BIOMETRICS 52, 1440–1449
December 1996

Tests of Independence for Bivariate Survival Data

Joanna H. Shih

Office of Biostatistics Research, National Heart, Lung, and Blood Institute,
Bethesda, Maryland 20892, U.S.A.

and

Thomas A. Louis

Division of Biostatistics, University of Minnesota,
Minneapolis, Minnesota 55455, U.S.A.

SUMMARY

We propose two test statistics based on the covariance process of the martingale residuals for testing independence of bivariate survival data. The first test statistic takes the supremum over time of the absolute value of the covariance process, and the second test statistic is a time-weighted summary of the process. We derive asymptotic properties of the two test statistics under the null hypothesis of independence. In addition, we derive the asymptotic distribution of the weighted test and construct optimal weights for contiguous alternatives to independence. Through simulations, we compare the performance of the proposed tests and the inner product of the Savage scores statistics of Clayton and Cuzick (1985, *Journal of the Royal Statistical Society, Series A* **148,** 82–108). These demonstrate that the supremum test is generally more powerful with comparatively little power loss relative to their test when Clayton's family alternative holds, and the weighted test is more powerful when the weight is appropriately chosen.

1. Introduction

Testing independence for bivariate survival data is important in many biomedical studies. For example, leukemia patients after bone marrow transplantation are at risk of acute graft versus host disease (AGVHD) and cytomegalovirus (CMV). The times to these two diseases may be correlated. For patients with diabetic retinopathy in both eyes, the times to blindness of the two eyes may be associated due to natural pairing.

Several researchers have developed tests of independence. Oakes (1982) developed a concordance test for independence in the presence of censoring. Other work has focused on semiparametric approaches in order to eliminate the parametric modeling of the marginal distributions. Cuzick (1982) derived the locally most powerful test in the Bhuchongkul model (Bhuchongkul, 1964). Clayton and Cuzick (1985) derived the score test in the gamma frailty model (Clayton, 1978).

Prentice and Self (1985) derived the score test in a general class of proportional frailty models. Dabrowska (1986) presented statistics that generalize the Spearman rank correlation and the log-rank correlation. These semiparametric approaches all lead to a similar test statistic which is the inner product of Savage scores.

We present two alternative test statistics which are based on the covariance process of the martingale residuals for the marginal distributions (Fleming and Harrington, 1991, p. 163). The first approach takes the supremum over time of the absolute value of the covariance process over time, and is designed to detect high association which may occur before the end of the study. The second approach is a time-weighted summary intended to weight early and late association differently. The choice of weight depends on the association structure of the failure times and analytic goals. In Section 2 we derive asymptotic properties of the two new test statistics under

Key words: Bivariate failure times; Covariance; Cross ratio; Martingale residuals.

the null hypothesis of independence. In Section 3 we derive the asymptotic distribution of the weighted test and construct optimal weights for a sequence of contiguous alternatives. In Section 4 we evaluate the performance of the proposed tests by simulations. In Section 5 we demonstrate the proposed methodology with a well-known matched failure time data and the Stanford heart transplant data.

2. Proposed Test Statistics

2.1 *Definition*

For $i = 1, \cdots, n$, let (T_{i1}, T_{i2}) denote the paired failure times and (C_{i1}, C_{i2}) the paired censoring times. Assume that (T_{i1}, T_{i2}) are random samples with continuous bivariate survival functions S and marginal survival functions S_1, S_2. Also assume that (C_{i1}, C_{i2}) are random samples with bivariate survival functions G and marginal survival functions G_1, G_2, and that (C_{i1}, C_{i2}) are independent of (T_{i1}, T_{i2}). If the two failure times are subject to independent censorship by a single censoring variable, namely $C_{i1} \equiv C_{i2}$, then $G_1 \equiv G_2$ and $G(t_1, t_2) = G_1(t_1 \vee t_2)$, where $a \vee b = \max(a, b)$. For $i = 1, \ldots, n$, $j = 1, 2$, we observe $X_{ij} = T_{ij} \wedge C_{ij}$, $\delta_{ij} = I[X_{ij} = T_{ij}]$. Let $N_{ij}(t) = I[X_{ij} \leq t, \delta_{ij} = 1]$ denote the counting process which indicates whether member j in pair i has experienced an event over time t. We assume that the intensity function for $N_{ij}(t)$ is given by $Y_{ij}(t)\, d\Lambda_j(t)$, where $Y_{ij}(t) = I[X_{ij} \geq t]$, indicating whether member j in pair i is at risk at time t, and Λ_j is the cumulative hazard function. Define the marginal martingale by $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)\, d\Lambda_j(u)$. For fixed t, $M_{ij}(t)$ is a square integrable martingale with respect to the filtration $\mathcal{F}_{ij}(t), t \geq 0$, which is the sub–$\sigma$-field generated by $\{N_{ij}(u), I[X_{ij} \leq u, \delta_{ij} = 0] : 0 \leq u \leq t\}$. Then the martingale residual is defined to be

$$\hat{M}_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)\, d\hat{\Lambda}_j(u), \qquad i = 1, \ldots, n,\ j = 1, 2.$$

The Nelson estimator of the cumulative hazard $\hat{\Lambda}_j$ is given by

$$\hat{\Lambda}_j(t) = \int_0^t \frac{d\bar{N}_j(u)}{\bar{Y}_j(u)},$$

where $\bar{Y}_j(t) = \Sigma_i Y_{ij}(t), \bar{N}_j(t) = \Sigma_i N_{ij}(t)$.

The martingale residual can be interpreted as the difference between the observed and predicted number of events over time t. Note that for each j and $t \geq 0$, $\mathrm{E}(M_{ij}(t)) = \Sigma_i \hat{M}_{ij}(t) = 0$.

The inner product of the Savage scores statistic proposed by Clayton and Cuzick (1985) and others can be expressed by

$$\begin{aligned} T_n &= \sum_i \{\delta_{i1} - \hat{\Lambda}_1(X_{i1})\}\{\delta_{i2} - \hat{\Lambda}_2(X_{i2})\} \\ &= \sum_i \int_0^{t_0} \int_0^{t_0} d\hat{M}_{i1}(u_1)\, d\hat{M}_{i2}(u_2), \end{aligned} \tag{1}$$

where t_0 is the length of the study period. The test statistic T_n measures the sample covariance of the martingale residuals at the last follow-up points for each pair. A large absolute value of T_n provides evidence that the two failure times are correlated.

The first new test statistic is given by

$$U_n = \sup_{0 \leq t \leq t_0} \left| \sum_i \int_0^t \int_0^t d\hat{M}_{i1}(u_1)\, d\hat{M}_{i2}(u_2) \right|. \tag{2}$$

At each time point t, the conditional covariance of the marginal martingale residuals is calculated given follow-up information over $[0, t]$, and U_n is the maximal absolute value of the covariance. The time-dependent covariance measures the association of the failure times for studies with two types of events or unit-specific pairing from each subject. In such studies, the follow-up time is the same for each pair, and the correlation of the marginal martingale residuals describes how the association of the paired events changes as the follow-up time increases.

The second new test statistic is given by

$$V_n = \sum_i \int_0^{t_0} \int_0^{t_0} W_n(u_1, u_2)\, d\hat{M}_{i1}(u_1)\, d\hat{M}_{i2}(u_2). \tag{3}$$

The weight $W_n(t_1, t_2)$ is assumed to converge to $w(t_1, t_2)$ uniformly over $[0, t_0] \times [0, t_0]$ and to be a bounded predictable process with respect to filtration $\mathcal{F}_{\boldsymbol{t}}$ for bivariate time $\boldsymbol{t} = (t_1, t_2)$, where $\mathcal{F}_{\boldsymbol{t}}$ has to be increasing with respect to natural partial order such that $s \leq \boldsymbol{t}$ if and only if $s_i \leq t_i, i = 1, 2$ (Andersen et al., 1992). Thus, $\mathrm{E}(W_n(t_1, t_2) \mid \mathcal{F}_{\boldsymbol{t}^- = (t_1^-, t_2^-)}) = W_n(t_1, t_2)$. The test statistic V_n can be written in terms of summations as

$$V_n = \sum_i \left[\delta_{i1}\delta_{i2}W_{ii} - \delta_{i1}\left\{ \sum_k I_{ki}^2 W_{ik} \frac{d\bar{N}_2(X_{k2})}{\bar{Y}_2(X_{k2})} \right\} - \delta_{i2}\left\{ \sum_k I_{ki}^1 W_{ki} \frac{d\bar{N}_1(X_{k1})}{\bar{Y}_1(X_{k1})} \right\} + \sum_k \sum_l I_{ki}^1 I_{li}^2 W_{kl} \frac{d\bar{N}_1(X_{k1})}{\bar{Y}_1(X_{k1})} \frac{d\bar{N}_2(X_{l2})}{\bar{Y}_2(X_{l2})} \right], \tag{4}$$

where $W_{ki} = W_n(X_{k1}, X_{i2})$ and $I_{ki}^j = I[X_{kj} \leq X_{ij}]$. Expression (4) is particularly useful for computing V_n. The test statistic V_n is analogous to the weighted logrank test statistic for testing equality of two marginal distributions. It equals T_n when $W_n(t_1, t_2) \equiv 1$. We may want to weight the data according to values of the marginal survival functions by setting $W_n(t_1, t_2) = g(\hat{S}_1(t_1^-), \hat{S}_2(t_2^-))$, where $\hat{S}_1, \hat{S}_2$ are the Kaplan–Meier estimates.

2.2 *Asymptotic Properties of U_n and V_n*

In this section we derive asymptotic properties of U_n and V_n under the null hypothesis of independence.

We first show that the difference between $n^{-1/2} \Sigma_i \hat{M}_{i1}(t)\hat{M}_{i2}(t)$ and $n^{-1/2} \Sigma_i M_{i1}(t)M_{i2}(t)$ converges to 0 in probability uniformly in $[0, t_0]$.

THEOREM 1. *For $0 \leq t \leq t_0$, the difference between $n^{-1/2} \Sigma_i \hat{M}_{i1}(t)\hat{M}_{i2}(t)$ and $n^{-1/2} \Sigma_i M_{i1}(t)M_{i2}(t)$ converges to 0 in probability uniformly in $[0, t_0]$.*

We give the proof in Appendix A.

We next show that $M_{i1}(t)M_{i2}(t)$ has mean zero and uncorrelated increments under independence, and $n^{-1/2} \Sigma M_{i1}(t)M_{i2}(t)$ converges to a Gaussian process on $[0, t_0]$.

THEOREM 2. *Under the null hypothesis of independence, $M_{i1}(t)M_{i2}(t)$ has mean zero and uncorrelated increments, and $n^{-1/2} \Sigma M_{i1}(t)M_{i2}(t)$ converges to a Gaussian process on $[0, t_0]$.*

We give the proof in Appendix B.

The variance of $n^{-1/2} \Sigma M_{i1}(t)M_{i2}(t)$, by the uncorrelated increments structure of $M_{i1}(t)M_{i2}(t)$, can be defined and estimated by

$$\sigma^2(t) = \int_0^t \int_0^t \mathrm{E}\{Y_1(u_1, u_2)\}\lambda_1(u_1)\lambda_2(u_2)\, du_1\, du_2$$

$$\hat{\sigma}^2(t) = \frac{1}{n} \sum_i \sum_k \sum_l I_{ki}^1(t) I_{li}^2(t) \frac{d\bar{N}_1(X_{k1})}{\bar{Y}_1(X_{k1})} \frac{d\bar{N}_2(X_{l2})}{\bar{Y}_2(X_{l2})},$$

where $Y_i(t_1, t_2) = I[X_{i1} \geq t_1, X_{i2} \geq t_2]$ and $I_{ki}^j(t) = I[X_{kj} \leq X_{ij} \wedge t]$. Thus, $n^{-1/2}U_n$ converges to $\sigma(t_0) \sup_{u \in A} |Z(u)|$, where $Z(.)$ is Brownian motion and $A = \{\sigma(t)/\sigma(t_0) : 0 \leq t \leq t_0\}$. For Brownian motion $Z(u)$, $p(\sup_{u \in [0,1]} Z(u) > y) = 2(1 - \Phi(y))$, where Φ is the standard normal distribution function. When y is large, $p(\sup_{u \in [0,1]} |Z(u)| > y)$ is almost identical to $2p(\sup_{u \in [0,1]} Z(u) > y) = 4p(Z(1) > y)$. Thus, as an approximation, we reject the null hypothesis when $U_n/\{n\hat{\sigma}^2(t_0)\}^{1/2}$ exceeds $\Phi^{-1}(1 - \alpha/4)$ at α significance level.

We next derive asymptotic properties of V_n. Since $W_n(t_1, t_2)$ is a bounded predictable process and converges to $w(t_1, t_2)$ uniformly on $[0, t_0] \times [0, t_0]$, using the arguments similar to those given in Appendix A, we can show that $n^{-1/2}V_n$ is asymptotically equivalent to $n^{-1/2} \Sigma \int\int w(u_1, u_2)\, dM_{i1}(u_1)\, dM_{i2}(u_2)$, which is a normalized sum of n independent and identically distributed random variables. By the central limit theorem, under independence, $n^{-1/2} \Sigma \int\int w(u_1, u_2)\, dM_{i1}(u_1)\, dM_{i2}(u_2)$ converges to normal with mean zero. Its variance can be defined and estimated by

$$\sigma^2 = \int_0^{t_0} \int_0^{t_0} \mathrm{E}\{Y_1(u_1, u_2)\} w^2(u_1, u_2) \lambda_1(u_1)\lambda_2(u_2)\, du_1\, du_2$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i \sum_k \sum_l I_{ki}^1 I_{li}^2 W_{kl}{}^2 \frac{d\bar{N}_1(X_{k1})}{\bar{Y}_1(X_{k1})} \frac{d\bar{N}_2(X_{l2})}{\bar{Y}_2(X_{l2})}.$$

For a two-sided test of size α, we reject the null hypothesis when $|V_n|/\{n\hat{\sigma}^2\}^{1/2}$ exceeds $\Phi^{-1}(1-\alpha/2)$.

3. Constructing Optimal Weights

The optimal choice of the weight function $w(t_1, t_2)$ depends on the underlying dependency structure of the bivariate failure times. Following Anderson, Louis, and Holm (1992), we consider a general bivariate survival function defined by

$$S(t_1, t_2) = e^{-\{\Lambda_1(t_1)+\Lambda_2(t_2)+A(t_1,t_2)\}}.$$

Here $A(t_1, t_2)$ defines the association structure. If T_1, T_2 are independent, then $A(t_1, t_2) = 0$. We consider contiguous alternatives H_{an}: $A(t_1, t_2) = O(n^{-1/2})$.

Oakes (1989) describes the association by the cross ratio function, defined by

$$\theta(t_1, t_2) = S(t_1, t_2) \frac{\partial^2 S(t_1, t_2)/\partial t_1 \partial t_2}{\partial S(t_1, t_2)/\partial t_1 \partial S(t_1, t_2)/\partial t_2}.$$

The cross ratio $\theta(t_1, t_2)$ is related to $A(t_1, t_2)$, and is equal to 1 when T_1, T_2 are independent, greater than 1 when T_1, T_2 are positively associated, and less than 1 when T_1, T_2 are negatively associated. Then H_{an} can be expressed by $\theta(t_1, t_2) = 1 + n^{-1/2} g(t_1, t_2)$, where g is a known function.

THEOREM 3. *Under the contiguous alternatives* H_{an}: $\theta(t_1, t_2) = 1 + n^{-1/2} g(t_1, t_2)$, $V_n/\{n\hat{\sigma}^2\}^{1/2}$ *converges to a normal distribution with unit variance and mean*

$$\frac{n^{1/2} \int\int w(t_1, t_2)(\theta(t_1, t_2) - 1) S(t_1, t_2) G(t_1, t_2) \lambda_1(t_1) \lambda_2(t_2)\, dt_1\, dt_2}{\int\int w^2(t_1, t_2) S(t_1, t_2) G(t_1, t_2) \lambda_1(t_1) \lambda_2(t_2)\, dt_1\, dt_2}. \tag{5}$$

We give the proof in Appendix C.

By Cauchy–Schwarz inequality, (5) is maximized when $w(t_1, t_2)$ is proportional to $\theta(t_1, t_2) - 1$.

4. Simulations

We carried out a Monte-Carlo study to compare the performance of T_n, U_n, V_n. For V_n we chose $W_n(t_1, t_2)$ to be $\hat{S}_1(t_1)\hat{S}_2(t_2)$, which is the empirical version of the optimal weight in Frank's family. Thus, V_n gives higher weight to early failures. We describe below three families of bivariate distributions from which the simulated data were generated.

Clayton's family. The bivariate survival function in Clayton's family (Clayton, 1978) takes the form

$$\{S_1(t_1)^{1-\gamma} + S_2(t_2)^{1-\gamma} - 1\}^{1/(1-\gamma)}, \qquad \gamma > 1. \tag{6}$$

Clayton's family models positive association and $\theta(t_1, t_2) = \gamma$. Since $\theta(t_1, t_2)$ is a constant, T_n is optimal for Clayton's family. We generated paired failure times with $\gamma = 1.22$ corresponding to Kendall tau $= 0.1$. We allow the two margins to be different and let the marginal hazards be $\lambda_1(t) = 1, \lambda_2(t) = 2$.

Frank's family. The bivariate survival function in Frank's family (Genest, 1987) has the representation

$$S(t_1, t_2) = \log_\gamma \left\{ 1 + \frac{(\gamma^{S_1(t_1)} - 1)(\gamma^{S_2(t_2)} - 1)}{\gamma - 1} \right\}, \tag{7}$$

where $\gamma > 0$, and $\log_\gamma$ denotes logarithm to the base γ. T_1, T_2 are positively associated when $\gamma < 1$, negatively associated when $\gamma > 1$, and independent when γ approaches 1. Here $\theta(t_1, t_2) = -s \log \gamma \{1 + \gamma^s/(1-\gamma^s)\}$, where $s = S(t_1, t_2)$. It is monotonically increasing with s and approximately linear (Shih, 1992). The local optimal weight for Frank's family alternative is $S_1(t_1)S_2(t_2)$. We chose $\gamma = 0.39$ with Kendall's tau $= 0.1$ approximately. The two margins are the same as used in Clayton's family. We used the algorithm of Genest (1987) to generate random pairs.

Mixture in Frank's family. We consider a mixture of two bivariate distributions in Frank's family. Kendall's tau is 0.1 in one distribution and -0.1 in the other distribution. Although association of opposite direction may not occur commonly in application, we use this scenario to illustrate the gain in power when examining the maximal association over time compared to examining the association at the end of the study. For the distribution of positive association, we let the marginal hazards be $\lambda_i(t) = 5, i = 1, 2$. For the distribution of negative association, let the marginal hazards be $\lambda_i(t) = 1$. Each pair has 50% probability of being drawn from either of the two distributions. Since the marginal hazards in the bivariate distribution with positive association are larger than those in the other, high positive association should occur early. However, when t approaches ∞, the negative association cancels the positive association. Thus, U_n is expected to perform better than T_n in detecting the early association.

We consider both no censoring and 30% censoring. To achieve 30% censoring, we let the two censoring variables be identically and independently distributed as uniform over (0, 2.3) for the first two families and (0, 1.7) for the mixture. We generated 5000 random samples of uncorrelated pairs to study the size of the test, and 1000 random samples from the three bivariate distributions above to study the power. In each sample, we estimated the marginal cumulative hazards by the Nelson estimator.

Table 1 presents the simulation results for the size of the test and empirical power for Clayton's family. The empirical Type I error of T_n and V_n is close to the nominal level, whereas U_n is slightly conservative. The empirical power of U_n is very close to the optimal level achieved by T_n. As expected, V_n is not as powerful as the other two tests for this scenario because the weight is not optimal. Table 2 presents the simulation results for Frank's family alternative. Because the weight chosen is optimal for Frank's family, V_n is superior to the other two tests. One referee points out that the test U_n generally performs better than T_n because the former also gives more weight to early failures. It is of interest that T_n appears to be less affected by censoring than the other two tests. Table 3 presents the simulation results for the mixture. In this scenario, U_n also performs better than T_n. Note that, under the 30% censoring scheme, a large portion of the pairs with negative association are censored. Therefore, the powers of T_n and U_n with censoring are higher

Table 1
Empirical type I *error, in per cent, of* 5000 *random samples and empirical power* (%) *of* 1000 *random samples from Clayton's family*

Significance (%)	n	Test	No censoring $\tau = 0$	No censoring $\tau = 0.1$	30% Censoring $\tau = 0$	30% Censoring $\tau = 0.1$
5	50	T_n	5.3	25.6	4.9	13.4
5	50	U_n	4.2	25.3	3.7	9.9
5	50	V_n	5.2	17.3	5.5	11.1
5	100	T_n	5.1	44.8	4.9	22.4
5	100	U_n	4.4	43.9	3.9	18.3
5	100	V_n	5.0	29.0	5.0	19.9
5	200	T_n	4.8	72.3	4.9	42.9
5	200	U_n	4.5	71.9	4.4	40.3
5	200	V_n	5.0	54.5	5.1	36.2
1	50	T_n	1.0	11.2	1.0	3.2
1	50	U_n	0.6	8.7	0.7	2.0
1	50	V_n	1.0	5.6	1.3	2.4
1	100	T_n	0.9	24.2	0.9	8.1
1	100	U_n	0.9	23.6	0.6	6.6
1	100	V_n	1.0	11.1	1.0	5.7
1	200	T_n	0.9	50.4	1.1	21.5
1	200	U_n	0.8	50.2	1.1	19.9
1	200	V_n	0.8	29.8	0.8	15.9

Table 2
Empirical power, in percent, of 1000 *random samples from Frank's family*

Significance (%)	n	Test	$\tau = 0.1$ No censoring	$\tau = 0.1$ 30% Censoring
5	50	T_n	11.6	11.5
5	50	U_n	11.9	9.5
5	50	V_n	18.8	16.2
5	100	T_n	18.1	19.2
5	100	U_n	19.4	18.1
5	100	V_n	34.4	25.7
5	200	T_n	33.2	36.7
5	200	U_n	40.1	37.0
5	200	V_n	57.0	46.9
1	50	T_n	3.2	2.6
1	50	U_n	3.0	2.4
1	50	V_n	5.1	4.6
1	100	T_n	6.3	5.3
1	100	U_n	6.6	4.4
1	100	V_n	13.7	9.5
1	200	T_n	14.5	14.6
1	200	U_n	17.6	15.1
1	200	V_n	32.6	23.3

than without censoring. It is not a surprise that V_n performs better than T_n because V_n gives heavy weight to early failures.

5. Examples

We first illustrate the proposed methods with the well-known skin graft data (Holt and Prentice, 1974). The days of survival of closely-matched and poorly-matched skin grafts on the same person

Table 3
Empirical power, in percent, of 1000 *random samples from a mixture of distributions in Frank's family*

Significance (%)	n	Test	$\tau = 0.1$ No censoring	$\tau = 0.1$ 30% Censoring
5	50	T_n	27.3	40.9
5	50	U_n	36.7	42.4
5	50	V_n	61.4	56.0
5	100	T_n	53.1	72.9
5	100	U_n	72.0	78.4
5	100	V_n	89.0	84.5
5	200	T_n	85.5	95.4
5	200	U_n	97.5	98.4
5	200	V_n	99.9	98.9
1	50	T_n	8.6	16.1
1	50	U_n	10.5	15.9
1	50	V_n	33.1	29.3
1	100	T_n	25.9	43.8
1	100	U_n	37.2	49.9
1	100	V_n	73.1	65.5
1	200	T_n	60.1	85.0
1	200	U_n	84.0	90.3
1	200	V_n	98.2	94.8

Table 4
Days of survival of closely and poorly matched skin grafts on the same person

Subject	4	5	7	8	9	10	11	12	13	15	16
Survival of close match	37	19	57*	93	16	22	20	18	63	29	60*
Survival of poor match	29	13	15	26	11	17	26	21	43	15	40

* Censored observation.

are listed in Table 4. Although the sample size is too small to draw conclusions based on asymptotic approximations, the data do demonstrate the advantage of the new tests. Here the standardized T_n equals 1.21 and the standardized U_n equals 1.64. The U_n test indicates mild evidence against independence. We consider two weight functions: $\hat{S}_1(t_1)\hat{S}_2(t_2)$ and $1/[\hat{\Lambda}_1(t_1)\hat{\Lambda}_2(t_2)]$. The first weight function ranges from 1 to 0 and the second from ∞ to 0. Thus, both weight functions give higher weights to early failures; however, the latter weight function gives higher weight to very early failures than the former one. With the weight fucntion $\hat{S}_1(t_1)\hat{S}_2(t_2)$, the standardized V_n equals 1.74, while with the other weight function, the standardized V_n equals 2.65. The result suggests that very early failures are more associated with pairing than late failures are.

The second example concerns the waiting time for a donor and posttransplant survival time for transplanted patients in the Stanford Heart Transplant Program. The data taken from Kalbfleisch and Prentice (1980) consist of 69 patients. The standardized T_n equals 1.09 and the standardized U_n equals 1.20. Both tests indicate little evidence against independence. With the weight function $\hat{S}_1(t_1)\hat{S}_2(t_2)$, the standardized V_n equals 1.97, giving strong evidence against the hypothesis of independence. However, with the weight function $1/[\hat{\Lambda}_1(t_1)\hat{\Lambda}_2(t_2)]$, the standardized V_n equals 1.28, providing little evidence against independence. The discrepancy of the conclusions drawn using the two weight functions suggests that early failures have different degrees of associaton, but sorting them out is difficult (Hougaard, 1995).

6. Conclusion

In this article we have proposed two tests of independence for bivariate survival data. The test U_n is generally more powerful than the conventional test T_n with comparatively little power loss relative to T_n under Clayton's family alternative, where late failures are most correlated. In contrast with U_n, the weighted test V_n is optimal upon the correct specification of the weight, but may lose power when the weight is not specified properly. Furthermore, the weight provides a useful tool in describing how the association of the two failure times changes over time. For example, since the optimal weight in Frank's family is $S_1(t_1)S_2(t_2)$, compared to Clayton's family, it indicates that high association information occurs earlier in Frank's family.

In this work, although we evaluate the supremum of the covariance process only along the diagonal of $t_1 = t_2$, the properties of U_n demonstrated in the simulations are promising. It is not clear that we could gain power by examining the covariance process over the whole plane. Complexity arises when deriving asymptotic properties of the covariance process on the plane. For instance, because multivariate time mostly is only partially ordered and not totally ordered, the covariance of the correlation process at two different pairs of time points may be identical, and, hence, we cannot base inference on Brownian motion. Consequently, we have to use Monte-Carlo methods such as the Bootstrap to draw inference. As the associate editor pointed out, we might indeed lose power because taking a supremum of a lager set does not give a more general function and might induce extra error. Also, there will be added computational complexity in the supremum analysis over the whole plane.

Acknowledgements

Partial support of this work is provided by a contract to the University of Minnesota from the National Institute of Allergy and Infectious Diseases, U.S.A. Part of the work was done while the first author was at the Division of Biostatistics, University of Minnesota. The authors thank the associate editor for comments.

Résumé

On propose deux statistiques pour tester l'indépendance de données de survie bivarié. Ces tests sont basés sur le processus de covariance des résidus martingales. La première statistique de test consiste à prendre le supremum de la valeur absolue du processus de covariance sur toutes les valeurs du temps. La deuxième statistique est un résumé de ce processus pondéré par le temps. Sous l'hypothèse d'indépendance, les propriétés asymptotiques des deux statistiques sont développées. De plus, on développe la distribution asymptotique du test pondéré et on construit des poids optimaux pour des alternatives contigus à l'indépendance. La performance des tests proposés ainsi que le produit interne des statistiques du score de Savage de Clayton et Cuzick (1985) sont comparés par simulations. Ces simulations montrent que le test du supremum est généralement plus puissant sans grande perte de puissance relative à leur test pour une famille d'alternatives de Clayton. Quand les poids sont choisis d'une manière appropriée, le test pondéré est plus puissant.

References

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

Anderson, J. E., Louis, T. A., and Holm, N. V. (1992). Time dependent association for bivariate survival distributions. *Journal of the American Statistical Association* **87,** 641–650.

Bhuchongkul, S. (1964). A class of nonparametric tests for independence in bivariate populations. *Annals of Mathematical Statistics* **35,** 138–149.

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

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65,** 141–151.

Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A* **148,** 82–108.

Cuzick, J. (1982). Rank tests for association with right censored data. *Biometrika* **69,** 351–364.

Dabrowska, D. M. (1986). Rank tests for independence for bivariate censored data. *The Annals of Statistics* **14,** 250–264.

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

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74,** 549–555.

Holt, J. D. and Prentice, R. L. (1974). Survival analysis in twin studies and matched pair experiments. *Biometrika* **61,** 17–30.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis* **1,** 255–273.

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

Oakes, D. (1982). A concordance test for independence in the presence of censoring. *Biometrics* **38,** 451–455.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84,** 487–493.

Prentice, R. L. and Self, S. (1985). Comments on "Multivariate generalizations of the proportional hazards model." *Journal of the Royal Statistical Society, Series A* **148,** 82–108.

Shih, J. H. (1992). Models and Analysis for Multivariate Failure Time Data. Ph.D. thesis, Division of Biostatistics, University of Minnesota, Minneapolis.

Received February 1995; *revised October* 1995 *and April* 1996; *accepted May* 1996.

Appendix A

Proof of Theorem 1

We first write $n^{-1/2}\,\Sigma\,\hat{M}_{i1}(t)\hat{M}_{i2}(t)$ as

$$n^{-1/2}\sum_i \hat{M}_{i1}(t)\hat{M}_{i2}(t) = n^{-1/2}\sum_i \Bigg\{ M_{i1}(t)M_{i2}(t) - M_{i1}(t)\int_0^t Y_{i2}(u)\,d(\hat{\Lambda}_2-\Lambda_2)(u)$$

$$- M_{i2}(t)\int_0^t Y_{i1}(u)\,d(\hat{\Lambda}_1-\Lambda_1)(u)$$

$$+ \int_0^t Y_{i1}(u_1)\,d(\hat{\Lambda}_1-\Lambda_1)(u_1)$$

$$\times \int_0^t Y_{i2}(u_1)\, d(\hat{\Lambda}_2 - \Lambda_2)(u_2) \Big\} . \tag{8}$$

We show that the next to last terms in (8) converge to zero in probability uniformly in $[0, t_0]$. For the second term,

$$\begin{aligned}
&\sup_{0 \le t \le t_0} \left| n^{-1/2} \sum_i M_{i1}(t) \int_0^t Y_{i2}(u)\, d(\hat{\Lambda}_2 - \Lambda_2)(u) \right| \\
&\le \sup_{0 \le (t_1, t_2, t) \le t_0} \left| n^{-1/2} \sum_i M_{i1}(t_1) \int_0^t Y_{i2}(t_2)\, d(\hat{\Lambda}_2 - \Lambda_2)(u) \right| \\
&= \sup_{0 \le (t_1, t_2, t) \le t_0} \left| n^{-1/2} \sum_i M_{i1}(t_1) Y_{i2}(t_2) \{\hat{\Lambda}_2(t) - \Lambda_2(t)\} \right| \\
&\le \sup_{0 \le (t_1, t_2) \le t_0} \left| n^{-1/2} \sum_i M_{i1}(t_1) Y_{i2}(t_2) \right| \sup_{0 \le t \le t_0} \left| \hat{\Lambda}_2(t) - \Lambda_2(t) \right| .
\end{aligned} \tag{9}$$

The first term in (9) converges to the supremum of an absolute zero-mean Gaussian process because the mean of $M_{i1}(t_1)Y_{i2}(t_2)$ equals 0, and the second term converges to 0 in probability uniformly in $[0, t_0]$ (Fleming and Harrington, 1991, p. 115). Thus, the second term in (8) converges to 0 in probability uniformly in $[0, t_0]$. In a similar fashion, the uniform convergence to 0 of the third term in (8) can be established.

Finally,

$$\begin{aligned}
&\sup_{0 \le t \le t_0} \left| n^{-1/2} \sum_i \int_0^t Y_{i1}(u_1)\, d(\hat{\Lambda}_1 - \Lambda_1)(u_1) \int_0^t Y_{i2}(u_1)\, d(\hat{\Lambda}_2 - \Lambda_2)(u_2) \right| \\
&\le \sup_{0 \le (t_1, t_2) \le t_0} \left| n^{-1} \sum_i Y_{i2}(t_1) Y_{i2}(t_2) \right| \sup_{0 \le t \le t_0} \left| n^{1/2}(\hat{\Lambda}_1(t) - \Lambda_1(t)) \right| \sup_{0 \le t \le t_0} \left| \hat{\Lambda}_2(t) - \Lambda_2(t) \right| .
\end{aligned}$$

The first term is bounded by 1. The second term converges to the supremum of an absolute zero-mean Gaussian process (Fleming and Harrington, 1991, p. 231). The third term converges to 0 in probability uniformly in $[0, t_0]$. Hence, the last term in (8) converges to 0 in probability uniformly in $[0, t_0]$.

Appendix B

Proof of Theorem 2

Since

$$\mathrm{E}[M_{i1}(t) M_{i2}(t)] = \int_0^t \int_0^t \mathrm{EE}[dM_{i1}(u_1)\, dM_{i2}(u_2) \mid \mathcal{F}_{\boldsymbol{u}-}],$$

and for $s < t$,

$$\mathrm{E}[M_{i1}(t)M_{i2}(t) - M_{i1}(s)M_{i2}(s) \mid \mathcal{F}_{\boldsymbol{s}=(s,s)}] = \mathrm{E}[\{M_{i1}(t) - M_{i1}(s)\}\{M_{i2}(t) - M_{i2}(s)\} \mid \mathcal{F}_{\boldsymbol{s}=(s,s)}],$$

it suffices to show $\mathrm{E}(dM_{i1}(t_1)\, dM_{i2}(t_2) \mid \mathcal{F}_{\boldsymbol{t}-}) = 0$.

Let the joint, conditional, and marginal hazard functions be defined by

$$\begin{aligned}
\lambda(t_1, t_2)\, dt_1\, dt_2 &= p(t_1 \le T_1 < t_1 + dt_1, t_2 \le T_2 < t_2 + dt_2 \mid T_1 \ge t_1, T_2 \ge t_2) \\
\lambda(t_1 \mid T_2 \ge t_2)\, dt_1 &= p(t_1 \le T_1 < t_1 + dt_1 \mid T_1 \ge t_1, T_2 \ge t_2) \\
\lambda(t_2 \mid T_1 \ge t_1)\, dt_2 &= p(t_2 \le T_2 < t_2 + dt_2 \mid T_1 \ge t_1, T_2 \ge t_2) \\
\lambda_j(t)\, dt &= d\Lambda_j(t), \qquad j = 1, 2.
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{E}(dM_{i1}(t_1)\, dM_{i2}(t_2) \mid \mathcal{F}_{\boldsymbol{t}-}) = Y_i(t_1, t_2)\{&\lambda(t_1, t_2) - \lambda(t_1 \mid T_2 \ge t_2)\lambda_2(t_2) \\
&- \lambda(t_2 \mid T_1 \ge t_1)\lambda_1(t_1) + \lambda_1(t_1)\lambda_2(t_2)\}\, dt_1\, dt_2,
\end{aligned} \tag{10}$$

where $Y_i(t_1, t_2) = I[X_{i1} \geq t_1, X_{i2} \geq t_2]$. The conditional expectation in (10) equals 0 because under independence $\lambda(t_1, t_2) = \lambda_1(t_1)\lambda_2(t_2)$, $\lambda(t_1 \mid T_2 \geq t_2) = \lambda_1(t_1)$, and $\lambda(t_2 \mid T_1 \geq t_1) = \lambda_2(t_2)$. It follows that $M_{i1}(t_1)M_{i2}(t_2)$ has uncorrelated increments and its expectation equals 0. By the central limit theorem, the asymptotic finite dimensional distributions of $n^{-1/2}\,\Sigma\, M_{i1}(t)M_{i2}(t)$ are normal. Since $n^{-1/2}\,\Sigma\, M_{i1}(t)M_{i2}(t)$ is tight on $D[0, t_0]$ (Billingsley, 1968), it follows that $n^{-1/2}\,\Sigma\, M_{i1}(t)M_{i2}(t)$ converges to a Gaussian process on $[0, t_0]$.

Appendix C

Proof of Theorem 3

Note that

$$
\begin{aligned}
\mathrm{E}[w(t_1, t_2)\, dM_{11}(t_1)\, dM_{21}(t_2) \mid \mathcal{F}_{t-}] = Y_1(t_1, t_2)w(t_1, t_2)\{&\lambda(t_1, t_2) - \lambda(t_1 \mid T_2 \geq t_2)\lambda_2(t_2) \\
&- \lambda(t_2 \mid T_1 \geq t_1)\lambda_1(t_1) \\
&+ \lambda_1(t_1)\lambda_2(t_2)\}\, dt_1\, dt_2. \qquad (11)
\end{aligned}
$$

Expressing (11) in terms of θ and Taylor expanding around $\theta = 1$, we can show that (11) is asymptotically equivalent to

$$
Y_1(t_1, t_2)w(t_1, t_2)\lambda_1(t_1)\lambda_2(t_2)(\theta(t_1, t_2) - 1)\, dt_1\, dt_2.
$$

Consequently,

$$
\begin{aligned}
&\operatorname{var}\left[\int_0^{t_0}\int_0^{t_0} w(u_1, u_2)\, dM_{11}(u_1)\, dM_{21}(u_2)\right] = \\
&\int_0^{t_0}\int_0^{t_0} \mathrm{E}[Y_1(u_1, u_2)]w^2(u_1, u_2)\lambda(u_1, u_2)\, du_1\, du_2 + O(n^{-1/2}).
\end{aligned}
$$

Because $A(t_1, t_2)$ is $O(n^{-1/2})$, $\lambda(t_1, t_2) - \lambda_1(t_1)\lambda_2(t_2)$ is $O(n^{-1/2})$. Thus, $V_n/\{n\hat{\sigma}^2\}^{1/2}$ is asymptotically equivalent to $S_n + R_n$, where

$$
S_n = \frac{n^{-1/2}\displaystyle\sum_i \int\!\!\int w(t_1, t_2)\{dM_{i1}(t_1)\, dM_{i2}(t_2) - Y_i(t_1, t_2)\lambda_1(t_1)\lambda_2(t_2)(\theta(t_1, t_2) - 1)\, dt_1\, dt_2\}}{\left\{n^{-1}\displaystyle\sum_i \int\!\!\int Y_i(t_1, t_2)w^2(t_1, t_2)\lambda(t_1, t_2)\, dt_1\, dt_2\right\}^{1/2}},
$$

$$
R_n = \frac{n^{-1/2}\displaystyle\sum_i \int\!\!\int Y_i(t_1, t_2)w(t_1, t_2)(\theta(t_1, t_2) - 1)\lambda_1(t_1)\lambda_2(t_2)dt_1dt_2}{\left\{n^{-1}\displaystyle\sum_i \int\!\!\int Y_i(t_1, t_2)w^2(t_1, t_2)\lambda_1(t_1)\lambda_2(t_2)\, dt_1dt_2\right\}^{1/2}}.
$$

The numerator of S_n is a normalized sum of n independent and identically distributed random variables with mean 0. The square of the denominator of S_n converges to the limiting variance of $\int\int w(t_1, t_2)\, dM_{11}(t_1)\, dM_{12}(t_2)$. Thus, by the law of large numbers and by the central limit theorem, S_n converges to the standard normal distribution. The second term R_n converges to (5) because $\Sigma_i\, Y_i(t_1, t_2)/n$ converges to $S(t_1, t_2)G(t_1, t_2)$.