BIOMETRICS 57, 369–375
June 2001

Nonparametric Tests for the Gap Time Distributions of Serial Events Based on Censored Data

D. Y. Lin

Department of Biostatistics, University of North Carolina,
CB 7400, McGavran-Greenberg Hall, Chapel Hill, North Carolina 27599-7400, U.S.A.
email: lin@bios.unc.edu

and

Zhiliang Ying

Department of Statistics, Hill Center, Busch Campus, Rutgers University,
Piscataway, New Jersey 08855, U.S.A.

SUMMARY. This article deals with the problem of comparing two populations with respect to the distribution of the gap time between two successive events when each subject can experience a series of events and when the event times are potentially right censored. Several families of nonparametric tests are developed, all of which allow arbitrary distributions and dependence structures for the serial events. The asymptotic and small-sample properties of the proposed tests are investigated. An illustration with data taken from a colon cancer study is provided. The related problem of testing the independence of two successive gap times is also studied.

KEY WORDS: Correlated failure times; Dependent censoring; Kaplan–Meier estimator; Log-rank statistic; Multiple events; Recurrent events; Survival analysis.

1. Introduction

In many longitudinal studies, each individual subject can potentially experience a series of events. These events may be repetitions of essentially the same phenomenon or may be events of entirely different natures. Medical examples of such serial events include episodic illness, such as multiple infections, and the progression of a disease through successive stages, such as HIV infection → AIDS → death. Other examples include the repeated breakdowns of a machinery in reliability testing and the employment/unemployment cycles in socioeconomic analysis. In such studies, the investigators are often interested in the gap time between two successive events, such as the time between the first and second infection episodes, the time from the development of AIDS to death, and the duration of employment.

Serial events are typically subject to right censoring in that the follow-up is terminated before the developments of the events of interest. For such censored data, the Kaplan–Meier estimator cannot be used to estimate the gap time distribution unless the gap time is independent of the preceding event time. Recently, Wang and Wells (1998), Lin, Sun, and Ying (1999), and Wang and Chang (1999) provided appropriate nonparametric estimators of the gap time distribution.

In some applications, it is desirable to compare the gap time distributions between two groups. For treating a recurrent symptom, e.g., one may be interested in comparing two treatments not only in terms of how long it takes for the treatment to become effective, as measured from the initiation of the treatment to the disappearance of the symptom, but also in terms of how long the treatment maintains its effectiveness, as measured from the disappearance of the symptom to the reappearance of the symptom. The existing censored-data statistics, such as the log-rank statistic, can be used to compare the first durations between the treatments but not the second durations. It is not obvious how to develop two-sample tests for the gap time distributions because the aforementioned nonparametric estimators are much more complicated than the Kaplan–Meier and Nelson–Aalen estimators.

In this article, we propose several classes of two-sample nonparametric statistics for comparing the gap time distributions. These statistics are based on a nonparametric estimator of the gap time distribution given by Lin et al. (1999) and are analogous to familiar censored-data statistics, such as weighted log-rank statistics and Pepe–Fleming (Pepe and Fleming, 1989) statistics. We describe the proposed statistics and their asymptotic properties in the next section. In Section 3, we report the results of some simulation studies. In Section 4, we revisit the colon cancer data previously analyzed by Lin et al. (1999). In Section 5, we discuss tests of independence for successive gap times and address some other issues.

2. Two-Sample Tests

2.1 *Preliminaries*

Suppose that one is interested in the gap time between the first and second events. All the results will apply to the gap time between any two successive events. We will adopt the basic notation introduced by Lin et al. (1999) and use the superscript (j) to denote the jth $(j = 1, 2)$ group. Specifically, for the jth group, let $Y_1^{(j)} < Y_2^{(j)}$ be the times to the first and second events measured from the study entry and let $C^{(j)}$ be the follow-up or censoring time, also measured from the study entry. Define $T_1^{(j)} = Y_1^{(j)}$ and $T_2^{(j)} = Y_2^{(j)} - Y_1^{(j)}$ $(j = 1, 2)$. We will at times suppress the superscript (j), in which case the statement pertains to both $j = 1$ and 2. We assume that C is independent of T_1 and T_2 but allow T_1 and T_2 to be correlated. We also allow $C^{(1)}$ and $C^{(2)}$ to have different distributions. It is important to note that the censoring time on T_2 is $C - T_1$, which is correlated with T_2 unless T_1 and T_2 are independent. Because of this dependent censoring, standard survival analysis methods, such as the Kaplan–Meier estimator and log-rank test, cannot be used to make inference about T_2.

As discussed in Lin et al. (1999), the marginal distribution of T_2 is not estimable unless the support of Y_2 is contained in that of C. Thus, Lin et al. considered the conditional distribution function

$$F(t \mid s) = \Pr(T_2 \le t \mid T_1 \le s).$$

If $\Pr(T_1 \le s_0)$ is large, then $F(t \mid s_0)$ provides a good approximation to the marginal distribution of T_2. Even if the marginal distribution is estimable, the conditional distribution may still be of interest. As demonstrated by Lin et al., it is possible to estimate $F(t \mid s)$ as long as $t + s \le \tau_c$, where $\tau_c = \sup\{t : \Pr(C \ge t) > 0\}$.

We wish to develop nonparametric statistics for comparing $F^{(1)}$ and $F^{(2)}$ between groups 1 and 2. Specifically, suppose that $\Pr(C^{(j)} \ge \tau) > 0$ $(j = 1, 2)$ for some constant τ and that $s_0 < \tau$ is another fixed constant. Then we are interested in testing the null hypothesis

$$\text{H}_0\colon F^{(1)}(t \mid s) = F^{(2)}(t \mid s) \qquad \text{for all } \{(t, s) : t + s \le \tau\},$$

or

$$\text{H}_0'\colon F^{(1)}(t \mid s_0) = F^{(2)}(t \mid s_0) \qquad \text{for all } t \le \tau - s_0.$$

Suppose that there are n_j subjects in the jth, $(j = 1, 2)$ group. Denote $n = n_1 + n_2$. The data consist of $(\tilde{Y}_{1i}^{(j)}, \tilde{Y}_{2i}^{(j)}, \delta_{1i}^{(j)}, \delta_{2i}^{(j)})$, $(i = 1, \ldots, n_j; j = 1, 2)$, where $\tilde{Y}_{ki}^{(j)} = Y_{ki}^{(j)} \wedge C_i^{(j)}$ and $\delta_{ki}^{(j)} = I(Y_{ki}^{(j)} \le C_i^{(j)})$ $(k = 1, 2; i = 1, \ldots, n_j; j = 1, 2)$. Here and in the sequel, $a \wedge b = \min(a, b)$, $a^+ = \max(a, 0)$, and $I(\cdot)$ is the indicator function.

Let $G^{(j)}(t) = \Pr(C^{(j)} > t)$ and let $\hat{G}^{(j)}$ be the Kaplan–Meier estimator of $G^{(j)}$ based on the data $(\tilde{Y}_{2i}^{(j)}, 1 - \delta_{2i}^{(j)})$ $(i = 1, \ldots, n_j)$. In addition, define

$$\hat{H}^{(j)}(s, t) = n_j^{-1} \sum_{i=1}^{n_j} \frac{I\left(\tilde{Y}_{1i}^{(j)} \le s, \tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} > t\right)}{\hat{G}^{(j)}(\tilde{Y}_{1i}^{(j)} + t)}, \quad j = 1, 2,$$

$\hat{H}^{(j)}(t \mid s) = \hat{H}^{(j)}(s, t)/\hat{H}^{(j)}(s, 0)$, and $\hat{F}^{(j)}(t \mid s) = 1 - \hat{H}^{(j)}(t \mid s)$. According to Lin et al. (1999), $F^{(j)}(t \mid s)$ can be consistently estimated by $\hat{F}^{(j)}(t \mid s)$. In the remainder of this section, we develop four families of test statistics based on $\hat{F}^{(j)}(t \mid s)$ $(j = 1, 2)$.

2.2 *Pepe–Fleming-Type Tests*

We first introduce a family of tests for comparing two gap time distributions that is analogous to the Pepe–Fleming (1989) tests for censored survival data. Let $\mu_n(dt, ds)$ be a measure on $[0, \tau] \times [0, \tau]$. The measure can be data dependent but is assumed to converge in probability to a nonrandom measure $\mu(dt, ds)$. Define

$$U_{\text{PF}} = \int_0^{\tau} \int_0^{\tau - s} \left\{\hat{F}^{(2)}(t \mid s) - \hat{F}^{(1)}(t \mid s)\right\} \mu_n(dt, ds). \quad (1)$$

Clearly, a test based on properly normalized U_{PF} will be consistent against any alternative $\text{H}_1\colon F^{(2)}(t \mid s) \ge F^{(1)}(t \mid s)$ or $F^{(2)}(t \mid s) \le F^{(1)}(t \mid s)$ for all $\{(t, s) : t + s \le \tau\}$, where the inequality is strict for some (t, s) associated with positive $\mu(dt, ds)$.

With flexible choices of μ_n, U_{PF} encompasses a rich class of statistics. We may take the measure to be the product of two marginal measures, i.e., $\mu_n(dt, ds) = \mu_{n,1}(dt)\mu_{n,2}(ds)$. To obtain appropriate statistics for testing H_0', we set $\mu_{n,2}(ds)$ to be a point mass, say one, at s_0. Then U_{PF} becomes

$$U_{\text{PF}}' = \int_0^{\tau - s_0} \left\{\hat{F}^{(2)}(t \mid s_0) - \hat{F}^{(1)}(t \mid s_0)\right\} \mu_{n,1}(dt)$$

or

$$U_{\text{PF}}' = \int_0^{\tau - s_0} W_n(t) \left\{\hat{F}^{(2)}(t \mid s_0) - \hat{F}^{(1)}(t \mid s_0)\right\} dt \quad (2)$$

with a suitable weight function $W_n(t)$. We denote U_{PF} by U_{PF}' in (2) to indicate that the latter is used for testing H_0'. If $T_1^{(j)}$ $(j = 1, 2)$ were nonrandom and equal to s_0, then U_{PF}' would reduce to the original Pepe–Fleming statistics. One possible choice of the weight function for U_{PF}' is

$$W_n(t) = \frac{n\hat{G}^{(1)}(s_0 + t)\hat{G}^{(2)}(s_0 + t)}{n_1\hat{G}^{(1)}(s_0 + t) + n_2\hat{G}^{(2)}(s_0 + t)}, \quad (3)$$

which is analogous to a weight function suggested by Pepe and Fleming (1989).

We show in Appendix 1 that, under H_0, $(n_1 n_2/n)^{1/2} U_{\text{PF}}$ converges in distribution to a normal random variable with mean zero and with a variance that can be consistently estimated by

$$\hat{V}_{\text{PF}} = \sum_{j=1}^{2} \frac{n - n_j}{n} \times \int \hat{\sigma}^{(j)}(t_1, t_2 \mid s_1, s_2)\mu_n(dt_1, ds_1)\mu_n(dt_2, ds_2), \quad (4)$$

where the integration is over $\{(s_1, s_2, t_1, t_2) : s_1 + t_1 \le \tau, s_2 + t_2 \le \tau\}$,

$$\hat{\sigma}^{(j)}(t_1, t_2 \mid s_1, s_2)$$

$$= \frac{n_j^{-1}}{\hat{H}^{(j)}(s_1,0)\hat{H}^{(j)}(s_2,0)} \times \sum_{i=1}^{n_j} \Bigg[\hat{D}_i^{(j)}(s_1,t_1)\hat{D}_i^{(j)}(s_2,t_2) - \frac{\left(1-\delta_{2i}^{(j)}\right)\hat{B}^{(j)}\left(s_1,t_1;\tilde{Y}_{2i}^{(j)}\right)}{\left\{n_j^{-1}\sum_{l=1}^{n_j} I\left(\tilde{Y}_{2l}^{(j)} \ge \tilde{Y}_{2i}^{(j)}\right)\right\}^2} \times \hat{B}^{(j)}\left(s_2,t_2;\tilde{Y}_{2i}^{(j)}\right)\Bigg],$$

$$\hat{D}_i^{(j)}(s,t) = \frac{\delta_{1i}^{(j)} I\left(\tilde{Y}_{1i}^{(j)} \le s\right)\hat{H}^{(j)}(t \mid s)}{\hat{G}^{(j)}\left(\tilde{Y}_{1i}^{(j)}\right)} - \frac{I\left(\tilde{Y}_{1i}^{(j)} \le s, \tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} > t\right)}{\hat{G}^{(j)}\left(\tilde{Y}_{1i}^{(j)} + t\right)},$$

$$\hat{B}^{(j)}(s,t;u) = \hat{H}^{(j)}(t \mid s)\left\{\hat{H}^{(j)}(s,0) - \hat{H}^{(j)}(u,0)\right\}^+ - \left\{\hat{H}^{(j)}(s,t) - \hat{H}^{(j)}(u-t,t)\right\}^+.$$

As a by-product, under H_0', $(n_1n_2/n)^{1/2}U_{PF}'$ is asymptotically zero-mean normal with consistent variance estimator

$$\hat{V}_{PF}' = \sum_{j=1}^{2} \frac{n-n_j}{n\, n_j\, \hat{H}^{(j)}(s_0,0)^2} \times \sum_{i=1}^{n_j} \Bigg(\delta_{1i}^{(j)} I\left(\tilde{Y}_{1i}^{(j)} \le s_0\right) \times \left[\int_0^{\tau-s_0} W_n(t) \times \left\{ \frac{\hat{H}^{(j)}(t \mid s_0)}{\hat{G}^{(j)}\left(\tilde{Y}_{1i}^{(j)}\right)} - \frac{I\left(\tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} > t\right)}{\hat{G}^{(j)}\left(\tilde{Y}_{1i}^{(j)} + t\right)}\right\} dt\right]^2 - \frac{\left(1-\delta_{2i}^{(j)}\right)}{\left\{n_j^{-1}\sum_{l=1}^{n_j} I\left(\tilde{Y}_{2l}^{(j)} \ge \tilde{Y}_{2i}^{(j)}\right)\right\}^2} \times \left\{\int_0^{\tau-s_0} W_n(t)\hat{B}^{(j)}\left(s_0,t;\tilde{Y}_{2i}^{(j)}\right) dt\right\}^2 \Bigg).$$

2.3 *Log-Rank-Type Tests*

Define $\Lambda^{(j)}(t \mid s) = -\log\{1 - F^{(j)}(t \mid s)\}$, which is analogous to the cumulative hazard function. We can estimate $\Lambda^{(j)}(t \mid s)$ consistently by $\hat{\Lambda}^{(j)}(t \mid s) = -\log \hat{H}^{(j)}(t \mid s)$. Let $\nu_{n,1}(ds)$ be a measure on $[0,\tau]$ and let $\nu_{n,2}(t \mid s)$ be a bivariate function such that, for each fixed s, $\nu_{n,2}(t \mid s)$ is left continuous and has bounded variation with respect to t. Suppose that $\nu_{n,2}(t \mid s) = 0$ if $t + s \ge \tau$. We assume that $\nu_{n,1}$ and $\nu_{n,2}$ converge in probability to nonrandom ν_1 and ν_2. The following class of statistics is a generalization of the familiar weighted log-rank statistics (Fleming and Harrington, 1991, pp. 107–108):

$$U_{LR} = \int_0^{\tau}\int_0^{\tau-s} \nu_{n,2}(t \mid s) \times \left\{\hat{\Lambda}^{(2)}(dt \mid s) - \hat{\Lambda}^{(1)}(dt \mid s)\right\}\nu_{n,1}(ds).$$

If $\nu_{n,1}(ds)$ is a point mass of one at s_0, then U_{LR} reduces to

$$U_{LR}' = \int_0^{\tau-s_0} \nu_{n,2}(t \mid s_0)\left\{\hat{\Lambda}^{(2)}(dt \mid s_0) - \hat{\Lambda}^{(1)}(dt \mid s_0)\right\}.$$

Naturally, normalized U_{LR} and U_{LR}' can be used to test H_0 and H_0', respectively.

If $T_1^{(j)}$ $(j = 1,2)$ were equal to s_0, then U_{LR}' would be asymptotically equivalent to the weighted log-rank statistics. The form of $\nu_{n,2}$ for U_{LR}' that corresponds to the popular unweighted log-rank statistic (Fleming and Harrington, 1991, p. 107) is

$$\nu_{n,2}(t \mid s_0) = \frac{\prod_{j=1}^{2}\sum_{i=1}^{n_j} I\left(\tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} \ge t\right)}{\sum_{j=1}^{2}\sum_{i=1}^{n_j} I\left(\tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} \ge t\right)}, \quad 0 \le t < \tau - s_0. \tag{5}$$

We show in Appendix 2 that, under H_0, $(n_1n_2/n)^{1/2}U_{LR}$ converges in distribution to a zero-mean normal random variable with a variance that can be consistently estimated by

$$\hat{V}_{LR} = \sum_{j=1}^{2}\frac{n-n_j}{n} \times \int \hat{\sigma}^{(j)}(t_1,t_2 \mid s_1,s_2) \times \frac{\nu_{n,2}(dt_1 \mid s_1)\nu_{n,1}(ds_1)}{\{1-\hat{F}(t_1 \mid s_1)\}\{1-\hat{F}(t_2 \mid s_2)\}} \times \nu_{n,2}(dt_2 \mid s_2)\nu_{n,1}(ds_2),$$

where the integration is over $\{(s_1,s_2,t_1,t_2) : s_1 + t_1 \le \tau, s_2 + t_2 \le \tau\}$ and $\hat{F}(t \mid s)$ is the estimator of the common $F(t \mid s)$ based on the pooled data from the two groups. It follows that, under H_0', $(n_1n_2/n)^{1/2}U_{LR}'$ is asymptotically zero-mean normal with consistent variance estimator

$$\hat{V}_{LR}' = \sum_{j=1}^{2}\frac{n-n_j}{n\, n_j\, \hat{H}^{(j)}(s_0,0)^2} \times \sum_{i=1}^{n_j}\Bigg(\delta_{1i}^{(j)} I\left(\tilde{Y}_{1i}^{(j)} \le s_0\right) \times \left[\int_0^{\tau-s_0}\left\{\frac{\hat{H}^{(j)}(t \mid s_0)}{\hat{G}^{(j)}\left(\tilde{Y}_{1i}^{(j)}\right)} - \frac{I\left(\tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} > t\right)}{\hat{G}^{(j)}\left(\tilde{Y}_{1i}^{(j)} + t\right)}\right\}\right.$$

$$\times \frac{\nu_{n,2}(dt \mid s_0)}{1-\hat{F}(t \mid s_0)}\Bigg]^2 - \frac{\left(1-\delta_{2i}^{(j)}\right)}{\left\{n_j^{-1}\sum_{l=1}^{n_j} I\left(\tilde{Y}_{2l}^{(j)} \geq \tilde{Y}_{2i}^{(j)}\right)\right\}^2} \times \left\{\int_0^{\tau-s_0} \frac{\hat{B}^{(j)}\left(s_0, t; \tilde{Y}_{2i}^{(j)}\right)\nu_{n,2}(dt \mid s_0)}{1-\hat{F}(t \mid s_0)}\right\}^2\Bigg).$$

As in the case of classical survival data, the Pepe–Fleming-type tests are sensitive to the alternatives that $F^{(2)}(t \mid s)$ is consistently greater (or less) than $F^{(1)}(t \mid s)$, while the log-rank-type tests are sensitive to the alternatives that $\Lambda^{(2)}(dt \mid s)$ is consistently greater (or less) than $\Lambda^{(1)}(dt \mid s)$. Thus, the choice between these two types of tests as well as the choices of μ_n, $\nu_{n,1}$, and $\nu_{n,2}$ depend on the alternatives against which one is most interested in detecting. Computationally, it is less demanding to implement the log-rank-type tests than the Pepe–Fleming-type tests, especially if one makes use of the alternative expression for $U_{\rm LR}$ given in (A.4).

2.4 *Kolmogorov–Smirnov- and Cramér-von Mises-Type Tests*

One can define Kolmogorov–Smirnov- and Cramér-von Mises-type statistics (Schumacher, 1984),

$$U_{\rm KS} = n^{1/2} \sup_{0\leq s\leq\tau}\ \sup_{0\leq t\leq\tau-s} g_n(t,s)\left|\hat{F}^{(2)}(t \mid s) - \hat{F}^{(1)}(t \mid s)\right|,$$

$$U_{\rm CV} = n\int_0^{\tau}\int_0^{\tau-s}\left\{\hat{F}^{(2)}(t \mid s) - \hat{F}^{(1)}(t \mid s)\right\}^2 \zeta_n(dt,ds),$$

where $g_n(t,s)$ is a positive bivariate function and $\zeta_n(dt,ds)$ is a bivariate measure defined on $[0,\tau]\times[0,\tau]$. A nice feature of $U_{\rm KS}$ and $U_{\rm CV}$ is that they are consistent against any departure from $\rm H_0$. The cutoff points for these test statistics are difficult to determine analytically even if s is fixed but can be obtained by computer simulation.

Specifically, let

$$R^{(j)}(t \mid s) = n_j^{-1}\sum_{i=1}^{n_j}\hat{Q}_i^{(j)}(t \mid s)Z_i^{(j)},$$

where $\hat{Q}_i^{(j)}$ $(i=1,\ldots,n_j; j=1,2)$ are defined in Appendix 3 and $Z_i^{(j)}$ $(i=1,\ldots,n_j; j=1,2)$ are independent standard normal random variables that are independent of the data $(\tilde{Y}_{1i}^{(j)}, \tilde{Y}_{2i}^{(j)}, \delta_{1i}^{(j)}, \delta_{2i}^{(j)})$ $(i=1,\ldots,n_j; j=1,2)$. Define

$$\tilde{U}_{\rm KS} = n^{1/2} \sup_{0\leq s\leq\tau}\ \sup_{0\leq t\leq\tau-s} g_n(t,s)\left|R^{(2)}(t \mid s) - R^{(1)}(t \mid s)\right|,$$

$$\tilde{U}_{\rm CV} = n\int_0^{\tau}\int_0^{\tau-s}\left\{R^{(2)}(t \mid s) - R^{(1)}(t \mid s)\right\}^2 \zeta_n(dt,ds).$$

We show in Appendix 3 that the null distributions of $U_{\rm KS}$ and $U_{\rm CV}$ can be approximated by those of $\tilde{U}_{\rm KS}$ and $\tilde{U}_{\rm CV}$, respectively. The latter distributions can be determined by repeatedly generating the normal random samples $Z_i^{(j)}$ $(i=1,\ldots,n_j; j=1,2)$ while holding the data $(\tilde{Y}_{1i}^{(j)}, \tilde{Y}_{2i}^{(j)}, \delta_{1i}^{(j)}, \delta_{2i}^{(j)})$ $(i=1,\ldots,n_j; j=1,2)$ at their observed values.

3. Numerical Studies

We conducted a series of simulation studies to evaluate the operating characteristics of the proposed two-sample tests in realistic settings. The gap times (T_1, T_2) had Gumbel's (1960) bivariate distribution function

$$F(t_1,t_2) = F_1(t_1)F_2(t_2)\left[1+\theta\{1-F_1(t_1)\}\{1-F_2(t_2)\}\right]$$

with exponential margins F_1 and F_2. The dependence parameter θ was set to either zero or one, corresponding to zero or 0.25 correlation between T_1 and T_2. Denote the hazard rates of $T_k^{(j)}$ $(k=1,2; j=1,2)$ by $\lambda_k^{(j)}$. We set $\lambda_1^{(1)} = \lambda_2^{(1)} = 1$ and $(\lambda_1^{(2)}, \lambda_2^{(2)}) = (1,1)$, $(1,2)$, or $(2,2)$. The follow-up times $C^{(j)}$ $(j=1,2)$ were uniform$[0,4]$ variables. Consequently, the censoring percentages for T_1 and T_2 were about 25 and 50% when $\lambda_1=\lambda_2=1$, 25 and 36% when $\lambda_1=1$ and $\lambda_2=2$, and 12 and 25% when $\lambda_1=\lambda_2=2$.

The studies focused on the Pepe–Fleming-type test $U'_{\rm PF}$ with W_n given in (3) and the log-rank-type test $U'_{\rm LR}$ with $\nu_{n,2}$ given in (5). We set $\tau=4$ and $s_0=1$, 2, or 2.5 and considered sample sizes of $n_1=n_2=50$ or 100. For each configuration of the simulation parameters, 10,000 data sets were generated to estimate the empirical sizes/powers of the two tests.

The findings of these studies are shown in Table 1. The log-rank-type test $U'_{\rm LR}$ maintains its size near the nominal level in all cases, while the Pepe–Fleming-type test $U'_{\rm PF}$ appears to be slightly anticonservative, especially when the sample sizes are small. The anticonservativeness of $U'_{\rm PF}$ is mainly due to the tail instabilities of $\hat{F}$. There are considerable power differences between the two types of tests. For example, $U'_{\rm LR}$ tends to be more powerful than $U'_{\rm PF}$ when s_0 and $\lambda_1^{(2)}$ are large. This is not a surprise because $U'_{\rm LR}$ with $\nu_{n,2}$ given in (5) is the most powerful against proportional hazards alternatives and because the difference between $F^{(2)}(t \mid s_0)$ and $F^{(1)}(t \mid s_0)$ tends to proportional hazards as $\Pr(T_1^{(j)} < s_0)$ $(j=1,2)$ approach one. While the power of $U'_{\rm LR}$ tends to increase as s_0 increases, the power of $U'_{\rm PF}$ appears to be higher at $s_0=2$ than at $s_0=2.5$. These phenomena are attributed to the fact that both the form of the alternative and the amount of data used for comparison vary with s_0.

Additional simulation studies were conducted to assess the sizes of the proposed tests under $\lambda_1^{(2)}=2$ and $\lambda_2^{(2)}=1$. In this situation, the marginal distributions between the two groups are the same whereas their conditional distributions are not unless T_1 and T_2 are independent or s_0 is large. The results revealed that, when $\theta=0$, the sizes of $U'_{\rm LR}$ and $U'_{\rm PF}$ are similar to those shown in Table 1, and when $\theta=1$, the sizes are reasonable if τ and s_0 are sufficiently large, say $\tau=5$ and $s_0=4$ or $\tau=7$ and $s_0=5$.

4. A Real Example

In a randomized clinical trial presented by Lin et al. (1999), 304 patients with resected colon cancer received adjuvant therapy with levamisole plus fluoraouracil while 315 were observed without therapy. The study lasted over 8 years. By the end of the study, 119 patients in the treated group had cancer recurrence, among whom 108 died, while in the untreated group, 177 patients had cancer recurrence, among whom 155 died. Figure 1c in Lin et al. revealed the somewhat

Table 1
Empirical sizes/powers for $U'_{\rm PF}$ and $U'_{\rm LR}$ under bivariate exponential models at the 0.05 nominal significance level

			$\lambda_1^{(2)} = \lambda_2^{(2)} = 1$		$\lambda_1^{(2)} = 1, \lambda_2^{(2)} = 2$		$\lambda_1^{(2)} = \lambda_2^{(2)} = 2$	
$n_1 = n_2$	θ	s_0	$U'_{\rm PF}$	$U'_{\rm LR}$	$U'_{\rm PF}$	$U'_{\rm LR}$	$U'_{\rm PF}$	$U'_{\rm LR}$
50	0	1.0	0.064	0.041	0.547	0.453	0.539	0.518
		2.0	0.064	0.048	0.580	0.545	0.542	0.605
		2.5	0.066	0.060	0.494	0.569	0.469	0.621
	1	1.0	0.065	0.045	0.524	0.437	0.369	0.315
		2.0	0.062	0.044	0.589	0.522	0.484	0.502
		2.5	0.065	0.056	0.515	0.551	0.440	0.555
100	0	1.0	0.056	0.040	0.845	0.789	0.853	0.852
		2.0	0.056	0.046	0.872	0.871	0.869	0.913
		2.5	0.058	0.054	0.809	0.869	0.764	0.907
	1	1.0	0.059	0.045	0.819	0.752	0.656	0.605
		2.0	0.060	0.051	0.869	0.842	0.805	0.833
		2.5	0.059	0.057	0.816	0.853	0.736	0.859

surprising result that the treated patients appeared to die faster than the untreated patients when the survival time was measured from cancer recurrence although therapy was effective in prolonging survival as measured from the time of randomization. It was not clear from the graphical display whether the result was statistically significant or not.

As mentioned in Lin et al. (1999), cancer recurrence normally took place within 5 years if at all and nearly 90% of the patients who suffered from cancer recurrence died within 3 years of recurrence. Thus, $U'_{\rm PF}$ or $U'_{\rm LR}$ with $\tau = 8$ and $s_0 = 5$ would be an appropriate test for the treatment difference with respect to the gap time between cancer recurrence and death. The normalized value of $U'_{\rm PF}$ with W_n given in (3) is 2.796, while that of $U'_{\rm LR}$ with $\nu_{n,2}$ given in (5) is 2.816. With either test, one would declare a significant treatment difference at the 0.01 level.

There are two plausible explanations for why the treated patients who get cancer tend to have shorter survival times following cancer than their untreated counterparts. One is the treatment toxicity. The other is the sieve phenomenon—the patients who get cancer even under treatment are the weakest of all the patients, whereas the control patients who get cancer may be relatively healthier, so the survival time distribution of cancer cases under treatment would correspond to the left tail of the survival time distribution of the control cancer cases.

5. Discussion

As demonstrated in the colon cancer example, the proposed nonparametric tests allow one to compare formally the gap time distributions between two treatment groups. Such comparisons provide valuable insights into the mechanisms of treatment actions. The results should be interpreted with caution and within the problem-specific knowledge. Since the subjects who have developed the first event are no longer comparable between the two treatment groups, it would be ill advised to judge the ultimate treatment benefit solely based on the gap time between the first and second events.

To avoid potential nonidentifiability, we recommend comparing the conditional distributions among the subjects with relatively short first event times, i.e., $T_1 \le s_0$. The choice of s_0 is somewhat arbitrary. In many applications, the conditional distributions are important in their own right and it may be of interest to consider various values of s_0. If the marginal distribution is of primary interest, then it is necessary to have a long follow-up period so that s_0 can be set to a sufficiently large value, say with $\Pr(T_1 \le s_0) > 99\%$. One exception is when T_1 and T_2 are only weakly correlated, in which case the conditional and marginal distributions are approximately the same.

As mentioned in Section 2, the proposed tests are directly applicable to the gap time between any two successive events. Furthermore, it is possible to assess simultaneously the treatment differences with respect to all the gap times of interest, along the lines of Wei, Lin, and Weissfeld (1989). The key to such simultaneous inference is to determine the joint distribution among the relevant test statistics. This is straightforward because the two-sample statistic for each gap time can be approximated by a sum of independent random variables, as evident from the appendices. Given these approximations, one can also construct appropriate tests for the simultaneous comparisons of multiple treatment groups.

It may be desirable to test the independence between two successive gap times, say T_1 and T_2, for each group. To this end, we consider the following class of statistics:

$$U_I = \int_0^{s_0} \int_0^{\tau - s_0} \left\{\hat{F}(t \mid s) - \hat{F}(t \mid s_0)\right\} \xi_n(dt, ds),$$

where ξ_n is a measure defined on $\{(t, s) : t + s \le \tau\}$. Under $\tilde{\rm H}_0$: $F(t \mid s) = F(t \mid s_0)$ for all $(s, t) \in [0, s_0] \times [\tau - s_0]$,

$$U_I = \int_0^{s_0} \int_0^{\tau - s_0} \left[\left\{\hat{F}(t \mid s) - F(t \mid s)\right\} - \left\{\hat{F}(t \mid s_0) - F(t \mid s_0)\right\}\right] \xi_n(dt, ds),$$

which is similar to the difference of two Pepe–Fleming-type statistics. Thus, it follows from the results of Section 2.2 and Appendix 1 that the asymptotic null distribution of $n^{1/2}U_I$ is zero-mean normal with consistent variance estimator

$$\int \left\{ \hat{\sigma}(t_1, t_2 \mid s_1, s_2) + \hat{\sigma}(t_1, t_2 \mid s_0, s_0) - 2\hat{\sigma}(t_1, t_2 \mid s_0, s_2) \right\} \times \xi_n(dt_1, ds_1)\xi_n(dt_2, ds_2),$$

where the integration is taken over $(s_1, s_2, t_1, t_2) \in [0, s_0] \times [0, s_0] \times [0, \tau - s_0] \times [0, \tau - s_0]$. Test statistics analogous to U_{LR}, U_{KS}, and U_{CV} can also be developed.

Acknowledgements

This work was supported by the U.S. National Institutes of Health, the National Science Foundation, and the National Security Agency. The authors are grateful to the reviewers for their helpful comments.

Résumé

L'objet de cet article est de traiter du problème de la comparaison entre deux populations, de la distribution du délai séparant deux événements successifs, lorsque les sujets peuvent voir apparaître une série d'événements distincts et que les délais de survenue de ces événements peuvent être censurés à droite. Plusieurs familles de tests non paramétriques sont développées, chacune autorise des distributions arbitraires et des structures de dépendance éventuelle entre les différents événements possibles. Les propriétés des tests proposés sont étudiées en situation asymptotique ou dans le cas de petits échantillons. Une illustration en est faite à partir des données d'une étude sur le cancer du colon. Le problème concomitant de l'étude de l'indépendance entre deux délais successifs est aussi étudié.

References

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

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

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80,** 557–572.

Lin, D. Y., Sun, W., and Ying, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* **86,** 59–70.

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.

Schumacher, M. (1984). Two-sample tests of Cramér-von Mises and Kolmogorov–Smirnov type for randomly censored data. *International Statistical Review* **52,** 263–281.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association* **94,** 146–153.

Wang, W.-J. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85,** 561–572.

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

Received April 1999. *Revised July* 2000.
Accepted September 2000.

Appendix 1

Asymptotic Properties of U_{PF}

The weak convergence of $\hat{F}^{(j)}(t \mid s)$ $(j = 1, 2)$ for fixed s was established by Lin et al. (1999). Here we extend their result by regarding $\hat{F}^{(j)}(t \mid s)$ as a two-dimensional process in t and s. Let $F_1^{(j)}(s) = \Pr(T_1^{(j)} \le s)$, $H^{(j)}(s, t) = \Pr(T_1^{(j)} \le s, T_2^{(j)} > t)$, and $H^{(j)}(t \mid s) = H^{(j)}(s, t)/H^{(j)}(s, 0)$. As shown in the appendix of Lin et al.,

$$n_j^{1/2}\left\{\hat{F}^{(j)}(t \mid s) - F^{(j)}(t \mid s)\right\} = \frac{n_j^{1/2}}{F_1^{(j)}(s)}\left[\left\{1 - F^{(j)}(t \mid s)\right\}\left\{\hat{H}^{(j)}(s, 0) - H^{(j)}(s, 0)\right\} - \left\{\hat{H}^{(j)}(s, t) - H^{(j)}(s, t)\right\}\right] + o_p(1). \tag{A.1}$$

Thus, by the weak convergence of $\hat{H}^{(j)}(s, t)$ given in the appendix of Lin et al., $n_j^{1/2}\{\hat{F}^{(j)}(t \mid s) - F^{(j)}(t \mid s)\}$ converges weakly to a zero-mean Gaussian process, the limiting covariance function at (t_1, s_1) and (t_2, s_2) being

$$\sigma^{(j)}(t_1, t_2 \mid s_1, s_2) = \frac{1}{F_1^{(j)}(s_1)F_1^{(j)}(s_2)} \times \left[\mathrm{E}\left\{D^{(j)}(s_1, t_1)D^{(j)}(s_2, t_2)\right\} - \int_0^\infty \frac{B^{(j)}(s_1, t_1; u)B^{(j)}(s_2, t_2; u)d\Lambda_c^{(j)}(u)}{\Pr\left(\tilde{Y}_2^{(j)} > u\right)}\right],$$

where

$$D^{(j)}(s, t) = \frac{\delta_1^{(j)} I\left(\tilde{Y}_1^{(j)} \le s\right) H^{(j)}(t \mid s)}{G^{(j)}\left(\tilde{Y}_1^{(j)}\right)} - \frac{I\left(\tilde{Y}_1^{(j)} \le s, \tilde{Y}_2^{(j)} - \tilde{Y}_1^{(j)} > t\right)}{G^{(j)}\left(\tilde{Y}_1^{(j)} + t\right)},$$

$$B^{(j)}(s, t; u) = H^{(j)}(t \mid s)\left\{F_1^{(j)}(s) - F_1^{(j)}(u)\right\}^+ - \left\{H^{(j)}(s, t) - H^{(j)}(u - t, t)\right\}^+,$$

and $\Lambda_c^{(j)}$ is the cumulative hazard function for $C^{(j)}$.

Under H_0: $F^{(2)}(t \mid s) = F^{(1)}(t \mid s)$ for all $\{(t, s) : t + s \le \tau\}$,

$$U_{\mathrm{PF}} = \int_0^\tau \int_0^{\tau - s} \left[\left\{\hat{F}^{(2)}(t \mid s) - F^{(2)}(t \mid s)\right\}\right.$$

$$-\left\{\hat{F}^{(1)}(t \mid s) - F^{(1)}(t \mid s)\right\}\Big] \mu_n(dt, ds).$$

Suppose that $n_j/n \to \rho_j > 0$ $(j = 1, 2)$ as $n \to \infty$. As mentioned in Section 2, μ_n converges in probability to μ. It then follows from the aforementioned weak convergence of $\hat{F}^{(j)}(t \mid s)$ $(j = 1, 2)$, together with a straightforward variance calculation, that $(n_1 n_2/n)^{1/2} U_{\mathrm{PF}}$ converges in distribution to a normal variable with mean zero and variance

$$V_{\mathrm{PF}} = \sum_{j=1}^{2} (1-\rho_j) \int \sigma^{(j)}(t_1, t_2 \mid s_1, s_2) \mu(dt_1, ds_1) \mu(dt_2, ds_2), \tag{A.2}$$

where the integration is over $\{(s_1, s_2, t_1, t_2) : s_1 + t_1 \leq \tau, s_2 + t_2 \leq \tau\}$. Replacing all the unknown quantities in (A.2) with their respective sample estimators, we obtain the variance estimator $\hat{V}_{\mathrm{PF}}$ given in (4). The consistency of $\hat{V}_{\mathrm{PF}}$ follows from the arguments given in the appendix of Lin et al. (1999).

Appendix 2

Asymptotic Properties of U_{LR}

It is easy to see that $\hat{\Lambda}^{(j)}(t \mid s)$ $(j = 1, 2)$ are right continuous in t. Thus, through the integration-by-parts formula (Fleming and Harrington, 1991, p. 320),

$$\int_0^{\tau - s} \nu_{n,2}(t \mid s) \hat{\Lambda}^{(j)}(dt \mid s) = \nu_{n,2}(t \mid s) \hat{\Lambda}^{(j)}(t \mid s)\Big|_0^{\tau - s} - \int_0^{\tau - s} \hat{\Lambda}^{(j)}(t \mid s) \nu_{n,2}(dt, s). \tag{A.3}$$

Because $\hat{\Lambda}^{(j)}(0 \mid s) = 0$ $(j = 1, 2)$ and $\nu_{n,2}(\tau - s \mid s) = 0$, (A.3) can be reexpressed as

$$\int_0^{\tau - s} \nu_{n,2}(t \mid s) \hat{\Lambda}^{(j)}(dt \mid s) = -\int_0^{\tau - s} \hat{\Lambda}^{(j)}(t \mid s) \nu_{n,2}(dt \mid s).$$

Therefore,

$$U_{\mathrm{LR}} = -\int_0^{\tau} \int_0^{\tau - s} \left\{\hat{\Lambda}^{(2)}(t \mid s) - \hat{\Lambda}^{(1)}(t \mid s)\right\} \times \nu_{n,2}(dt \mid s) \nu_{n,1}(ds). \tag{A.4}$$

By the mean value theorem,

$$\hat{\Lambda}^{(2)}(t \mid s) - \hat{\Lambda}^{(1)}(t \mid s) = -\frac{\hat{F}^{(2)}(t \mid s) - \hat{F}^{(1)}(t \mid s)}{1 - \tilde{F}(t \mid s)},$$

where $\tilde{F}(t \mid s)$ lies between $\hat{F}^{(2)}(t \mid s)$ and $\hat{F}^{(1)}(t \mid s)$. Thus, (A.4) can be written as

$$U_{\mathrm{LR}} = \int_0^{\tau} \int_0^{\tau - s} \left\{\hat{F}^{(2)}(t \mid s) - \hat{F}^{(1)}(t \mid s)\right\} \times \frac{\nu_{n,2}(dt \mid s) \nu_{n,1}(ds)}{1 - F(t \mid s)} + o_p(n^{-1/2}), \tag{A.5}$$

where $F(t \mid s)$ is the common value of $F^{(j)}(t \mid s)$ $(j = 1, 2)$ under H_0. Since (A.5) is in the form of the Pepe–Fleming-type statistic with

$$\mu_n(dt, ds) = \frac{\nu_{n,2}(dt \mid s) \nu_{n,1}(ds)}{1 - F(t \mid s)},$$

the desired convergence as well as the variance formula follows from Appendix 1.

Appendix 3

Asymptotic Properties of U_{KS} and U_{CV}

Define

$$\Psi_i^{(j)}(s, t) = \frac{I\left(\tilde{Y}_{1i}^{(j)} \leq s, \tilde{Y}_{2i}^{(j)} - \tilde{Y}_{1i}^{(j)} > t\right)}{G^{(j)}\left(\tilde{Y}_{1i}^{(j)} + t\right)} - H^{(j)}(s, t) + \int_0^{\infty} \frac{\left\{H^{(j)}(s, t) - H^{(j)}(u - t, t)\right\}^+}{\Pr\left(\tilde{Y}_2^{(j)} > u\right)} dM_i^{(j)}(u),$$

where

$$M_i^{(j)}(t) = I\left(C_i^{(j)} \leq t \wedge Y_{2i}^{(j)}\right) - \int_0^t I\left(\tilde{Y}_{2i}^{(j)} \geq u\right) d\Lambda_c^{(j)}(u).$$

In view of (A.3) of Lin et al. (1999) and (A.1) of Appendix 1,

$$\hat{F}^{(j)}(t \mid s) - F^{(j)}(t \mid s) = n_j^{-1} \sum_{i=1}^{n_j} Q_i^{(j)}(t \mid s) + o_p\left(n^{-1/2}\right),$$

where

$$Q_i^{(j)}(t \mid s) = \frac{\left[\left\{1 - F^{(j)}(t \mid s)\right\} \Psi_i^{(j)}(s, 0) - \Psi_i^{(j)}(s, t)\right]}{F_1^{(j)}(s)}.$$

Note that

$$\mathrm{E}\left\{Q_1^{(j)}(t_1 \mid s_1) Q_1^{(j)}(t_2 \mid s_2)\right\} = \sigma^{(j)}(t_1, t_2 \mid s_1, s_2).$$

Let $\hat{Q}_i^{(j)}$ be obtained from $Q_i^{(j)}$ by replacing all the unknown parameters with their respective sample estimators. By the arguments given in the appendix of Lin et al. (1999), $n_j^{-1} \times \Sigma_{i=1}^{n_j} \hat{Q}_i^{(j)}(t_1 \mid s_1) \hat{Q}_i^{(j)}(t_2 \mid s_2)$ converges almost surely to $\sigma^{(j)}(t_1, t_2 \mid s_1, s_2)$. It then follows from the arguments given in Appendix 1 of Lin, Wei, and Ying (1993) that $n^{1/2}\{\hat{F}^{(j)}(t \mid s) - F^{(j)}(t \mid s)\}$ has the same asymptotic distribution as $n^{1/2} R^{(j)}(t \mid s)$. Consequently, the null distributions of U_{KS} and U_{CV} are asymptotically the same as those of $\tilde{U}_{\mathrm{KS}}$ and $\tilde{U}_{\mathrm{CV}}$, respectively.