ADAPTIVE WEIGHTED LOG-RANK TESTS FOR GROUP SEQUENTIAL CLINICAL TRIALS

Department of Biostatistics & Medical Informatics

Technical Report 144

By Michael R. Kosorok

University of Wisconsin at Madison

Abstract. Over the last two decades, two-sample weighted log-rank statistics have been extensively and effectively used in group sequential clinical trials with time-to-event outcomes. Because the power of these statistics depends on the relationship between the chosen weight function and the actual unknown alternative hypothesis, choosing the best weight function for a given application has remained a challenging issue. Kosorok and Lin (1999) demonstrate that one can form adaptive statistics sensative to a broad collection of alternative hypotheses by considering families of weighted log-rank statistics as indexed by both the weight function and the time scale and then computing continuous functionals of the resulting function indexed stochastic processes. In this paper, we extend these results to the group sequential clinical trial setting with staggered entry of patients. We also demonstrate the new methodology on the Beta-Blocker Heart Attack Trial data.

Key words and phrases. Censored data, Clinical trials, Counting processes, Empirical processes, Martingales.

1 Introduction

Over the last two decades, two-sample weighted log-rank statistics have been extensively and effectively used in group sequential clinical trials with time-to-event outcomes. Because the power of these statistics depends on the relationship between the chosen weight function and the actual unknown alternative hypothesis, choosing the best weight function for a given application has remained a challenging issue. Kosorok and Lin (1999) demonstrate that one can form adaptive two-sample test statistics sensative to a broad collection of alternative hypotheses by considering families of weighted log-rank statistics as indexed by both the weight function and the time scale and then computing continuous functionals of the resulting function indexed stochastic processes. Lin and Kosorok (1999) extend this idea to the general class of Jones and Crowley (1990) which permits evaluating the effect of continuous covariates on survival time. In the present paper, we extend the two-sample results of Kosorok and Lin to the group sequential clinical trial setting with staggered entry of patients. New asymptotic theory is needed to establish the validity of a conditional bootstrap used for constructing critical boundaries. We also demonstrate the new methodology on the Beta-Blocker Heart Attack Trial (hereafter abbreviated BHAT) data.

The Beta-Blocker Heart Attack Trial was one of the first randomized, double-blind, placebo controlled clinical trials to use group sequential interim monitoring. Accrual for the study began June 19, 1978, and continued through October 2, 1981, when the O'Brien-Fleming critical boundary (see O'Brien and Fleming, 1979) was crossed by the log-rank statistic assessing the effect of propranolol hydrochloride on time to death after myocardial infarction. This stopping time was 8.5 months earlier than the scheduled end of the study on June 19, 1982 (Beta-Blocker Heart Attack Trial Research Group, 1981 and 1982; DeMets, Hardy, Friedman, and Lan, 1984). Kosorok and Lin (1999) show that the treatment effect occurs very early on for patients in the BHAT and that the log-rank can be substantially improved upon by using a statistic sensative to early differences.

Since the precise nature of the treatment effect is usually not known before beginning a study, maximin efficiency robust tests (Gastwirth, 1985), or adaptive statistics, such as the maximum statistics of Fleming and Harrington (1991, chapter 7) and Tarone (1981) and the function indexed statistics of Kosorok and Lin, are needed to ensure sensativity to a spectrum of possible alternative hypotheses. Kosorok and Lin also give a detailed review and simulation study comparing these statistics, indicating that certain adaptive statistics, such as Renyi-type supremums (Fleming and Harrington, 1991; and Fleming, Harrington, and O'Sullivan, 1987; Gill, 1980) or function indexed statistics, are the most powerful in many settings. A challenge with implementing these adaptive statistics in group sequential clinical trials is that the limiting joint distributions are not Gaussian after taking supremums, and thus constructing appropriate critical boundaries is not possible with existing methodology.

In what follows, we obtain the necessary distribution theory by using Hoffmann-Jørgensen-Dudley (hereafter abbreviated HJD) weak convergence as outlined in van der Vaart and Wellner (1996) via Pollard's central limit theorem (hereafter abbreviated PCLT) for empirical processes (theorem 10.6, Pollard, 1990). This theorem applies to triangular arrays of independent (within rows) but not identically distributed stochastic processes. This level of generality is needed because the clinical trial design or individual censoring patterns may result in non-identically distributed observations (see, Wei, 1978; and definition 3.1.1, Fleming and Harrington, 1991). This approach to weak convergence is different than the approach taken in either Kosorok and Lin (1999) or Lin and Kosorok (1999). The approach taken in Kosorok and Lin uses classical weak convergence (Billingsly, 1968) in a separable topology closely related to the Skorohod topology. In contrast, Lin and Kosorok (1999) utilize HJD weak convergence but not PCLT for both the statistics of interest and an unconditional multiplier bootstrap but were unable to obtain consistency of the more practical conditional multiplier bootstrap. Lin and Kosorok's approach also required the use of a complex adaptation of Gill's (1980) random time transformation method for handling ties in the failure times. Advantages of the new approach based on HJD weak convergence in the uniform topology and on PCLT include: ties in failure times can be easily accommodated, the uniform topology is simpler to work with in applications, and the results can be readily extended to establish consistency of the conditional multiplier bootstrap for use in conducting inference.

In section 2, the statistical setting and proposed function indexed statistics are introduced. Weak convergence of the proposed statistics in the group sequential setting with staggered entry is demonstrated in section 3. Except for a few special cases, the complexity of the limiting distribution of these statistics precludes obtaining critical regions through analytical means. A conditional multiplier bootstrap can be used for inference. The implementation and theory for this bootstrap are presented in section 4. Practical Monte Carlo methods for obtaining critical regions and conducting hypothesis testing in group sequential clinical trials are described in section 5. The performance of the proposed methods in the BHAT clinical trial are examined in section 6. The paper concludes with a brief discussion in section 7.

2 The statistical setting

A general introduction to the properties of weighted log-rank statistics for right-censored survival data can be found in Fleming and Harrington (1991) and Andersen, Borgan, Gill, and Keiding (1992). Application to group sequential trials with staggered entry of patients is given in Tsiatis (1982). Let $n = n_1 + n_2$, where n_l is the sample size for group l, $l = 1, 2$. Assume that there are J interim looks. For each interim look $j = 1 \ldots J$, define the function indexed weighted log-rank statistic

$$Z_j^n(f,t) \equiv n^{-1/2} \int_0^t f\{b_j^n(s)\} \frac{\overline{Y}_{1j}(s)\overline{Y}_{2j}(s)}{\overline{Y}_{1j}(s)+\overline{Y}_{2j}(s)} \left\{ \frac{d\overline{N}_{1j}(s)}{\overline{Y}_{1j}(s)} - \frac{d\overline{N}_{2j}(s)}{\overline{Y}_{2j}(s)} \right\},$$

where $\overline{N}_{lj}(s)$ and $\overline{Y}_{lj}(s)$ are the number of observed events and number at risk, respectively, at time s for individuals in group l ($l = 1, 2$) and for look j, and where f and

b_j^n are defined as follows. For some positive integer r, $f : [0,1]^r \mapsto [0,1]$ is the function index and is an element of a compact set $\boldsymbol{H}$ satisfying:

ASSUMPTION 1 *$\boldsymbol{H}$ is closed and, for all $f \in \boldsymbol{H}$, all first partial derivatives have L_2-norm bounded by some $K < \infty$, in the following sense:*

$$\max_{\{\rho_1,\ldots,\rho_r\}\in\{0,1\}^r} \left(\int_{[0,1]^r} \left[\prod_{l=1}^{r} \left(\frac{\partial}{\partial x_l} \right)^{\rho_l} f(x_1,\ldots,x_r) \right]^2 dx_1 \cdots dx_r \right)^{1/2} \leq K.$$

Also, $b_j^n : [0,\infty) \mapsto [0,1]^r$ is predictable with respect to the filtration $\mathcal{F}^{nj}$ based on data available at the j'th look, $j = 1 \ldots J$, $n \geq 1$. As we show later, many familiar classes of weighted log-rank statistics fit into the above framework.

The statistics we will actually use are continuous mappings of the variance-standardized $\tilde{Z}_j^n(f,t) \equiv \left\{ \tilde{V}_{jj}^n(f,f,\infty) \right\}^{-1/2} Z_j^n(f,t)$, where

$$\begin{aligned} \tilde{V}_{jk}^n(f,g,t) \quad &\equiv \quad n^{-1} \int_0^t f\{b_j^n(x)\} g\{b_k^n(x)\} \frac{\overline{Y}_{1j}(x)\overline{Y}_{2j}(x)}{\overline{Y}_{1j}(x)+\overline{Y}_{2j}(x)} \\ &\quad \times \left\{ 1 - \frac{\Delta\overline{N}_{1k}(x)+\Delta\overline{N}_{2k}(x)-1}{\overline{Y}_{1k}(x)+\overline{Y}_{2k}(x)-1} \right\} \frac{d\overline{N}_{1k}(x)+d\overline{N}_{2k}(x)}{\overline{Y}_{1k}(x)+\overline{Y}_{2k}(x)}, \end{aligned}$$

for $1 \leq j \leq k \leq J$ and $f, g \in \boldsymbol{H}$. $\tilde{V}^n$ is one of the standard covariance estimators used for weighted log-rank statistics (see Chapter 3, Fleming and Harrington, 1991).

The way survival data arises in the group sequential setting is as follows. For patient $i = 1 \ldots n_l$ in group $l = 1, 2$, let U_{il} denote the calendar time of entry, T_{il} denote the failure time measured from study entry, and C_{il} the time to loss of follow-up measured from study entry. Let $A_1 < A_2 < \ldots < A_J$ be fixed calendar times for interim analyses. Then the observed data at the j'th interim analysis are (X_{ilj}, δ_{ilj}), where $X_{ijl} \equiv T_{il} \wedge ([U_{il} - A_j] \vee 0) \wedge C_{il}$ and $\delta_{ilj} \equiv I\{T_{ij} \leq ([U_{il} - A_j] \vee 0) \wedge C_{il}\}$, $i = 1 \ldots n_l$, $l = 1, 2$, and where $I\{B\}$ is the indicator of B while $x \vee y$ and $x \wedge y$ are the maximum and minimum, respectively, of x and y. Accordingly, $Y_{ilj}(t) \equiv I\{X_{ilj} \geq t\}$, $N_{ilj}(t) \equiv I\{X_{ilj} \leq t, \delta_{ilj} = 1\}$, $\overline{Y}_{lj}(t) \equiv \sum_{i=1}^{n_l} Y_{ilj}(t)$, $\overline{N}_{lj}(t) \equiv \sum_{i=1}^{n_l} N_{ilj}(t)$, and

$$\mathcal{F}_t^{nj} \equiv \sigma\{Y_{ilj}(s), N_{ilj}(s), 0 \leq s \leq t, i = 1 \ldots n_l, l = 1, 2\};$$

where $\sigma\{B\}$ is the smallest σ-field making all of B measurable, and where we will denote $\mathcal{F}^{nj} \equiv \sigma\{\cup_{t\geq 0}\mathcal{F}_t^{nj}\}$. For the above data, we assume the following model (with two submodels) which is an extension of model 1 of Kosorok (1999) to the group sequential setting:

MODEL 1 *For the above data, assume that* $\lim_{n\to\infty} n_l/n = a_l \in (0,1)$, *for* $l = 1,2$. *For* $j = 1\ldots J$ *and* $l = 1,2$, *assume* $([U_{il} - A_j] \vee 0) \wedge C_{il}$ *is stochastically independent of* T_{il} *with distribution* L^n_{ilj} *such that*

$$\sup_{t\in[0,\infty]} \left| n_l^{-1} \sum_{i=1}^{n_l} L^n_{ilj}(t) - L_{lj}(t) \right| \to 0,$$

as $n \to \infty$. *Assume also that one of the following submodels holds:*

(L) For some $\phi : [0,\infty] \mapsto [-c_\phi, c_\phi]$, *where* $0 < c_\phi < \infty$, *some finite* β, *and some baseline hazard* $\Lambda_0(t)$,

$$d\Lambda_1^n(t) = \frac{\exp\left[\frac{\beta\phi(t)}{2\sqrt{n}}\right] d\Lambda_0(t)}{1 + \left(\exp\left[\frac{\beta\phi(t)}{2\sqrt{n}}\right] - 1\right) \Delta\Lambda_0(t)}$$

and

$$d\Lambda_2^n(t) = \frac{\exp\left[-\frac{\beta\phi(t)}{2\sqrt{n}}\right] d\Lambda_0(t)}{1 + \left(\exp\left[-\frac{\beta\phi(t)}{2\sqrt{n}}\right] - 1\right) \Delta\Lambda_0(t)},$$

with $\Lambda_0(0) = 0$.

(F) For $j = 1,2$, $d\Lambda_j^n(\cdot) = d\Lambda_j(\cdot)$ *for all* $n \geq 1$, *with* $\Lambda_1(0) = \Lambda_2(0) = 0$.

REMARK 1 *For each* n, *submodel (L) of model 1 is the discrete proportional hazards alternative and can be reexpressed as*

$$\frac{d\Lambda_1^n(t)/\left(1 - \Delta\Lambda_1^n(t)\right)}{d\Lambda_2^n(t)/\left(1 - \Delta\Lambda_2^n(t)\right)} = \exp\left[\frac{\beta\phi(t)}{\sqrt{n}}\right],$$

which becomes the proportional hazards alternative when Λ_0 *is continuous. This submodel is a convenient choice for comparing Pitman efficiencies of weighted log-rank statistics.*

Remark 2 *Under the assumptions of model 1, standard probability results imply*

$$\sup_{t\in[0,\infty]} \left|n^{-1}\overline{Y}_{lj}(t) - \pi_{lj}(t)\right| \to 0$$

almost surely, as $n \to \infty$*, for some non-increasing* $\pi_{lj} : [0,\infty] \mapsto [0,1]$*,* $l = 1,2$*,* $j = 1\dots J$*.*

Let $\boldsymbol{I}_j \equiv \{t : \pi_{1j}(t) \wedge \pi_{2j}(t) > 0\}$. We make the additional assumption:

Assumption 2 *For* $j = 1\dots J$*,* b_j^n *is* $\mathcal{F}^{nj}$*-predictable for each* $n \geq 1$ *and there exists a function* $b_j : [0,\infty] \mapsto [0,1]^r$ *such that for each closed subinterval of* $\boldsymbol{I}_j$*,* $\boldsymbol{I}' \subset \boldsymbol{I}_j$*,* $\sup_{s\in\boldsymbol{I}'} \left\|b_j^n(s) - b_j(s)\right\| \to 0$ *in probability, as* $n \to \infty$*, where* $\|\cdot\|$ *is the uniform norm and* r *is given in assumption 1.*

Remark 3 *Two examples which satisfy assumption 2 for* $r = 1$ *are the pooled Kaplan-Meier estimator,* $\hat{S}_p^j$*, based on data from both groups available at interim look* j*, and the pooled at-risk estimator* $\hat{\pi}_p^j(t) \equiv (n+1)^{-1}\left[\overline{Y}_{1j}(t) + \overline{Y}_{2j}(t)\right]$*.*

Many well known families of weighted log-rank statistics can be remolded into the above function indexing framework by proper choice of $\boldsymbol{H}$ and $\{b_j^n, j = 1\dots J, n \geq 1\}$. Consider, for example, the large $G^{\rho,\gamma}$ family of weighted log-rank statistics (Harrington and Fleming, 1982), with weights of the form $\{\hat{S}_p^j(t-)\}^\rho\{1 - \hat{S}_p^j(t-)\}^\gamma$. Subsets of the structure $\{G^{\rho,\gamma} : \rho \in \{0\} \cup [\epsilon_1, \tau_1], \gamma \in \{0\} \cup [\epsilon_2, \tau_2]\}$, where $0 < \epsilon_l \leq \tau_l < \infty$, $l = 1,2$, can be made into a function indexed family by letting $r = 2$, taking $b_j^n : [0,\infty) \mapsto [0,1]^2$ to be the map $t \mapsto \left(\left\{\hat{S}_p^j(t-)\right\}^{\epsilon_1}, \left\{1 - \hat{S}_p^j(t-)\right\}^{\epsilon_2}\right)$, and by setting

$$\boldsymbol{H} = \left\{f(x,y) = x^a y^b : a \in \{0\} \cup [1, \tau_1/\epsilon_1], b \in \{0\} \cup [1, \tau_2/\epsilon_2]\right\}.$$

Kosorok and Lin (1999) demonstrate for $J = 1$ that statistics such as $\sup_{f\in\boldsymbol{H}} \left|\tilde{Z}_1^n(f,\infty)\right|$ and $\sup_{f\in\boldsymbol{H}} \sup_{t\in[0,\infty]} \left|\tilde{Z}_1^n(f,\infty)\right|$ can be quite sensative to broad collections of alternative hypotheses.

3 Weak convergence

In this section, we will determine the limiting distribution of $\tilde{Z}_j^n(f,t)$ as a process on $\boldsymbol{D} \equiv [\ell^\infty(\boldsymbol{H} \times [0,\infty])]^J$ under submodel (L) of model 1, where, for a space G, $\ell^\infty(G)$ is the metric space of uniformly bounded real functions on G endowed with the uniform metric. Before giving the main result, we need to define the mean function

$$\mu_j(f,t) \equiv \beta \int_0^t f(b_j(x))\phi(x) \left\{ \frac{a_1 a_2 \pi_{1j}(x)\pi_{2j}(x)}{a_1\pi_{1j}(x) + a_2\pi_{2j}(x)} \right\} [1 - \Delta\Lambda_0(x)]\, d\Lambda_0(x)$$

and the covariance function

$$W_{jk}(f,g,t) \equiv \int_0^t f(b_j(x))g(b_k(x)) \left\{ \frac{a_1 a_2 \pi_{1j}(x)\pi_{2j}(x)}{a_1\pi_{1j}(x) + a_2\pi_{2j}(x)} \right\} [1 - \Delta\Lambda_0(x)]\, d\Lambda_0(x),$$

where $1 \le j \le k \le J$, $f \in \boldsymbol{H}$, and $t \in [0,\infty]$.

The following is the main result of this section:

Theorem 1 *Assume model 1, submodel (L), and assumptions 1 and 2 hold. Then, provided* $\inf_{j\in\{1\ldots J\}, f\in\boldsymbol{H}} W_{jj}(f,f,\infty) > 0$, $\tilde{Z}^n \equiv \left\{\tilde{Z}_1^n, \ldots, \tilde{Z}_J^n\right\}$ *converges HJD weakly in the uniform topology on* $\boldsymbol{D}$ *to a tight Gaussian process* $Z \equiv \{Z_1, \ldots, Z_J\}$, *with mean function*

$$\mathbb{E}[Z_j(f,t)] = [W_{jj}(f,f,\infty)]^{-1/2} \mu_j(f,t) \tag{1}$$

and covariance function

$$\operatorname{cov}[Z_j(f,s),\ Z_k(g,t)] = [W_{jj}(f,f,\infty)W_{kk}(g,g,\infty)]^{-1/2} W_{jk}(f,g,s\wedge t),$$

where $1 \le j \le k \le J$, $f,g \in \boldsymbol{H}$, *and* $s,t \in [0,\infty]$.

Remark 4 *For* $j = 1\ldots J$, *the noncentrality parameter given in (1) has the form* $\beta\xi_j(f,t)$, *where*

$$\xi_j^2(f,\infty) \equiv \frac{\left\{\int_0^\infty f(b_j(x))\,\phi(x) \left\{\frac{a_1a_2\pi_{1j}(x)\pi_{2j}(x)}{a_1\pi_{1j}(x)+a_2\pi_{2j}(x)}\right\} \{1 - \Delta\Lambda_0(x)\}\, d\Lambda_0(x)\right\}^2}{\int_0^\infty \{f(b_j(x))\}^2 \left\{\frac{a_1a_2\pi_{1j}(x)\pi_{2j}(x)}{a_1\pi_{1j}(x)+a_2\pi_{2j}(x)}\right\} \{1 - \Delta\Lambda_0(x)\}\, d\Lambda_0(x)}.$$

$\xi_j^2(f,\infty)$ can be used to compare the Pitman ARE of weighted log-rank statistics. In this case, weighted log-rank statistics with asymptotic weight $f(b_j(\cdot)) \propto \phi(\cdot)$ will have maximum Pitman ARE for the specified (L) submodel (when $\beta \neq 0$) over all weighted log-rank statistics at interim look j.

When $\boldsymbol{H}$ is finite, the proof of theorem 1 follows from standard martingale results; however, new techniques are required when $\boldsymbol{H}$ is infinite as permitted by assumption 1. The following theorem is needed to help establish tightness over $\boldsymbol{H}$:

THEOREM 2 *For each $n \geq 1$, let*

$$X_n(f,\cdot) \equiv \int_0^{(\cdot)} f(b^n(s))dW_n(s), \tag{2}$$

where $b^n : [0,\infty) \mapsto [0,1]^r$ for some positive integer r and $f \in \boldsymbol{H}$, with $\boldsymbol{H}$ satisfying assumption 1. Suppose we also have the following conditions for each $n \geq 1$, where $\tilde{\mathcal{F}}^n = \{\tilde{\mathcal{F}}_t^n, t \geq 0\}$ is a filtration:

(i) W_n is a cadlag (right-continuous with left-hand limits) $\tilde{\mathcal{F}}^n$-adapted process on $[0,u]$, for some $u \in (0,\infty]$, such that

(i.a) W_n is locally of bounded variation on $[0,u]$ and $W_n(0)=0$, $\forall n \geq 1$, and

(i.b) $\exists C < \infty$ such that for any $\tilde{\mathcal{F}}^n$-predictable h^n, with $|h^n| \leq c$, there exists a nonnegative $\tilde{\mathcal{F}}^n$-adapted right-continuous submartingale $G_n(h^n,t)$ such that

$$\left\{\int_0^t h^n(s)dW_n(s)\right\}^2 \leq G_n(h^n,t)$$

$\forall t \in [0,u]$ and $\mathbb{E}[G_n(h^n,u)] \leq c^2C$, $\forall n \geq 1$.

(ii) $\forall n \geq 1$, b^n is an $\tilde{\mathcal{F}}^n$-predictable process mapping from $[0,u]$ to $[0,1]^r$.

Then there exists a sequence of non-negative random variables $\{U_n, n \geq 1\}$—depending stochastically only on $\{G_n, n \geq 1\}$—such that $\sup_{n\geq 1}\mathbb{E}[U_n] < \infty$ and

$$\sup_{t\in[0,u]} |X_n(f,t) - X_n(g,t)| \leq \|f-g\|^{1/4}U_n,$$

for every $f,g \in \boldsymbol{H}$ and all $n \geq 1$, where $\|\cdot\|$ is the uniform norm on $\boldsymbol{H}$.

Proof. $\boldsymbol{H}$ can be shown by standard functional analysis methods to be compact with respect to the uniform metric. Each $f \in \boldsymbol{H}$ can also be shown to have a Fourier representation

$$f(\cdot) = \sum_{\boldsymbol{\rho} \in N_\infty^r} f_{(\boldsymbol{\rho})} \Psi_{\boldsymbol{\rho}}(\cdot),$$

where $\boldsymbol{\rho} \equiv \{\rho_1, \ldots, \rho_r\}$, N_∞^r is the set of all multi-indexes composed of r nonnegative integers,

$$\begin{aligned}
\Psi_{\boldsymbol{\rho}}(\boldsymbol{x}) &\equiv \prod_{l=1}^{r} \psi_{\rho_l}(x_l), \text{ and, for } l = 1 \ldots r, \\
\psi_l(x) &\equiv \begin{cases} \sqrt{2} \sin\left[\left(\frac{2l+1}{2}\right)\pi x\right], & \text{if } l > 0, \\ 1, & \text{if } l = 0. \end{cases}
\end{aligned}$$

One can also show that for some positive constants $c_1, c_2 < \infty$, depending only on r and K,

$$\sum_{\boldsymbol{\rho} \in N_\infty^r} \left(f_{(\boldsymbol{\rho})} - g_{(\boldsymbol{\rho})}\right)^2 \leq c_1 \|f - g\|^2$$

and

$$\sum_{\boldsymbol{\rho} \in N_\infty^r} \frac{\left(f_{(\boldsymbol{\rho})} - g_{(\boldsymbol{\rho})}\right)^2}{\lambda_{(\boldsymbol{\rho})}^2} \leq c_2,$$

where $\lambda_{(\boldsymbol{\rho})} \equiv \prod_{l=1}^{r} \lambda_{\rho_l}$, $\lambda_i = [(2i+1)\pi/2]^{-1}$ for $i > 0$, and $\lambda_0 = 1$. For additional details on the properties of $\boldsymbol{H}$, see Kosorok (1999).

For all $n \geq 1$ and $f, g \in \boldsymbol{H} \cup \{0\}$, we have

$$\begin{aligned}
&\sup_{t \in [0,u]} |X_n(f,t) - X_n(g,t)| \\
&= \sup_{t \in [0,u]} \left| \sum_{\boldsymbol{\rho} \in N_\infty^r} \left[f_{(\boldsymbol{\rho})} - g_{(\boldsymbol{\rho})}\right] \int_0^t \Psi_{(\boldsymbol{\rho})}\left(b^n(s)\right) dW_n(s) \right| \\
&\leq \sup_{t \in [0,u]} \left[\left(\sum_{\boldsymbol{\rho} \in N_\infty^r} \frac{\left\{f_{(\boldsymbol{\rho})} - g_{(\boldsymbol{\rho})}\right\}^2}{\lambda_{(\boldsymbol{\rho})}^{3/2}} \right)^{1/2} \left(\sum_{\boldsymbol{\rho} \in N_\infty^r} \lambda_{(\boldsymbol{\rho})}^{3/2} \left\{ \int_0^t \Psi_{(\boldsymbol{\rho})}\left(b^n(s)\right) dW_n(s) \right\}^2 \right)^{1/2} \right]
\end{aligned}$$

$$
\begin{aligned}
&\leq \left(\sum_{\boldsymbol{\rho} \in N_\infty^r} \frac{\left\{ f_{(\boldsymbol{\rho})} - g_{(\boldsymbol{\rho})} \right\}^2}{\lambda_{(\boldsymbol{\rho})}^{3/2}} \right)^{1/2} \sup_{t \in [0,u]} \left(\sum_{\boldsymbol{\rho} \in N_\infty^r} \lambda_{(\boldsymbol{\rho})}^{3/2} G_n \left\{ \Psi_{(\boldsymbol{\rho})} \left(b^n(\cdot) \right), t \right\} \right)^{1/2} \\
&\leq \|f - g\|^{1/4} U_n,
\end{aligned}
$$

where

$$
U_n \equiv c_1^{1/8} c_2^{3/8} \sup_{t \in [0,u]} \left(\sum_{\boldsymbol{\rho} \in N_\infty^r} \lambda_{(\boldsymbol{\rho})}^{3/2} G_n \left\{ \Psi_{(\boldsymbol{\rho})} \left(b^n(\cdot) \right), t \right\} \right)^{1/2},
$$

by the Cauchy-Schwarz and Holder inequalities.

By conditions (i) and (ii) of theorem 2, $\mathbb{E}\left[G_n \left\{ \Psi_{(\boldsymbol{\rho})} \left(b^n(\cdot) \right), t \right\} \right] \leq 2^r C_y$, since $\sup_{t \in [0,1]} |\psi_l(t)| \leq \sqrt{2}$ for all $l \geq 0$. Also note that

$$
\sum_{\boldsymbol{\rho} \in N_\infty^r} \lambda_{(\boldsymbol{\rho})}^{3/2} = \left(\sum_{l=0}^{\infty} \psi_l^{3/2} \right)^r
$$

and

$$
\sum_{l=0}^{\infty} \lambda_l^{3/2} \leq 1 + \pi^{-3/2} \int_0^\infty \left(\frac{2}{2x+1} \right)^{3/2} dx \leq 2.
$$

Hence $\mathbb{E}[|U_n|] \leq \sqrt{\mathbb{E}[U_n^2]} \leq 2^{r+1} c_1^{1/8} c_2^{3/8} \sqrt{C_y}$ by proposition 2.16(b) of Chapter 2 of Ethier and Kurtz (1986).□

Proof of theorem 1. We will first obtain weak convergence on $[\ell^\infty(\boldsymbol{h} \times [0,\infty])]^J$, for all finite subsets $\boldsymbol{h} \equiv \{f_1, \ldots, f_m\} \subset \boldsymbol{H}$. While theorem 4.2.1 of Gill (1980) would work for this purpose, we present an alternative proof based on empirical process techniques which facilitates establishing consistency of the multiplier bootstrap presented later in section 4. Fix $\boldsymbol{h} \subset \boldsymbol{H}$. Note that

$$
\begin{aligned}
Z_j^n(f,t) &= n^{-1/2} \int_0^t f\{b_j^n(s)\} \frac{\overline{Y}_{1j}(s)\overline{Y}_{2j}(s)}{\overline{Y}_{1j}(s) + \overline{Y}_{2j}(s)} \left\{ \frac{d\overline{M}_{1j}(s)}{\overline{Y}_{1j}(s)} - \frac{d\overline{M}_{2j}(s)}{\overline{Y}_{2j}(s)} \right\} \\
&\quad + n^{-1} \int_0^t f\{b_j^n(s)\} \frac{\overline{Y}_{1j}(s)\overline{Y}_{2j}(s)}{\overline{Y}_{1j}(s) + \overline{Y}_{2j}(s)} n^{1/2} \left\{ d\Lambda_1^n(s) - d\Lambda_2^n(s) \right\} \\
&\equiv \hat{B}_j^n(f,t) + \hat{\mu}_j^n(f,t),
\end{aligned}
$$

where $\overline{M}_{lj}(t) \equiv \sum_{i=1}^{n_l} M_{ilj}(t)$ and $M_{ilj}(t) \equiv N_{ilj}(t) - \int_0^t Y_{ilj}(s) d\Lambda_l^n(s)$, $i = 1 \ldots n_l$, $l = 1, 2$, and $j = 1 \ldots J$. If we define

$$B_j^n(f,t) = n^{-1/2} \int_0^t f\{b_j(s)\} \frac{a_1 a_2 \pi_{1j}(s) \pi_{2j}(s)}{a_1 \pi_{1j}(s) + a_2 \pi_{2j}(s)} \left\{ \frac{d\overline{M}_{1j}(s)}{a_1 \pi_{1j}(s)} - \frac{d\overline{M}_{2j}(s)}{a_2 \pi_{2j}(s)} \right\},$$

then

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in [0,\infty]} \left| \hat{B}_j^n(f,t) - B_j^n(f,t) \right| \right] &\leq \sqrt{\mathbb{E}\left[\sup_{t \in [0,\infty]} \left| \hat{B}_j^n(f,t) - B_j^n(f,t) \right|^2 \right]} \\ &\leq 2\sqrt{\mathbb{E}\left[D_{1j}^n + D_{2j}^n \right]}, \end{aligned}$$

where, for $l = 1, 2$ and $l' \equiv 2 - l + 1$,

$$\begin{aligned} D_{lj}^n &\equiv \int_0^\infty \left\{ f\{b_j^n(s)\} \frac{\overline{Y}_{l'j}(s)}{\overline{Y}_{1j}(s) + \overline{Y}_{2j}(s)} - f\{b_j(s)\} \frac{a_{l'} \pi_{l'j}(s)}{a_1 \pi_{1j}(s) + a_2 \pi_{2j}(s)} \right\}^2 \\ &\quad \times \frac{\overline{Y}_{lj}(s)}{n} \left[1 - \Delta \Lambda_l^n(s) \right] d\Lambda_l^n(s), \end{aligned}$$

by proposition 2.16(b) of Ethier and Kurtz (1986). However,

$$\mathbb{E}\left[D_{lj}^n \right] \leq \mathbb{E}\left[\int_0^\infty \left\{ \frac{\overline{Y}_{l'j}(s)}{\overline{Y}_{1j}(s) + \overline{Y}_{2j}(s)} - \frac{a_{l'} \pi_{l'j}(s)}{a_1 \pi_{1j}(s) + a_2 \pi_{2j}(s)} \right\}^2 \frac{d\overline{N}_{lj}(s)}{n} \right] \to 0,$$

as $n \to \infty$, by standard probability results and the properties of model 1. Standard probability results also verify that $\sup_{t \in [0,\infty]} \left| \hat{\mu}_j^n(f,t) - \mu_j(f,t) \right| \to 0$ in probability, as $n \to \infty$. Thus weak convergence of $\left\{ Z_j^n(f,\cdot), j = 1 \ldots J, f \in \boldsymbol{H} \right\}$, after subtracting the mean process, is equivalent to weak convergence of $\left\{ B_j^n(f,\cdot), j = 1 \ldots J, f \in \boldsymbol{H} \right\}$.

Now, let $\boldsymbol{T} \equiv \{(j,k,t) : 1 \leq j \leq J, 1 \leq k \leq m, t \in [0,\infty]\}$, and, for $\boldsymbol{t} = (j,k,t)$, let $\check{\jmath}(\boldsymbol{t}) \equiv j$, $\check{k}(\boldsymbol{t}) \equiv k$ and $\check{t}(\boldsymbol{t}) \equiv t$ be functions which extract the three components of $\boldsymbol{t}$. Also, for $\boldsymbol{t} \in \boldsymbol{T}$, define $\boldsymbol{B}_n(\boldsymbol{t}) \equiv B_{\check{\jmath}(\boldsymbol{t})}^n(f_{\check{k}(\boldsymbol{t})}, \check{t}(\boldsymbol{t}))$. Note that $\boldsymbol{B}_n(\boldsymbol{t}) \equiv \sum_{i=1}^n \zeta_{ni}(\boldsymbol{t})$, where, for $i = 1 \ldots n_1$,

$$\zeta_{ni}(\boldsymbol{t}) = n^{-1/2} \int_0^{\check{t}(\boldsymbol{t})} f_{\check{k}(\boldsymbol{t})} \left(b_{\check{\jmath}(\boldsymbol{t})}(s) \right) \frac{a_2 \pi_{2\check{\jmath}(\boldsymbol{t})}(s)}{a_1 \pi_{1\check{\jmath}(\boldsymbol{t})}(s) + a_2 \pi_{2\check{\jmath}(\boldsymbol{t})}(s)} dM_{i1\check{\jmath}(\boldsymbol{t})}(s),$$

and, for $i = n_1 + 1 \ldots n$,

$$\zeta_{ni}(\boldsymbol{t}) = -n^{-1/2} \int_0^{\check{t}(\boldsymbol{t})} f_{\check{k}(\boldsymbol{t})} \left(b_{\check{\jmath}(\boldsymbol{t})}(s) \right) \frac{a_1 \pi_{1\check{\jmath}(\boldsymbol{t})}(s)}{a_1 \pi_{1\check{\jmath}(\boldsymbol{t})}(s) + a_2 \pi_{2\check{\jmath}(\boldsymbol{t})}(s)} dM_{(i-n_1),2\check{\jmath}(\boldsymbol{t})}(s).$$

The triangular array $\{\zeta_{ni}, i = 1 \ldots n, n \geq 1\}$ is separable and hence sufficiently measurable by lemmas 1 and 2 of Kosorok (2002) to permit application of the PCLT, provided the remaining conditions of that theorem are satisfied.

First note that each ζ_{ni} is the difference between two nonnegative monotone processes in t with total envelope $F_{ni} \equiv n^{-1/2}\left(1 + \Lambda_i^{n\bullet}(T_{ni})\right)$, where $T_{n1}, \ldots, T_{nn}$ are independent failure times, with the integrated hazard for T_{ni} being $\Lambda_i^{n\bullet}$, and where $\Lambda_i^{n\bullet} = \Lambda_1^n$ for $i = 1 \ldots n_1$ and $\Lambda_i^{n\bullet} = \Lambda_2^n$ for $i = n_1 + 1 \ldots n$. Hence $\{\zeta_{ni}\}$ is manageable with envelopes $\{F_{ni}\}$. Thus, by lemma 2 in the appendix, $\mathbb{E}[F_{ni}^3] \leq 28 n^{-3/2}$, and the Lindeberg conditions for the envelopes are therefore satisfied. Now, for any $\boldsymbol{s}, \boldsymbol{t} \in \boldsymbol{T}$ with $\check{\jmath}(\boldsymbol{s}) \leq \check{\jmath}(\boldsymbol{t})$,

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{B}_n(\boldsymbol{s})\boldsymbol{B}_n(\boldsymbol{t})] \quad = \quad & \int_0^{\check{t}(\boldsymbol{s}) \wedge \check{t}(\boldsymbol{t})} f_{\check{k}(\boldsymbol{s})}\left\{b_{\check{\jmath}(\boldsymbol{s})}(x)\right\} f_{\check{k}(\boldsymbol{t})}\left\{b_{\check{\jmath}(\boldsymbol{t})}(x)\right\} \\
& \times \left(\frac{a_2 \pi_{2\check{\jmath}(\boldsymbol{s})}(x)}{a_1 \pi_{\check{\jmath}(\boldsymbol{s})}(x) + a_2 \pi_{2\check{\jmath}(\boldsymbol{s})}(x)}\right) \\
& \times \left(\frac{a_2 \pi_{2\check{\jmath}(\boldsymbol{t})}(x)}{a_1 \pi_{1\check{\jmath}(\boldsymbol{t})}(x) + a_2 \pi_{2\check{\jmath}(\boldsymbol{t})}(x)}\right) \frac{\overline{Y}_{1\check{\jmath}(\boldsymbol{s})}(x)}{n} d\Lambda_1^n(x) \\
& + \int_0^{\check{t}(\boldsymbol{s}) \wedge \check{t}(\boldsymbol{t})} f_{\check{k}(\boldsymbol{s})}\left\{b_{\check{\jmath}(\boldsymbol{s})}(x)\right\} f_{\check{k}(\boldsymbol{t})}\left\{b_{\check{\jmath}(\boldsymbol{t})}(x)\right\} \\
& \times \left(\frac{a_1 \pi_{1\check{\jmath}(\boldsymbol{s})}(x)}{a_1 \pi_{1\check{\jmath}(\boldsymbol{s})}(x) + a_2 \pi_{2\check{\jmath}(\boldsymbol{s})}(x)}\right) \\
& \times \left(\frac{a_1 \pi_{1\check{\jmath}(\boldsymbol{t})}(x)}{a_1 \pi_{1\check{\jmath}(\boldsymbol{t})}(x) + a_2 \pi_{2\check{\jmath}(\boldsymbol{t})}(x)}\right) \frac{\overline{Y}_{2\check{\jmath}(\boldsymbol{s})}(x)}{n} d\Lambda_2^n(x) \\
\rightarrow \quad & W_{\check{\jmath}(\boldsymbol{s})\check{\jmath}(\boldsymbol{t})}\left(f_{\check{k}(\boldsymbol{s})}, g_{\check{k}(\boldsymbol{t})}, \check{t}(\boldsymbol{s}) \wedge \check{t}(\boldsymbol{t})\right)
\end{aligned} \tag{3}
$$

in probability, as $n \to \infty$, by lemma 1 in the appendix and standard arguments. Since (3) is right continuous in the time arguments and bounded, all of the conditions of the PCLT are satisfied. Thus $\boldsymbol{B}_n$ converges HJD weakly in the uniform topology on $\ell^\infty(\boldsymbol{T})$ to a tight Gaussian process with covariance given by the last term on the right-hand-side of (3).

Now we will use theorem 2 to establish tightness of $Z^n \equiv \{Z_1^n, \ldots, Z_J^n\}$. First note, for $j = 1 \ldots J$, that Z_j^n has the form given in (2) with $W_n(t) = \hat{B}_j^n(1,t) + \hat{\mu}_j^n(1,t)$ and

$\tilde{\mathcal{F}}^n = \mathcal{F}^{nj}$. The only difficult condition of theorem 2 to satisfy in this case is (i.b). However, for any $\mathcal{F}^{nj}$-predictable h^n with $|h^n| \leq c$,

$$\mathbb{E}\left[\left\{\int_0^t h^n(s)\hat{B}_j^n(1, ds)\right\}^2\right] \leq 2c^2,$$

for all $t \in [0, \infty]$ by standard martingale arguments. After some computation, it can be shown that model 1 implies $\sqrt{n}\,|d\Lambda_1^n(t) - d\Lambda_2^n(t)| \leq c' d\Lambda_2^n(t)$ for some $c' < \infty$ depending only on β and c_ϕ. Thus

$$\mathbb{E}\left[\left\{\int_0^t h^n(s)\hat{\mu}_j^n(1, ds)\right\}^2\right] \leq (cc')^2\, \mathbb{E}\left[\{\Lambda_2^n(T_{12})\}^2\right] \leq 2\,(cc')^2,$$

by lemma 2, and all of the conditions of theorem 2 are thereby satisfied. Thus, for every $\epsilon, \eta > 0$, there exists a $\delta > 0$ such that

$$\mathbb{P}\left[\sup_{1\leq j\leq J}\ \sup_{f,g\in \boldsymbol{H}:\|f-g\|\leq\delta}\ \sup_{t\in[0,\infty]}\left|Z_j^n(f,t) - Z_j^n(g,t)\right| > \epsilon\right] < \eta$$

for all $n \geq 1$. Thus, by theorem 1.5.6 of van der Vaart and Wellner (1996), Z^n converges HJD weakly, in the uniform topology on $\boldsymbol{D}$, to a tight Gaussian process $Z^{(0)} \equiv \{Z_1^{(0)}, \ldots, Z_J^{(0)}\}$ with limiting mean function $\mathbb{E}\left[Z_j^{(0)}(f,t)\right] = \mu_j(f,t)$ and limiting covariance $\operatorname{cov}\left[Z_j^{(0)}(f,s),\ Z_k^{(0)}(g,t)\right] = W_{jk}(f,g,s\wedge t)$, for $1 \leq j \leq k \leq n$, $f, g \in \boldsymbol{H}$, and $s, t \in [0, \infty]$.

The proof will be complete, by the version of Slutsky's lemma given as Example 1.4.7 in van der Vaart and Wellner, if we can establish uniform consistency of $\tilde{V}_{jk}^n(f,g,t)$. For any finite subset $\boldsymbol{h} \subset \boldsymbol{H}$, uniform consistency in probability over $1 \leq j \leq k \leq J$, $f, g \in \boldsymbol{h}$, and $t \in [0,\infty]$ follows from standard martingale results. Now we will apply theorem 2 with $W_n(t) = W_n^\bullet(t, h_n^\bullet)$, where

$$W_n^\bullet(t, h_n^\bullet) = \int_0^t h_n^\bullet(s)\frac{\overline{Y}_{1j}(s)\overline{Y}_{2j}(s)}{\overline{Y}_{1j}(s) + \overline{Y}_{2j}(s)}\left\{1 - \frac{\Delta\overline{N}_{1k}(s) + \Delta\overline{N}_{2k}(s) - 1}{\overline{Y}_{1k}(s) + \overline{Y}_{2k}(s) - 1}\right\}\frac{d\overline{N}_{1k}(s) + d\overline{N}_{2k}(s)}{\overline{Y}_{1k}(s) + \overline{Y}_{2k}(s)}$$

and $h_n^\bullet$ is any $\mathcal{F}^{nk}$-predictable process with $|h_n^\bullet| \leq 1$. Since for any $\mathcal{F}^{nk}$-predictable h^n with $|h^n| \leq c$,

$$\left\{\int_0^t h^n(x)W_n^\bullet(dt, h_n^\bullet)\right\}^2 \leq \left\{\int_0^t h^n(x)W_n^\bullet(dt, 1)\right\}^2 \leq c^2,$$

all of the conditions of theorem 2 are satisfied for $W_n(\cdot) = W_n^\bullet(\cdot, h_n^\bullet)$ with

$$G_n(h^n, t) = \int_0^t h^n(x) W_n^\bullet(dt, 1).$$

Thus, for any $f_1, f_2, g_1, g_2 \in \boldsymbol{H}$,

$$\begin{aligned}
\sup_{t\in[0,\infty]} \left|\tilde{V}_{jk}^n(f_1, g_1, t) - \tilde{V}_{jk}^n(f_2, g_2, t)\right| &\leq \sup_{t\in[0,\infty]} \left|\tilde{V}_{jk}^n(f_1, g_1, t) - \tilde{V}_{jk}^n(f_1, g_2, t)\right| \\
&\quad + \sup_{t\in[0,\infty]} \left|\tilde{V}_{jk}^n(f_1, g_2, t) - \tilde{V}_{jk}^n(f_2, g_2, t)\right| \\
&\leq \|g_1 - g_2\|^{1/4} U_n + \|f_1 - f_2\|^{1/4} U_n,
\end{aligned}$$

where $\sup_{n\geq 1} \mathbb{E}[U_n] < \infty$, $1 \leq j \leq k \leq n$, and where U_n does not depend on f_1, f_2, g_1, g_2. Thus we have uniform consistency of $\tilde{V}^n$ and the proof is complete.□

4 Monte Carlo critical regions

To obtain Monte Carlo critical values, let $\boldsymbol{\chi}_n \equiv \{\chi_{il}, i = 1 \ldots n_l, l = 1, 2\}$ be a vector of independent and identically distribued mean zero variance 1 real random variables generated independent of the data (for example, these could be standard normals). We will use multiplier bootstraps $\hat{Z}^n$ to approximate the null distribution of $\tilde{Z}^n$, where "null" means we subtract out terms having nonzero expectation, and $\hat{Z}_j^n(f,t) \equiv \left[V_{jj}^n(f, f, \infty)\right]^{-1/2} \check{Z}_j^n(f, t)$,

$$\begin{aligned}
\check{Z}_j^n(f,t) &\equiv n^{-1/2} \left(\sum_{i=1}^{n_1} \chi_{i1} \int_0^t f\{b_j^n(x)\} \frac{\overline{Y}_{2j}(x)}{\overline{Y}_{1j}(x) + \overline{Y}_{2j}(x)} d\hat{M}_{i1j}(x) \right. \\
&\quad \left. + \sum_{i=1}^{n_2} \chi_{i2} \int_0^t f\{b_j^n(x)\} \frac{\overline{Y}_{1j}(x)}{\overline{Y}_{1j}(x) + \overline{Y}_{2j}(x)} d\hat{M}_{i2j}(x) \right),
\end{aligned}$$

$$
\begin{aligned}
V_{jk}^n(f,g,t) &\equiv n^{-1}\left[\sum_{i=1}^{n_1}\left\{\int_0^t f\{b_j^n(x)\}\frac{\overline{Y}_{2j}(x)}{\overline{Y}_{1j}(x)+\overline{Y}_{2j}(x)}d\hat{M}_{i1j}(x)\right\}\right.\\
&\quad\times\left\{\int_0^t g\{b_k^n(x)\}\frac{\overline{Y}_{2k}(x)}{\overline{Y}_{1k}(x)+\overline{Y}_{2k}(x)}d\hat{M}_{i1k}(x)\right\}\\
&\quad+\sum_{i=1}^{n_2}\left\{\int_0^t f\{b_j^n(x)\}\frac{\overline{Y}_{1j}(x)}{\overline{Y}_{1j}(x)+\overline{Y}_{2j}(x)}d\hat{M}_{i2j}(x)\right\}\\
&\quad\left.\times\left\{\int_0^t g\{b_k^n(x)\}\frac{\overline{Y}_{1k}(x)}{\overline{Y}_{1k}(x)+\overline{Y}_{2k}(x)}d\hat{M}_{i2k}(x)\right\}\right],\quad\text{and}\\
\hat{M}_{ilj}(t) &\equiv N_{ilj}(t)-\int_0^t Y_{ilj}(s)\frac{d\overline{N}_{lj}(s)}{\overline{Y}_{lj}(s)},
\end{aligned}
$$

for $i=1\ldots n_l$, $l=1,2$, and $1\leq j\leq k\leq J$. V^n is an alternative to $\tilde{V}^n$, for estimating the covariance of weighted log-rank statistics, which is also consistent for W under submodel (L) of model 1 but which performs better under submodel (F).

Remark 5 *For either submodel (L) or (F), arguments similar to those used in the proof of theorem 1 can be combined with lemma 1 to establish the uniform consistency in outer probability of V^n for V, where*

$$
\begin{aligned}
(4)\qquad V_{jk}(f,g,t) &\equiv \int_0^t f\left\{b_j(x)\right\}g\left\{b_k(x)\right\}\\
&\quad\times\left(\frac{a_1a_2\pi_{1j}(x)\pi_{2j}(x)}{a_1\pi_{1j}(x)+a_2\pi_{2j}(x)}\right)\left(\frac{a_1a_2\pi_{1k}(x)\pi_{2k}(x)}{a_1\pi_{1k}(x)+a_2\pi_{2k}(x)}\right)\\
&\quad\times\left[\frac{\{1-\Delta\Lambda_1(x)\}d\Lambda_1(x)}{a_1\pi_{1k}(x)}+\frac{\{1-\Delta\Lambda_2(x)\}d\Lambda_2(x)}{a_2\pi_{2k}(x)}\right],
\end{aligned}
$$

for $1\leq j\leq k\leq J$.

Let θ_j^n be the univariate statistic based on a continuous mapping ψ of $\tilde{Z}_j^n$ (such as the supremum over the function index) for interim look j and let $\hat{\theta}_j^n=\psi\left(\hat{Z}_j^n\right)$ be a multiplier bootstrap of the same statistic, $j=1\ldots J$. Critical regions can then be computed for interim analysis $k\leq J$, conditional on not reaching critical values for any interim look $j<k$, based on the α-spending function of choice and on the conditional distribution of the vector $\{\hat{\theta}_1^n,\ldots,\hat{\theta}_k^n\}$ given the data (obtained by generating

many replicates of $\boldsymbol{\chi}_n$). The choice of standard normal multipliers yields a method of inference similar in spirit to that used by Su and Wei (1991).

We now establish the validity of this conditional multiplier bootstrap. For a metric space $\{D, d\}$, let $BL_1(D)$ be the space of real valued functions on D with Lipshitz norm bounded by 1, ie., for any $f \in BL_1(D)$, $\sup_{x \in D} |f(x)| \leq 1$ and $|f(x) - f(y)| \leq d(x, y)$ for all $x, y \in D$. In our setting, d will always be the uniform metric. As described in the discussion following theorem 1.12.2 of van der Vaart and Wellner (1996), we know that X_n converges HJD weakly to X on D, where X is Borel Measurable and separable, if and only if $\sup_{f \in BL_1(D)} |\mathbb{E}^* [f(X_n)] - \mathbb{E}[f(X)]| \to 0$ as $n \to \infty$, where superscript $*$ denotes outer expectation. One of the appealing features of this approach to weak convergence is that it nicely resolves measurability problems for conditional bootstraps. In the following theorem, let $\mathbb{E}_\chi$ denote taking expectation over the multipliers $\boldsymbol{\chi}_n$ conditional on the data, and let $\boldsymbol{D} \equiv [\ell^\infty (\boldsymbol{H} \times [0, \infty])]^J$ as in theorem 1.

THEOREM 3 *Assume model 1 and assumptions 1 and 2 hold with* $\inf_{j \in \{1 \ldots J\}, f \in \boldsymbol{H}} V_{jj}(f, f, \infty) > 0$. *Let* $\boldsymbol{\chi}_n \equiv \{\chi_{il}, i = 1 \ldots n_l, l = 1, 2\}$ *be a vector of independent and identically distribued mean zero variance 1 real random variables generated independent of the observed data. Then* $\hat{Z}^n \equiv \left\{\hat{Z}_1^n, \ldots, \hat{Z}_J^n\right\}$ *is asymptotically measurable with*

$$\sup_{h \in BL_1(\boldsymbol{D})} \left| \mathbb{E}_\chi \left[h\left(\hat{Z}^n\right)\right] - \mathbb{E}[h(Z^\bullet)] \right| \quad \to \quad 0 \tag{5}$$

in outer probability, as $n \to \infty$*; where* $Z^\bullet \equiv \{Z_1^\bullet, \ldots, Z_J^\bullet\}$*, and*

(i) *under submodel (L),* $Z^\bullet = Z$*, with* Z *as defined in theorem 1, and*

(ii) *under submodel (F),* $Z^\bullet$ *is a tight, mean zero Gaussian process with covariance*

$$\operatorname{cov}\left[Z_j^\bullet(f, s),\ Z_k^\bullet(g, t)\right] = [V_{jj}(f, f, \infty) V_{kk}(g, g, \infty)]^{-1/2} V_{jk}(f, g, s \wedge t).$$

Thus, under both submodels, the conditional multiplier bootstrap $\hat{Z}^n$ *is consistent for the limiting null distribution of* $\tilde{Z}^n$.

Proof. Assume the data from model 1 exists on the probability space $\{\Omega, \mathcal{B}, \mathcal{P}\}$ and that $\{\boldsymbol{\chi}_n, n \geq 1\}$ exists on an independent probability space $\{\Omega_\chi, \mathcal{B}_\chi, \mathcal{P}_\chi\}$. We will first justify the lack of any outer expectation in (5). Fix $h \in BL_1(\boldsymbol{D})$. Clearly $h(Z^\bullet)$ is measurable, under either submodel, because $Z^\bullet$ is Borel Measurable, as verified in theorem 1 for submodel (L) and as we prove below for submodel (F). Now fix $\omega \in \Omega$ and $\boldsymbol{c}, \boldsymbol{d} \in \mathbb{R}^n$, with $\boldsymbol{c} = (c_{11}, \ldots, c_{n_1 1}, c_{12} \ldots, c_{n_2 2})$ and $\boldsymbol{d} = (d_{11}, \ldots, d_{n_1 1}, d_{12}, \ldots, d_{n_2 2})$, and denote $\boldsymbol{Z}_c^n$ to be $\{\hat{Z}_1^n, \ldots, \hat{Z}_J^n\}$ after replacing $\boldsymbol{\chi}_n$ with $\boldsymbol{c}$, and similarly define $\boldsymbol{Z}_d^n$ for replacing $\boldsymbol{\chi}_n$ with $\boldsymbol{d}$. Then

$$
\begin{aligned}
|h(\boldsymbol{Z}_c^n) - h(\boldsymbol{Z}_d^n)|^2 \ \leq \ & \left[\sum_{l=1}^{2}\sum_{i=1}^{n_l}(c_{il} - d_{il})^2\right] \times \left[\inf_{f \in \boldsymbol{H}} V_{11}^n(f, f, \infty)\right]^{-1} n^{-1} \\
& \times \sup_{1 \leq j \leq J} \sup_{t \in [0,\infty], f \in \boldsymbol{H}} \left(\sum_{i=1}^{n_1}\left[\int_0^t f\{b_j^n(x)\}\frac{\overline{Y}_{2j}(x)}{\overline{Y}_{1j}(x) + \overline{Y}_{2j}(x)} d\hat{M}_{i1j}(x)\right]^2\right. \\
& \left. + \sum_{i=1}^{n_2}\left[\int_0^t f\{b_j^n(x)\}\frac{\overline{Y}_{1j}(x)}{\overline{Y}_{1j}(x) + \overline{Y}_{2j}(x)} d\hat{M}_{i2j}(x)\right]^2\right),
\end{aligned}
$$

implying that the set $\{\boldsymbol{c} \in \mathbb{R}^n : h(\boldsymbol{Z}_c^n) \leq \alpha\}$ is the closure of a countable set for all $\alpha \in \mathbb{R}$. Thus $h(\hat{Z}^n)$ is a Borel measurable random variable on $\{\Omega_\chi, \mathcal{B}_\chi, \mathcal{P}_\chi\}$.

For $j = 1 \ldots J$, $f \in \boldsymbol{h} \equiv \{f_1, \ldots, f_m\} \subset \boldsymbol{H}$, and $t \in [0, \infty]$, define

$$
\begin{aligned}
B_{\chi j}^n(f, t) \ \equiv \ & n^{-1/2}\int_0^t f\{b_j(x)\}\left(\frac{a_1 a_2 \pi_{1j}(x)\pi_{2j}(x)}{a_1\pi_{1j}(x) + a_2\pi_{2j}(x)}\right) \\
& \times \left\{\frac{\sum_{i=1}^{n_1}\chi_{i1} dM_{i1j}(x)}{a_1\pi_{1j}(x)} - \frac{\sum_{i=1}^{n_2}\chi_{i2} dM_{i2j}(x)}{a_2\pi_{2j}(x)}\right\}
\end{aligned}
$$

and $\tilde{\mathcal{F}}_t^{nj} \equiv \sigma\left\{\mathcal{F}_t^{nj}, \chi_{il}, i = 1 \ldots n_l, l = 1, 2\right\}$, with $\tilde{\mathcal{F}}^{nj} \equiv \sigma\left\{\cup_{t \geq 0}\mathcal{F}_t^{nj}\right\}$. Using proposition 2.16(b) of Ethier and Kurtz (1986) as in the proof of theorem 1, we have that

$$
\mathbb{E}\left[\sup_{t \in [0,\infty]}\left|\check{Z}_j^n(f, t) - B_{\chi j}^n(f, t)\right|\right] \leq 2\sqrt{2\mathbb{E}\left[D_{1j}^n + D_{2j}^n + \hat{D}_{1j}^n + \hat{D}_{2j}^n\right]},
$$

where $\mathbb{E}\left[D_{lj}^n\right] \to 0$, as $n \to \infty$, as demonstrated in the proof of theorem 1, and where

$$
\hat{D}_{lj}^n \equiv \int_0^\infty f^2\left\{b_j^n(x)\right\}\left(\frac{\overline{Y}_{l'j}(x)}{\overline{Y}_{1j}(x) + \overline{Y}_{2j}(x)}\right)^2\left(\frac{\sum_{i=1}^{n_l}\chi_{il}Y_{ilj}(x)}{\overline{Y}_{lj}(x)}\right)^2\frac{\overline{Y}_{lj}(x)}{n}\left[1 - \Delta\Lambda_1^n(x)\right] d\Lambda_1^n(x)
$$

arises from the approximation of M_{ilj} with $\hat{M}_{ilj}$, $i = 1 \ldots n_l$. However, standard martingale arguments establish that $\mathbb{E}\left[\hat{D}^n_{lj}\right] \to 0$, as $n \to \infty$, for $l = 1, 2$ and $j = 1 \ldots J$.

Now let $\boldsymbol{T}$, $\check{j}$, $\check{k}$, $\check{t}$, and $\{\zeta_{ni}, i = 1 \ldots n, n \geq 1\}$ be as defined in the proof of theorem 1; and, for $\boldsymbol{t} \in \boldsymbol{T}$, let

$$\hat{\boldsymbol{B}}_n(\boldsymbol{t}) \equiv \left[V_{\check{j}(\boldsymbol{t})\check{j}(\boldsymbol{t})}(f_{\check{k}(\boldsymbol{t})}, f_{\check{k}(\boldsymbol{t})}, \infty)\right]^{-1/2} \sum_{i=1}^{n} \chi'_{ni}\zeta_{ni}(\boldsymbol{t}),$$

where $\chi'_{ni} = \chi_{i1}$ for $i = 1 \ldots n_1$ while $\chi'_{ni} = \chi_{(i-n_1),1}$ for $i = n_1 + 1 \ldots n$. In the proof of theorem 1, we established that $\sum_{i=1}^{n} \zeta_{ni}(\cdot)$ satisfies the measurability conditions lemmas 1 and 2 of Kosorok (2002) and the PCLT conditions for submodel (L). The validity of the PCLT for submodel (F) follows similarly. Thus, because V is bounded below, all of the conditions of theorem 2 of Kosorok (2002) are satisfied for $\hat{\boldsymbol{B}}_n$—under both submodels—and hence $\hat{\boldsymbol{B}}_n$ is asymptotically measurable and

$$\sup_{h \in BL_1(\ell^\infty(\boldsymbol{T}))} \left|\mathbb{E}_\chi[h(\hat{\boldsymbol{B}}_n)] - \mathbb{E}[h(\boldsymbol{B})]\right| \;\to\; 0, \tag{6}$$

where $\boldsymbol{B}$ is a tight, mean zero Gaussian process with the same covariance as $\left\{Z^\bullet_{\check{j}(\boldsymbol{t})}(f_{\check{k}(\boldsymbol{t})}, \check{t}(\boldsymbol{t})),\right.$ $\left.\boldsymbol{t} \in \boldsymbol{T}\right\}$, and where the distribution of $Z^\bullet$ depends on the submodel as described in the statement of theorem 3.

As mentioned in remark 5, $V^n_{jk}(f, g, t)$ is uniformly consistent in outer probability for $V_{jk}(f, g, t)$ over $1 \leq j \leq k \leq J$, $f, g \in \boldsymbol{H}$, and $t \in [0, \infty]$ for both submodels, where, for submodel (L), $\Lambda_1 = \Lambda_2 = \Lambda_0$ and $V_{jk}(f, g, t) = W_{jk}(f, g, t)$. Now, for each $\delta > 0$, let $\Pi_\delta : \boldsymbol{H} \mapsto \boldsymbol{H}$ be a map which selects a closest element of a finite δ-mesh of $\boldsymbol{H}$, $\boldsymbol{H}_\delta$, with the property $\sup_{f \in \boldsymbol{H}} \inf_{g \in \boldsymbol{H}_\delta} \|f - g\| \leq \delta$. For $z \equiv \{z_1, \ldots, z_J\} \in \ell^\infty(\boldsymbol{D})$, let $z \circ \Pi_\delta \equiv \{z_1 \circ \Pi_\delta, \ldots, z_J \circ \Pi_\delta\}$, where $(z_j \circ \Pi_\delta)(f, t) = z_j(\Pi_\delta(f), t)$, $j = 1 \ldots J$. What we have proven in the previous few paragraphs is that, for every $\delta > 0$, $\hat{Z}^n \circ \Pi_\delta$ is asymptotically measurable and

$$\sup_{h \in BL_1(\boldsymbol{D})} \left|\mathbb{E}_\chi\left[h\left(\hat{Z}^n \circ \Pi_\delta\right)\right] - \mathbb{E}\left[h\left(Z^\bullet \circ \Pi_\delta\right)\right]\right| \to 0$$

in outer probability, as $n \to \infty$, for both submodels.

Now, for $j = 1 \ldots J$ and after some computation, we have for any $\tilde{\mathcal{F}}^{nj}$-predictable h^n with $|h^n| \le c$,

$$\mathbb{E}\left[\left\{\int_0^t h^n(x)\check{Z}_j^n(1, dx)\right\}^2\right] \le 2c^2,$$

for all $t \in [0, \infty]$ by standard martingale arguments combined with the specified properties of $\boldsymbol{\chi}_n$. Thus theorem 2 implies that for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\mathbb{E}\left[\sup_{1\le j\le J}\ \sup_{f,g\in \boldsymbol{H}:\|f-g\|\le\delta}\ \sup_{t\in[0,\infty]}\left|\check{Z}_j^n(f,t) - \check{Z}_j^n(g,t)\right|\right] < \epsilon$$

for all $n \ge 1$ and for both submodels. Thus, by the uniform consistency in outer probability of V^n and the fact that $V_{jk}(f, g, t)$ is bounded below and continuous in f and g, we have that for any sequence $\{r_n\}$ going to zero,

$$\sup_{h\in BL_1(\boldsymbol{D})}\left|\mathbb{E}_\chi\left[h\left(\hat{Z}^n\right)\right] - \mathbb{E}_\chi\left[h\left(\hat{Z}^n\circ\Pi_{r_n}\right)\right]\right| \to 0$$

in outer probability, as $n \to \infty$. These results also imply that $\hat{Z}^n$ is asymptotically measurable and converges weakly, in the uniform topology on $\boldsymbol{D}$, to the tight Gaussian process $Z^\bullet$. Hence also

$$\lim_{\delta\downarrow 0}\ \sup_{h\in BL_1(\boldsymbol{D})}\left|\mathbb{E}\left[h\left(Z^\bullet\right)\right] - \mathbb{E}\left[h\left(Z^\bullet\circ\Pi_\delta\right)\right]\right| \to 0,$$

and the proof is complete.□

5 Application to clinical trials

Assume that the total number of looks J is fixed in advance. As in section 4, let θ_j^n be the univariate statistic based on a selected continuous mapping ψ of $\tilde{Z}_j^n$ for interim look j, and let $\hat{\theta}_j^n = \psi\left(\hat{Z}_j^n\right)$ be a Monte Carlo realization of the same statistic, $j = 1 \ldots J$.

Suppose that an α-spending function has been selected and a method for determining information time at each look has been chosen (see Lan and DeMets, 1983). Critical regions can then be computed for interim analysis k, conditional on not reaching a critical value for any interim look $j < k$, based on this α-spending function and on the conditional distribution of the vector $\{\hat{\theta}_1^n, \ldots, \hat{\theta}_k^n\}$ given the data (obtained via the multiplier bootstrap of section 4). Specifically, let α_j, $j = 1 \ldots J$, be the computed type I error to be spent at each look such that the experiment-wise type I error is $\alpha = \alpha_1 + \cdots + \alpha_J$.

At the time of the first interim analysis, the critical region for the first look, $C_1 \subset \mathbb{R}$, is computed from α_1 and the Monte Carlo distribution of $\hat{\theta}_1^n$. If $\theta_1^n \in C_1$, the clinical trial is stopped and the appropriate alternative hypothesis is accepted. If $\theta_1^n \notin C_1$, then the trial continues to look 2. In general, if the trial has reached look $k \geq 2$, ie., $\theta_j^n \notin C_j$ for all $j < k$, then C_k is determined from the Monte Carlo distribution of $\{\hat{\theta}_1^n, \ldots, \hat{\theta}_k^n\}$ conditional on the observed data by solving $\hat{\mathrm{P}}_M \left\{ \hat{\theta}_k^n \in C_k, \hat{\theta}_j^n \notin C_j, j = 1 \ldots k-1 \right\} = \alpha_k$, where $\hat{\mathrm{P}}_M$ is the empirical probability measure of $\{\hat{\theta}_1^n, \ldots, \hat{\theta}_k^n\}$ based on a large number M of multiplier bootstraps.

An interesting simulation study comparing power between several function indexed statistics to which the proposed methodology is applicable, based on the $G^{\rho,\gamma}$ family of weighted log-rank statistics mentioned earlier, is found in Kosorok and Lin (1999). These simulation studies also demonstrate that type I error is preserved with these statistics for moderate and large sample sizes. Of particular interest are the adaptive statistics

$$\sup_{f \in \boldsymbol{H}} \left| \tilde{Z}_j^n(f, \infty) \right|, \quad \text{and} \tag{7}$$

$$\sup_{f \in \boldsymbol{H}} \sup_{t \geq 0} \left| \tilde{Z}_j^n(f, t) \right|. \tag{8}$$

An additional statistic which holds promise for being sensative to broad stochastic ordering alternatives is

$$\sup_{f \in \boldsymbol{H}} \left| \sup_{t \geq 0} \tilde{Z}_j^n(f,t) + \inf_{t \geq 0} \tilde{Z}_j^n(f,t) \right| . \tag{9}$$

We will examine the proposed methodology applied to these statistics in the next section. We should also point out that the methodology we have developed enables, for the first time as far as we are aware, the use of Renyi-type supremum log-rank statistics (Fleming and Harrington, 1991; Fleming, Harrington, and O'Sullivan, 1987; and Gill, 1980) in group sequential clinical trials with staggered entry of patients.

6 The BHAT example

We now apply the proposed methods to the BHAT clinical trial data described in the introduction. Seven interim analyses were initially planned at six month intervals. It is difficult to precisely duplicate the actual interim monitoring procedure, because the relevant publications do not provide sufficient detail and because lags in data acquisition that existed in the original study are no longer traceable, but we will be able to come as close as possible. The months and years of the six interim analyses actually performed (a seventh was never performed because the study stopped early) are May 1979, October 1979, March 1980, October 1980, April 1981, and October 1981 (on the 2nd, to be exact). We will assume that these analyses occurred on the 19th of each month (the anniversary date), except for the last month wherein the analysis occurred on the 2nd: the timing, in months, of these analyses were then rounded to 11, 16, 21, 28, 34, and 39.5, respectively. The information time was then computed by dividing each of these times by the total planned duration of the study, which was 48 months (DeMets, Hardy, Friedman, and Lan, 1984). The amount of α spent at each look, based on the O'Brien-Fleming spending function and an overall type I error of 0.05, was computed using the group sequential boundary program of Reboussin, DeMets, Kim, and Lan (2000). The resulting α spent at each look was

0.00001, 0.00020, 0.00120, 0.00527, 0.00880, 0.01148, and 0.02304 (the seventh is never used because the study stopped after look 6).

A collection of statistics of the form (7), (8), and (9) are computed at each look, and then the Monte Carlo method of section 4 is used to compute critical values. In this manner, we can determine when the trial (hypothetically) would have been stopped had the statistics studied been used. 100,000 Monte Carlo replications were used to obtain the critical boundaries using the procedure outlined in section 5. The specific statistics we looked at are all based on the $G^{\rho,\gamma}$ family of weights as defined at the end of section 2. Specifically, for a given collection of ρ's and γ's (which collection we refer to in this section as the "function index set"), we evaluate (7), (8), and (9) for $\boldsymbol{H}$ as defined at the end of section 2.

Table 1 gives the stopping looks (the look at which the statistic first crossed the critical boundary) for each of the statistics considered. The function index in the left column gives the set of ρ's and γ's maximized over. Sets of the form $\{\cdot\}$ consist of a single point, while sets of the form $[0, \cdot]$ indicate intervals. For example, $\{0, 0\}$ denotes maximization over the single point $\rho = 0$, $\gamma = 0$, which corresponds to the weight associated with the log-rank statistic. On the other hand, $[0, 4] \times \{0\}$ denotes ranging ρ over the interval $[0, 4]$ while holding $\gamma = 0$. Intervals for ρ are maximized over a resolution of 0.2 units, e.g., the interval $[0, 1]$ consists of the points 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. Intervals for γ are maximized over a resolution of 0.5 units. These approximations greatly simplify computation, and, as discussed in Kosorok and Lin, 1999, this kind of approximation is quite close to the targeted continuous maximization.

We see in table 1 that the supremum type statistics usually stop before the non-supremum statistics, and that having the "right" weight function allows dramatically early stopping (look 4 is almost a full year before look 6). This dramatically early stopping is even more impressive when considering how conservative the O'Brien-Fleming α-spending function is. What is even more impressive is that the adaptive statistic $G(\sup + \inf)$, maximized over $\rho \in [0, 10]$ while holding $\gamma = 0$, also stops at look 4. This clearly indicates that the proposed adaptive statistics can offer a dramatic increase in power, even when the precise optimal weight function is unknown ahead of time. Although the $G(\text{sup})$ statistic is also very successful in this context, a serious concern with this statistic is its sensativity to stochastic crossing alternatives which may not be desirable in a clinical trial context (see Kosorok and Lin, 1999).

Table 2 gives the actual statistics and critical boundaries at each look for several of the most interesting statistics from table 1. The observed values of these statistics and the associated critical boundaries are given in figure 1. The two statistics featured on the lower right-hand-side of this figure both result in the study stopping at look 4 (October 19, 1980), almost a year before the study actually stopped (October 2, 1981) based on the log-rank statistic (displayed in the top left plot).

7 Discussion

While the data analysis of the BHAT study indicates that the proposed adaptive statistics could potentially result in substantial time, mortality, and expense savings in group sequential clinical trials, there are a number of issues that merit further investigation before these methods should be used routinely in clinical trials. One of these issues is that a reasonable range of weights (the "function index set") needs to be selected in advance of any planned study. While the results of section 6 indicate that these sets can be quite large, adding unnecessary sensitivity to late differences in treatment effect on survival can result in a decrease in power. This problem is apparent for the statistics given in the bottom third of table 1, which appear to be overly sensative to late differences. This is obviously a consequence of the strong early treatment effect in the BHAT study. On the other hand, Kosorok and Lin (1999) show that sensitivity to late differences is merited in certain settings.

Another potentially useful but technically challenging extension of the results of this paper would be to allow sequential monitoring of adaptive statistics in continuous time. Unfortunately, the tail instability problem (see Bilias, Gu, and Ying, 1997; Lai and Ying, 1988) is extremely difficult as a consequence of the function indexing.

A Appendix

LEMMA 1 *Let $T < \infty$ be a random failure time with integrated hazard Λ such that $\Lambda(0) = 0$, and let C_1 and C_2 be two random right-censoring times independent of T with $C_1 \leq C_2$. Also let H_1 and H_2 be bounded left-continuous functions on $[0,\infty]$ with right-hand limits. Then, for all $s,t \in [0,\infty]$,*

$$\mathbb{E}\left[\int_0^s H_1(x)dM_1(x)\int_0^t H_2(x)dM_2(x)\right] = \int_0^{s\wedge t} H_1(x)H_2(x)\mathbb{E}\left[Y_1(x)\right]\left[1-\Delta\Lambda(x)\right]d\Lambda(x),$$

where $M_j(t) \equiv N_j(t) - \int_0^t Y_j(x)d\Lambda(x)$, $N_j(t) \equiv I\{T \leq t, T \leq C_j\}$ and $Y_j(t) \equiv I\{T \wedge C_j \geq t\}$, $j = 1,2$.

REMARK 6 *This lemma is a necessary link in establishing the independent increment structure of survival statistics in group sequential clinical trials with staggered entry (see also Slud, 1984; Gu and Lai, 1991; Scharfstein, Tsiatis and Robins, 1997).*

Proof of lemma 1. Let $\mathcal{F}_{1t} \equiv \sigma\left\{Y_1(s+), N_1(s), 0 \leq s \leq t\right\}$ with $\mathcal{F}_1 \equiv \sigma\{\mathcal{F}_{1t}, t \geq 0\}$, and let $\mathcal{F}_{2t} \equiv \sigma\left\{\mathcal{F}_{1s}, Y_2(s+), N_2(s), 0 \leq s \leq t\right\}$ with $\mathcal{F}_2 \equiv \sigma\{\mathcal{F}_{2t}, t \geq 0\}$. For notational clarity, we will suppress the time arguments in the integrals. Now,

$$\begin{aligned}
\mathbb{E}\left[\int_0^s H_1 dM_1 \int_0^t H_2 dM_2\right] &= \mathbb{E}\left[\int_0^{s\wedge t} H_1 dM_1 \int_0^{s\wedge t} H_2 dM_2\right] \\
&\quad + \mathbb{E}\left[\int_{s\wedge t}^s H_1 dM_1 \int_{s\wedge t}^t H_2 dM_2\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_0^{s\wedge t} H_1 dM_1 \int_{s\wedge t}^t H_2 dM_2 \middle| \mathcal{F}_{2,s\wedge t}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_{s\wedge t}^s H_1 dM_1 \int_0^{s\wedge t} H_2 dM_2 \middle| \mathcal{F}_{2,s\wedge t}\right]\right] \\
&= \mathbb{E}\left[\int_0^{s\wedge t} H_1 dM_1 \int_0^{s\wedge t} H_2 dM_2\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_{s\wedge t}^s H_1 dM_1 \int_0^{s\wedge t} H_2 dM_2 \middle| \mathcal{F}_{2,s\wedge t}\right]\right].
\end{aligned}$$

But, in spite of the fact that $\mathcal{F}_{1,s\wedge t} \subset \mathcal{F}_{2,s\wedge t}$, one can show that, in this setting,

$$\mathbb{E}\left[\int_{s\wedge t}^{s} H_1 dM_1 \middle| \mathcal{F}_{2,s\wedge t}\right] = \mathbb{E}\left[\int_{s\wedge t}^{s} H_1 dM_1 \middle| \mathcal{F}_{1,s\wedge t}\right] = 0.$$

Also,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^{s\wedge t} H_1 dM_1 \int_0^{s\wedge t} H_2 dM_2 \middle| \mathcal{F}_{1,s\wedge t}\right] \\
&\quad = \int_0^{s\wedge t} H_1 dM_1 \left(\int_0^{s\wedge t} H_2 dM_1 + \mathbb{E}\left[\int_0^{s\wedge t} H_2 \left(dN_1 - dN_2 - [Y_1 - Y_2] d\Lambda\right) \middle| \mathcal{F}_{1,s\wedge t}\right]\right),
\end{aligned}$$

but the expectation on the far right-hand-side can be shown to be equal to 0 by the properties of counting processes; and the desired results follow after taking an additional expectation and by the fact that the predictable variation of M_1^2 at time t is $\int_0^t Y_1(s)[1 - \Delta\Lambda(s)] d\Lambda(s)$.□

LEMMA 2 *For a nonnegative failure time T with integrated hazard Λ, $\mathbb{E}[\Lambda^r(T)] \leq r!$ for any nonnegative integer r.*

Proof. Let F be the distribution function associated with Λ and let $S = 1 - F$. Then

$$\begin{aligned}
\int_0^t \Lambda^r(s) dF(s) &= \Lambda^r(t)F(t) - \int_0^t F(s-) d\Lambda^r(s) \\
&= \int_0^t S(s-) d\Lambda^r(s) - \int_0^t d\Lambda^r(s) - \Lambda^r(t)S(t) + \Lambda^r(t) \\
&\leq \int_0^t S(s-) d\Lambda^r(s) \\
&\leq r \int_0^t S(s-)\Lambda^{r-1}(s) d\Lambda(s),
\end{aligned}$$

by integration by parts combined with the fact that for nonnegative reals a and b, $(a+b)^r - a^r \leq r(a+b)^{r-1}b$. However,

$$\int_0^t S(s-)\Lambda^{r-1}(s) d\Lambda(s) = \int_0^t \Lambda^{r-1}(s) dF(s),$$

since $\Lambda(t) = \int_0^t S^{-1}(s-) dF(s)$ by definition, and the result follows.□

ACKNOWLEDGEMENTS

The author was partially supported by grant CA75142 from the U.S. National Cancer Institute.

REFERENCES

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding (1992). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Beta-Blocker Heart Attack Trial Research Group (1981). Beta Blocker Heart Attack Trial: Design features. *Controlled Clinical Trials* **2**, 275–285.

Beta-Blocker Heart Attack Trial Research Group (1982). A randomize trial of propranolol in patients with acute myocardial infraction. *Journal of the American Medical Association* **247**, 1707–1714.

Y. Bilias, M. Gu, and Z. Ying (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Annals of Statistics* **25**, 662–682.

P. Billingsley (1968). *Convergence of Probability Measures.* New York: Wiley.

D. L. DeMets, R. Hardy, L. M. Friedman, and K. K. G. Lan (1984). Statistical aspects of early termination in the Beta-Blocker Heart Attack Trial. *Controlled Clinical Trials* **5**, 362–372.

S. N. Ethier and T. G. Kurtz (1986). *Markov Processes.* New York: Wiley.

T. R. Fleming and D. P. Harrington (1991). *Counting Processes and Survival Analysis.* New York: Wiley.

T. R. Fleming, D. P. Harrington, and M. O'Sullivan (1987). Supremum versions of the log-rank and generalized Wilcoxon statistics. *Journal of the American Statistical Association* **82**, 312–320.

J. L. Gastwirth (1985). The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *Journal of the American Statistical Association* **80**, 380–384.

R. D. Gill (1980). *Censoring and Stochastic Integrals*. Volume 124 of *Mathematical Centre Tracts*. Amsterdam: Mathematisch Centrum.

M. G. Gu and T. L. Lai (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Annals of Statistics* **19**, 1403–1433.

D. P. Harrington and T. R. Fleming (1982). A class of rank test procedures for censored survival data. *Biometrika* **69**, 133–143.

M. P. Jones and J. Crowley (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Annals of Statistics* **18**, 1203–1220.

M. R. Kosorok (1999). Families of weighted log-rank statistics as function-indexed stochastic processes. Technical Report 98, Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, Wisconsin.

M. R. Kosorok (2002). Bootstraps of sums of independent but not identically distributed stochastic processes. Technical Report 133, Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, Wisconsin.

M. R. Kosorok and C.-Y. Lin (1999). The versatility of function-indexed weighted log-rank statistics. *Journal of the American Statistical Association* **94**, 320–332.

T. L. Lai and Z. Ying (1988). Stochastic integral of empirical-type processes with applications to censored regression. *Journal of Multivariate Analysis* **27**, 334–358.

K. K. G. Lan and D. L. DeMets (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659–663.

C.-Y. Lin and M. R. Kosorok (1999). A general class of function indexed nonparametric tests for survival analysis. *Annals of Statistics* **27**, 1722–1744.

P. C. O'Brien and T. R. Fleming (1979). A multiple testing procedure for clinical trials. *Biometrics* **33**, 549–556.

D. Pollard (1990). *Empirical Processes: Theory and Applications*. Hayward, CA: IMS.

D. M. Reboussin, D. L. DeMets, K.-M. Kim, and K. K. G. Lan (2000). Computations for group sequential boundaries using the Lan-DeMets spending function method. *Biometrics*, To appear.

D. O. Scharfstein, A. A. Tsiatis, and J. M. Robins (1997). Semiparametric efficiency and its implication on the design and analysis of group sequential studies. *Journal of the American Statistical Association* **92**, 1342–1350.

E. V. Slud (1984). Sequential linear rank tests for two-sample censored survival data. *Annals of Statistics* **12**, 551–571.

J. Q. Su and L. J. Wei (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association* **86**, 420–426.

R. E. Tarone (1981). On the distribution of the maximum of the log-rank statistic and the modified Wilcoxon statistic. *Biometrics* **37**, 79–85.

A. A. Tsiatis (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association* **77**, 855–861.

A. W. van der Vaart and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

L. J. Wei (1978). The adaptive biased coin design for sequential experiments. *Annals of Statistics* **6**, 92–100.

Table 1: Stopping looks (from 1–6) for function indexed statistics applied to the BHAT clinical trial.

		Test Statistics	
Function Index Set	G	$G(\sup)$	$G(\sup+\inf)$
$\{0,0\}$	6	5	5
$\{1,0\}$	6	5	5
$\{4,0\}$	5	4	4
$\{10,0\}$	5	4	4
$\{20,0\}$	4	4	4
$\{40,0\}$	4	4	5
$[0,1]\times\{0\}$	6	5	5
$[0,4]\times\{0\}$	5	4	5
$[0,10]\times\{0\}$	5	4	4
$[0,20]\times\{0\}$	5	4	5
$[0,40]\times\{0\}$	5	5	5
$[0,1]\times[0,1]$	6	5	5
$[0,4]\times[0,1]$	6	5	5
$[0,10]\times[0,1]$	6	5	5
$[0,20]\times[0,1]$	5	5	5
$[0,40]\times[0,1]$	5	5	5

Table 2: Values of several statistics and their associated critical boundaries (in parentheses below) at each interim look. An "$*$" denotes the first time the boundary is crossed.

		Interim Look					
Function Index Set	Statistical Test	1	2	3	4	5	6
$\{0,0\}$	G:	1.73	2.26	2.36	2.27	2.37	2.92*
		(4.71)	(3.65)	(3.18)	(2.75)	(2.49)	(2.32)
$\{0,0\}$	$G(\sup)$:	1.87	2.58	3.12	2.88	3.06*	3.08
		(4.71)	(3.80)	(3.34)	(2.93)	(2.69)	(2.51)
$\{0,0\}$	$G(\sup+\inf)$:	1.59	2.36	2.94	2.82	3.00*	3.02
		(4.71)	(3.74)	(3.30)	(2.88)	(2.60)	(2.41)
$[0,10]\times\{0\}$	G:	1.75	2.33	2.59	2.63	2.73*	3.16
		(4.77)	(3.71)	(3.24)	(2.83)	(2.56)	(2.39)
$[0,10]\times\{0\}$	$G(\sup)$:	1.87	2.58	3.17	3.06*	3.24	3.37
		(4.77)	(3.83)	(3.40)	(3.01)	(2.76)	(2.58)
$[0,10]\times\{0\}$	$G(\sup+\inf)$:	1.59	2.36	2.96	2.97*	3.15	3.29
		(4.77)	(3.78)	(3.36)	(2.95)	(2.68)	(2.50)

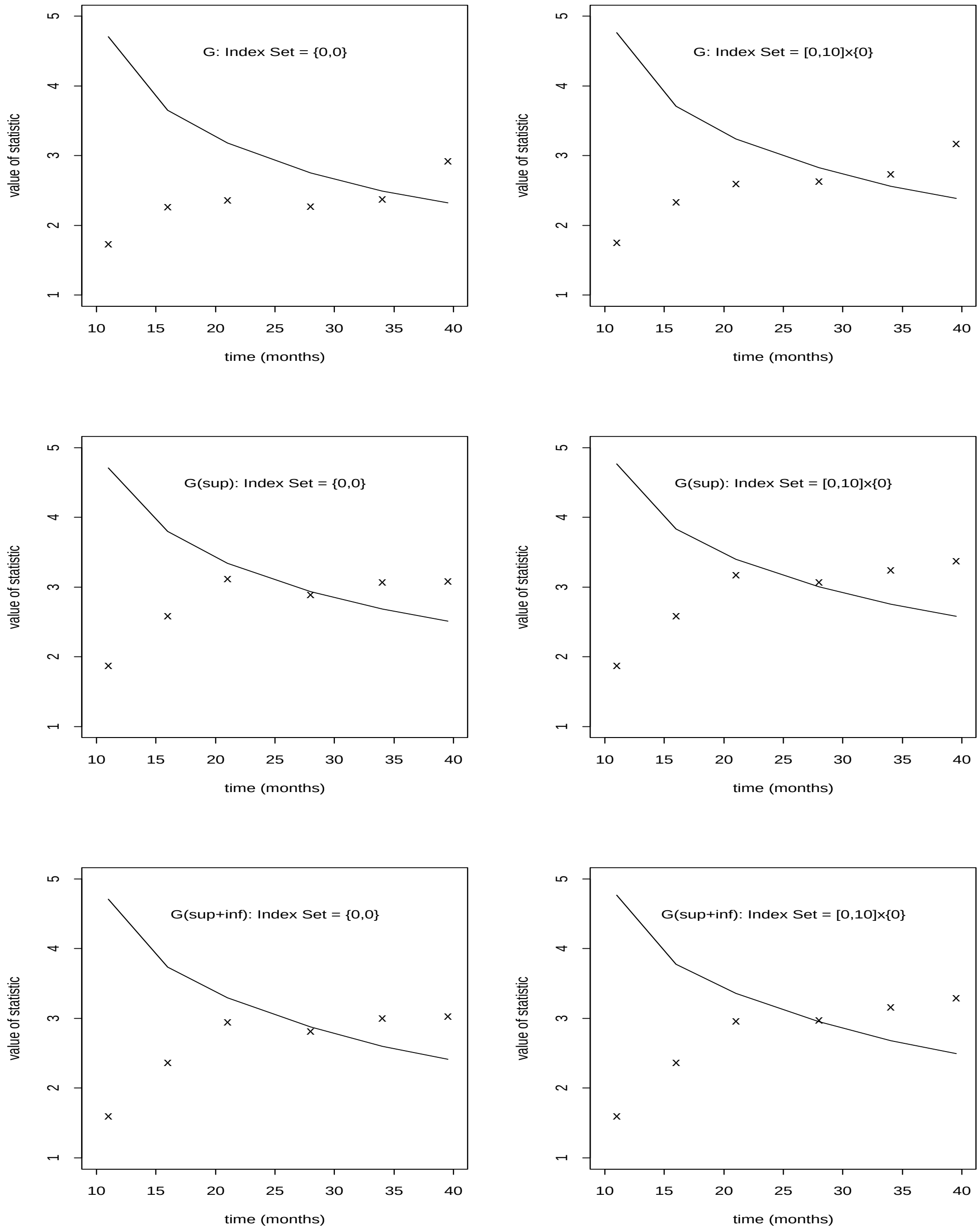

Figure 1: Interim analysis procedures applied to several different statistics. The type of statistic and function index set used are given inside the plot. The observed values of the statistics are marked with an "x" while solid lines give the critical boundaries.