Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 76 (2006) 304–309
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Resolving the tail instability in weighted log-rank statistics for clustered survival data

Michael R. Kosorok[a,*,1], Ronald E. Gangnon[b,2]

[a]Departments of Statistics and Biostatistics and Medical Informatics, University of Wisconsin– Madison, 1300 University Avenue, Madison, Wisconsin 53706-1532, USA
[b]Department of Biostatistics and Medical Informatics, University of Wisconsin– Madison, 610 N. Walnut Street, Madison, Wisconsin 53726-2397, USA

Received 28 October 2004
Available online 7 September 2005

Abstract

In this note, we consider weighted log-rank statistics applied to clustered survival data with variable cluster sizes and arbitrary treatment assignments within clusters. Specifically, we verify that the contribution over the time interval for which the risk set proportion is arbitrarily small (the so-called "tail instability") is asymptotically negligible. These results were claimed but not proven by Gangnon and Kosorok [2004. Sample-size formula for clustered survival data using weighted log-rank statistics. Biometrika 91, 263–275.] who developed sample size formulas in this context. The main difficulty is that standard martingale methods cannot be used on account of the dependencies within clusters, and new methods are required.

Keywords: Clustered data; Local alternative; Log-rank statistic; Martingales; Multivariate survival data; Tail instability

1. Introduction

This paper considers weighted log-rank statistics applied to clustered survival data with variable cluster sizes and arbitrary treatment assignments within clusters. Clustered survival data arises when the time to a single type of event is assessed on two or more distinct, similar units within a common subject. In ophthalmology, for example, time to moderate visual loss could be assessed separately on both eyes of a person. Multivariate survival data, in contrast, arises when distinct events or repeated events occur to the same subject. For example, in cardiology, time to myocardial infarction and time to stroke could both be assessed on the same person.

*Corresponding author. Tel.: +1 608 263 6797; fax: +1 608 265 5579.
E-mail address: kosorok@biostat.wisc.edu (M.R. Kosorok).
[1]Supported in part by National Cancer Institute Grant CA75142.
[2]Supported in part by a contract with Novartis Pharma AG.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.08.026

A key difference between clustered survival data and multivariate survival data is that a common marginal hazards model is likely to be appropriate for clustered survival data but inappropriate for multivariate survival data. Gangnon and Kosorok (2004) use this principle to construct valid test procedures based on weighted log-rank statistics which do not require knowledge of the dependency structure within clusters. These results are obtained under a contiguous sequence of local alternative hypotheses so that sample size formulas can be derived. They claim, but do not prove, that the contribution over the time interval for which the risk set proportion is arbitrarily small (the so-called "tail instability") is asymptotically negligible.

The main contribution of the present paper is to verify Gangnon and Kosorok's claim. Since events within a cluster are dependent, standard martingale methods for independent, right censored data can no longer be used. A related tail instability issue is addressed by Kosorok (2002) who studies bivariate survival function estimation over a region where the minimum risk set size c_n goes to infinity, as $n \to \infty$, but $c_n/n \to 0$. A fundamental difference between this earlier paper and the present paper is that the risk set in the current paper is allowed to be as small as a single observation. Thus the techniques in Kosorok (2002) are not applicable, and new methods are required. The background of the problem, including the data set-up, assumptions and statistical tests, are presented in Section 2. The main results and proofs are given in Section 3.

2. Background

2.1. The data set-up and assumptions

The data set-up and assumptions are the same as those given in Gangnon and Kosorok (2004). The observed data $\{(X_{ijk}, \delta_{ijk}), k = 1, \ldots, m_{ij}, j = 1, 2, i = 1, \ldots, n\}$ consist of n independent clusters, two treatments, and m_{ij} individuals within cluster i and treatment j; m_{ij} may be zero. $X_{ijk} = T_{ijk} \wedge C_{ijk}$ and $\delta_{ijk} = \mathbf{1}\{X_{ijk} = T_{ijk}\}$, where T_{ijk} is a time-to-event of interest, C_{ijk} is a right censoring time, $x \wedge y$ denotes the minimum of x and y, and $\mathbf{1}\{A\}$ denotes the indicator of A.

We assume that $\{T_{ijk}, k = 1, \ldots, m_{ij}\}$ and $\{C_{ijk}, k = 1, \ldots, m_{ij}\}$ are independent within each cluster and treatment combination, $j = 1, 2$ and $i = 1, \ldots, n$. Failure and censoring times may be otherwise dependent within clusters. Although the distribution functions involved may depend on n, we will sometimes suppress this dependence for clarity. Let $\tilde{\pi}_j^n(t) \equiv n^{-1}\sum_{i=1}^n \sum_{k=1}^{m_{ij}} E\mathbf{1}\{C_{ijk} \geqslant t\}$ be the average number of individuals per cluster assigned to treatment j not yet censored at time $t-$, where we define any summation from 1 to m_{ij} to be zero if $m_{ij} = 0$. We also assume that $\tilde{\pi}_j^n$ converges uniformly to $\tilde{\pi}_j$, $j = 1, 2$, and that cluster sizes are bounded, i.e., $0 \leqslant m_{ij} \leqslant m_0 < \infty$, $i = 1, \ldots, n$, $j = 1, 2$, and that $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n m_{ij} \in (0, m_0]$. Let $\tilde{I} \equiv \{t \geqslant 0 : \tilde{\pi}_1(t) \wedge \tilde{\pi}_2(t) > 0\}$ denote the interval of observation permitted by censoring.

Since the purpose of Gangnon and Kosorok was to develop sample size formula, a contiguous sequence of models for the failure times was used. For each sample-size $n \geqslant 1$, we assume that the marginal distributions of failure times are identical within treatment $j = 1, 2$, with integrated hazard Λ_j^n having the following properties: for $j = 1, 2$, $\sup_{t\in\tilde{I}} |\mathrm{d}\Lambda_j^n(t)/\mathrm{d}\Lambda_0(t) - 1| \to 0$ and $\sup_{t\in\tilde{I}} |\sqrt{n}\{\mathrm{d}\Lambda_1^n(t)/\mathrm{d}\Lambda_2^n(t) - 1\} - \phi(t)\{1 + \eta(t)\}| \to 0$, as $n \to \infty$, for some cumulative hazard Λ_0 with corresponding survival function S_0, where ϕ is either cadlag, i.e. right continuous with left-hand limits, or caglad, i.e. left-continuous with right-hand limits, with bounded total variation, and where η is bounded with nonzero values only at the jump points of S_0, where there may be ties in the failure times. It was shown in Gangnon and Kosorok that both proportional hazards and proportional odds local alternatives satisfy the above requirements, and, moreover, that it is necessary to assume $\sup_{t\in\tilde{I}} \Delta\Lambda_0(t) < 1$ to ensure that η is bounded.

We will use counting process notation (Fleming and Harrington, 1991; Andersen et al., 1993) throughout the paper. Define the at-risk processes $Y_{ijk}(t) \equiv \mathbf{1}\{X_{ijk} \geqslant t\}$ and $\overline{Y}_j \equiv \sum_{i=1}^n \sum_{k=1}^{m_{ij}} Y_{ijk}$, and let $\pi_j^n(t) = n^{-1}\sum_{i=1}^n \sum_{k=1}^{m_{ij}} E[Y_{ijk}(t)]$. Under the above assumptions, standard arguments yield $\sup_{t\in[0,\infty]} |n^{-1}\overline{Y}_j(t) - \pi_j(t)| \to 0$ almost surely, as $n \to \infty$, where $\pi_j(t) \equiv \tilde{\pi}_j(t)S_0(t-)$, $j = 1, 2$. We write $I \equiv \{t \geqslant 0 : \pi_1(t) \wedge \pi_2(t) > 0\}$ and $\tau \equiv \sup I$, and also assume that $\sup_{t\in\{[0,\infty]-I\}} \pi_1^n(t) \wedge \pi_2^n(t) = 0$ for all n large enough.

2.2. The test statistics and tail instability

The weighted log-rank test for clustered data proposed by Gangnon and Kosorok (2004) is

$$H_n \equiv n^{-1/2} \int_0^\infty \hat{U}_n(s) \frac{\overline{Y}_1(s)\overline{Y}_2(s)}{\overline{Y}_1(s)+\overline{Y}_2(s)} \left\{ \frac{\mathrm{d}\overline{N}_1(s)}{\overline{Y}_1(s)} - \frac{\mathrm{d}\overline{N}_2(s)}{\overline{Y}_2(s)} \right\},$$

where $N_{ijk}(t) \equiv \mathbf{1}\{X_{ijk} \leqslant t, \delta_{ijk} = 1\}$, $k = 1, \ldots, m_{ij}$, $i = 1, \ldots, n$, and $\overline{N}_j \equiv \sum_{i=1}^n \sum_{k=1}^{m_{ij}} N_{ijk}$, $j = 1, 2$, are the counting processes of observed events, and where $\hat{U}_n \geqslant 0$ is either cadlag or caglad with uniformly bounded total variation. We assume that $\sup_{t\in K} |\hat{U}_n(t) - U(t)| \to 0$ in probability, for some function U and every closed subinterval $K \subset I$. A discussion of several useful choices of $\hat{U}_n$ is given by Gangnon and Kosorok (2004).

Now let

$$M_{ijk}(t) \equiv N_{ijk}(t) - \int_0^t Y_{ijk}(s)\,\mathrm{d}\Lambda_j^n(s), \quad \hat{M}_{ijk}(t) \equiv N_{ijk}(t) - \int_0^t Y_{ijk}(s) \frac{\mathrm{d}\overline{N}_j(s)}{\overline{Y}_j(s)},$$

$k = 1, \ldots, m_{ij}$, and let

$$\overline{M}_{ij\cdot} \equiv \sum_{k=1}^{m_{ij}} M_{ijk}, \quad \check{M}_{ij\cdot} \equiv \sum_{k=1}^{m_{ij}} \hat{M}_{ijk}, \quad \overline{M}_{\cdot j\cdot} = \sum_{i=1}^{n} \overline{M}_{ij\cdot},$$

$j = 1, 2$, $i = 1, \ldots, n$. Define also

$$\sigma_n^2(t) \equiv n^{-1} \sum_{i=1}^n E\left[\int_0^t U(s) \left\{ \frac{\pi_2^n(s)}{\pi_1^n(s)+\pi_2^n(s)}\,\mathrm{d}\overline{M}_{i1\cdot}(s) - \frac{\pi_1^n(s)}{\pi_1^n(s)+\pi_2^n(s)}\,\mathrm{d}\overline{M}_{i2\cdot}(s) \right\} \right]^2$$

and

$$\hat{\sigma}_n^2(t) \equiv n^{-1} \sum_{i=1}^n \left[\int_0^t \hat{U}_n(s) \left\{ \frac{\overline{Y}_2(s)}{\overline{Y}_1(s)+\overline{Y}_2(s)}\,\mathrm{d}\check{M}_{i1\cdot}(s) - \frac{\overline{Y}_1(s)}{\overline{Y}_1(s)+\overline{Y}_2(s)}\,\mathrm{d}\check{M}_{i2\cdot}(s) \right\} \right]^2.$$

The main asymptotic results in Gangnon and Kosorok (2004) are the following two theorems:

Theorem 1. *Under the stated model assumptions, and provided* $\lim_{n\to\infty} \sigma_n^2(\infty) = \sigma^2 < \infty$, H_n *converges in distribution to a normal random variable with mean* μ *and variance* σ^2, *where*

$$\mu \equiv \int_0^\infty U(s)\phi(s)\{1+\eta(s)\} \frac{\pi_1(s)\pi_2(s)}{\pi_1(s)+\pi_2(s)}\,\mathrm{d}\Lambda_0(s).$$

Theorem 2. *Under the conditions of Theorem* 1, $\hat{\sigma}_n^2(\infty) \to \sigma^2$ *in probability, as* $n \to \infty$.

The proofs given by Gangnon and Kosorok (2004) hinge on the claim that, when $\tau \notin I$, the following results hold for any increasing sequence $\{t_n\} \in I$ with $t_n \uparrow \tau$:

$$n^{-1/2} \int_{(t_n,\tau)} U(s) \frac{\pi_{j'}^n(s)}{\pi_1^n(s)+\pi_2^n(s)}\,\mathrm{d}\overline{M}_{\cdot j\cdot}(s) = \mathrm{o}_p(1), \tag{1}$$

$$n^{-1/2} \int_{(t_n,\tau)} \hat{U}_n(s) \frac{\overline{Y}_{j'}(s)}{\overline{Y}_1(s)+\overline{Y}_2(s)}\,\mathrm{d}\overline{M}_{\cdot j\cdot}(s) = \mathrm{o}_p(1), \tag{2}$$

$$n^{-1} \sum_{i=1}^n E\left[\int_{(t_n,\tau)} U \frac{\pi_{j'}^n}{\pi_1^n+\pi_2^n}\,\mathrm{d}\overline{M}_{ij\cdot} \right]^2 = \mathrm{o}(1), \tag{3}$$

$$n^{-1} \sum_{i=1}^n \left(\int_{(t_n,\tau)} \hat{U}_n \frac{\overline{Y}_{j'}}{\overline{Y}_1+\overline{Y}_2}\,\mathrm{d}\check{M}_{ij\cdot} \right)^2 = \mathrm{o}_p(1), \tag{4}$$

where $j' \equiv 3-j$ for $j = 1, 2$. This claim was made without proof. The issue of whether these results hold or not is the "tail instability" problem. In the case of independent failure times (clusters of size 1), standard martingale techniques can be used to verify these results. The fact that we have dependence within clusters in the current set-up precludes this possibility and new methods are required.

3. Main results

The following theorem and its proof are the main results of this paper:

Theorem 3. *Under the conditions of Theorem* 1, *results* (1)–(4) *hold when* $\tau \notin I$.

Before giving the proof of this theorem, we need the following three lemmas:

Lemma 1. *Suppose the processes* $\{A_n : [s,t] \mapsto \mathbb{R},\ n \geqslant 1\}$ *are all either cadlag or caglad and the processes* $\{B_n : [s,t] \mapsto \mathbb{R}, n \geqslant 1\}$ *are cadlag. Suppose for a given sequence* $\{t_n\} \in [s,t)$, *that* $\sup_{x \in (t_n,t)} |A_n(x)| + |B_n(x)| = O_p(1)$, $\sup_{s \in (t_n,t)} |B_n(s) - B_n(t_n)| = o_p(1)$ *and* $V_{(t_n,t)}[A_n] = O_p(1)$, *where* $V_C[f]$ *is the total variation of the function* $f : C \mapsto \mathbb{R}$ *over the interval* $C \subset \mathbb{R}$. *Then* $\int_{(t_n,t)} A_n(x)\,\mathrm{d}B_n(x) = o_p(1)$.

Proof. Assume that A_n is caglad and let A_n^+ be the right continuous version of A_n. By integration by parts, we have that $\int_{(t_n,t)} A_n(x)\,\mathrm{d}B_n(x) = A_n(t)\{B_n(t-) - B_n(t_n)\} - \int_{(t_n,t)} \{B_n(x) - B_n(t_n)\}\,\mathrm{d}A_n^+(x)$, and the result follows. A similar argument can be used when A_n is cadlag. □

Lemma 2. *For a failure time* T *with integrated hazard* Λ, $E\Lambda^r(T) \leqslant r!$ *for any integer* $r \geqslant 0$.

Proof. Let F be the distribution function associated with Λ and let $S = 1 - F$. Then $\int_0^t \Lambda^r(s)\,\mathrm{d}F(s) = \Lambda^r(t)F(t) - \int_0^t F(s-)\,\mathrm{d}\Lambda^r(s) = \int_0^t S(s-)\,\mathrm{d}\Lambda^r(s) - \int_0^t \mathrm{d}\Lambda^r(s) - \Lambda^r(t)S(t) + \Lambda^r(t) \leqslant \int_0^t S(s-)\,\mathrm{d}\Lambda^r(s) \leqslant r\int_0^t S(s-)\Lambda^{r-1}(s)\,\mathrm{d}\Lambda(s)$, by integration by parts combined with the fact that, for nonnegative reals a and b, $(a+b)^r - a^r \leqslant r(a+b)^{r-1}b$. However, $\int_0^t S(s-)\Lambda^{r-1}(s)\,\mathrm{d}\Lambda(s) = \int_0^t \Lambda^{r-1}(s)\,\mathrm{d}F(s)$, since $\Lambda(t) = \int_0^t S^{-1}(s-)\,\mathrm{d}F(s)$ by definition, and the result follows. □

Lemma 3. *Let* f, g *be nonincreasing functions on* $[a,b]$ *such that* $g > 0$ *and* $0 \leqslant f \leqslant g^p$ *for some* $1 < p \leqslant 2$. *Then*

$$V_{[a,b]}[f/g] \leqslant \frac{p}{p-1}[f^{(p-1)/p}(a) - f^{(p-1)/p}(b)] + \frac{1}{p-1}[g^{p-1}(a) - g^{p-1}(b)].$$

Proof. This is part (i) of Lemma 3 of Gu and Lai (1991), and the proof is given therein. □

Proof of Theorem 3. Although the quantities involved are not martingales on account of the dependence within clusters, we can still use a special martingale construction to establish the desired results. Without loss of generality, assume that $m_{ij} = m_0$ for all $i = 1, \ldots, n$. This can be accomplished by adding counting and at-risk processes that are zero everywhere as needed. Let $\{e_i, i \geqslant 1\}$ be an infinite sequence of independent and identically distributed vectors, independent of the data, where $\{e_i(k), k = 1, \ldots, m_0\}$ is a permutation of the labels $k = 1, \ldots, m_0$ with all possible permutations being equally likely, for all $i \geqslant 1$. For each $k = 1, \ldots, m_0$, define $N^*_{ijk} = N_{ije_i(k)}$, $Y^*_{ijk} = Y_{ije_i(k)}$ and $M^*_{ijk} = M_{ije_i(l)}$, $j = 1,2$. For $j = 1,2$ and $k = 1, \ldots, m_0$, the processes M^*_{ijk}, $i = 1, \ldots, n$, are now zero-mean $\mathscr{F}_t^{n,jk}$-martingales, where $\mathscr{F}_t^{n,jk} = \sigma\{N^*_{ijk}(s), Y^*_{ijk}(s+), s \in [0,t], 1 \leqslant i \leqslant n\}$ and $\sigma\{A\}$ denotes the smallest σ-field making all of A measurable. We now have that

$$n^{-1/2}\int_{(t_n,\tau)} U(s)\,\frac{\pi^n_{j'}(s)}{\pi^n_1(s) + \pi^n_2(s)}\,\mathrm{d}\overline{M}_{\cdot j\cdot}(s) = \sum_{k=1}^{m_0} n^{-1/2}\int_{(t_n,\tau)} U(s)\,\frac{\pi^n_{j'}(s)}{\pi^n_1(s) + \pi^n_2(s)}\,\mathrm{d}\overline{M}^*_{jk}(s),$$

$$n^{-1/2}\int_{(t_n,\tau)} \hat{U}_n(s)\,\frac{\overline{Y}_{j'}(s)}{\overline{Y}_1(s) + \overline{Y}_2(s)}\,\mathrm{d}\overline{M}_{\cdot j\cdot}(s) = \sum_{k=1}^{m_0} n^{-1/2}\int_{(t_n,\tau)} \hat{U}_n(s)\,\frac{\overline{Y}_{j'}(s)}{\overline{Y}_1(s) + \overline{Y}_2(s)}\,\mathrm{d}\overline{M}^*_{jk}(s),$$

where $\overline{M}^*_{jk} \equiv \sum_{i=1}^n M^*_{ijk}$ are $\mathscr{F}_t^{n,jk}$-martingales, $k = 1, \ldots, m_0$, $j = 1,2$. Note that we have not made any additional assumptions about the data; we are only using randomized relabeling as a device for proof.

Fix $j \in \{1,2\}$ and $k \in \{1,\ldots,m_0\}$. A standard application of Lenglart's inequality (see Theorem 3.4.1 of Fleming and Harrington (1991)) gives

$$n^{-1/2}\int_{(t_n,\tau)} U(s)\,\frac{\pi^n_{j'}(s)}{\pi^n_1(s)+\pi^n_2(s)}\,\mathrm{d}\overline{M}^*_{jk}(s) = \mathrm{o}_p(1),$$

and thus (1) is established. Now note that

$$\begin{aligned} & n^{-1/2}\int_{(t_n,\tau)} \hat{U}_n(s)\,\frac{\overline{Y}_{j'}(s)}{\overline{Y}_1(s)+\overline{Y}_2(s)}\,\mathrm{d}\overline{M}^*_{jk}(s) \\ &\quad = n^{-1/2}\int_{(t_n,\tilde{T}_n]} \hat{U}_n(s)\left[\frac{\overline{Y}_{j'}(s)\{\overline{Y}^*_{jk}(s)/n\}^{1/4}}{\overline{Y}_1(s)+\overline{Y}_2(s)}\right]\frac{\mathrm{d}\overline{M}^*_{jk}(s)}{\{\overline{Y}^*_{jk}(s)/n\}^{1/4}}, \end{aligned}$$

where $\overline{Y}^*_{jk} \equiv \sum_{i=1}^n Y^*_{ijk}$ and $\tilde{T}_n = \tau \wedge \sup\{t \in I : \overline{Y}_1(t)+\overline{Y}_2(t)>0\}$. The predictable compensator for the submartingale

$$\left[n^{-1/2}\int_0^t \left\{\frac{\overline{Y}^*_{jk}(s)}{n}\right\}^{-1/4}\mathrm{d}\overline{M}^*_{jk}(s)\right]^2$$

is

$$\int_0^t \left\{\frac{\overline{Y}^*_{jk}(s)}{n}\right\}^{1/2}[1-\Delta\Lambda^n_j(s)]\,\mathrm{d}\Lambda^n_j(s),$$

which, for $0 \leqslant s < t < \tau$, satisfies

$$\begin{aligned} E\int_s^t \left\{\frac{\overline{Y}^*_{jk}(s)}{n}\right\}^{1/2}[1-\Delta\lambda^n_j(s)]\,\mathrm{d}\Lambda^n_j(s) &\leqslant \int_s^t \{S^n_j(s-)\}^{1/2}\,\mathrm{d}\Lambda^n_j(s) \\ &\leqslant 2[\{S^n_j(s)\}^{1/2} - \{S^n_j(t)\}^{1/2}]. \end{aligned}$$

Hence

$$\sup_{t\in I}\left|n^{-1/2}\int_0^t \left\{\frac{\overline{Y}^*_{jk}(s)}{n}\right\}^{-1/4}\mathrm{d}\overline{M}^*_{jk}(s)\right| = \mathrm{O}_p(1)$$

and

$$\sup_{s\in(t_n,\tilde{T}_n)}\left|n^{-1/2}\int_{(t_n,s)} \left\{\frac{\overline{Y}^*_{jk}(s)}{n}\right\}^{-1/4}\mathrm{d}\overline{M}^*_{jk}(s)\right| = \mathrm{o}_p(1).$$

By Lemma 3 above,

$$\begin{aligned} V_{(t_n,\tilde{T}_n]}\left[\frac{\overline{Y}_{j'}(s)\{\overline{Y}^*_{jk}(s)/n\}^{1/4}}{\overline{Y}_1(s)+\overline{Y}_2(s)}\right] &\leqslant 5V_{(t_n,\tilde{T}_n]}\left[\{\overline{Y}_{j'}(s)/n\}^{1/5}\left\{\frac{\overline{Y}^*_{jk}(s)}{n}\right\}^{1/20}\right] \\ &\quad + 4V_{(t_n,\tilde{T}_n]}\left[\left\{\frac{\overline{Y}_1(s)+\overline{Y}_2(s)}{n}\right\}^{1/4}\right], \end{aligned}$$

and thus, by application of Lemma 1 above, we obtain

$$n^{-1/2}\int_{(t_n,\tau_j)}\left[\frac{\overline{Y}_{j'}(s)\{\overline{Y}^*_{jk}(s)/n\}^{1/4}}{\overline{Y}_1(s)+\overline{Y}_2(s)}\right]\frac{\mathrm{d}\overline{M}^*_{jk}(s)}{\{\overline{Y}^*_{jk}(s)/n\}^{1/4}} = \mathrm{o}_p(1).$$

By a second application of Lemma 1, we are also able to resolve the inclusion of the $\hat{U}_n$ term because of its bounded total variation. Thus (2) is established.

Next, we have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} E\left[\int_{(t_n,\tau)} U\,\frac{\pi_{j'}^n}{\pi_1^n+\pi_2^n}\,\mathrm{d}\overline{M}_{ij\cdot}\right]^2 &\leqslant m_0 n^{-1}\sum_{i=1}^{n} E\left[\sum_{k=1}^{m_{ij}}\left(\int_{(t_n,\tau)} U\,\frac{\pi_{j'}^n}{\pi_1^n+\pi_2^n}\,\mathrm{d}M_{ijk}(s)\right)^2\right] \\ &\leqslant c\{S_j^n(t_n)-S_j^n(\tau_j-)\} \\ &\to 0, \end{aligned}$$

in probability, for some positive constant $c<\infty$, and thus (3) is established.

Finally,

$$n^{-1}\sum_{i=1}^{n}\left(\int_{(t_n,\tau)} \hat{U}_n\,\frac{\overline{Y}_{j'}}{\overline{Y}_1+\overline{Y}_2}\,\mathrm{d}\check{M}_{ij\cdot}\right)^2 \leqslant c' n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{m_{ij}}\left[\left(\int_{(t_n,\tau)}\mathrm{d}N_{ijk}\right)^2+\left\{\int_{(t_n,\tau)} Y_{ijk}\,\frac{\mathrm{d}\overline{N}_j}{\overline{Y}_j}\right\}^2\right] \tag{5}$$

for some $c'<\infty$ not depending on n. However,

$$n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{m_{ij}}\left(\int_{(t_n,\tau)}\mathrm{d}N_{ijk}\right)^2 = n^{-1}\int_{(t_n,\tau)}\mathrm{d}\overline{N}_j(t) = \mathrm{o}_p(1).$$

Moreover,

$$n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{m_{ij}}\left\{\int_{(t_n,\tau)} Y_{ijk}\,\frac{\mathrm{d}\overline{N}_j}{\overline{Y}_j}\right\}^2 \leqslant n^{-1}m_0\sum_{i=1}^{n}\sum_{k=1}^{m_0}\sum_{l=1}^{m_0}\left\{\int_{(t_n,\tau)} Y_{ijk}^{*}\,\frac{\mathrm{d}\overline{N}_{jl}^{*}}{\overline{Y}_{jl}^{*}}\right\}^2, \tag{6}$$

where $\overline{N}_{jl}^{*}\equiv\sum_{i=1}^{n}N_{ijl}^{*}$, $l=1,\ldots,m_0$. Now fix $l\in\{1,\ldots,m_0\}$ and note that

$$E\left(\int_{(t_n,\tau)} Y_{ijk}^{*}\,\frac{\mathrm{d}\overline{N}_{jl}^{*}}{\overline{Y}_{jl}^{*}}\right)^2 \leqslant 2E\left(\int_{(t_n,\tau)}\mathrm{d}N_{ijl}^{*}\right)^2 + 2E\left(\int_{(t_n,\tau)} Y_{ijk}^{*}\,\frac{\mathrm{d}\overline{N}_{(i)jl}^{*}}{\overline{Y}_{(i)jl}^{*}}\right)^2, \tag{7}$$

where $\overline{N}_{(i)jl}^{*}\equiv\sum_{r\neq i}N_{rjl}^{*}$ and $\overline{Y}_{(i)jl}^{*}\equiv\sum_{r\neq i}Y_{rjl}^{*}$; but

$$(7)\leqslant 2E\left(\int_{(t_n,\tau)}\mathrm{d}N_{ijl}^{*}\right) + 4E\left(\int_{(t_n,\tau)} Y_{ijk}^{*}\,\frac{\mathrm{d}\overline{M}_{(i)jl}^{*}}{\overline{Y}_{(i)jl}^{*}}\right)^2 + 4E\left(\int_{(t_n,\tau)} Y_{ijk}^{*}\,\mathrm{d}\Lambda_j^n\right)^2, \tag{8}$$

where $\overline{M}_{(i)jl}^{*}\equiv\sum_{r\neq i}M_{rjl}^{*}$. By independence and the martingale construction we now have that

$$\begin{aligned} (8) &\leqslant 6[F_j^n(\tau-)-F_j^n(t_n)] + 4[F_j^n(\tau-)-F_j^n(t_n)]^{1/2}\{E[\Lambda_j^n(T_j^n)]^4\}^{1/2} \\ &\equiv K_{jn}, \end{aligned}$$

where T_j^n is a failure time with integrated hazard Λ_j^n and $K_{jn}\to 0$, as $n\to\infty$, by Lemma 2. Since l was arbitrary, the forgoing arguments yield that the expectation of the left-hand-side of (6) is $\leqslant m_0^3 K_{jn}$, and we have proven (4). □

References

Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N.K., 1993. Statistical Models Based on Counting Processes. Springer, Berlin, New York.

Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York.

Gangnon, R.E., Kosorok, M.R., 2004. Sample-size formula for clustered survival data using weighted log-rank statistics. Biometrika 91, 263–275.

Gu, M.G., Lai, T.L., 1991. Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. Ann. Statist. 19, 1403–1433.

Kosorok, M.R., 2002. On global consistency of a bivariate survival estimator under univariate censoring. Statist. Probab. Lett. 56, 439–446.