C_\ell(x) S_{\bullet k}(x)\mathrm{d}\Lambda_{\bullet k}(x),$$

$$V_{kk'}(t, s \mid \tau_\ell, \tau_{\ell'}) = a_{1\ell\wedge\ell'}a_{2\ell\wedge\ell'}\sqrt{\frac{\gamma_\ell \wedge \gamma_{\ell'}}{\gamma_\ell \vee \gamma_{\ell'}}} \times \begin{cases} \displaystyle\int_0^{t\wedge s} W^{(\ell)}_k(x) W^{(\ell')}_k(x) C_{\ell\wedge\ell'}(x) S_{\bullet k}(x)\mathrm{d}\Lambda_{\bullet k}(x) & \text{if } k = k', \\ \displaystyle\int_0^t \int_0^s W^{(\ell)}_k(x) W^{(\ell')}_{k'}(y) C_{\ell\wedge\ell'}(x \vee y) A_\bullet(\mathrm{d}x, \mathrm{d}y) & \text{if } k \neq k', \end{cases}$$

where we have $\Lambda_{\bullet k}(x) = \Lambda_{1k}(x) = \Lambda_{2k}(x)$, $S_{\bullet k}(x) = S_{1k}(x) = S_{2k}(x)$ and $A_\bullet(x, y) = A_1(x, y) = A_2(x, y)$ because $\psi_1 = \psi_2 = 1$. The result given for the single endpoint (e.g., Andersen et al. 1993, X.2) is that the asymptotic correlation between group-sequential weighted logrank statistics

$$\mathrm{Corr}\,[Z_k^*(\tau_\ell), Z_k^*(\tau_{\ell'})] = \frac{V_{kk}(\tau_\ell, \tau_{\ell'} \mid \tau_\ell, \tau_{\ell'})}{\sqrt{V_{kk}(\tau_\ell, \tau_\ell \mid \tau_\ell, \tau_\ell) V_{kk}(\tau_{\ell'}, \tau_{\ell'} \mid \tau_{\ell'}, \tau_{\ell'})}},$$

reduces to $\sqrt{R\mathcal{I}_{k\ell\wedge\ell'} / R\mathcal{I}_{k\ell\vee\ell'}}$ when the null hypothesis is true ($\psi_1 = \psi_2 = 1$) and $W_k^{(\ell)}(s)$ is independent of ℓ such as $W_k^{(\ell)}(s) = 1$. Theorem 1 describes that the correlation of $(\widehat{Z}_k(\tau_\ell), \widehat{Z}_{k'}(\tau_{\ell'}))$ including between different endpoints for $k, k' = 1, 2$ and $1 \le \ell \le \ell' \le L$ can be approximated by

$$\mathrm{Corr}\,[Z_k^*(\tau_\ell), Z_{k'}^*(\tau_{\ell'})] = \frac{V_{kk'}(\tau_\ell, \tau_{\ell'} \mid \tau_\ell, \tau_{\ell'})}{\sqrt{V_{kk}(\tau_\ell, \tau_\ell \mid \tau_\ell, \tau_\ell) V_{k'k'}(\tau_{\ell'}, \tau_{\ell'} \mid \tau_{\ell'}, \tau_{\ell'})}},$$

which is $\{\mathbb{1}(k = k') + \mathbb{1}(k \neq k')\rho_Z(\tau_\ell, \tau_{\ell'})\}\sqrt{R\mathcal{I}_{k\ell'} / R\mathcal{I}_{k\ell}}$ if the null hypothesis is true and $W_k^{(\ell)}(s)$ is independent of ℓ where

$$\rho_Z(\tau_\ell, \tau_{\ell'}) = \frac{V_{kk'}(\tau_\ell, \tau_{\ell'} \mid \tau_\ell, \tau_{\ell'})}{\sqrt{V_{kk}(\tau_\ell, \tau_{\ell'} \mid \tau_\ell, \tau_{\ell'}) V_{k'k'}(\tau_\ell, \tau_{\ell'} \mid \tau_\ell, \tau_{\ell'})}}.$$

4 Application to group-sequential design

We provide an application to the group-sequential design based on the result discussed in Sect. 3. As a motivating example, consider a major HIV treatment trial within the AIDS Clinical Trials Group, “A Phase III Randomized Comparative Study of Three Non-Nucleoside Reverse Transcriptase Inhibitor (NNRTI)-Sparing Antiretroviral Regimens for Treatment-Naïve HIV-1-Infected Volunteers (The ARDENT Study: Atazanavir, Raltegravir, or Darunavir with Emtricitabine/Tenofovir for Naïve Treatment)” (Lennox et al. 2014). The planned total sample size of 1800 (equally-sized groups) was calculated for the paired comparison of the three regimens with respect to the two co-primary endpoints: “virologic failure” and “regimen failure due to tolerability”, not taking into account the potential correlation, with 3% inflation to the adjustment for interim monitoring, under the study duration of 96 weeks after enrollment of the last subject, where the two failures are non-fatal. The study had (i) a power of 0.90 to establish non-inferiority in the risk reduction of virologic failure with the non-inferiority margin of 10% at $\alpha = 0.0125$ for one-sided test, assuming the virologic failure rate of 25% at 96 weeks, and (ii) a power of 0.85 to detect a 10% difference in regimen failure at $\alpha = 0.025$ for two-sided test, assuming the regimen failure rate of 45% at 96 weeks.

For illustrative purposes, suppose that the objective of the ARDENT trial was to test for a two intervention superiority on both co-primary endpoints (OC1: virologic failure, OC2: regimen failure). The allocation ratios are assumed to be constant across analyses ($a_{j1} = \cdots = a_{jL}$) and are not changed arbitrarily during trial as the arbitrary choices may effect on the Type I error and power. The significance level of 2.5% ($\alpha = 0.025$) is allocated to each endpoint using a one-sided logrank test in a group-sequential setting, where the group sizes at each analysis are equal ($a_{1\ell} = 0.5$), the survival rate at 96 weeks is assumed to be 75% and

85% for OC1, and 55% and 65% for OC2, in the control and test intervention groups, respectively ($S_{11}(96) = 0.75$, $S_{21}(96) = 0.85$; $S_{12}(96) = 0.55$, $S_{22}(96) = 0.65$). Two analyses are planned: the first at $\tau_1 = 48 + \tau_A$ and the final at $\tau_2 = 96 + \tau_A$ ($L = 2$). Letting $\boldsymbol{\psi}^* = (\psi_1^*, \psi_2^*)$ be the hazard ratios of interest as a true vector of $\boldsymbol{\psi} = (\psi_1, \psi_2)$, typical exponential assumptions lead to $\boldsymbol{\psi}^* \doteq (0.565, 0.721)$ based on $S_{2k}(t) = S_{1k}(t)^{\psi_k^*}$ for ARDENT study.

A superiority clinical trial with two event-time outcomes (OC1 and OC2) as "co-primary" endpoints is often designed to evaluate if the test intervention is superior to the control on both outcomes. For two co-primary endpoints, the testing procedure is to test the union $\mathrm{H}_0^{\mathrm{cp}} = \mathrm{H}_{01} \cup \mathrm{H}_{02}$ of two individual nulls against the alternative $\mathrm{H}_1^{\mathrm{cp}} = \mathrm{H}_{11} \cap \mathrm{H}_{12}$. For simplicity, suppose that the proportional hazards hypothesis $\psi_1(t) \equiv \psi_1$ and $\psi_2(t) \equiv \psi_2$, and single null hypothesis $\mathrm{H}_{0k}: \psi_k = 1$ is tested versus $\mathrm{H}_{1k}: \psi_k < 1$ at the significance level α for each k. When evaluating a joint effect on both endpoints within the context of group-sequential designs, one decision-making framework associated with hypothesis testing is to reject $\mathrm{H}_0^{\mathrm{cp}}$ if statistical significance of the test intervention relative to the control intervention is achieved for both of the endpoints at any interim analysis until the final analysis not necessarily simultaneously (Asakura et al. 2014; Hamasaki et al. 2015). The power corresponding to the decision-making framework at $\boldsymbol{\psi} = \boldsymbol{\psi}^*$ is

$$\begin{aligned} 1 - \beta &= P\left(\left\{\bigcup_{\ell=1}^{L}\{Z_{1\ell}(\tau_\ell) > c_{1\ell}(\alpha)\}\right\} \bigcap \left\{\bigcup_{\ell=1}^{L}\{Z_{2\ell}(\tau_\ell) > c_{2\ell}(\alpha)\}\right\}\right) \\ &= 1 - P\left(\bigcap_{\ell=1}^{L}\{Z_{1\ell}(\tau_\ell) \le c_{1\ell}(\alpha)\}; \psi_1 = \psi_1^*\right) \\ &\quad - P\left(\bigcap_{\ell=1}^{L}\{Z_{2\ell}(\tau_\ell) \le c_{2\ell}(\alpha)\}; \psi_2 = \psi_2^*\right) \\ &\quad + P\left(\bigcap_{\ell=1}^{L}\{Z_{1\ell}(\tau_\ell) > c_{1\ell}(\alpha) \bigcap Z_{2\ell}(\tau_\ell) > c_{2\ell}(\alpha)\}; \boldsymbol{\psi} = \boldsymbol{\psi}^*\right), \end{aligned} \tag{6}$$

where $c_{k\ell}(\alpha)$ is the critical boundary at the ℓth analysis for the kth outcome, specified and determined in advance using any group-sequential methods, as if the two endpoints were a single primary endpoint, ignoring the other endpoint, analogously to the single endpoint case. Note that only the marginal results of Theorem 1 are required for the standardized internal times $R\mathcal{I}_{k\ell}$, where $R\mathcal{I}_{k\ell}$ does not depend on the correlation between OC1 and OC2 in the situation where both outcomes are non-fatal. Once $R\mathcal{I}_{k\ell}$, $k = 1, 2$ are determined, then the critical boundaries can be calculated using the group-sequential methods to control an overall Type I error rate in each marginal. Using the result of Theorem 1 that the distribution of $\widehat{\boldsymbol{Z}}$ can be approximated by that of $\boldsymbol{Z}^*$ under a large sample size, the power (6) can be approximately calculated as

$$1-\beta = 1 - \int_{-\infty}^{c_{11}^*} \cdots \int_{-\infty}^{c_{1L}^*} f_L(z_{11}, \ldots, z_{1L}; \boldsymbol{R}_{11}) \mathrm{d}z_{11} \cdots \mathrm{d}z_{1L} - \int_{-\infty}^{c_{21}^*} \cdots \int_{-\infty}^{c_{2L}^*} f_L(z_{21}, \ldots, z_{2L}; \boldsymbol{R}_{22}) \mathrm{d}z_{21} \cdots \mathrm{d}z_{2L} + \int_{-\infty}^{c_{11}^*} \cdots \int_{-\infty}^{c_{1L}^*} \int_{-\infty}^{c_{21}^*} \cdots \int_{-\infty}^{c_{2L}^*} f_{2L}(z_{11}, \ldots, z_{1L}, z_{21}, \ldots, z_{2L}; \boldsymbol{R}) \mathrm{d}z_{11} \cdots \mathrm{d}z_{1L} \mathrm{d}z_{21} \cdots \mathrm{d}z_{2L}, \tag{7}$$

where $f_m(\cdot; \boldsymbol{A})$ is m-variate normal density function with zero mean vector and variance-covariance matrix $\boldsymbol{A}$, $\boldsymbol{R}$ is the correlation matrix given by

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{pmatrix} = \boldsymbol{S}^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{S}^{-\frac{1}{2}},$$

$\boldsymbol{S} = \mathrm{diag}(\sigma_{1111}, \sigma_{1122}, \ldots, \sigma_{11LL}, \sigma_{2211}, \sigma_{2222}, \ldots \sigma_{22LL})$, the integration limits $c_{k\ell}^*$ are

$$c_{k\ell}^* = \frac{1}{\sqrt{\sigma_{kk\ell\ell}}} \{ c_{k\ell}(\alpha) - \sqrt{\gamma_\ell n_L} \mu_{k\ell} \}, \ k = 1, 2; \ell = 1, \ldots, L,$$

and recall $n_\ell = \gamma_\ell n_L$.

Returning to the ARDENT study, let $\tau_A = 0$ similarly to the manner assumed by Lennox et al. (2014). Although τ_A is not zero in fact, this selection of τ_A provides a conservative result and is reasonable in practice because of difficulty of estimating the feasible entry period. Two fixed analysis times are $(\tau_1, \tau_2) = (48, 96)$, where the censoring distribution (3) under $\tau_A = 0$ is simplified to

$$C_1(t) = \begin{cases} 1, & 0 \le t \le \tau_1 = 48 \\ 0, & \tau_1 < t \end{cases}, \quad C_2(t) = \begin{cases} 1, & 0 \le t < \tau_2 = 96 \\ 0, & \tau_2 < t \end{cases}. \tag{8}$$

We select the weight function of $W_k^{(\ell)}(s) = 1$ corresponding to the logrank statistic. Under these configurations with the exponential marginal assumption, we calculate $R\mathcal{I}_{k\ell}$ whose values are 0.5314 and 0.5669 at 48 weeks for the OC1 and OC2, and then determine $c_{k\ell}(\alpha)$ by the O'Brien-Fleming-type function (O'Brien and Fleming 1979) using the Lan-DeMets error-spending method (Lan and DeMets 1983), as shown in Table 1 including the Pocock-type boundary (Pocock 1977). The power (7) is then calculated, given the settings of the joint survival functions $S_j(t, s)$ and the correlations ρ_j between the OC1 and OC2. We use the copula model (5) to identify the joint survival distribution $S_j(t, s)$. In particular, we utilize the Clayton copula (late time-dependency) (Clayton 1976) and the Gumbel copula (early time-dependency) (Hougaard 1986), that is, we set, under the Clayton copula,

$$S_j(t, s) = (\mathrm{e}^{\theta_j \lambda_{j1} t} + \mathrm{e}^{\theta_j \lambda_{j2} s} - 1)^{-1 / \theta_j}$$

and, under the Gumbel copula,

$$S_j(t,s) = \exp\left(-\{(\lambda_{j1}t)^{1/\theta_j} + (\lambda_{j2}s)^{1/\theta_j}\}^{\theta_j}\right),$$

the mariginal hazard rates are given by $\lambda_{1k} = -\log S_{1k}(96)/96$ and $\lambda_{2k} = \lambda_{1k}\psi_k^*$, and the association parameter θ_j is determined by the value of ρ_j (see Sugimoto et al. (2013) for more details). For simplicity, we set the correlations as $\rho_1 = \rho_2 \equiv \rho$ and consider $\rho = 0, 0.1, \ldots, 0.9$ and 0.95. Based on (7), the total maximum sample size (MSS) required for the final analysis is the smallest integer n_L which provides (7) not less than the desired power at the prespecified $\boldsymbol{\psi} = \boldsymbol{\psi}^*$. For example, using the method with above parameter configuration and setting, for $\rho = 0$, $\boldsymbol{R}_{11}$, $\boldsymbol{R}_{22}$, and $\boldsymbol{R}_{12}$ are approximately calculated by

$$\boldsymbol{R}_{11} = \begin{pmatrix} 1, & 0.7260 \\ 0.7260, & 1 \end{pmatrix}, \quad \boldsymbol{R}_{22} = \begin{pmatrix} 1, & 0.7507 \\ 0.7507, & 1 \end{pmatrix}$$
$$\text{and} \quad \boldsymbol{R}_{12} = \begin{pmatrix} 0, & 0 \\ 0, & 0 \end{pmatrix},$$

respectively, and for $\rho = 0.8$

$$\boldsymbol{R}_{11} = \begin{pmatrix} 1, & 0.7260 \\ 0.7260, & 1 \end{pmatrix}, \quad \boldsymbol{R}_{22} = \begin{pmatrix} 1, & 0.7507 \\ 0.7507, & 1 \end{pmatrix}$$
$$\text{and} \quad \boldsymbol{R}_{12} = \begin{pmatrix} 0.2159, & 0.1569 \\ 0.1622, & 0.3341 \end{pmatrix},$$

respectively. Once the MSS is computed, the maximum event number (MEN) d_{kL} is calculated using $d_{kL} = n_L P_{kL}$ (event), where $P_{k\ell}$(event) is the probability that the event of the kth outcome occurs on the time interval $(0, \tau_\ell]$ and can be calculated, for example, based on Collett (2003) or Sugimoto et al. (2017, Appendix B). Also, the average event number (AEN) $\bar{d}_k$ is calculated using hypothetical reference values, similarly to Asakura et al. (2014), by

$$\bar{d}_k = \sum_{\ell=1}^{L-1} d_{k\ell} P_\ell(\text{stop}) + d_{kL}\left(1 - \sum_{\ell=1}^{L-1} P_\ell(\text{stop})\right)$$

where $d_{k\ell} = n_\ell P_{k\ell}$(event), and P_ℓ(stop) is the stopping probability as defined by the frequency of crossing the critical boundaries at the ℓth interim analysis under the true values $\boldsymbol{\psi}^*$ of the intervention effects. The AEN can provide information regarding the number of events anticipated in a group-sequential design in order to reach a decision point.

Table 2 summarizes the MSS, MEN and AEN, and empirical power for the late time-dependent association. The empirical power under the calculated MSS achieves the targeted power. First of all, we can see that the group-sequential design provides a quite smaller AEN than fixed-sample design does in every case, which is preferred in terms of costs saving. As expected, the MSS decreases with higher positive correlation, but the reduction is small. Power and sample size is less impacted by the correlation than the hazard ratio. The MSS is nearly determined by the hazard ratio closer to 1 and it does not vary with the correlation

when one hazard ratio is relatively smaller (or larger) than the other. Similarly there is little difference in the MEN between the group-sequential and fixed-sample designs. Based on these results, for the ARDENT study, the MSS is nearly determined by OC2. We only describe the result assuming the late time-dependent association. Similar patterns are observed in the case of an early time-dependent association (Gumbel copula), where the design planning results under the Gumbel copula is provided in Table B.1 of Appendix B. Also, we provide the considerations and results about Type I error rates control in Appendix B (Tables B.2 and B.3).

In this illustration, the interim analysis was planned to be conducted at the prespecified calendar times as participants are recruited in calendar time. On the other hand, one may design a survival trial based on information fraction as interim summary statistics depend on the amount of information available. For example, the first analysis is planned when 50% of the maximum event numbers for one endpoint has been observed. The proposed method can be applied to information fraction as well. Table B.4 in Appendix B summarizes the statistics required for information-based designs including the corresponding calendar time, variance, and information fraction for one endpoint relative to information fraction for the other endpoint.

5 Discussion

A single primary endpoint may or may not provide a comprehensive picture of the important effects of the intervention. For this reason, many investigators prefer to design clinical trials with more than one primary endpoint (Dmitrienko et al. 2009). Multiple primary endpoints offer an attractive design feature as they capture a more complete characterization of the effect of an intervention on short and long term outcomes. For example, the Ambassador trial (NCT03244384) was designed to test the effect of pembrolizumab on overall survival and disease-free survival in patients with bladder cancer. In addition, it is common in oncology trials to use two primary endpoints to study the effect of treatment in different patient populations. For example, SWOG S0819 (Herbst et al. 2018) was designed to test the effect of cetuximab plus chemotherapy on overall survival in all patients with lung cancer and to study the impact of the combination therapy on progression-free survival in patients who were EGFR positive. However, for both multiple primary and co-primary endpoints, it is non-trivial to control the Type I and Type II errors when the endpoints are correlated. Evaluating an impact of the correlations among the endpoints is important, in design and analysis of clinical trials with multiple endpoints. Although methodologies to address continuous or binary endpoints in fixed-sample designs are well-developed, methodologies for event-time endpoints are limited (Halabi 2012; Rauch et al. 2016), especially in a group-sequential setting.

In this paper, we discuss a basic theory and method for group-sequential design in clinical trials with two non-fatal event-time outcomes. We present the asymptotic form and computing method of the variance-covariance function for the two sets of group-sequential weighted logrank statistic, which is fundamental for determining the information fraction for each outcome and for evaluating the probability of rejecting the null hypotheses. Several authors have developed many methods for group-sequential designs. However, in the context

of comparing co-primary or multiple endpoints between groups, the form of the asymptotic variance-covariance matrix has not been provided based on the data correlation structure among two event-times. The description of the multivariate central limit theorem and the Cramer-Wald device by some authors did not clearly provide the asymptotic form of the variance-covariance matrix and the connection with a martingale approach, which cause challenges when calculating a power and the required sample size for a trial design. Although the covariance form similar to Theorem 1 has been reported in Murray (2000) and Andrei and Murray (2005), their contexts are different from ours and in paired logrank statistics on the same time axis. When two martingale components with event-time outcomes are correlated on different time axes, it is difficult to directly apply the standard martingale theory for survival analysis. We overcome these difficulties by deriving the two-dimensional Volterra integral equation using the discrete Ito formula (Jacod and Shiryaev 2003) within a martingale approach, which is provided in Appendix A as the proof of Theorem 1. From the simulation result, the asymptotic distribution of Theorem 1 works well in most practical situations as long as the event rate or sample size is not so small.

We apply the asymptotic result to group-sequential methodology for monitoring both or one of the event-time outcomes, when the trial is designed to evaluate a joint effect on both outcomes. There are several advantages for our developed methods. First, they provide an approach to determine the information and information fraction for two event-time outcomes. Second, these methods present the opportunity of evaluating the relationship between two event-time endpoints and how it impacts the decision-making for rejecting the null hypothesis, in terms of the Type I error, power, sample size and number of events. Finally, the methods provide insights on how to optimally choose a strategy for monitoring two event-time endpoints. We outline the method for calculating the probability, sample size, and number of events for the method, and illustrate the methods using a clinical trial example in HIV. Under a calculated total maximum sample size for a joint effect on two outcomes, the monitoring method achieves the targeted power and adequately controls the Type I error. The empirical Type I error rate was evaluated using Monte-Carlo simulation, and the methods presented here are valid in other practical situations. The objectives of the methods are to incorporate the correlation between the two event-time outcomes in power, Type I error evaluation and sample size calculation and to investigate how they behave as the correlation varies. The strength and shape of the association may be estimated from external or internal pilot data, but are usually unknown.

We discuss the situation where both event-time outcomes are non-fatal. Sugimoto et al. (2017) discussed the fixed-sample design when one event is fatal, and when both are fatal. An extension of their work to a group-sequential setting will require an extensive study to modify the variance-covariance structure of the group-sequential logrank statistics in order to handle dependent censoring. Research on group-sequential designs under such situations is an important area for future studies.

Acknowledgements

We thank one reviewer and the Associate Editor for their comments. Research reported in this publication was supported by JSPS KAKENHI grant numbers JP17K00054 and JP17K00069, the Project Promoting Clinical Trials for Development of New Drugs (18lk0201061h0002/18lk0201061h0202) from the Japan Agency for Medical

Research and Development (AMED) and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number UM1AI068634. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix

A Proof of Theorem 1

Let $M_{ik}^{(\ell)}(t) = N_{ik}^{(\ell)}(t) - \int_0^t Y_{ik}^{(\ell)}(x) d\Lambda_{g_i k}(x)$ and $\{\mathcal{F}_{k,t}^{(\ell)} : t \geq 0\}$ be a standard filtration generated from the history through time t for the kth outcome and the ℓth analysis ($\mathcal{F}_{k,t}^{(\ell)}$ is the smallest σ-algebra generated by $\{N_{ik}^{(\ell)}(x), N_{ik}^{C(\ell)}(x) : 0 \leq x \leq t, i = 1, \cdots, n_\ell\}$, where $N_{ik}^{C(\ell)}(t) = \mathbb{1}\{T_{ik}^{(\ell)} \leq t, \Delta_{ik}^{(\ell)} = 0\}$ is a censoring counting process). As is well-known, $M_{ik}^{(\ell)}(t)$ has the $\mathcal{F}_{k,t}^{(\ell)}$-martingale property. We discuss the asymptotic behavior using the decomposition of the weighted logrank process $U_k^{(\ell)}(t) = \hat{m}_k^{(\ell)}(t) + n_\ell^{-\frac{1}{2}} \mathcal{M}_k^{(\ell)}(t)$ from the definition of $U_k^{(\ell)}$, where

$$\mathcal{M}_k^{(\ell)}(t) = \int_0^t \hat{H}_k^{(\ell)}(x) n_\ell^{\frac{1}{2}} \sum_{i=1}^{n_L} \mathrm{d}\widetilde{M}_{ik}^{(\ell)}(x) = \int_0^t \hat{H}_k^{(\ell)}(x) n_\ell^{\frac{1}{2}} \mathrm{d}\widetilde{M}_{\bullet k}^{(\ell)}(x),$$
$$\mathrm{d}\widetilde{M}_{\bullet k}^{(\ell)}(x) = \frac{\mathrm{d}\overline{M}_{2k}^{(\ell)}(x)}{\overline{Y}_{2k}^{(\ell)}(x)} - \frac{\mathrm{d}\overline{M}_{1k}^{(\ell)}(x)}{\overline{Y}_{1k}^{(\ell)}(x)}, \quad \mathrm{d}\overline{M}_{jk}^{(\ell)}(x) = \sum_{i=1}^{n_\ell} \mathbb{1}\{g_i = j\} \mathrm{d}M_{ik}^{(\ell)}(x),$$
$$\mathrm{d}\widetilde{M}_{ik}^{(\ell)}(x) = \mathbb{1}\{i \leq n_\ell\} \left\{ \frac{\mathbb{1}\{g_i = 2\}}{\overline{Y}_{2k}^{(\ell)}(x)} - \frac{\mathbb{1}\{g_i = 1\}}{\overline{Y}_{1k}^{(\ell)}(x)} \right\} \mathrm{d}M_{ik}^{(\ell)}(x),$$

and $\mathcal{M}_k^{(\ell)}(t)$ is $\mathcal{F}_{k,t}^{(\ell)}$-martingale because $\hat{H}_k^{(\ell)}(t)$ is $\mathcal{F}_{k,t}^{(\ell)}$-predictable.

Let $\hat{\boldsymbol{Z}}^* = (\hat{Z}_1^*(\tau_1), \ldots, \hat{Z}_1^*(\tau_L), \hat{Z}_2^*(\tau_1), \ldots, \hat{Z}_2^*(\tau_L))^{\mathrm{T}}$ and let $\hat{Z}_k^*(\tau_\ell)$ be $\hat{Z}_k(\tau_\ell)$ whose denominator is replaced by the limit version,

$$\hat{Z}_k^*(\tau_\ell) = n_\ell^{\frac{1}{2}} \frac{U_k^{(\ell)}(\tau_\ell)}{\sqrt{V_{kk}^{0(\ell)}(\tau_\ell)}} = n_\ell^{\frac{1}{2}} \hat{\mu}_{k\ell} + \xi_{k\ell} \mathcal{M}_k^{(\ell)}(\tau_\ell),$$

where we write $\xi_{k\ell} = 1 / \sqrt{V_{kk}^{0(\ell)}(\tau_\ell)}$ for simplicity. The distribution of $\hat{\boldsymbol{Z}} - D_{\boldsymbol{n}} \hat{\boldsymbol{\mu}}$ is asymptotically equivalent to

$$\hat{\boldsymbol{Z}}^* - D_{\boldsymbol{n}} \hat{\boldsymbol{\mu}} = \left(\xi_{11} \mathcal{M}_1^{(1)}(\tau_1), \ldots, \xi_{1L} \mathcal{M}_1^{(L)}(\tau_L), \xi_{21} \mathcal{M}_2^{(1)}(\tau_1), \ldots, \xi_{2L} \mathcal{M}_2^{(L)}(\tau_L) \right)^{\mathrm{T}}$$

because the dominated convergence theorem works by the convergence of $\widehat{V}_{kk}^{0(\ell)}(\tau_\ell) \xrightarrow{P} V_{kk}^{0(\ell)}(\tau_\ell)$ uniformly on $\ell = 1,\ldots, L$ as $n_L \geq \cdots \geq n_1 \to \infty$. We find it necessary to study the covariance of $\mathscr{M}_k^{(\ell)}$s for characterizing the distribution of $\widehat{\boldsymbol{Z}}^* - D_n\widehat{\boldsymbol{\mu}}$.

In the proof hereafter, it is sufficient to consider the case of $L = 2$. As a function related to the characteristic function of $\mathscr{M}_k^{(\ell)}(t)$, define

$$G_k^{(\ell)}(t) = \exp\left(\boldsymbol{i} z_{k\ell}\mathscr{M}_k^{(\ell)}(t) + \frac{z_{k\ell}^2}{2}\langle \mathscr{M}_k^{(\ell)}, \mathscr{M}_k^{(\ell)}\rangle(t)\right)$$

for a real non-zero $z_{k\ell}$ and $\boldsymbol{i} = \sqrt{-1}$, where $\langle m_1, m_2\rangle$ denotes a predictable covariance process for two martingales m_1 and m_2. In this case we have

$$\langle \mathscr{M}_k^{(\ell)}, \mathscr{M}_k^{(\ell')}\rangle(t) = n_\ell^{\frac{1}{2}} n_{\ell'}^{\frac{1}{2}} \int_0^t \widehat{H}_k^{(\ell)}(x)\widehat{H}_k^{(\ell')}(x)\left\{\frac{\mathrm{d}\Lambda_{1k}(x)}{\bar{Y}_{1k}^{(\ell\vee\ell')}(x)} + \frac{\mathrm{d}\Lambda_{2k}(x)}{\bar{Y}_{2k}^{(\ell\vee\ell')}(x)}\right\},$$

following the standard martingale theory of survival analysis (see Fleming and Harrington (1991)). The consistency of $\hat{S}_{jk}^{(\ell)}$, the Glivenko-Cantelli theorem, and Conditions 1 and 3 imply $\sup_{0\leq x\leq \tau_\ell} | \widehat{H}_k^{(\ell)}(x) - H_k^{(\ell)}(x) | \xrightarrow{P} 0$ and

$$\sup_{0\leq x\leq \tau_\ell} \left| \widehat{H}_k^{(\ell)}(x) / n_\ell^{-1}\bar{Y}_{jk}^{(\ell)}(x) - h_{jk}^{(\ell)}(x) \right| \xrightarrow{P} 0 \text{ as } n_\ell \to \infty, \tag{9}$$

where

$$h_{jk}^{(\ell)}(x) = \frac{H_k^{(\ell)}(x)}{a_{j\ell}y_{jk}^{(\ell)}(x)} = W_k^{(\ell)}(x)\frac{a_{j'\ell}S_{j'k}(x_-)}{S_{\bullet k}^{(\ell)}(x_-)}, \quad j' = 3 - j,$$

and note that $0 \leq H_k^{(\ell)}(x) < \infty$ for $x \in [0, \tau_\ell]$, $H_k^{(\ell)}(x) = 0$ for $\tau_\ell < x$ and $0 \leq h_{jk}^{(\ell)}(x) < \infty$ for all x. The univariate asymptotic result provides $E(\mathrm{e}^{\boldsymbol{i} z_{k\ell}\mathscr{M}_k^{(\ell)}(t)}) \to \exp(-\frac{z_{k\ell}^2}{2}V_{kk}(t, t \mid \tau_\ell, \tau_\ell))$ as $n_\ell \to \infty$, which corresponds to the following convergences,

$$E(G_k^{(\ell)}(t)) \to 1 \text{ and } \langle \mathscr{M}_k^{(\ell)}, \mathscr{M}_k^{(\ell')}\rangle(t) \xrightarrow{P} V_{kk}(t, t \mid \tau_\ell, \tau_{\ell'})$$

(Nishiyama 2011). For different $k \equiv k'$, it is difficult to show joint normality with correlation between $\mathscr{M}_k^{(\ell)}$ and $\mathscr{M}_{k'}^{(\ell)}$ with standard martingale theory of counting processes (Fleming and Harrington 1991; Andersen et al. 1993). However, we overcome the challenge applying Ito's

formula. The discrete Ito's formula (Jacod and Shiryaev 2003; Huang and Strawderman 2006) provides the decomposition of $G_k^{(\ell)}(t)$,

$$\begin{aligned} G_k^{(\ell)}(t) - 1 = & \sum_{j=1,2} \int_0^t G_k^{(\ell)}(x-)\widetilde{H}_{jk}^{\mathrm{a}(\ell)}(x)\mathrm{d}\overline{M}_{jk}^{(\ell)}(x) \\ & + \sum_{j=1,2} \int_0^t G_k^{(\ell)}(x-)\widetilde{H}_{jk}^{(\ell)}(x)\bar{Y}_{jk}^{(\ell)}(x)\mathrm{d}\Lambda_{jk}(x), \end{aligned} \tag{10}$$

where, with $\boldsymbol{i}_1 = -\boldsymbol{i}$ and $\boldsymbol{i}_2 = \boldsymbol{i}$,

$$\widetilde{H}_{jk}^{\mathrm{a}(\ell)}(x) = \exp\left(\boldsymbol{i}_j z_{k\ell} \frac{\sqrt{n_\ell}\widehat{H}_k^{(\ell)}(x)}{\bar{Y}_{jk}^{(\ell)}(x)}\right) - 1,$$

$$\widetilde{H}_{jk}^{(\ell)}(x) = \exp\left(\boldsymbol{i}_j z_{k\ell} \frac{\sqrt{n_\ell}\widehat{H}_k^{(\ell)}(x)}{\bar{Y}_{jk}^{(\ell)}(x)}\right) - 1 - \boldsymbol{i}_j z_{k\ell} \frac{\sqrt{n_\ell}\widehat{H}_k^{(\ell)}(x)}{\bar{Y}_{jk}^{(\ell)}(x)} + \frac{z_{k\ell}^2}{2}\left(\frac{\sqrt{n_\ell}\widehat{H}_k^{(\ell)}(x)}{\bar{Y}_{jk}^{(\ell)}(x)}\right)^2.$$

The expectation of the right-hand side of (10) converges to zero as $n_\ell \to \infty$, because

$$\begin{aligned} & E\left(\int_0^t G_k^{(\ell)}(x-)\widetilde{H}_{jk}^{\mathrm{a}(\ell)}(x)\mathrm{d}\overline{M}_{jk}^{(\ell)}(x)\right) = 0 \\ \text{and } & E\left(\int_0^t G_k^{(\ell)}(x-)\widetilde{H}_{jk}^{(\ell)}(x)\bar{Y}_{jk}^{(\ell)}(x)\mathrm{d}\Lambda_{jk}(x)\right) \to 0 \end{aligned} \tag{11}$$

by the martingale property of $\overline{M}_{jk}^{(\ell)}$ and the Lindeberg condition, respectively. In fact, using the integrable martingale property of $G_k^{(\ell)}(x_-)$ and the well-known inequality

$$|\exp(\boldsymbol{i}c) - 1 - \boldsymbol{i}c + \frac{1}{2}c^2| \le \mathbb{1}\{|c| \le \varepsilon\}|c|^3 + \mathbb{1}\{|c| > \varepsilon\}|c|^2$$

for any real c, the latter result of (11) is obtained as

$$\begin{aligned} & E\left(\int_0^t \left| G_k^{(\ell)}(x_-)\widetilde{H}_{jk}^{(\ell)}(x) \right| \bar{Y}_{jk}^{(\ell)}(x)\mathrm{d}\Lambda_{jk}(x)\right) \\ & \le \exp\left(\frac{z_{k\ell}^2}{2}\langle \mathcal{M}_k^{(\ell)}, \mathcal{M}_k^{(\ell)}\rangle(t)\right) \times \\ & \quad \left\{ E\left(\int_0^t |c_{jk\ell}(x)|^3 \mathbb{1}\{|c_{jk\ell}(x)| \le \varepsilon\}\bar{Y}_{jk}^{(\ell)}(x)\mathrm{d}\Lambda_{jk}(x)\right) \right. \\ & \quad \left. + E\left(\int_0^t |c_{jk\ell}(x)|^2 \mathbb{1}\{|c_{jk\ell}(x)| > \varepsilon\}\bar{Y}_{jk}^{(\ell)}(x)\mathrm{d}\Lambda_{jk}(x)\right)\right\} \to 0 \end{aligned}$$

as $n_\ell \to \infty$, where ε is an arbitrary positive number, $c_{jk\ell}(x) = z_{k\ell}\sqrt{n_\ell}\widehat{H}_k^{(\ell)}(x) / \bar{Y}_{jk}^{(\ell)}(x)$ and we have $\sqrt{n_\ell}c_{jk\ell}(x) \xrightarrow{P} z_{k\ell}h_{jk}^{(\ell)}(x)$ uniformly on $(0, \tau_d]$ from (9). Hence, we have

$$E((G_1^{(\ell)}(t)-1)(G_2^{(\ell')}(s)-1)) \to E(G_1^{(\ell)}(t)G_2^{(\ell')}(s))-1 \tag{12}$$

as $n_\ell\, n_{\ell'} \to \infty$ by the univariate results of $E(G_k^{(\ell)}(t)) \to 1$, while using the formula (10) we can also find

$$\begin{aligned} &E((G_1^{(\ell)}(t)-1)(G_2^{(\ell')}(s)-1)) \to \\ &\sum_{j=1,2}\int_0^t\int_0^s E\Big(G_1^{(\ell)}(x_-)G_2^{(\ell')}(y_-)\widetilde{H}_{j1}^{\mathrm{a}(\ell)}(x)\widetilde{H}_{j2}^{\mathrm{a}(\ell')}(y)\mathrm{d}\overline{M}_{j1}^{(\ell)}(x)\mathrm{d}\overline{M}_{j2}^{(\ell')}(y)\Big) \end{aligned} \tag{13}$$

as $n_\ell\, n_{\ell'} \to \infty$. Similarly to showing the latter result of (11), with asymptotic equality, we can replace the terms $\mathrm{e}^{\boldsymbol{i}j(c_{j1\ell}(x)+c_{j2\ell'}(y))}$ and $\mathrm{e}^{\boldsymbol{i}jc_{jk\ell}(\cdot)}$ included in (13) by $1+\boldsymbol{i}_j\{c_{j1\ell}(x)+c_{j2\ell'}(y)\}-\frac{1}{2}\{c_{j1\ell}(x)+c_{j2\ell'}(y)\}^2$ and $1+\boldsymbol{i}_jc_{jk\ell}(\cdot)-\frac{1}{2}c_{jk\ell}(\cdot)^2$, respectively. In fact, we can show that

$$\begin{aligned} \widetilde{H}_{j1}^{\mathrm{a}(\ell)}(x)\widetilde{H}_{j2}^{\mathrm{a}(\ell')}(y) &= \mathrm{e}^{\boldsymbol{i}j(c_{j1\ell}(x)+c_{j2\ell'}(y))}-\mathrm{e}^{\boldsymbol{i}jc_{j1\ell}(x)}-\mathrm{e}^{\boldsymbol{i}jc_{j2\ell'}(y)}+1 \\ &= -c_{j1\ell}(x)c_{j2\ell'}(y)+o_P(1/\sqrt{n_\ell n_{\ell'}}) \end{aligned}$$

from the convergence result of $\sqrt{n_\ell}c_{jk\ell}(x)$. Hence, we have

$$\sqrt{n_\ell n_{\ell'}}\widetilde{H}_{j1}^{\mathrm{a}(\ell)}(x)\widetilde{H}_{j2}^{\mathrm{a}(\ell')}(y) \xrightarrow{P} -z_{1\ell}z_{2\ell'}h_{j1}^{(\ell)}(x)h_{j2}^{(\ell')}(y) \tag{14}$$

as $n_\ell\, n_{\ell'} \to \infty$, so that we can apply this result to (13). Also, similar to Prentice and Cai (1992) and Sugimoto et al. (2013, 2017), we can show

$$\begin{aligned} \frac{1}{\hat{a}_{j\ell\wedge\ell'}n_{\ell\wedge\ell'}}E\Big(\iint \mathrm{d}\overline{M}_{j1}^{(\ell)}(x)\mathrm{d}\overline{M}_{j2}^{(\ell')}(y)\Big) &= E\Big(\iint \mathrm{d}M_{i1}^{(\ell)}(x)\mathrm{d}M_{i2}^{(\ell')}(y)\mid g_i=j\Big) \\ &= \iint C_{\ell\wedge\ell'}(x\wedge y)A_j(\mathrm{d}x,\mathrm{d}y)\,. \end{aligned}$$

For simplicity, let $\phi(t,s)=E(G_1^{(\ell)}(t)G_2^{(\ell')}(s))$. From (12), (13), (14), $\hat{\gamma}_\ell \xrightarrow{P} \gamma_\ell$, $\hat{a}_{j\ell} \xrightarrow{P} a_{j\ell}$ (Conditions 1-2) and the dominated convergence theorem, we have the integral equation for $\phi(t,s)$ under $n_\ell\, n_{\ell'} \to \infty$,

$$\begin{aligned} \phi(t,s)-1 = &-z_{1\ell}z_{2\ell'}\sqrt{\frac{\gamma_{\ell\wedge\ell'}}{\gamma_{\ell\vee\ell'}}}\times \\ &\int_0^t\int_0^s \phi(x_-,y_-)\sum_{j=1}^2 a_{j\ell\wedge\ell'}h_{j1}^{(\ell)}(x)h_{j2}^{(\ell')}(y)C_{\ell\wedge\ell'}(x\wedge y)A_j(\mathrm{d}x,\mathrm{d}y)\,. \end{aligned} \tag{15}$$

Similarly to bivariate survival function (Dabrowska 1988), the two-dimensional Volterra integral equation

$$\phi(t,s) = 1 + \int_0^t \int_0^s \phi(x_-, y_-) b_{12}(\mathrm{d}x, \mathrm{d}y) \text{ with } \phi(t,0) = \phi(0,s) = 1$$

is solved as $\phi(t,s) = \exp[\int_0^t \int_0^s \{b_{12}(\mathrm{d}x, \mathrm{d}y) - b_1(\mathrm{d}x, y) b_2(x, \mathrm{d}y)\}]$, where

$$b_1(\mathrm{d}x, y) = \phi(\mathrm{d}x, y) \,/\, \phi(x_-, y_-) \text{ and } b_2(x, \mathrm{d}y) = \phi(x, \mathrm{d}y) \,/\, \phi(x_-, y_-)\,.$$

However, note that it is difficult to obtain $b_k(x, y)$, $k = 1, 2$ by directly differentiating (15) because of including the expectation of non-differentiable $M_{i1}^{(\ell)}(x)$ and $M_{i2}^{(\ell')}(y)$. Alternatively, we can use the formula (10) again for the purpose, so that by the discussion similar to obtaining (15), as $n_\ell\, n_{\ell'} \to \infty$, we have

$$\begin{aligned} \int \phi(\mathrm{d}x, y) &= \int \left\{ E\left(\mathrm{d}G_1^{(\ell)}(x) \mathrm{d}G_2^{(\ell')}(y_-)\right) + E\left(\mathrm{d}G_1^{(\ell)}(x) G_2^{(\ell')}(y_-)\right) \right\} \\ &\to \int \phi(x_-, y_-) E\left(\sum_j \widetilde{H}_{j1}^{\mathrm{a}(\ell)}(x) \mathrm{d}\overline{M}_{j1}^{(\ell)}(x) \right) = 0\,. \end{aligned}$$

This yields $\int\int b_1(\mathrm{d}x, y) b_2(x, \mathrm{d}y) = 0$. Hence, the solution of (15) is

$$\phi(t,s) = \exp\left(- z_{1\ell} z_{2\ell'} V_{12}(t, s \mid \tau_\ell, \tau_{\ell'}) \right).$$

Therefore, if $E(\iint \mathrm{d}\overline{M}_{j1}^{(\ell)}(x) \mathrm{d}\overline{M}_{j2}^{(\ell')}(y)) \neq 0$, the correlation between the two martingales works, which results in $E(G_1^{(\ell)}(t) G_2^{(\ell')}(s)) \neq 1$ but concludes

$$E(G_1^{(\ell)}(t) G_2^{(\ell')}(s)) \phi(t,s)^{-1} \to 1 \text{ as } n_L \geq \cdots \geq n_1 \to \infty\,.$$

In summary, these results provide that the characteristic function of marginal martingale vector $(\mathscr{M}_k^{(\ell)}(t), \mathscr{M}_{k'}^{(\ell')}(s))^{\mathrm{T}}$ converges to that of bivariate normal distribution as

$$\begin{aligned} &E\left(\mathrm{e}^{\mathbf{i} z_{k\ell} \mathscr{M}_k^{(\ell)}(t) + \mathbf{i} z_{k'\ell'} \mathscr{M}_{k'}^{(\ell')}(s)} \right) \\ &\to \exp\left(- \frac{1}{2} z_{k\ell}^2 V_{kk}(t, s \mid \tau_\ell, \tau_\ell) - z_{k\ell} z_{k'\ell'} V_{kk'}(t, s \mid \tau_\ell, \tau_{\ell'}) - \frac{1}{2} z_{k'\ell'}^2 V_{k'k'}(t, s \mid \tau_{\ell'}, \tau_{\ell'}) \right) \\ &= \begin{cases} \exp\left(- 2 z_{k\ell}^2 V_{kk}(t, s \mid \tau_\ell, \tau_\ell) \right) & \text{if } k = k', \ell = \ell', \\ \exp\left(- \frac{1}{2} \{ z_{k\ell} V_{kk}(t, s \mid \tau_\ell, \tau_\ell)^{1/2} + z_{k\ell'} V_{kk}(t, s \mid \tau_{\ell'}, \tau_{\ell'})^{1/2} \}^2 \right) & \text{if } k = k', \ell \neq \ell', \\ \text{same as the above form} & \text{otherwise.} \end{cases} \end{aligned}$$

A replication of the similar discussion provides that $(\mathscr{M}_1^{(1)}(t), \mathscr{M}_1^{(2)}(t), \mathscr{M}_2^{(1)}(t), \mathscr{M}_2^{(2)}(t))$ converges in distribution to a multivariate normal distribution with zero mean vector and covarince matrix

$$\begin{pmatrix} V_{11}(t,s \mid \tau_1,\tau_1), \\ V_{11}(t,s \mid \tau_2,\tau_1), V_{11}(t,s \mid \tau_2,\tau_2), \\ V_{21}(t,s \mid \tau_1,\tau_1), V_{21}(t,s \mid \tau_1,\tau_2), V_{22}(t,s \mid \tau_1,\tau_1), \\ V_{21}(t,s \mid \tau_2,\tau_1), V_{12}(t,s \mid \tau_2,\tau_2), V_{22}(t,s \mid \tau_2,\tau_1), V_{22}(t,s \mid \tau_2,\tau_2) \end{pmatrix}.$$

These results lead imidiately to the convergence of $\widehat{\boldsymbol{Z}}^* - \boldsymbol{D_n}\widehat{\boldsymbol{\mu}}$ in distibution to $\boldsymbol{Z}^* - \boldsymbol{D_n}\boldsymbol{\mu}$, as summarized in Theorem 1. □

B Some additional results

Table 2 of Sect. 4 displays the results obtained under the assumption of a late time-dependent association (Clayton copula) for the joint survival distribution of the two event-time outcomes. The users may be interested in how the results change if the other types of dependency between two outcomes are assumed. In Table B.1, we provide results from the design stage calculated under the same assumptions as Table 2 except that the joint survival distribution is replaced by an early time-dependent association (Gumbel copula). The pattern of the results of MSS, MEN and AEN under Gumbel copula are quite similar to Table 2, but, as the correlation is higher, their reduction rates from the values at zero correlation are slightly larger than those under Clayton copula.

Table B.1

Sample sizes, number of events, and empirical powers in a group-sequential trial with two co-primary outcomes under an early time-dependent association (Gumbel copula).

Corr. ρ_j	*FSS	Group-sequential design					Empirical power (%)			
			MEN		AEN		Both EP	At least one EP	Single EP	
		MSS	OC1	OC2	OC1	OC2			OC1	OC2
0.0	830	835	168	335	141	293	80.6	99.3	95.3	84.5
0.1	824	829	166	332	139	290	80.6	99.0	95.2	84.4
0.2	818	823	165	330	138	289	80.4	98.6	95.2	83.9
0.3	812	817	164	327	137	286	80.5	98.1	94.9	83.7
0.4	805	810	163	325	136	285	80.5	97.7	94.8	83.4
0.5	799	804	161	322	134	282	80.6	97.3	94.6	83.3
0.6	792	797	160	319	133	280	80.3	96.7	94.2	82.7
0.7	786	791	159	317	132	279	80.6	96.3	94.2	82.7
0.8	780	785	158	315	132	277	80.3	96.0	94.1	82.2
0.9	776	781	157	313	131	276	80.7	95.8	94.2	82.3
0.95	775	780	157	313	131	276	80.4	95.6	94.0	82.0

* FSS:Sample sizes required for fixed-sample design.

This table is created under the same settings and descriptions as those of Table 2 except the association between two outcomes OC1 and OC2. The joint survival distribution is modeled using the Gumbel copula which provides an early time-dependent association.

As indicated by one referee, an important matter of concern is how the Type I error rates are controlled or not. In fact, the proposed design method is based on asymptotic results. To answer such a problem, we evaluate the behavior of the actual Type I error rates under sample sizes calculated by the proposed methods. Using ARDENT study, we consider three settings of (ψ_1, ψ_2) = (1.0, 1.0), (0.565, 1.0) and (1.0, 0.721) (both null hypotheses and the two marginals) under the same configurations as Sect. 4, and we confirm the behavior via Monte-Carlo simulation with 1,000,000 runs. For the simulation, a trial ended at the planned follow-up duration. When the observed numbers were larger than the planned ones, the critical value at the final analysis was recalculated based on

$$1 - P(Z_{k1} < c_{k1}, \ldots, Z_{kL} < \tilde{c}_{kL} \mid H_{0k}) = \alpha_k,$$

where $\tilde{c}_{kL}$ is the critical value at the final analysis, recalculated such that the above equation is satisfied to control the Type I error adequately if the planned numbers are different from the observed ones.

Tables B.2 and B.3 show the results of the actual Type I error rates, which are corresponding to the situations under null hypotheses of Tables 2 and B.1 under Clayton and Gumbel copulas, respectively. Where the columns “Both” and “ALO” give the probabilities to reject two null hypotheses of OC1 and OC2 jointly (Both) and at least one (ALO), respectively, and “OC1” and “OC2” provide the probabilities to reject two single hypotheses of OC1 and OC2, respectively. We observe that the results of “Joint” are well controlled at the nominal error rate 2.5% in the three cases. Those of “ALO” are less than 2 × 2.5% only at both null hypotheses and reflect the effect of multiplicity using two times testing. Also, the results of “OC1” and “OC2’ are well controlled at the nominal Type I error rate in three cases. Therefore, our method works well in controlling the nominal Type I error rate under the calculated sample size.

Table B.2

Simulation assessment: Probability of rejecting null hypothesis under Clayton copula.

Corr. ρ_j	MSS	(ψ_1, ψ_2) = (0.565, 1.0)				(ψ_1, ψ_2) = (1.0, 0.721)				(ψ_1, ψ_2) = (1.0.1.0)			
		Both	ALO	OC1	OC2	Both	ALO	OC1	OC2	Both	ALO	OC1	OC2
0.0	835	2.39	95.4	95.3	2.50	2.10	85.0	2.49	84.6	0.06	4.91	2.48	2.49
0.1	833	2.39	95.4	95.3	2.50	2.15	84.8	2.50	84.5	0.08	4.94	2.50	2.52
0.2	832	2.42	95.3	95.3	2.51	2.20	84.7	2.51	84.4	0.08	4.91	2.49	2.51
0.3	831	2.44	95.3	95.2	2.51	2.24	84.6	2.50	84.4	0.10	4.92	2.51	2.52
0.4	829	2.44	95.2	95.2	2.50	2.27	84.6	2.51	84.3	0.12	4.90	2.50	2.52
0.5	827	2.43	95.2	95.1	2.48	2.30	84.3	2.50	84.2	0.15	4.87	2.53	2.49
0.6	825	2.47	95.1	95.1	2.51	2.33	84.3	2.49	84.2	0.18	4.80	2.46	2.51
0.7	821	2.48	95.0	95.0	2.51	2.38	84.0	2.49	83.9	0.23	4.80	2.51	2.52
0.8	816	2.52	94.9	94.9	2.53	2.45	83.8	2.52	83.7	0.30	4.69	2.49	2.51
0.9	806	2.51	94.7	94.4	2.52	2.48	83.3	2.51	83.3	0.45	4.50	2.48	2.47

Corr. ρ_j	MSS	$(\psi_1, \psi_2) = (0.565, 1.0)$				$(\psi_1, \psi_2) = (1.0, 0.721)$				$(\psi_1, \psi_2) = (1.0.1.0)$			
		Both	ALO	OC1	OC2	Both	ALO	OC1	OC2	Both	ALO	OC1	OC2
0.95	797	2.51	94.5	94.5	2.51	2.46	82.8	2.47	82.8	0.62	4.39	2.51	2.50

Table B.3

Simulation assessment: Probability of rejecting null hypothesis under Gumbel copula.

Corr. ρ_j	MSS	$(\psi_1, \psi_2) = (0.565, 1.0)$				$(\psi_1, \psi_2) = (1.0, 0.721)$				$(\psi_1, \psi_2) = (1.0.1.0)$			
		Both	ALO	OC1	OC2	Both	ALO	OC1	OC2	Both	ALO	OC1	OC2
0.0	835	2.40	95.4	95.3	2.51	2.11	84.9	2.51	84.5	0.06	4.94	2.49	2.51
0.1	829	2.44	95.3	95.2	2.50	2.26	84.5	2.50	84.2	0.12	4.89	2.51	2.50
0.2	823	2.46	95.0	95.0	2.49	2.36	84.1	2.50	84.0	0.18	4.84	2.51	2.51
0.3	817	2.47	94.9	94.9	2.49	2.42	83.8	2.50	83.8	0.27	4.75	2.50	2.52
0.4	810	2.51	94.8	94.8	2.52	2.47	83.6	2.51	83.5	0.35	4.63	2.48	2.50
0.5	804	2.52	94.6	94.6	2.52	2.50	83.2	2.52	83.2	0.47	4.55	2.53	2.50
0.6	797	2.51	94.5	94.5	2.51	2.49	82.8	2.49	82.8	0.57	4.42	2.50	2.49
0.7	791	2.48	94.3	94.3	2.48	2.52	82.6	2.52	82.6	0.69	4.27	2.49	2.47
0.8	785	2.52	94.1	94.1	2.52	2.49	82.3	2.50	82.3	0.82	4.17	2.50	2.49
0.9	781	2.51	94.0	94.0	2.51	2.50	82.1	2.50	82.1	0.95	4.06	2.49	2.52
0.95	780	2.52	94.0	94.0	2.52	2.51	82.1	2.51	82.1	0.99	4.05	2.51	2.54

Table 1 of Sect. 4 displays the planning information for a group-sequential design at the fixed analysis time points (48 and 96 weeks) considered in ARDENT trial. Other group-sequential designs based on selected information fractions can be constructed. Table B.4 displays the planning information for a group-sequential design for information fractions of 0.5 and 1.0.

Table B.4

Variance, calendar time and information fraction corresponding to the other endpoint's information fraction

Endpoint	Variance, corresponding calendar time and information fraction		1st analysis	Final analysis
Virologic failure (OC1)	information fraction		0.5	1.0
	Corresponding Calendar time (week)		45.5	96.0
	OC1	$V_{11}^{0(\ell)}(\tau_\ell)$	0.0252	0.0499
	OC2 (Regimen failure)	$V_{22}^{0(\ell)}(\tau_\ell)$	0.0539	0.0998
		Corresponding information fraction	0.5400	1.0
Regimen failure (OC2)	information fraction		0.5	1.0
	Corresponding Calendar time (week)		42.0	96.0

Endpoint	Variance, corresponding calendar time and information fraction		1st analysis	Final analysis
	OC1 (Virologic failure)	$V_{11}^{0(\ell)}(\tau_\ell)$	0.0233	0.0499
		Corresponding information fraction	0.4675	1.0
	OC2	$V_{22}^{0(\ell)}(\tau_\ell)$	0.0502	0.0998

References

Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer-Verlag, New York.

Andrei A-C, Murray S (2005) Simultaneous group sequential analysis of rank-based and weighted Kaplan-Meier tests for paired censored survival data. Biometrics 61:715–720. [PubMed: 16135022]

Asakura K, Hamasaki T, Sugimoto T, Hayashi K, Evans SR, Sozu T (2014) Sample size determination in group-sequential clinical trials with two co-primary endpoints. Stat Med 33:2897–2913. [PubMed: 24676799]

Clayton DG (1976) A model for association in bivariate life tablesand its application in epidemiological studies of familial tendencyin chronic disease. Biometrika 65:141–151.

Collett D (2003) Modelling survival data in medical research, 2nd edn. Chapman & Hall/CRC, Boca Raton.

Cook RJ, Farewell VT (1994) Guidelines for monitoring efficacy and toxicity responses in clinical trials. Biometrics 50:1146–1152. [PubMed: 7786995]

Dabrowska DM (1988) Kaplan-Meier estimate on the plane. Ann Stat 16:1475–1489.

Dmitrienko A, Tamhane AC, Bretz F (2009) Multiple Testing Problems in Pharmaceutical Statistics. Chapman & Hall/CRC, Boca Raton.

Fleming TR, Harrington DP (1991) Counting process and survival analysis. John Wiley & Sons, New York.

Glimm E, Mauer W, Bretz F (2009) Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med 29:219–228.

Gombay E (2008) Weighted logrank statistics in sequential tests. Sequential Anal 27:97–104.

Gordon LKK, Lachin JM (1990) Implementation of group sequential logrank tests in a maximum duration trial. Biometrika 46:759–770.

Gu MG, Lai TL (1991) Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. Ann Stat 19:1403–1433.

Halabi S (2012) Adjustment on the type I error rate for a clinical trial monitoring for both intermediate and primary endpoints. J Biom Biostat 7:15. [PubMed: 24466469]

Herbst RS, Redman MW, Kim ES, Semrad TJ, Bazhenova L, Masters G, Oettel K, Guaglianone P, Reynolds C, Karnad A, Arnold SM, Varella-Garcia M, Moon J, Mack PC, Blanke CD, Hirsch FR, Kelly K, Gandara DR (2018) Cetuximab plus carboplatin and paclitaxel with or without bevacizumab versus carboplatin and paclitaxel with or without bevacizumab in advanced NSCLC (SWOG S0819): a randomised, phase 3 study. Lancet Oncol 19: 101–114. [PubMed: 29169877]

Hamasaki T, Asakura K, Evans SR, Sugimoto T, Sozu T (2015) Group-sequential strategies in clinical trials with multiple co-primary endpoints. Stat Biopharm Res 7:36–54. [PubMed: 25844122]

Hougaard P (1986) A class of multivariate failure time distribution. Biometrika 73:671–678.

Hsu L, Prentice RL (1996) On assessing the strength of dependency between failure time variables. Biometrika 83:491–506.

Huang X, Strawderman RL (2006) A note on the Breslow survival estimator. J Nonparametr Stat 18:45–56.

Hung HMJ, Wang SJ, O'Neill RT (2007) Statistical considerations for testing multiple endpoints in group sequential or adaptive clinical trials. J Biopharm Stat 17:1201–1210. [PubMed: 18027226]

Jacod J, Shiryaev AN (2003) Limit theorems for stochastic processes, 2nd edn. Springer-Verlag, Berlin-Heidelberg.

Jennison C, Turnbull BW (2000) Group sequential methods with applications to clinical trials. Chapman & Hall/CRC, Boca Raton.

Jung S-H (2008) Sample size calculation for the weighted rank statistics with paired survival data. Stat Med 27:3350–3365. [PubMed: 18205148]

Kosorok MR, Shi Y, DeMets DL (2004) Design and analysis of group-sequential clinical trials with multiple primary endpoints. Biometrics 60:134–145. [PubMed: 15032783]

Lai TL, Shih M-C (2004) Power, sample size and adaptation considerations in the design of group sequential clinical trials. Biometrika 91:507–528.

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

Lennox JL, Landovitz RJ, Ribaudo HJ, Ofotokun I, Na LH, Godfrey C, Kuritzkes DR, Sagar M, Brown TT, Cohn SE, McComsey GA, Aweeka F, Fichtenbaum CJ, Presti RM, Koletar SL, Haas DW, Patterson KB, Benson CA, Baugh BP, Leavitt RY, Rooney JF, Seekins D, Currier JS (2014) A phase III comparative study of the efficacy and tolerability of three non-nucleoside reverse transcriptase inhibitor-sparing antiretroviral regimens for Treatment-naïve HIV-1-infected volunteers: A randomized, controlled trial. Annals of Internal Medicine 161:461–471. [PubMed: 25285539]

Lin DY, Shen L, Ying Z, Breslow NE (1996) Group sequential designs for monitoring survival probabilities. Biometrics 52:1033–1041. [PubMed: 8805766]

Lin DY (1991) Nonparametric sequential testing in clinical trials with incomplete multivariate observations. Biometrika 78:123–131.

Murray S (2000) Nonparametric rank-based methods for group sequential monitoring of paired censored survival data. Biometrics 54:984–990.

Nishiyama Y (2011) Statistical analysis by the theory of martingales. Kindaikagakusha, Tokyo. (in Japanese)

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

Pocock ST, Geller NL, Tsiatis AA (1987) The analysis of multiple endpoints in clinical trials. Biometrics 43:487–498. [PubMed: 3663814]

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

Prentice RL, Cai J (1992) Covariance and survivor function estimation using censored multivariate failure time data. Biometrika 79:495–512.

Rauch G, Schüler S, Wirths M, Stefan E, Kieser M (2016) Adaptive designs for two candidate primary time-to-event endpoints. Stat Biopharm Res 8:207–216.

Slud EV, Wei LJ (1982) Two-sample repeated significance tests based on the modified Wilcoxon statistic. J Am Stat Assoc 77: 862–868.

Sugimoto T, Hamasaki T, Sozu T, Evans SR (2017) Sizing clinical trials when comparing bivariate time- to-event outcomes. Stat Med 36:1363–1382. [PubMed: 28120524]

Sugimoto T, Sozu T, Hamasaki T, Evans SR (2013) A logrank test-based method for sizing clinical trials with two co-primary time-to-event endpoints. Biostatistics 14:409–421. [PubMed: 23307913]

Tamhane AC, Mehta CR, Liu L (2010) Testing a primary and secondary endpoint in a group sequential design. Biometrics 66:1174–1184. [PubMed: 20337631]

Tamhane AC, Wu Y, Mehta C (2012) Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. Stat Med 31:2027–2040. [PubMed: 22729929]

Tang DI, Gnecco C, Geller NL (1989) Design of group sequential clinical trials with multiple endpoints. J Am Stat Assoc 84:776–779.

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

Tsiatis AA (1982) Group sequential methods for survival analysis with staggered entry In Survival Analysis (eds., Crowley J and Johnson RA), Hayward, California: IMS Lecture Notes, 257–268.

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

Wei LJ, Lachin JM (1984) Two-sample asymptotically distribution-free tests for imcomplete multivariate observations. J Am Stat Assoc 79:653–661.

Wu J, Xiong X (2017) Group-sequential survival trial design and monitoring using the log-rank test. Stat Biopharm Res 9:35–43. [PubMed: 28966722]

Yin G (2012) Clinical trial design: Bayesian and frequentist adaptive methods. John Wiley & Sons, New York.

Table 1

Calculated information *fractions* and the corresponding O'Brien-Fleming-type (OF) and Pocock-type (PC) critical boundaries.

Anal-ysis #	Calen-dar Time	OC1			OC2		
		Information Fraction	OF-type Bound	PC-type Bound	Information Fraction	OF-type Bound	PC-type Bound
1	48	0.5314	2.8616	2.1390	0.5669	2.7576	2.1200
2	96	1.0000	1.9718	2.2110	1.0000	1.9761	2.2215

Table 2

Sample sizes, number of events, and empirical powers in a group-sequential trials with two co-primary outcomes (Clayton copula).

Corr ρ_j	*FSS	Group-sequential design					Empirical power (%)			
			MEN		AEN		Both EP	At least one EP	Single EP	
		MSS	OC1	OC2	OC1	OC2			OC1	OC2
0.0	830	835	168	335	141	293	80.6	99.3	95.3	84.6
0.1	829	833	167	334	140	292	80.5	99.2	95.2	84.5
0.2	827	832	167	333	140	291	80.4	99.2	95.2	84.4
0.3	826	831	167	333	140	291	80.7	99.1	95.3	84.5
0.4	824	829	166	332	139	291	80.5	99.0	95.2	84.3
0.5	822	827	166	331	139	290	80.6	99.0	95.1	84.3
0.6	820	825	166	331	139	290	80.6	99.0	95.1	84.2
0.7	816	821	165	329	138	288	80.5	98.5	95.0	83.9
0.8	811	816	164	327	137	287	80.5	98.2	95.0	83.7
0.9	801	806	162	323	136	284	80.3	97.5	94.6	83.2
0.95	792	797	160	319	134	280	80.4	96.8	94.4	82.8

*FSS:Sample sizes required for fixed-sample design.

The trial is designed to evaluate if an intervention is superior to the control with respect to both virologic (OC1) and regimen failure (OC2) with 80% power at the 2.5% significance level of a one-sided logrank test, where two analyses are planned at fixed calendar times of 48 and 96 weeks. For both outcomes, the critical boundaries are determined using the Lan-DeMets error-spending method with the O'Brien-Fleming type function. The bivariate exponential distribution is modeled using the Clayton copula. Empirical power is calculated using 100,000 repetitions. The marginal powers for OC1 and OC2 are calculated under a calculated maximum sample size.