BIOMETRICS 56, 554–562
June 2000

Nonparametric Analysis of Recurrent Events and Death

Debashis Ghosh and D. Y. Lin*

Department of Biostatistics, Box 357232, University of Washington, Seattle, Washington 98195, U.S.A.
**email:* danyu@biostat.washington.edu

SUMMARY. This article is concerned with the analysis of recurrent events in the presence of a terminal event such as death. We consider the mean frequency function, defined as the marginal mean of the cumulative number of recurrent events over time. A simple nonparametric estimator for this quantity is presented. It is shown that the estimator, properly normalized, converges weakly to a zero-mean Gaussian process with an easily estimable covariance function. Nonparametric statistics for comparing two mean frequency functions and for combining data on recurrent events and death are also developed. The asymptotic null distributions of these statistics, together with consistent variance estimators, are derived. The small-sample properties of the proposed estimators and test statistics are examined through simulation studies. An application to a cancer clinical trial is provided.

KEY WORDS: Competing risks; Counting process; Cumulative incidence; Empirical process; Multiple events; Terminal event.

1. Introduction

In many clinical and epidemiological studies, subjects can potentially experience recurrent events. Examples include transient ischemic attacks in patients with cerebrovascular disease (Hobson et al., 1993), repeated seizures in epileptic patients (Albert, 1991), tumor recurrences in cancer patients (Byar, 1980), and recurrent opportunistic infections in HIV patients (Li and Lagakos, 1997). Often, the recurrence of serious events, such as tumors and opportunistic infections, is associated with an elevated risk of death so that the subject dies during the course of the study.

As an example, let us consider the well-known bladder tumor clinical trial conducted by the Veterans Administration Co-operative Urological Research Group (Byar, 1980). A total of 117 patients with superficial bladder tumors were randomly assigned to placebo, pyridoxine (vitamin B_6), or thiotepa. The goal of the study was to determine the effectiveness of pyridoxine and thiotepa in reducing the rate of tumor recurrence. By the end of follow-up, there were 87, 57, and 45 recurrences among the 48, 31, and 38 patients in the placebo, pyridoxine, and thiotepa groups, respectively. Some patients died before any tumor recurrence, while others died after recurrences. There were 11, 7, and 12 observed deaths in the placebo, pyridoxine, and thiotepa groups, respectively.

Many authors (e.g., Andersen and Gill, 1982; Wei, Lin, and Weissfeld, 1989; Pepe and Cai, 1993; Lawless and Nadeau, 1995) have developed methods for analyzing recurrent events. All these methods, however, assume that recurrent events are not terminated by death during the study. Some efforts have been made in recent years to account for death in the analysis of recurrent events. In particular, Cook and Lawless (1997) studied the rate function of recurrence conditional on survival. Li and Lagakos (1997) modeled the cause-specific hazard function (Kalbfleisch and Prentice, 1980, p. 167) for each recurrence by treating death as the competing risk. One could also model the intensity function of the recurrence process within the framework of multivariate counting processes (Andersen et al., 1993, Section II.4.1). Neither the cause-specific hazard function nor the intensity function characterizes the subject's ultimate recurrence experience.

In this article, we develop one- and two-sample nonparametric methods based on the marginal mean of the cumulative number of recurrent events over time. This quantity, to be referred to as the mean frequency function, incorporates the fact that subjects who die cannot experience any further recurrent events. Cook and Lawless (1997) suggested an estimator for this function but did not study its properties. As will be shown later, there is an intimate connection between the mean frequency function and the cumulative incidence function for the classical competing risks data (Kalbfleisch and Prentice, 1980, p. 168).

The course of this article is as follows. In Section 2, we study the nonparametric estimation of the mean frequency function. In Section 3, we develop nonparametric statistics for comparing two mean frequency functions. The issue of combining information about recurrent events and death is also addressed. In Section 4, we report the results of some simulation studies. In Section 5, the proposed methods are applied to the aforementioned bladder tumor clinical trial. Some concluding remarks are given in Section 6.

2. One-Sample Estimation

Let D denote the survival time and $N^*(t)$ denote the number of recurrent events that occur in the time interval $[0, t]$. Naturally, subjects who die cannot experience any further re-

current events so that $N^*(\cdot)$ does not jump after D. Let C be the follow-up or censoring time, which is assumed to be independent of both D and $N^*(\cdot)$. Note, however, that we make no assumptions regarding the dependence among the recurrent events or the dependence between $N^*(\cdot)$ and D. Because of censoring, D and $N^*(\cdot)$ may not be fully observed. Instead, one observes $\{N(\cdot), X, \delta\}$, where $N(t) = N^*(t \wedge C)$, $X = D \wedge C$, and $\delta = I(D \leq C)$. Here and in the sequel, $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. The observations $\{N_i(\cdot), X_i, \delta_i\}$ $(i = 1, \ldots, n)$ are assumed to be independent replicates of $\{N(\cdot), X, \delta\}$. For technical reasons, we assume that there exists a constant τ such that $\Pr(X \geq \tau) > 0$.

The mean frequency function is defined as $\mu(t) = \mathrm{E}\{N^*(t)\}$. This is the marginal expected number of recurrent events up to t per subject, acknowledging the fact that no subject can experience any further recurrent event after death. This quantity has a clear clinical meaning and is interpretable regardless of what the distributions of D and $N^*(\cdot)$ are.

It is easy to see that

$$\mu(t) = \int_0^t S(u)dR(u), \tag{2.1}$$

where $S(t) = \Pr(D \geq t)$ and $dR(t) = \mathrm{E}\{dN^*(t) \mid D \geq t\}$. This representation suggests an obvious estimator. Specifically, let $\hat{S}(t)$ be the Kaplan–Meier estimator of $S(t)$ based on (X_i, δ_i) $(i = 1, \ldots, n)$ and let $\hat{R}(t)$ be the Nelson–Aalen estimator for $R(t)$ based on the recurrence data, i.e.,

$$\hat{R}(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{\sum_{j=1}^{n} Y_j(u)},$$

where $Y_i(t) = I(X_i \geq t)$. Then $\mu(t)$ is estimated by

$$\hat{\mu}(t) = \int_0^t \hat{S}(u)d\hat{R}(u).$$

This estimator was mentioned briefly by Cook and Lawless (1997). These authors did not study the asymptotic properties of $\hat{\mu}(t)$, although they discussed the variance estimation for Nelson–Aalen-type estimators.

In the absence of death, $\hat{\mu}(t)$ reduces to $\hat{R}(t)$. In the presence of death, $\hat{\mu}(t) < \hat{R}(t)$ for t greater than the first observed death. Thus, the Nelson–Aalen estimator (treating death as a censoring event for recurrences) generally overestimates the mean frequency function whether or not the death and recurrence processes are independent. If the event of interest can occur at most one time such that $N^*(t)$ is a one-jump counting process, then $\hat{\mu}(t)$ reduces to the estimator of the cumulative incidence function for the event of interest, with death acting as a competing risk (Gray, 1988; Pepe and Mori, 1993; Lin, 1997).

Let $N_i^D(t) = I(X_i \leq t, \delta_i = 1)$ and $\bar{Y}(t) = \Sigma_{i=1}^n Y_i(t)$. Also let $\hat{\Lambda}^D(t) = \Sigma_{i=1}^n \int_0^t \bar{Y}^{-1}(u)dN_i^D(u)$, $\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(u)d\hat{R}(u)$, and $\hat{M}_i^D(t) = N_i^D(t) - \int_0^t Y_i(u)d\hat{\Lambda}^D(u)$. Note that $\hat{\Lambda}^D(t)$ is the Nelson–Aalen estimator for the cumulative hazard function of death $\Lambda^D(t)$. We show in Appendix A.1 that $n^{1/2}\{\hat{\mu}(t) - \mu(t)\}$ $(0 \leq t \leq \tau)$ converges weakly to a mean-zero Gaussian process whose covariance function at (s, t) can be consistently estimated by

$$\hat{\xi}(s, t) = n^{-1} \sum_{i=1}^{n} \hat{\Psi}_i(s)\hat{\Psi}_i(t),$$

where

$$\hat{\Psi}_i(t) = \int_0^t \frac{\hat{S}(u)d\hat{M}_i(u)}{\bar{Y}(u)/n} - \hat{\mu}(t) \int_0^t \frac{d\hat{M}_i^D(u)}{\bar{Y}(u)/n} + \int_0^t \frac{\hat{\mu}(u)d\hat{M}_i^D(u)}{\bar{Y}(u)/n}. \tag{2.2}$$

For any fixed time point t, the asymptotic normality of $\hat{\mu}(t)$, along with the consistent variance estimator $\hat{\xi}(t, t)$, allows one to construct pointwise confidence intervals for $\mu(t)$. Since $\mu(t)$ is a nonnegative function, we consider the transformed variable $n^{1/2}[\log\{\hat{\mu}(t)\} - \log\{\mu(t)\}]$, which has a distribution asymptotically equal to that of $n^{1/2}\{\hat{\mu}(t) - \mu(t)\}/\mu(t)$ for $\mu(t) > 0$. With the logarithmic transformation, an approximate $(1 - \alpha)$ pointwise confidence interval for $\mu(t)$ is given by

$$\hat{\mu}(t)e^{\pm n^{-1/2} z_{\alpha/2} \hat{\xi}^{1/2}(t,t)/\hat{\mu}(t)},$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentile of the standard normal distribution.

3. Two-Sample Testing

In this section, we consider the comparison of the mean frequency functions for recurrent events between two treatment groups. In many situations, one is primarily interested in the recurrent events, and death is treated as a nuisance. In other situations, especially when the death rates are high, it may be desirable to examine evidence from both endpoints. These two types of situations are dealt with in Sections 3.1 and 3.2, respectively.

3.1 *Comparing Mean Frequency Functions*

The null hypothesis to be tested is that the mean frequency functions are the same between the two treatment groups. For the classical competing risks problem, nonparametric statistics for testing the equality of two cumulative incidence functions have been proposed by Gray (1988), Pepe and Mori (1993), and Lin (1997). Perhaps the most appealing of these statistics is that of Gray because, in the case when there is no competing risk, the Gray test reduces to the familiar log-rank test for survival data. On the other hand, the Pepe–Mori test may be more powerful against alternatives in which the cumulative incidence for one group is consistently higher than that of the other. In this section, analogs of the Gray and Pepe–Mori tests are developed for the mean frequency function.

Let $\mu_1(t)$ and $\mu_2(t)$ denote the mean frequency functions for groups 1 and 2. The hypothesis to be tested is H_0: $\mu_1(t) = \mu_2(t)$ for all t. Let n_j be the number of subjects in the jth group $(j = 1, 2)$ and let $n = n_1 + n_2$. Also, let $\hat{\mu}_j(t)$ be the estimator of the mean frequency function for the jth group. As in the one-sample case, a constant τ_j is assumed for the jth group. Denote $\tau = \tau_1 \wedge \tau_2$. According to the results of Section 2, $n_j^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\}$ $(0 \leq t \leq \tau_j)$ converges weakly to a mean-zero Gaussian process with a covariance function

that can be consistently estimated by

$$\hat{\xi}_j(s,t) = n_j^{-1} \sum_{i=1}^{n_j} \hat{\Psi}_{ji}(s)\hat{\Psi}_{ji}(t), \qquad j = 1, 2,$$

where $\hat{\Psi}_{ji}$ is an obvious modification of $\hat{\Psi}_i$ given in (2.2) for the jth group. These results provide the theoretical basis for constructing the desired two-sample statistics.

The analog of the Gray statistic for the mean frequency function is given by

$$Q_{\mathrm{LR}} = \int_0^\tau \hat{K}_{\mathrm{LR}}(t) d\{\hat{\mu}_1(t) - \hat{\mu}_2(t)\},$$

where $\hat{K}_{\mathrm{LR}}(t)$ is a weight function based on the observed data. We will refer to Q_{LR} as the generalized weighted log-rank statistic. If there is no death, then Q_{LR} reduces to the statistic of Cook, Lawless, and Nadeau (1996). The weight function analogous to that of the familiar log-rank statistic is

$$\hat{K}_{\mathrm{LR}}(t) = \frac{\bar{Y}_1(t)\bar{Y}_2(t)}{\bar{Y}_1(t) + \bar{Y}_2(t)} \frac{n}{n_1 n_2}, \tag{3.1}$$

where $\bar{Y}_j(t)$ represents the number of subjects in the jth group who are at risk at time t.

Our extension of the Pepe–Mori statistic is given by

$$Q_{\mathrm{GT}} = \int_0^\tau \hat{K}_{\mathrm{GT}}(t)\{\hat{\mu}_1(t) - \hat{\mu}_2(t)\}dt,$$

where $\hat{K}_{\mathrm{GT}}(t)$ is a weight function generally different from $\hat{K}_{\mathrm{LR}}(t)$. Clearly, Q_{GT} with a positive $\hat{K}_{\mathrm{GT}}$ will be consistent against the alternatives H_1: $\mu_1(t) \geq \mu_2(t)$ or $\mu_1(t) \leq \mu_2(t)$ with strict inequality for some $t \in [0, \tau]$. We will refer to Q_{GT} as the generalized t-statistic because it reduces to the conventional t-statistic if there is a single event and no censoring. The weight function $\hat{K}_{\mathrm{GT}}$ needs to satisfy certain technical characteristics (Pepe and Fleming, 1989). One suitable choice is

$$\hat{K}_{\mathrm{GT}}(t) = \frac{n\hat{\mathrm{H}}_1(t)\hat{\mathrm{H}}_2(t)}{n_1\hat{\mathrm{H}}_1(t) + n_2\hat{\mathrm{H}}_2(t)}, \tag{3.2}$$

where $\hat{\mathrm{H}}_j(\cdot)$ is the Kaplan–Meier estimator for the survival function of the censoring time in the jth group ($j = 1, 2$).

We show in Appendix A.2 that, under H_0: $\mu_1(t) = \mu_2(t)$ for all t, both $(n_1n_2/n)^{1/2}Q_{\mathrm{LR}}$ and $(n_1n_2/n)^{1/2}Q_{\mathrm{GT}}$ converge in distribution to mean-zero normal random variables with variances that can be consistently estimated by

$$\hat{\sigma}^2_{\mathrm{LR}} = \frac{n_2}{nn_1} \sum_{i=1}^{n_1} \left\{ \int_0^\tau \hat{K}_{\mathrm{LR}}(t) d\hat{\Psi}_{1i}(t) \right\}^2 + \frac{n_1}{nn_2} \sum_{i=1}^{n_2} \left\{ \int_0^\tau \hat{K}_{\mathrm{LR}}(t) d\hat{\Psi}_{2i}(t) \right\}^2 \tag{3.3}$$

and

$$\hat{\sigma}^2_{\mathrm{GT}} = \frac{n_2}{nn_1} \sum_{i=1}^{n_1} \left\{ \int_0^\tau \hat{K}_{\mathrm{GT}}(t) \hat{\Psi}_{1i}(t) dt \right\}^2 + \frac{n_1}{nn_2} \sum_{i=1}^{n_2} \left\{ \int_0^\tau \hat{K}_{\mathrm{GT}}(t) \hat{\Psi}_{2i}(t) dt \right\}^2, \tag{3.4}$$

respectively. In the absence of death, (3.3) reduces to the unpooled variance formula of Cook et al. (1996).

3.2. *Combining Recurrent Events and Death*

In many applications, it is of interest to assess the treatment efficacy with respect to both the recurrent events and survival. For the former endpoint, Q_{LR} or Q_{GT} developed in the previous section can be used; for the latter endpoint, the weighted log-rank statistic or the Pepe–Fleming statistic can be used. Although it is sometimes justifiable to perform separate tests on the two endpoints without adjusting for the effect of multiple comparisons, it is often desirable to perform simultaneous inference on both endpoints. We discuss below how to test simultaneously the umbrella null hypothesis H_0^*: $\mu_1(t) = \mu_2(t)$ and $\Lambda_1^D(t) = \Lambda_2^D(t)$ for all t, where Λ_j^D denotes the cumulative hazard function of death for the jth group.

The weighted log-rank statistic for comparing the two survival distributions can be written as

$$Q_D = \int_0^\tau \hat{K}_{\mathrm{LR}}(t) d\left\{ \hat{\Lambda}_1^D(t) - \hat{\Lambda}_2^D(t) \right\},$$

where $\hat{\Lambda}_j^D(\cdot)$ is the Nelson–Aalen estimator of $\Lambda_j^D(\cdot)$ ($j = 1, 2$). We show in Appendix A.2 that, under H_0^*, the random vector $\mathbf{T} = (n_1n_2/n)^{1/2}(Q_{\mathrm{LR}}, Q_D)'$ converges in distribution to a mean-zero bivariate normal vector with a covariance matrix that can be consistently estimated by $\hat{\boldsymbol{\Sigma}}$. Thus, the quadratic form $Q_{\mathbf{T}} = \mathbf{T}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{T}$ is asymptotically chi-squared with 2 d.f.

We can construct a 1 d.f. test by taking a linear combination of Q_{LR} and Q_D. Specifically, define the combined statistic $Q_{\mathrm{WCT}} = pQ_{\mathrm{LR}} + (1-p) \times Q_D$ ($0 \leq p \leq 1$). We show in Appendix A.2 that, under H_0^*, $(n_1n_2/n)^{1/2}Q_{\mathrm{WCT}}$ converges in distribution to a mean-zero normal random variable with a variance consistently estimated by $\hat{\sigma}^2_{\mathrm{WCT}}$. Thus, the test statistic $\tilde{Q}_{\mathrm{WCT}} = (n_1n_2/n)^{1/2}Q_{\mathrm{WCT}}/\hat{\sigma}_{\mathrm{WCT}}$ is asymptotically standard normal under H_0^*. The choice of p depends on the relative importance of the two endpoints as well as on the anticipated treatment effects on the two endpoints. It can be any value and may even be data dependent; however, the decision on how to choose p must be made before examining the data. In the sequel, we set $p = 0.5$ and denote the resulting test statistic by $\tilde{Q}_{\mathrm{CT}}$. The choice of p is further discussed in Section 5.

We can also test the umbrella hypothesis H_0^* sequentially against the one-sided alternatives that the treatment is beneficial, along the lines of Wei et al. (1989). To be specific, let $\tilde{Q}_{\mathrm{LR}}$ and $\tilde{Q}_D$ denote the standardized values of Q_{LR} and Q_D. Without loss of generality, assume that positive Q_{LR} and Q_D correspond to beneficial treatment effects. Suppose that $\tilde{Q}_{\mathrm{LR}} \geq \tilde{Q}_D$. Let (V_1, V_2) be a bivariate normal vector with mean zero, unit variances, and with a correlation that is equal to the estimated covariance between $\tilde{Q}_{\mathrm{LR}}$ and $\tilde{Q}_D$ based on $\hat{\boldsymbol{\Sigma}}$. Define the subhypotheses H_{10}: $\mu_1(t) = \mu_2(t)$ and H_{20}: $\Lambda_1^D(t) = \Lambda_2^D(t)$. One rejects H_{10} if $\Pr\{\max(V_1, V_2) \geq \tilde{Q}_{\mathrm{LR}}\} \leq \alpha$; if H_{10} is rejected, then one rejects H_{20} if $\Pr(V_2 \geq \tilde{Q}_D) \leq \alpha$. The tests will be performed in the reverse order if $\tilde{Q}_{\mathrm{LR}} < \tilde{Q}_D$. It can be shown that the overall type I error rate for this multiple testing procedure is α. The procedure for testing two-sided alternatives can be developed in a similar fashion.

Each of the three proposed tests maintains a proper overall type I error rate and enables one to make a single probability statement regarding the effects of treatment on both the recurrent events and survival. The combined test tends to be more powerful than the quadratic-form test, the sequential test, and the separate tests (without adjusting for the effects of multiple comparisons) if the treatment effects on the two endpoints are similar but is not powerful if the two treatment effects are in the opposite directions. Neither the quadratic form or the combined test indicates with respect to which endpoint the treatment effect is significant when the umbrella H_0^* is rejected. By contrast, the sequential testing procedure enables one to determine on which endpoint the treatment has a beneficial effect.

4. Simulation Studies

Numerical studies were conducted to assess the finite-sample properties of the statistics proposed in the previous sections. We first examined the one-sample estimator $\hat{\mu}(\cdot)$. The survival time and the gap times for recurrent events were all exponential random variables. To create the dependence among these variables, we used Gumbel (1960)'s bivariate exponential distribution with correlation ρ. To be specific, ρ is the correlation between the first recurrent time and survival time as well as the common correlation between any two successive gap times. We set $\rho = 0, 0.0625, 0.125$, and 0.25. Marginally, the survival time was exponential with rate $\lambda^D = 0.25$, while each gap time was exponential with rate $\lambda = 1$. The censoring time was a uniform$[0, 10]$ random variable, which resulted in approximately two observed events per subject. We set $n = 50$ and 100. For each simulation setting, 10,000 samples were generated. The summary statistics from these studies are presented in Table 1. The estimator $\hat{\mu}(t)$ is practically unbiased for every t. The standard error estimator has little bias and the confidence intervals have reasonable coverage probabilities, at least for $n = 100$.

We also assessed the behavior of the generalized t-statistic Q_{GT}, the generalized weighted log-rank statistic Q_{LR}, and the combined statistic Q_{CT}. The weight functions given in (3.1) and (3.2) were used. The survival time and recurrent event times were generated in a manner similar to the one-sample case. For the jth group, the survival time was exponential with rate λ_j^D, while the gap times were exponential with rate λ_j. The censoring distribution was the same as in the one-sample case. We considered $n_1 = n_2 = 50$ and 100. For each setting, 2000 simulation samples were generated. The value of τ was the last observed event time. The results are displayed in Table 2.

All three statistics maintain their type I errors around the nominal level. The generalized t-statistic and log-rank statistic have similar power characteristics. When there is a difference in recurrence but no difference in survival, the use of the combined statistic results in a loss in power as compared

Table 1
Simulation summary statistics for $\hat{\mu}(\cdot)$

			$n = 50$				$n = 100$			
ρ	t	$\mu(t)$	$\mathrm{E}\{\hat{\mu}(t)\}$	SE	SEE	CP	$\mathrm{E}\{\hat{\mu}(t)\}$	SE	SEE	CP
0.0000	0.5	0.510	0.509	0.097	0.095	0.931	0.510	0.068	0.068	0.941
	1.0	0.956	0.954	0.127	0.122	0.939	0.957	0.089	0.089	0.943
	1.5	1.346	1.347	0.144	0.139	0.938	1.346	0.100	0.100	0.943
	2.0	1.692	1.690	0.152	0.146	0.936	1.691	0.106	0.105	0.943
	2.5	1.990	1.991	0.156	0.150	0.938	1.991	0.109	0.107	0.943
	3.0	2.254	2.253	0.157	0.150	0.937	2.253	0.109	0.108	0.942
0.0625	0.5	0.507	0.505	0.096	0.095	0.934	0.507	0.068	0.067	0.941
	1.0	0.944	0.945	0.128	0.125	0.936	0.944	0.090	0.089	0.942
	1.5	1.321	1.319	0.141	0.142	0.940	1.323	0.100	0.101	0.944
	2.0	1.652	1.649	0.153	0.155	0.938	1.653	0.107	0.106	0.948
	2.5	1.940	1.941	0.168	0.165	0.937	1.939	0.117	0.116	0.944
	3.0	2.189	2.187	0.179	0.176	0.939	2.188	0.121	0.119	0.940
0.1250	0.5	0.507	0.506	0.093	0.097	0.931	0.507	0.067	0.067	0.941
	1.0	0.940	0.942	0.122	0.117	0.935	0.939	0.091	0.090	0.944
	1.5	1.309	1.309	0.150	0.146	0.935	1.309	0.103	0.102	0.948
	2.0	1.628	1.630	0.161	0.156	0.933	1.627	0.112	0.111	0.947
	2.5	1.900	1.901	0.169	0.163	0.938	1.900	0.118	0.117	0.944
	3.0	2.136	2.138	0.175	0.170	0.935	2.136	0.120	0.118	0.942
0.2500	0.5	0.512	0.512	0.094	0.092	0.933	0.512	0.066	0.065	0.940
	1.0	0.944	0.945	0.125	0.123	0.935	0.944	0.088	0.087	0.943
	1.5	1.304	1.303	0.145	0.141	0.941	1.303	0.101	0.100	0.944
	2.0	1.604	1.602	0.158	0.153	0.942	1.603	0.110	0.109	0.946
	2.5	1.857	1.855	0.166	0.161	0.941	1.856	0.116	0.115	0.943
	3.0	2.073	2.071	0.171	0.165	0.939	2.072	0.119	0.117	0.941

Note: $\mathrm{E}\{\hat{\mu}(t)\}$, empirical mean of $\hat{\mu}(t)$; SE, empirical standard error of $\hat{\mu}(t)$; SEE, empirical mean of the standard error estimates; CP, coverage probability of the 95% confidence interval.

Table 2
Empirical sizes/powers of Q_{GT}, Q_{LR}, and Q_{CT}

			$n_1 = n_2 = 50$			$n_1 = n_2 = 100$		
ρ	λ_2^D	λ_2	Q_{GT}	Q_{LR}	Q_{CT}	Q_{GT}	Q_{LR}	Q_{CT}
0.0000	0.25	1.0	0.057	0.057	0.050	0.049	0.051	0.049
	0.25	1.5	0.737	0.726	0.702	0.944	0.939	0.915
	0.25	1.2	0.251	0.239	0.202	0.339	0.331	0.306
	0.50	1.5	0.155	0.127	0.884	0.288	0.284	0.992
	0.50	2.0	0.762	0.726	1.000	0.968	0.942	1.000
	0.30	1.5	0.589	0.559	0.834	0.851	0.826	0.989
0.0625	0.25	1.0	0.053	0.057	0.057	0.051	0.050	0.051
	0.25	1.5	0.768	0.785	0.749	0.935	0.948	0.910
	0.25	1.2	0.238	0.249	0.206	0.375	0.364	0.330
	0.50	1.5	0.145	0.178	0.871	0.239	0.268	0.989
	0.50	2.0	0.636	0.815	0.999	0.948	0.974	1.000
	0.30	1.5	0.571	0.612	0.845	0.871	0.889	0.984
0.1250	0.25	1.0	0.058	0.056	0.057	0.048	0.049	0.048
	0.25	1.5	0.775	0.795	0.756	0.965	0.975	0.926
	0.25	1.2	0.248	0.259	0.210	0.395	0.389	0.351
	0.50	1.5	0.165	0.186	0.891	0.299	0.311	0.991
	0.50	2.0	0.655	0.845	1.000	0.939	0.950	1.000
	0.30	1.5	0.611	0.645	0.885	0.883	0.899	0.992
0.2500	0.25	1.0	0.055	0.057	0.055	0.052	0.050	0.048
	0.25	1.5	0.825	0.833	0.799	1.000	1.000	0.942
	0.25	1.2	0.252	0.257	0.219	0.438	0.448	0.422
	0.50	1.5	0.139	0.115	0.861	0.210	0.218	0.992
	0.50	2.0	0.776	0.745	0.994	0.928	0.952	1.000
	0.30	1.5	0.636	0.647	0.783	0.898	0.900	0.970

Note: The survival time and the gap times for recurrent events were drawn from the bivariate exponential distribution (Gumbel, 1960) with rates λ_j^D and λ_j for the jth group; $\lambda_1^D = 0.25$ and $\lambda_1 = 1$.

to the statistics for recurrent events alone. However, when there are differences in both endpoints, the combined statistic tends to be more powerful.

In response to a referee's suggestion, we evaluated the sizes of the proposed tests when $\mu_1(t) = \mu_2(t)$ but the survival experiences are different between the two groups. The survival functions for the two groups, S_1 and S_2, and the corresponding recurrence rates (conditional on survival), dR_1 and dR_2, were chosen to satisfy the random-effects models,

$$S_2(t \mid \psi) = S_1(t \mid \psi)^\theta$$
$$dR_2(t \mid \psi) = S_1(t \mid \psi)^{1-\theta} dR_1(t \mid \psi),$$

where ψ is a gamma random variable with mean one and variance σ^2. For nonzero σ^2, the recurrence times and survival time are positively correlated. We set $S_1(t \mid \psi) = e^{-0.2\psi t}$ and $R_1(t \mid \psi) = \psi t$. The censoring time was a uniform$[0, 7]$ random variable, which yielded an average of about two observed recurrences per subject. For each simulation setting, 2000 samples were generated. The results are given in Table 3. Again, the proposed tests maintain their sizes around the nominal level, at least for moderately large samples.

Table 3
Empirical sizes of Q_{GT} and Q_{LR} when $\mu_1(\cdot) = \mu_2(\cdot)$ but $S_1(\cdot) \neq S_2(\cdot)$

		n_1=n_2=50		n_1=n_2=100	
θ	σ^2	Q_{GT}	Q_{LR}	Q_{GT}	Q_{LR}
1.0	0.00	0.060	0.059	0.048	0.049
	0.25	0.060	0.059	0.049	0.048
	0.50	0.058	0.059	0.050	0.050
	1.00	0.059	0.059	0.049	0.051
1.2	0.00	0.058	0.061	0.049	0.051
	0.25	0.061	0.057	0.051	0.050
	0.50	0.062	0.060	0.049	0.050
	1.00	0.060	0.058	0.050	0.048
1.4	0.00	0.060	0.058	0.050	0.050
	0.25	0.058	0.062	0.050	0.048
	0.50	0.062	0.058	0.050	0.051
	1.00	0.058	0.059	0.050	0.050
1.6	0.00	0.060	0.059	0.049	0.049
	0.25	0.060	0.060	0.049	0.049
	0.50	0.062	0.061	0.049	0.051
	1.00	0.059	0.059	0.050	0.049

Note: The survival time and recurrence event times were generated from the random-effects models $S_2(t \mid \psi) = S_1(t \mid \psi)^\theta$, $dR_2(t \mid \psi) = S_1(t \mid \psi)^{1-\theta} dR_1(t \mid \psi)$, where ψ is a gamma random variable with mean one and variance σ^2.

5. A Real Example

We now apply the methods developed in Sections 2 and 3 to the bladder tumor study described in Section 1. We will focus

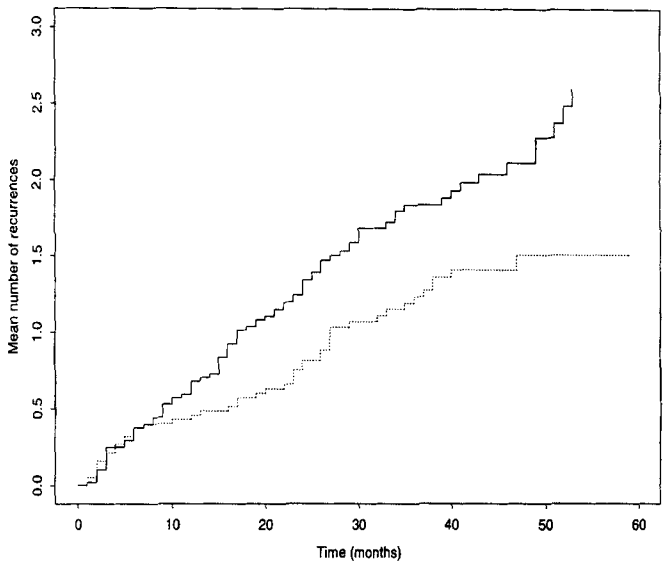

Figure 1. Estimated mean number of recurrences by treatment group: solid line, placebo; dashed line, thiotepa.

on the comparison between thiotepa and placebo. Thiotepa is a member of the class of alkylating agents, which were among the first anticancer drugs used. Alkylating agents are highly reactive and bind to certain chemical groups found in nucleic acids. These compounds inhibit proper synthesis of DNA and RNA, which leads to apoptosis or cell death. However, since alkylating agents cannot discriminate between cancerous and normal cells, both types of cells will be affected by this therapy. For example, normal cells can become cancerous due to alklyating agents. Thus, thiotepa is a highly cytotoxic compound and can potentially have adverse effects. Consequently, the effects of thiotepa on cancer recurrence and death are not obvious.

Figure 1 displays the estimated mean frequency functions of tumor recurrence for the placebo and thiotepa groups, while Figure 2 shows the estimates and the pointwise 95% confidence limits for the placebo group. There appears to be no early difference between the two treatment groups, but over time, patients on thiotepa have a lower frequency of recurrences than the placebo patients.

For the two-sample testing, we set τ to be the last observed event time and use weight functions (3.1) and (3.2). The observed value of $\tilde{Q}_{\mathrm{LR}}$ is 2.140, corresponding to a two-sided p-value of 0.030; the observed value of $\tilde{Q}_{\mathrm{GT}}$ is 1.958, yielding a two-sided p-value of 0.050. The observed value of $\tilde{Q}_{\mathrm{CT}}$ is 1.904, which yields a two-sided p-value of 0.056. The reason that the combined statistic is less significant than $\tilde{Q}_{\mathrm{LR}}$ and $\tilde{Q}_{\mathrm{GT}}$ is because the survival distributions in the two groups are very similar, as shown in Figure 3. The observed value of $\tilde{Q}_D$ is 0.894. The observed value of the quadratic form $Q_{\mathbf{T}}$ is 4.65, yielding a two-sided p-value of 0.098, which is less significant than the $\tilde{Q}_{\mathrm{CT}}$ test.

The main treatment difference appears to be that the patients on thiotepa have fewer recurrences than those on placebo. This is borne out by using the sequential testing procedure. In this case, $(\tilde{Q}_{\mathrm{LR}}, \tilde{Q}_D) = (2.140, 0.894)$ and the estimated correlation between $\tilde{Q}_{\mathrm{LR}}$ and $\tilde{Q}_D$ is 0.523. Since $\tilde{Q}_{\mathrm{LR}} > \tilde{Q}_D$, we test H_{10}: $\mu_1(t) = \mu_2(t)$ first and H_{20}: $\Lambda_1^D(t) = \Lambda_2^D(t)$ second. By numerical integration, $\Pr\{\max(V_1, V_2) \geq 2.140\} = 0.030$ and $\Pr(V_2 \geq 0.894) = 0.190$. Thus, at the one-sided significance level of 0.05 or two-sided significance level of 0.10, we reject H_{10} in favor of thiotepa and fail to reject H_{20}, as expected.

6. Discussion

In this article, we develop nonparametric methods for making inferences about the mean function of the recurrent events in

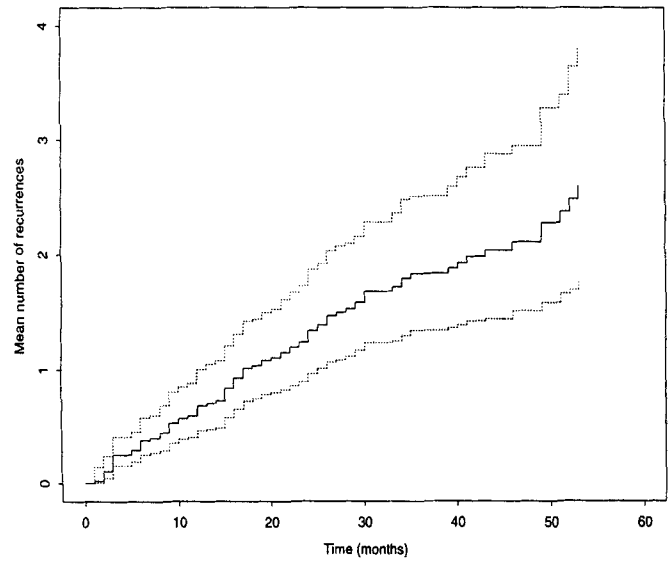

Figure 2. Estimated mean frequency function and associated 95% pointwise confidence limits for the placebo group: solid line, estimated mean frequency function; dotted line, 95% pointwise confidence limits.

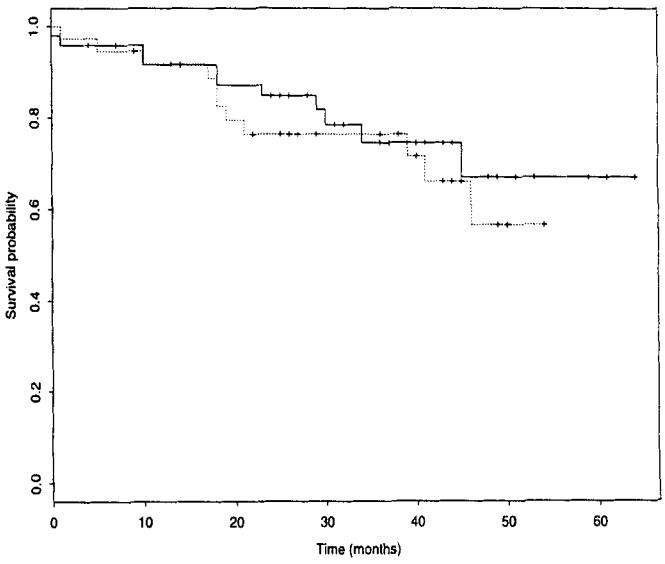

Figure 3. Estimated survival distributions by treatment group: solid line, placebo; dashed line, thiotepa.

the presence of death. If the event cannot recur, then the proposed methods reduce to those of Gray (1988), Pepe and Mori (1993), and Lin (1997) for the cumulative incidence function of competing risks. On the other hand, if there is no death, then the proposed methods reduce to those of Pepe and Cai (1993), Lawless and Nadeau (1995), Cook et al. (1996), and others for recurrent events. Thus, the proposed methodology merges the existing literature on recurrent events with that of competing risks.

A truncation time τ has been assumed for the two-sample statistics. One can extend τ to ∞ by setting the weight function in the statistics to zero for t greater than τ. As shown by the simulation results, one can take τ to be the last observed event time so that all the data are used in the analysis.

For censored survival data, the asymptotic efficiency of the weighted log-rank statistic has been well studied (Gill, 1980, Chapter 5). It does not seem possible to find an optimal weight for Q_{LR}. This was noted by Gray (1988) in the special case of the cumulative incidence function. Given the weight functions, however, one can chose an optimal value of p that maximizes the power of Q_{WCT}, along the lines of Wei et al. (1989).

In this article, we have concentrated on the one- and two-sample problems. It would also be desirable to develop regression methods. One could extend the approach of Cheng, Fine, and Wei (1998) by modeling $\lambda^D(t)$ and $dR(t)$ with proportional hazards and proportional rates models, respectively. It might be more appealing to model $\mu(t)$ directly. Both of these approaches are currently under investigation.

Acknowledgements

This research was supported by the U.S. National Institutes of Health grants GM47845 and AI29168. The authors thank the reviewers for their helpful comments.

Résumé

L'analyse d'événements récurrents en présence d'un événement terminal tel que le décès fait l'objet de cet article. Nous considérons une fonction de fréquence moyenne, définie comme la moyenne marginale du nombre cumulé d'événements récurrents au cours du temps. Un estimateur non paramétrique simple de cette quantité est présenté. Il est montré que cet estimateur, correctement normé, converge vers un processus gaussien de moyenne zéro, et dont la fonction de covariance est facilement estimable. Des tests non paramétriques permettant de comparer deux fonctions de fréquence moyenne et de prendre en compte simultanément les événements récurrents et les décès sont aussi proposés. Les distributions asymptotiques de ces statistiques sous l'hypothèse nulle, ainsi que des estimateurs de variance consistants sont calculés. Des études de simulation permettent d'examiner les propriétés des estimateurs proposés et des statistiques de tests en cas de petits échantillons. Une application aux données d'un essai clinique en cancérologie est présentée.

References

Albert, P. S. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47,** 1371–1381.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10,** 1100–1120.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Byar, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: Comparisons of placebo, pyridoxine and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology,* M. Pavone-Macaluso, P. H. Smith, and F. Edsmyr (eds), 363–370. New York: Plenum.

Cheng, S. C., Fine, J. P., and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* **54,** 219–228.

Cook, R. J. and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* **16,** 911–924.

Cook, R. J., Lawless, J. F., and Nadeau, C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52,** 557–571.

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

Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Amsterdam: The Mathematical Centre.

Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics* **16,** 1141–1154.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association* **55,** 698–707.

Hobson, R. W., Weiss, D. G., Fields, W. S., Goldstone, J., Moore, W. S., Towne, J. B., Wright, C. B., and the Veterans Affairs Cooperative Study Group. (1993). Effect of carotid endarterectomy for asymptomatic carotid stenosis. *New England Journal of Medicine* **328,** 221–227.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

Li, Q. and Lagakos, S. (1997). Use of the Wei–Lin–Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16,** 925–940.

Lin, D. Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine* **16,** 901–910.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85,** 605–618.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88,** 811–820.

Pepe, M. S. and Fleming, T. R. (1989). Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data. *Biometrics* **45,** 497–507.

Pepe, M. S. and Mori, N. (1993). Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* **12,** 737–751.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Received November 1998. *Revised July* 1999. *Accepted August* 1999.

APPENDIX

Derivations of Asymptotic Results

A.1. One-Sample Estimator

By the definitions of μ and $\hat{\mu}$,

$$n^{1/2}\{\hat{\mu}(t)-\mu(t)\} = n^{1/2}\int_0^t \hat{S}(u)d\left\{\hat{R}(u)-R(u)\right\} + n^{1/2}\int_0^t \left\{\hat{S}(u)-S(u)\right\}dR(u). \quad (A.1)$$

Due to the asymptotic equivalence of $\hat{S}(\cdot)$ and $e^{-\hat{\Lambda}^D(\cdot)}$, the second term on the right-hand side of (A.1) is

$$n^{1/2}\int_0^t \left\{\hat{S}(u)-S(u)\right\}dR(u) = -n^{1/2}\int_0^t \left\{\hat{\Lambda}^D(u)-\Lambda^D(u)\right\}d\mu(u)+o_P(1). \quad (A.2)$$

For $t \le \max_i X_i$,

$$n^{1/2}\left\{\hat{\Lambda}^D(t)-\Lambda^D(t)\right\} = n^{-1/2}\sum_{i=1}^n \int_0^t \frac{dM_i^D(u)}{\bar{Y}(u)/n}, \quad (A.3)$$

where $M_i^D(t) = N_i^D(t) - \int_0^t Y_i(u)d\Lambda^D(u)$. It then follows from the martingale central limit theorem (Fleming and Harrington, 1991, Theorem 5.3.5) that

$$n^{1/2}\left\{\hat{\Lambda}^D(t)-\Lambda^D(t)\right\} = n^{-1/2}\sum_{i=1}^n \int_0^t \frac{dM_i^D(u)}{\pi(u)}+o_P(1), \quad 0 \le t \le \tau, \quad (A.4)$$

where $\pi(t) = \Pr(X \ge t)$. Integrating the right-hand side of (A.2) by parts and using representation (A.4), we get that

$$n^{1/2}\int_0^t \left\{\hat{S}(u)-S(u)\right\}dR(u) = -n^{-1/2}\sum_{i=1}^n \left\{\mu(t)\int_0^t \frac{dM_i^D(u)}{\pi(u)} - \int_0^t \frac{\mu(u)dM_i^D(u)}{\pi(u)}\right\} + o_P(1). \quad (A.5)$$

By the consistency of the Kaplan–Meier estimator (Fleming and Harrington, 1991, p. 115), the first term on the right-hand side of (A.1) is

$$n^{1/2}\int_0^t \hat{S}(u)d\left\{\hat{R}(u)-R(u)\right\} = n^{1/2}\int_0^t S(u)d\{\hat{R}(u)-R(u)\}+o_P(1). \quad (A.6)$$

For $t \le \max_i X_i$,

$$n^{1/2}\{\hat{R}(t)-R(t)\} = n^{-1/2}\sum_{i=1}^n \int_0^t \frac{dM_i(u)}{\bar{Y}(u)/n}, \quad (A.7)$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(u)dR(u)$. Although the right-hand side of (A.7) takes a similar form to that of (A.3), its asymptotic properties are much more difficult to derive because, unlike the $M_i^D(\cdot)$'s, the $M_i(\cdot)$'s are not martingales, so the martingale central limit theorem is not applicable. Using some tools from the empirical process theory, however, we can show that

$$n^{1/2}\{\hat{R}(t)-R(t)\} = n^{-1/2}\sum_{i=1}^n \int_0^t \frac{dM_i(u)}{\pi(u)}+o_P(1), \quad 0 \le t \le \tau, \quad (A.8)$$

which implies that

$$n^{1/2}\int_0^t \hat{S}(u)d\left\{\hat{R}(u)-R(u)\right\} = n^{-1/2}\sum_{i=1}^n \int_0^t \frac{S(u)}{\pi(u)}dM_i(u)+o_P(1). \quad (A.9)$$

The arguments leading from (A.7) to (A.8) are very delicate and are thus omitted here. The interested reader is referred to the Appendix of Lin, Wei, and Ying (1998) for a similar proof in a related context.

In view of (A.1), (A.5), and (A.9), we have

$$n^{1/2}\{\hat{\mu}(t)-\mu(t)\} = n^{-1/2}\sum_{i=1}^n \Psi_i(t)+o_P(1), \quad (A.10)$$

where

$$\Psi_i(t) = \int_0^t \frac{S(u)}{\pi(u)}dM_i(u) - \mu(t)\int_0^t \frac{dM_i^D(u)}{\pi(u)} + \int_0^t \frac{\mu(u)dM_i^D(u)}{\pi(u)}.$$

For any fixed t, $\Psi_i(t)$ $(i = 1,\ldots,n)$ are i.i.d. mean-zero random variables. Thus, the finite-dimensional normality of (A.10) follows from the classical central limit theorem. By the arguments given in the proof of Theorem 1 of Lin et al. (1998), we can show that (A.10) is tight and thus converges weakly to a zero-mean Gaussian process. The limiting covariance function at the pair of time points (s,t) is given by $\xi(s,t) = \mathrm{E}\{\Psi_1(s)\Psi_1(t)\}$. It is natural to estimate $\xi(s,t)$ by $\hat{\xi}(s,t) = n^{-1}\sum_{i=1}^n \hat{\Psi}_i(s)\hat{\Psi}_i(t)$, where $\hat{\Psi}_i(t)$ is obtained by replacing all the unknown parameters in $\Psi_i(t)$ with their respective sample estimators, as given in (2.2). The consistency of $\hat{\xi}$ follows from the arguments used in the proof of Theorem 3 of Lin et al. (1998).

A.2. Two-Sample Statistics

Let $\rho_j = \lim_{n\to\infty} n_j/n$ $(j = 1,2)$. Assume that $\hat{K}_{\mathrm{LR}}(t)$ converges to a bounded function $K_{\mathrm{LR}}(t)$ uniformly on $[0,\tau]$. Then under H_0: $\mu_1(t) = \mu_2(t)$ for all t,

$$\left(\frac{n_1n_2}{n}\right)^{1/2} Q_{\mathrm{LR}} = \left(\frac{n_1n_2}{n}\right)^{1/2}$$

$$\times \left[\int_0^\tau K_{\mathrm{LR}}(t) d\{\hat{\mu}_1(t) - \mu_1(t)\} - \int_0^\tau K_{\mathrm{LR}}(t) d\{\hat{\mu}_2(t) - \mu_2(t)\} \right] + o_P(1). \quad \text{(A.11)}$$

According to (A.10),

$$n_j^{1/2}\{\hat{\mu}_j(t) - \mu_j(t)\} = n_j^{-1/2} \sum_{i=1}^{n_j} \Psi_{ji}(t) + o_P(1), \qquad j = 1, 2,$$

where $\Psi_{ji}(t)$ is the expression of $\Psi_i(t)$ for the jth group ($j = 1, 2$). Thus, the right-hand side of (A.11) is equal to

$$\left(\frac{n_2}{n}\right)^{1/2} n_1^{-1/2} \sum_{i=1}^{n_1} \int_0^\tau K_{\mathrm{LR}}(t) d\Psi_{1i}(t) - \left(\frac{n_1}{n}\right)^{1/2} n_2^{-1/2} \sum_{i=1}^{n_2} \int_0^\tau K_{\mathrm{LR}}(t) d\Psi_{2i}(t) + o_P(1), \quad \text{(A.12)}$$

which is essentially the difference of two independent sums of i.i.d. mean-zero random variables. It then follows from the classical central limit theorem that $(n_1 n_2/n)^{1/2} Q_{\mathrm{LR}}$ converges to a mean-zero normal random variable with variance

$$\sigma^2_{\mathrm{LR}} = \rho_2 \mathrm{E}\left\{ \int_0^\tau K_{\mathrm{LR}}(t) d\Psi_{11}(t) \right\}^2 + \rho_1 \mathrm{E}\left\{ \int_0^\tau K_{\mathrm{LR}}(t) d\Psi_{21}(t) \right\}^2.$$

The estimator $\hat{\sigma}^2_{\mathrm{LR}}$ given in (3.3) is obtained from σ^2_{LR} by replacing all the unknown quantities with their empirical counterparts. The consistency of $\hat{\sigma}^2_{\mathrm{LR}}$ follows from the kind of arguments given in the proof of Theorem 3 of Lin et al. (1998).

In a similar manner, we can show that, under H_0, the generalized t-statistic $(n_1 n_2/n)^{1/2} Q_{\mathrm{GT}}$ converges to mean-zero normal random variable with variance

$$\sigma^2_{\mathrm{GT}} = \rho_2 \mathrm{E}\left\{ \int_0^\tau K_{\mathrm{GT}}(t) \Psi_{11}(t) dt \right\}^2 + \rho_1 \mathrm{E}\left\{ \int_0^\tau K_{\mathrm{GT}}(t) \Psi_{21}(t) dt \right\}^2,$$

where K_{GT} is the limit of $\hat{K}_{\mathrm{GT}}$. A consistent estimator for σ^2_{GT} is given in (3.4).

We now derive the asymptotic distributions of $\mathbf{T}$ and Q_{WCT} under H_0^*: $\mu_1(t) = \mu_2(t)$ and $\Lambda_1^D(t) = \Lambda_2^D(t)$ for all t. By the martingale central limit theorem (Fleming and Harrington, 1991, Theorem 5.3.5),

$$\left(\frac{n_1 n_2}{n}\right)^{1/2} Q_D = \left(\frac{n_2}{n}\right)^{1/2} n_1^{-1/2} \sum_{i=1}^{n_1} \int_0^\tau K_{\mathrm{LR}}(t) \frac{dM_{1i}^D(t)}{\pi_1(t)} - \left(\frac{n_1}{n}\right)^{1/2} n_2^{-1/2} \sum_{i=1}^{n_2} \int_0^\tau K_{\mathrm{LR}}(t) \frac{dM_{2i}^D(t)}{\pi_2(t)} + o_P(1), \quad \text{(A.13)}$$

where $M_{ji}^D(t)$ and $\pi_j(t)$ are the expressions of $M_i^D(t)$ and $\pi(t)$ for the jth group ($j = 1, 2$). Because both (A.12) and the right-hand side of (A.13) are sums of i.i.d. mean-zero random variables, the classical multivariate central limit theorem implies that, under H_0^*, $\mathbf{T}$ converges in distribution to a bivariate normal vector with mean zero and covariance matrix $\boldsymbol{\Sigma} = \rho_2 \mathrm{E}(\mathbf{V}_{11}\mathbf{V}_{11}') + \rho_1 \mathrm{E}(\mathbf{V}_{21}\mathbf{V}_{21}')$, where $\mathbf{V}_{ji} = [\int_0^t K_{\mathrm{LR}}(t) d\Psi_{ji}(t), \int_0^t \{K_{\mathrm{LR}}(t)/\pi_j(t)\} dM_{ji}^D(t)]'$ ($i = 1, \ldots, n_j$; $j = 1, 2$). It also follows from (A.12) and (A.13) that

$$\left(\frac{n_1 n_2}{n}\right)^{1/2} Q_{\mathrm{WCT}} = \left(\frac{n_2}{n}\right)^{1/2} n_1^{-1/2} \sum_{i=1}^{n_1} U_{1i} - \left(\frac{n_1}{n}\right)^{1/2} n_2^{-1/2} \sum_{i=1}^{n_2} U_{2i} + o_P(1),$$

where $U_{ji} = \int_0^\tau K_{\mathrm{LR}}(t)\{p\, d\Psi_{ji}(t) + (1-p)\, dM_{ji}^D(t)/\pi_j(t)\}$ ($i = 1, \ldots, n_j; j = 1, 2$). Because the U_{ji}'s are i.i.d. mean-zero random variables, the classical central limit theorem implies that Q_{WCT} converges to a mean-zero normal random variable with variance $\sigma^2_{\mathrm{WCT}} = \rho_2 \mathrm{E}(U_{11}^2) + \rho_1 \mathrm{E}(U_{21}^2)$. Replacing all the unknown quantities in $\boldsymbol{\Sigma}$ and σ^2_{WCT} with their respective empirical counterparts yields consistent estimators $\hat{\boldsymbol{\Sigma}}$ and $\hat{\sigma}^2_{\mathrm{WCT}}$.