Biometrika (2001), **88,** 4, *pp.* 907–919

Printed in Great Britain

On semi-competing risks data

By J. P. FINE
Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.
fine@stat.wisc.edu

H. JIANG
Department of Biostatistics and Center for Biostatistics in AIDS Research, Harvard University, Boston, Massachusetts 02115, U.S.A.
jiangh@circe.harvard.edu

and R. CHAPPELL
Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.
chappell@stat.wisc.edu

Summary

We consider a variation of the competing risks problem in which a terminal event censors a non-terminal event, but not vice versa. The joint distribution of the events is formulated via a gamma frailty model in the upper wedge where data are observable (Day et al., 1997), with the marginal distributions unspecified. An estimator for the association parameter is obtained from a concordance estimating function. A novel plug-in estimator for the marginal distribution of the non-terminal event is shown to be uniformly consistent and to converge weakly to a Gaussian process. The assumptions on the joint distribution outside the upper wedge are weaker than those usually made in competing risks analyses. Simulations demonstrate that the methods work well with practical sample sizes. The proposals are illustrated with data on morbidity and mortality in leukaemia patients.

Some key words: Clayton model; Dependent censoring; Martingale; Product limit estimator; Robustness; U-statistic.

1. Introduction

1·1. *Motivation*

Medical research frequently yields multiple event times. The times may consist of a terminal event, such as death, and non-terminal events, including landmarks of the disease process. The practical concern for clinicians is patient survival, suggesting an analysis based on the distribution of the terminal event, or the cause-specific hazard or cumulative incidence functions. However, there is often scientific interest in the distribution of a non-terminal event, unconditionally on censoring, which may be informative. The quantity is controversial but is meaningful to many researchers.

In cancer trials, death from intercurrent disease complicates the analysis of the biological efficacy of treatment. The goal is estimation of disease-free survival in the absence of failure types other than cancer. Consider a study of leukaemia patients receiving allogeneic bone marrow transplants (Copelan et al., 1991). A relapse causes mortality, but graft versus host disease may lead to death before relapse. Out of 137 patients, there were 42 relapses, 40 preceding deaths. Of the 41 deaths without relapse, the majority were attributable to graft versus host disease. The distribution of time without relapse corresponds to a setting where death from graft versus host disease is preventable. This counterfactual interpretation is appropriate if the mechanism by which graft versus host disease death is prevented has no side-effect altering the relapse process. The relapse distribution and the ability of relapse to predict death may be important to investigators.

1·2. *Notation and model*

Let X and Y be failure times, possibly dependent, with continuous joint survivor function $F(x, y) = \text{pr}(X > x, Y > y)$. The random variable Y may censor X, but not vice versa, if for instance Y is death from any cause and X is relapse. The marginal distribution of X, $F(x, 0) = \text{pr}(X > x)$, corresponds to a setting where Y may exist but no longer censors X. There is also a censoring time C independent of both X and Y, such as administrative loss to follow-up. Define $Z = \min(X, Y)$, $\eta_x = I(X < Y)$, $\eta_y = I(Y < C)$, $R = \min(Y, C)$, $\eta_z = I(Z < C)$ and $S = \min(Z, C)$, where $I(.)$ is the indicator function. The observed data are n independent and identically distributed realisations of $(\eta_y, R, \eta_z, S, \eta_z \eta_x)$. Denote these 'semi-competing risks' data by $\{(\eta_{yi}, R_i, \eta_{zi}, S_i, \eta_{zi}\eta_{xi}), i = 1, \ldots, n\}$.

If only the time and type of the first event are recorded, that is η_y and R are disregarded, then the data may be analysed in the traditional competing risks paradigm. The nonidentifiability of the joint distribution of (X, Y) is established in the nonparametric setting (Tsiatis, 1975). Much effort has gone into bounding the marginals (Peterson, 1976; Slud & Rubinstein, 1983; Zheng & Klein, 1995). Under restrictive conditions, estimation is possible (Emoto & Matthews, 1990; Moeschberger, 1974). However, the methods are complex and have assumptions that are wholly untestable.

With semi-competing risks data, the distribution of (X, Y) is nonparametrically identifiable in the upper wedge where $X < Y$ while the marginal survivor function of X is not. To show this, we can extend Lin & Ying (1993) to estimate $F(x, y)$ nonparametrically for $x \leqslant y$. Let $\hat{G}(y)$ be the Kaplan–Meier estimator of $G(y) = \text{pr}(C > y)$ based on $[\{R_i, (1 - \eta_{yi})\}, i = 1, \ldots, n]$. The estimator $\hat{F}(x, y) = \{\hat{G}(y)\}^{-1}\{n^{-1}\sum_i I(S_i > x, R_i > y)\}$ is uniformly consistent for $F(x, y)$ for $0 \leqslant x \leqslant y \leqslant \tau$, where $\text{pr}(S > \tau) > v > 0$, with v fixed. Observe that $F(x, 0)$ is not estimable without a model.

In many cases, the dependence structure is of biological interest, and not a nuisance in the problem specification. When the events are positively correlated, it is natural to posit the gamma frailty model (Clayton, 1978; Hougaard, 1986). Since $F(x, y)$ is only identified when $X < Y$, we define the model on the upper wedge (Day et al., 1997). For $\theta \geqslant 1$ and $0 \leqslant x \leqslant y \leqslant \infty$, $F(x, y)$ equals

$$\{F_x(x)^{1-\theta} + F_y(y)^{1-\theta} - 1\}^{1/(1-\theta)}, \tag{1}$$

where the functions F_x and F_y meet the definition for survivor functions. This is equivalent to the predictive (Oakes, 1989) hazard ratio being constant. That is, for $x \leqslant y$, $\lambda(y \mid \{x\})/\lambda(y \mid (x, \infty]) = \theta$, where

$$\lambda(y \mid A) = \lim_{\varepsilon \to 0} d\{\text{pr}(Y < y + \varepsilon \mid Y \geqslant y, X \in A)\}/d\varepsilon, \quad A \subseteq (0, \infty).$$

When $\theta = 1$, X and Y are independent on the upper wedge and $F_x(x) = \text{pr}(X > x \mid Y > x)$.

Since $F(x, y)$ need not follow the model (1) on the lower wedge, θ may not have the usual relationship with Kendall's τ, namely $\tau = (\theta - 1)/(\theta + 1)$. The joint density on the upper wedge, $f_{x \leqslant y}(x, y)$, is

$$\theta\{F_x(x)^{1-\theta} + F_y(y)^{1-\theta} - 1\}^{(2\theta-1)/(1-\theta)} F_x(x)^{-\theta} F_y(y)^{-\theta} f_x(x) f_y(y), \tag{2}$$

where $f_x(x) = -d\{F_x(x)\}/dx$ and $f_y(y) = -d\{F_y(y)\}/dy$. The joint density on the lower wedge, $f_{y<x}(x, y)$, is unspecified but satisfies

$$\{F_x(x)^{1-\theta} + F_y(y)^{1-\theta} - 1\}^{1/(1-\theta)} = \int_y^\infty \int_x^t f_{x \leqslant y}(s, t)\, ds\, dt + \int_y^\infty \int_t^\infty f_{y<x}(s, t)\, ds\, dt$$

for $x \leqslant y$. Note that there are an infinite number of $f_{y<x}$ which solve the integral equations. A trivial solution is when the density (2) holds for $y < x$. A difficult technical issue is finding other $f_{y<x}$ which yield model (1).

The interpretation of F_x as the distribution of X requires that $F_x(x) = F(x, 0)$. A class of joint distributions having the property follows. Let $F(x, y)$ equal (1) for $x \leqslant y$ and equal $L\{F_x(x), F_y(y)\}$ for $y < x$, where $L(t_1, t_2) = \text{pr}(u_1 \leqslant t_1, u_2 \leqslant t_2)$ and (u_1, u_2) is a vector of $\text{Un}(0, 1)$ variates with unspecified joint distribution. If $L(t_1, t_2) = (t_1^{1-\theta} + t_2^{1-\theta} - 1)^{1/(1-\theta)}$, then (1) holds for all x, y. In general, if $L\{F_x(x), F_y(y)\}$ equals (1) for $x = y$, then a valid $F(x, y)$ results and has the same F_x and F_y and possibly different dependence structures on the two wedges. The predictive hazard ratio is constant on the upper wedge but may not be constant for $x > y$. The model on the lower wedge is unverifiable but is less restrictive than assuming an explicit form for $f_{y<x}$, as in competing risks analyses (Emoto & Matthews, 1990; Moeschberger, 1974). The copula L is nonparametric and X and Y may be dependent on the lower wedge. Note that $F_y(y)$ accords with $F(0, y) = \text{pr}(Y > y)$, regardless of $f_{y<x}$.

1·3. *Outline*

Maximisation of a pseudolikelihood derived from consistent estimators of F_x and F_y gives a consistent estimator for θ (Genest et al., 1995; Shih & Louis, 1995). With semi-competing risks data, one might employ the Kaplan–Meier estimators $\hat{F}_x$ and $\hat{F}_y$ computed with $\{(S_i, \eta_{zi}\eta_{xi}), i = 1, \ldots, n\}$ and $\{(R_i, \eta_{yi}), i = 1, \ldots, n\}$. However, $\hat{F}_x$ does not generally converge to F_x as $n \to \infty$. In fact, without an estimator for θ, a consistent estimator for F_x may not exist.

Day et al. (1997) modified Clayton (1978) to obtain an estimator for θ but the limiting distribution was not provided. In § 2, we adapt Oakes' (1982, 1986) closed-form estimator, which involves neither F_x nor F_y. The asymptotic properties are derived in the Appendix. A simple test of fit is described for checking model (1) on the support of the data.

Estimation of F_x has yet to be resolved. In § 3, we introduce a plug-in estimator which uses the estimator of θ from § 2. It also uses $\hat{F}_y$ and $\hat{F}_z$, the product limit estimator for F_z, the survivor function of Z, based on $\{(S_i, \eta_{zi}), i = 1, \ldots, n\}$. The estimator is robust to the joint density for $x > y$. That is, it estimates F_x regardless of the model on the lower wedge. If $F_x = F(x, 0)$, then it provides a consistent estimator for the marginal distribution. A weak convergence result is outlined in the text, with details in the Appendix. In § 4, simulations show that the methods perform well in moderately sized samples and are indeed insensitive to the region where $X > Y$. Data from the leukaemia study are analysed in § 5 and remarks about model specification and interpretation conclude in § 6.

2. INFERENCE FOR THE DEPENDENCE STRUCTURE

Consider two independent pairs (X_i, Y_i) and (X_j, Y_j). Let

$$\Delta_{ij} = I\{(X_i - X_j)(Y_i - Y_j) > 0\}.$$

In the semi-competing risks setting, Δ_{ij} is only computable when $\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij}$, where $\tilde{X}_{ij} = \min(X_i, X_j)$, $\tilde{Y}_{ij} = \min(Y_i, Y_j)$ and $\tilde{C}_{ij} = \min(C_i, C_j)$. If model (1) is assumed, then the expectation of Δ_{ij} conditional on $\tilde{X}_{ij} < \tilde{Y}_{ij}$ is $\theta_0/(1+\theta_0)$, where θ_0 is the true value of θ, assumed bounded. This occurs because the predictive hazard ratio equals θ_0 in the upper wedge.

The following extends Oakes' (1982, 1986) estimators. Let $D_{ij} = I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij})$ and define

$$\tilde{S}_{ij} = \min(\tilde{X}_{ij}, \tilde{Y}_{ij}, \tilde{C}_{ij}), \quad \tilde{R}_{ij} = \min(\tilde{Y}_{ij}, \tilde{C}_{ij}), \quad U(\theta) = \sum_{i<j} W(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij}\{\Delta_{ij} - \theta/(1+\theta)\}.$$

The weight $W(u, v)$ is a random function satisfying $\sup_{u,v}|W(u, v) - \tilde{W}(u, v)| \to 0$ in probability, where $\tilde{W}$ is deterministic and bounded for (u, v) in the support of $(\tilde{S}_{ij}, \tilde{R}_{ij})$. The solution to $U(\theta) = 0$ is

$$\hat{\theta} = \frac{\sum_{i<j} W(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij} \Delta_{ij}}{\sum_{i<j} W(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij}(1 - \Delta_{ij})}.$$

When $W = 1$, an unweighted concordance estimator obtains. A useful weight function is

$$W_{a,b}^{-1}(x, y) = n^{-1} \sum_{i=1}^{n} I\{S_i \geqslant \min(a, x), R_i \geqslant \min(b, y)\},$$

where a and b are constants. With $a = b = \infty$, $\hat{\theta}$ is analogous to the weighted estimator in Oakes (1986). The values a and b may be selected to dampen $W_{a,b}(x, y)$ for 'large' x and y.

It is easy to show that, as $n \to \infty$, $n^{-2}\{U(\theta) - \tilde{U}(\theta)\}$ vanishes uniformly for θ in a neighbourhood of θ_0, where $\tilde{U}$ is U with W replaced by $\tilde{W}$. Thus, $\hat{\theta}$ has the same limit as $\tilde{\theta}$, the root of $\tilde{U}(\theta)$. Now, $E\{E(\Delta_{ij} | \tilde{S}_{ij}, \tilde{R}_{ij}) | D_{ij} = 1\} = \theta_0/(1+\theta_0)$, implying that $E\{\tilde{U}(\theta_0)\} = 0$. The strong law of large numbers for U-statistics and a continuous mapping theorem give that $\tilde{\theta}$ is strongly consistent for θ_0. Thus, $\hat{\theta}$ is strongly consistent.

In the Appendix we show that $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ has a limiting normal distribution with variance Σ which is consistently estimated by $\hat{\Sigma} = \hat{I}^{-2}\hat{J}$, where

$$\hat{I} = n^{-2} \sum_{i<j} W(\tilde{S}_{ij}, \tilde{R}_{ij}) D_{ij}(1+\hat{\theta})^{-2},$$

$$\hat{J} = 2n^{-3} \sum_{k<l<m} (\hat{Q}_{kl}\hat{Q}_{km} + \hat{Q}_{kl}\hat{Q}_{lm} + \hat{Q}_{lm}\hat{Q}_{km}),$$

and $\hat{Q}_{kl} = W(\tilde{S}_{kl}, \tilde{R}_{kl}) D_{kl}\{\Delta_{kl} - \hat{\theta}/(1+\hat{\theta})\}$. A test of independence of X and Y may be constructed via a confidence interval for θ. If a $100(1-2\alpha)$ interval $(0 < \alpha < 0{\cdot}5)$ excludes 1, then the null $\theta = 1$ is rejected. A score test can be based on $\{n^{-3/2}U(1)\}(\hat{J})^{-1/2}$, where $\hat{Q}_{kl}$ is computed with 1 in place of $\hat{\theta}$.

Since inferences about F_x rely on copula (1) on the upper wedge, it would be helpful to assess this formulation. Unlike with competing risks data, model checking is possible. We propose a goodness-of-fit statistic based on the distance between two estimators from $U(\theta)$ with different weights (Shih, 1998). Under misspecification, the estimators may converge to distinct values and the test rejects with probability one.

Let $W_i = W_{a_i,b_i}$, let U_i be U with W_i in place of W, and let $\hat{\theta}_i$ be the corresponding estimator, $i = 1, 2$. In the Appendix, we show that when the copula is specified correctly $n^{\frac{1}{2}}(\hat{\theta}_1 - \hat{\theta}_2)$ is asymptotically normal with variance that is consistently estimated by

$$\hat{\Gamma} = 2n^{-3} \sum_{k<l<m} (\hat{Q}^*_{kl}\hat{Q}^*_{km} + \hat{Q}^*_{kl}\hat{Q}^*_{lm} + \hat{Q}^*_{lm}\hat{Q}^*_{km}),$$

where $\hat{Q}^*_{kl} = \hat{I}_1^{-1}\hat{Q}_{1kl} - \hat{I}_2^{-1}\hat{Q}_{2kl}$, and $\hat{I}_i$ and $\hat{Q}_{ikl}$ are $\hat{I}$ and $\hat{Q}_{kl}$ with W replaced by W_i, for $i = 1, 2$. For a 2α-level test, the critical region is $n^{\frac{1}{2}}|\hat{\theta}_1 - \hat{\theta}_2|\hat{\Gamma}^{-\frac{1}{2}} > \psi_{1-\alpha}$, where ψ_q is the qth quantile of the standard normal distribution.

The power of the test depends on $W_i(x, y)$, for $i = 1, 2$. Shih (1998) described a principle for choosing the weights so that the difference $\hat{\theta}_1 - \hat{\theta}_2$ is accentuated. One weight should result in a precise estimator if the true model for $F(x, y)$ is given by (1), for example $W_{\infty,\infty}(x, y)$. The other should emphasise (x, y) regions where the model is hypothesised to fit poorly. If the alternative has nonmonotone predictive hazard ratio when $x \leqslant y$, a good choice might be $W_{\infty,\infty}\{h_1(x), h_2(y)\}$, where $h_1(.)$ and $h_2(.)$ are nonmonotone functions.

3. Estimation of F_x

The survivor function $F_z(t)$ is $\text{pr}(X > t, Y > t)$, which equals

$$\{F_x(t)^{1-\theta} + F_y(t)^{1-\theta} - 1\}^{1/(1-\theta)}$$

under model (1) with $x = y = t$. Some algebra gives $F_x(t) = g\{F_z(t), F_y(t), \theta\}$, where $g(a, b, c) = (a^{1-c} - b^{1-c} + 1)^{1/(1-c)}$. This suggests the estmator $\hat{F}_x = g(\hat{F}_z, \hat{F}_y, \hat{\theta})$, where $\hat{\theta}$ is the root of $U(\theta)$, and $\hat{F}_z$ and $\hat{F}_y$ are the product limit estimators for F_z and F_y. When X and Y are assumed independent, $g(\hat{F}_z, \hat{F}_y, 1)$ reduces to the Kaplan–Meier estimator based on $\{(S_i, \eta_{xi}\eta_{zi}), i = 1, \ldots, n\}$.

Recall that $\hat{\theta}$ is strongly consistent for θ_0. Since Z and Y are subject only to independent censoring by C, $\hat{F}_z(t)$ and $\hat{F}_y(t)$ are strongly consistent for $F_z(t)$ and $F_y(t)$, uniformly for $t \in [0, \tau]$ (Fleming & Harrington, 1991, Ch. 6). Since g has bounded derivatives, a continuous mapping theorem gives the uniform strong convergence of $\hat{F}_x(t)$ to $F_x(t)$. This implies that, if $\theta = 1$ in (1) for $x < y$, then the usual Kaplan–Meier estimator is consistent for F_x.

In the Appendix, we show that $n^{\frac{1}{2}}\{\hat{F}_x(t) - F_x(t)\}$ converges weakly to a Gaussian process for $t \in [0, \tau]$. The covariance function

$$\sigma(s, t) = \text{cov}[n^{\frac{1}{2}}\{\hat{F}_x(s) - F_x(s)\}, n^{\frac{1}{2}}\{\hat{F}_x(t) - F_x(t)\}]$$

is given in (A7). A consistent estimator is

$$\hat{\sigma}(s, t) = n^{-3} \sum_{k<l<m} \{\hat{V}_{kl}(s)\hat{V}_{km}(t) + \hat{V}_{lm}(s)\hat{V}_{km}(t) + \hat{V}_{kl}(s)\hat{V}_{lm}(t) + \hat{V}_{lm}(s)\hat{V}_{kl}(t) + \hat{V}_{km}(s)\hat{V}_{kl}(t) + \hat{V}_{lm}(s)\hat{V}_{kl}(t)\},$$

where

$$\hat{V}_{ij}(t) = -g_1\{\hat{F}_z(t), \hat{F}_y(t), \hat{\theta}\}\hat{F}_z(t)\int_0^t \hat{\pi}_z(u)^{-1}\{d\hat{M}_{zi}(u) + d\hat{M}_{zj}(u)\}$$
$$- g_2\{\hat{F}_z(t), \hat{F}_y(t), \hat{\theta}\}\hat{F}_y(t)\int_0^t \hat{\pi}_y(u)^{-1}\{d\hat{M}_{yi}(u) + d\hat{M}_{yj}(u)\}$$
$$+ g_3\{\hat{F}_z(t), \hat{F}_y(t), \hat{\theta}\}\hat{I}^{-1}\hat{Q}_{ij},$$

g_1, g_2 and g_3 are defined in (A5)–(A6), $\hat\pi_z$ and $\hat\pi_y$ appear in (A3),

$$\hat M_{zi}(t) = I(S_i \leqslant t, \eta_{zi} = 1) - \int_0^t I(S_i \geqslant u)\, d\hat\Lambda_z(u),$$

$$\hat M_{yi}(t) = I(R_i \leqslant t, \eta_{yi} = 1) - \int_0^t I(R_i \geqslant u)\, d\hat\Lambda_y(u),$$

and $\hat\Lambda_z$ and $\hat\Lambda_y$ are Nelson–Aalen estimators for $-\log(F_z)$ and $-\log(F_y)$. To construct confidence intervals for F_x, we consider $n^{\frac{1}{2}}[m\{\hat F_x(t) - m\{F_x(t)\}]$, where m is an invertible and differentiable function. The function is chosen to bound the intervals in $[0, 1]$ and to stabilise the variance. Let $\dot m(x) = d\{m(x)\}/dx$. If we apply the δ-method, a $(1-2\alpha)$ interval for $F_x(t)$ has endpoints $m^{-1}[m\{\hat F_x(t)\} \pm n^{-\frac{1}{2}} \dot m\{\hat F_x(t)\} \hat\sigma(t, t)^{\frac{1}{2}} \psi_{1-\alpha}]$.

The inferences use a U-statistic approximation to $\hat F_x$. In simulations, we find that $\hat\sigma(t, t)$ may underestimate the variance. To boost performance, we consider

$$\hat\sigma^*(s, t) = \hat\sigma(s, t) + n^{-3} \sum_{i<j} \hat V_{ij}(s) \hat V_{ij}(t).$$

The extra summation is negligible as $n \to \infty$, but is nonnegative when $s = t$. It arises from the exact finite sample variance of the U-statistic equivalent for the estimator. Simulations show that $\hat\sigma^*$ is more reliable in small samples.

The estimator $\hat F_x$ is a step-function. The changes are at the observed values of X and Y at which $\hat F_z(t)^{1-\hat\theta} - \hat F_y(t)^{1-\hat\theta}$ jumps. In finite samples, $\hat F_z(t)$ may be greater than $\hat F_y(t)$, although $F_z(t) \leqslant F_y(t)$, for all t. Also, $\hat\theta$ may be less than one. This means that $\hat F_x(t)$ may not be monotone or may not be well defined. In contrast, the Kaplan–Meier estimator decreases at each S with $\eta_z \eta_x = 1$. The difficulties arise in estimating probabilities in the tail of F_x with heavy censoring of X by Y.

To address the instability, we restrict inferences to the interval $[0, t^*]$, where

$$t^* \leqslant \max\{s : \hat F_z(u)^{1-\hat\theta} - \hat F_y(u)^{1-\hat\theta} > -1, 0 \leqslant \hat F_x(u) \leqslant 1, u \leqslant s\}.$$

For $t \leqslant t^*$, define the monotone estimator $\hat F_x^*(t) = \min_{s \leqslant t}\{\hat F_x(s)\}$. This estimator accepts $\hat F_x(t)$ if it satisfies the monotonicity constraint. If not, it carries forward the smallest value of $\hat F(s)$ for $s \leqslant t$. Since $\hat F_x$ is uniformly consistent, so too is $\hat F_x^*$. We conjecture that $n^{\frac{1}{2}}\{\hat F_x^*(t) - F_x(t)\}$ and $n^{\frac{1}{2}}\{\hat F_x(t) - F_x(t)\}$ have the same limiting distribution. In Monte Carlo experiments, the adjustment improves upon $\hat F_x(t)$ on average.

4. Numerical studies

We began by generating n (X, Y) pairs with model (1) on both wedges. The marginals F_x and F_y were unit exponentials with $\mathrm{pr}(Y < X) = 0{\cdot}5$ for all θ. The independent censoring time C was a Un(0, 5) variate, giving 20% censoring of Y. For each combination of $\theta = 1$, 2 or 3 and $n = 100$ or 200, 1000 datasets were simulated. Since the procedures are invariant to rank-preserving transformations of the timeline, the following results hold for $h(X)$, $h(Y)$ and $h(C)$, where $h(.)$ is any monotone increasing function.

We studied $U(\theta)$ with $W = 1$ and with a and b in $W_{a,b}$ from § 2 equal to $x_{0{\cdot}95}$ and $y_{0{\cdot}95}$, the 95th percentiles of the observed X and Y values, respectively. In Table 1, we report the means of $\hat\theta$, Ave, and $n^{-1}\hat\Sigma$, AveVar, the empirical variance of $\hat\theta$, EmpVar, and the coverage probability of the nominal 0·95 interval for θ, Cov95. In all cases, $\hat\theta$ performs well. The bias is small, decreasing as n increases. The estimator $n^{-1}\hat\Sigma$ and the empirical

Table 1. *Numerical studies. Comparison of the weighted and unweighted estimators for θ. Taking $(a, b) = (0, 0)$ corresponds to $W = 1$*

θ	(a, b)	Ave	EmpVar	AveVar	Cov95	Ave	EmpVar	AveVar	Cov95
			$n = 100$				$n = 200$		
1	(0, 0)	1·02	0·054	0·059	0·948	1·01	0·028	0·029	0·945
	$(x_{0\cdot95}, y_{0\cdot95})$	1·03	0·037	0·046	0·953	1·01	0·017	0·023	0·974
2	(0,0)	2·07	0·205	0·212	0·945	2·03	0·100	0·103	0·942
	$(x_{0\cdot95}, y_{0\cdot95})$	2·07	0·160	0·179	0·961	2·03	0·075	0·089	0·968
3	(0, 0)	3·12	0·478	0·461	0·938	3·06	0·210	0·223	0·950
	$(x_{0\cdot95}, y_{0\cdot95})$	3·12	0·382	0·393	0·956	3·06	0·173	0·196	0·961

Ave, empirical mean; EmpVar, empirical variance; AveVar, model-based variance; Cov95, empirical coverage probability.

variance agree and the intervals behave properly. As expected, the weighted estimator is more precise than the unweighted (Oakes, 1986).

We computed $\hat{F}_x^*$ using the two estimators of θ. Confidence intervals were based on $\hat{\sigma}^*$ and $m(x) = \log\{x/(1-x)\}$. The findings are similar. Results for the weighted version are shown in Table 2 at various quantiles of F_x, including the percentage of valid estimates, PerVal. The estimator is almost always unbiased. The quantity $\hat{\sigma}^*$ is a good approximation to the variance. The coverage is generally close to 0·95 and improves with larger samples. There are problems when $F_x(t) = 0{\cdot}1$, but these diminish with $n = 200$.

To highlight the potential for bias, we calculated the naive Kaplan–Meier estimator. The average value and the empirical variance are given in Table 2. Under independence, the new estimator and the product limit estimator are both accurate. However, since θ is estimated and not fixed at 1, $\hat{F}_x^*$ may be somewhat less efficient. Of course, the naive estimator may severely overestimate the probabilities when $\theta > 1$.

Next, we simulated with radically different gamma frailty densities for $f_{x\leqslant y}$ and $f_{y<x}$, and a joint survivor function on the upper wedge which is approximately gamma with (F_x, F_y, θ) matching $f_{x\leqslant y}$. In the following, F_x may not equal $F(x, 0)$. These mixed wedge examples illustrate pitfalls in interpreting F_x as the distribution of X.

A difficulty in constructing such examples is that arbitrary densities with the form in (2) will not generally produce a valid $F(x, y)$. Even when $F(x, y)$ is valid, it will not generally have the form in (1) for $x \leqslant y$. The problem is that (1) $\Rightarrow$ (2) in the upper wedge, but the converse is not true. We consider $f_{x\leqslant y}$ with $(F_x = F_y = F^u, \theta = \theta^u)$ in (2) and $f_{y<x}$ with $(F_x = F_y = F^l, \theta = \theta^l)$ in (2). Denote by $\tilde{F}^u(x, y)$ (1) with $(F_x = F_y = F^u, \theta = \theta^u)$ and by $\tilde{F}^l(x, y)$ (1) with $(F_x = F_y = F^l, \theta = \theta^l)$. The joint distribution implied by $f_{x\leqslant y}$ and $f_{y<x}$, $F(x, y)$, exists and equals $\{\tilde{F}^l(y, y) - \tilde{F}^u(y, y)\}/2 + \tilde{F}^u(x, y)$ for $x \leqslant y$. This means that the absolute error in approximating $\tilde{F}^u(x, y)$ by $F(x, y)$ is $|\tilde{F}^l(y, y) - \tilde{F}^u(y, y)|/2$. A similar error formula applies to $\tilde{F}^l(x, y)$ on the lower wedge, with x replacing y.

For $\{F^u = \exp(-3t), \theta^u = 1\}$ and $\{F^l = \exp(-5{\cdot}25t), \theta^l = 12\}$,

$$\sup_{x\leqslant y} |F(x, y) - \tilde{F}^u(x, y)| = \sup_{x>y} |F(x, y) - \tilde{F}^l(x, y)| \leqslant 0{\cdot}015.$$

Thus, $F(x, y)$ is a good approximation to gamma models with distinct (F_x, F_y, θ) on each wedge and $F^l \simeq F(x, 0)$. Since misspecification is slight, $\hat{\theta}$ and $\hat{F}_x^*$ should be almost unbiased for θ^u and F^u.

Two hundred (X, Y) pairs were generated from the model. Based on 1000 replications, the average values of the weighted and unweighted estimators of θ^u were 1·06 and 1·04.

Table 2. *Numerical studies. Comparison of $\hat{F}_x^*$ and the Kaplan–Meier estimator*

$F_x/(t)$	n	$\hat{F}_x^*(t)$ Ave	EmpVar	AveVar	PerVal	Cov95	Kaplan–Meier Ave	EmpVar
				$\theta=1$				
0·9	100	0·90	0·102	0·084	1·00	0·942	0·90	0·102
	200	0·90	0·048	0·044	1·00	0·946	0·90	0·044
0·7	100	0·70	0·314	0·250	1·00	0·930	0·70	0·260
	200	0·70	0·160	0·130	1·00	0·926	0·70	0·137
0·5	100	0·50	0·547	0·490	1·00	0·937	0·50	0·423
	200	0·50	0·281	0·260	1·00	0·934	0·50	0·221
0·3	100	0·30	0·865	0·941	1·00	0·953	0·30	0·578
	200	0·30	0·449	0·518	1·00	0·965	0·30	0·291
0·1	100	0·09	1·720	1·020	0·99	0·419	0·11	0·828
	200	0·10	0·922	0·689	1·00	0·662	0·10	0·410
				$\theta=2$				
0·9	100	0·90	0·116	0·096	1·00	0·935	0·91	0·090
	200	0·90	0·053	0·048	1·00	0·946	0·91	0·044
0·7	100	0·70	0·397	0·292	1·00	0·906	0·74	0·250
	200	0·70	0·185	0·152	1·00	0·928	0·73	0·116
0·5	100	0·50	0·608	0·490	1·00	0·921	0·58	0·372
	200	0·50	0·292	0·260	1·00	0·947	0·58	0·176
0·3	100	0·31	0·640	0·672	1·00	0·950	0·42	0·449
	200	0·30	0·325	0·336	1·00	0·962	0·42	0·291
0·1	100	0·13	1·513	9·242	0·96	0·898	0·24	0·740
	200	0·12	0·462	0·608	1·00	0·956	0·23	0·348
				$\theta=3$				
0·9	100	0·90	0·130	0·102	1·00	0·935	0·91	0·096
	200	0·90	0·058	0·053	1·00	0·954	0·90	0·044
0·7	100	0·70	0·422	0·325	1·00	0·912	0·75	0·240
	200	0·70	0·203	0·168	1·00	0·922	0·76	0·102
0·5	100	0·51	0·578	0·476	1·00	0·931	0·62	0·336
	200	0·50	0·281	0·240	1·00	0·936	0·64	0·160
0·3	100	0·32	0·689	0·624	1·00	0·945	0·47	0·449
	200	0·31	0·260	0·260	1·00	0·959	0·47	0·221
0·1	100	0·11	1·145	14·592	0·90	0·921	0·27	0·723
	200	0·11	0·436	3·168	0·98	0·949	0·27	0·348

Ave, empirical mean; EmpVar, empirical variance; AveVar, model-based variance; PerVal, percentage of valid estimates; Cov95, empirical coverage probability.
EmpVar and AveVar are multiplied by 100.

When the simulations were redone with $F^u = F^l = \exp(-3t)$ and $\theta^u = \theta^l = 1$, the averages were 1·01 and 1·01. This shows that $\hat{\theta}$ is robust to F^l and θ^l. In Table 3, averages for $\hat{F}_x^*$ with the weighted estimator of θ^u and for the Kaplan–Meier estimator are given at four quantiles. Since $\theta^u = 1$, both are roughly consistent for F^u, regardless of whether or not $F^u = F^l$.

As a final test, we switched the models on the lower and upper wedges. The averages of the weighted and unweighted estimators for θ^u were 11·68 and 11·57. For $F^u = F^l =$

Table 3. *Numerical studies. Average values of $\hat{F}_x^*$ and the Kaplan–Meier estimator with different gamma frailty densities on the lower and upper wedges*

Model					$F^u(t)$			
F^u	F^l	θ^u	θ^l	Method	0·8	0·6	0·4	0·2
$\exp(-3t)$	$\exp(-5{\cdot}25t)$	1	12	New	0·80	0·60	0·41	0·21
				KM	0·80	0·60	0·41	0·22
$\exp(-3t)$	$\exp(-3t)$	1	1	New	0·80	0·60	0·40	0·20
				KM	0·80	0·60	0·40	0·20
$\exp(-5{\cdot}25t)$	$\exp(-3t)$	12	1	New	0·82	0·60	0·39	0·18
				KM	0·87	0·75	0·61	0·42
$\exp(-5{\cdot}25t)$	$\exp(-5{\cdot}25t)$	12	12	New	0·80	0·60	0·40	0·19
				KM	0·87	0·75	0·61	0·43

Methods: New, $\hat{F}_x^*$; KM, Kaplan–Meier.

$\exp(-5{\cdot}25t)$ and $\theta^u = \theta^l = 12$, the averages were 12·17 and 12·18. This confirms the insensitivity of $U(\theta)$ to the lower wedge. In Table 3, we see that $\hat{F}_x^*$ estimates F^u, while the Kaplan–Meier curve does not. This happens because X and Y are dependent on the upper wedge.

5. Leukaemia data

We analyse the data from Copelan et al. (1991). A complete dataset is available in Klein & Moeschberger (1997, p. 464). Let time to relapse be X and let time to death from any cause be Y, with the origin at transplantation. Estimates, with standard errors in parentheses, for θ using $U(\theta)$ with $W_{0,0}$ and $W_{\infty,\infty}$ are $\hat{\theta}_u = 8{\cdot}79$ (2·15) and $\hat{\theta}_w = 8{\cdot}61$ (2·15). Based on these estimates, the goodness-of-fit statistic is 0·47, with p-value 0·64; the model fits the data. Both estimates indicate that relapse is highly predictive of death.

We computed $\hat{F}_x^*$ employing $\hat{\theta}_w$, as well as the Kaplan–Meier estimator using $\{(S_i, \eta_{xi}\eta_{zi}), i = 1, \ldots, n\}$. These are shown in Fig. 1, along with the 0·95 confidence intervals from the new estimator with $\hat{\sigma}^*$ and $m(x) = \log\{x/(1-x)\}$. Note that the naive nonparametric estimator is uniformly above the upper 0·95 limit. Since the association between death and relapse is substantial, it is not surprising that the estimates are very different.

In this analysis, interpreting the estimand for $\hat{F}_x^*$ as the relapse distribution is debatable. The difficulty is partly conceptual. Since relapse can never occur after death, the distribution of (X, Y) on the lower wedge is somewhat arbitrary. There is no basis for assessing the relevance of the models in § 1 where $F_x = F(x, 0)$ and of those in § 4 where $F_x \neq F(x, 0)$. Situations in which $X > Y$ can occur but is not observed are more natural in the sense that there is a real meaning of $F(x, y)$ on the lower wedge. For example, in clinical trials, drop-out may informatively censor a non-terminal event which may occur subsequent to drop-out.

The mathematics of using semi-competing risks data to identify F_x with $F(x, 0)$ is the same in all cases. The issue is that neither $\hat{F}_x^*$ nor the Kaplan–Meier estimator uses information in the region where $X > Y$. Hence, these quantities do not have estimands on the lower wedge without extra assumptions. The appropriateness of particular conditions is a topic for substantive scientific argument.

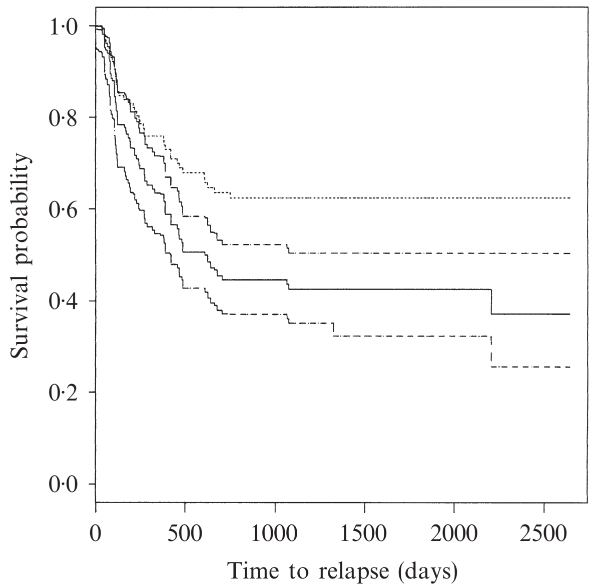

Fig. 1. Leukaemia data. Estimates of the survivor function for time to relapse. Solid line is the point estimate from $\hat{F}_x^*$, dashed lines are the limits of the corresponding 0·95 intervals, and the dotted line is the Kaplan–Meier estimate.

6. Remarks

We showed that with semi-competing risks data F_x is estimable on the upper wedge under a gamma frailty specification. The closed-form estimators $\hat{\theta}$ and $\hat{F}_x$ are intuitively appealing and should be highly efficient. It would be worthwhile to investigate whether or not there are other procedures which are computationally simple and yield more precise estimators.

We reiterate that $\hat{F}_x$ is insensitive to the joint density for $x > y$. Hence, our approach is fundamentally different from a sensitivity analysis (Peterson, 1976) which gives bounds for probabilities on the lower wedge. The parameters in model (1) are clearly defined in the region of the observable data. Whether or not $\hat{F}_x$ can be used to derive bounds for $F(x, 0)$ without any assumptions on the lower wedge is an important open question. However, it is beyond the scope of this paper. One justification for estimating F_x is that it is useful as an intermediate step in estimating identifiable aspects of $F(x, y)$ for $x \leqslant y$.

Of course, there is great interest in the interpretation of F_x as the marginal distribution of X. The nonidentifiable nonparametric assumptions on the lower wedge giving $F_x = F(x, 0)$ are weaker than those in traditional competing risks analyses. Evaluating their plausibility requires careful consideration of the mechanisms generating the data and the implications for the joint distribution on the lower wedge. It would be useful to explore this issue in various applications.

A special feature of the model (1) is that the conditional expectation of Δ_{ij} given $D_{ij} = 1$ only involves θ, which can be estimated separately from F_x and F_y. This means there is enough information to test the parametric assumption without estimating the marginals. Another statistic, which is omnibus, is

$$\sup_{0 \leqslant x \leqslant y \leqslant \tau} |\hat{F}(x, y) - \{\hat{F}_x(x)^{1-\hat{\theta}} + \hat{F}_y(y)^{1-\hat{\theta}} - 1\}^{1/(1-\hat{\theta})}|.$$

Thus, the model for $F(x, y)$ is fully testable for $x \leqslant y$.

The simulations with mixed wedge models indicate that, if model (1) is a reasonable approximation to the true joint survivor function for $x \leqslant y$, then the plug-in estimator is nearly unbiased for F_x. A formal study of the robustness of $\hat{F}_x$ to non-gamma models on the upper wedge is a topic for future research.

The Kaplan–Meier estimator does not have a straightforward estimand when X and Y are dependent on the upper wedge, while $\hat{F}_x$ does. If $\theta = 1$ in model (1), then both estimators are consistent for $\text{pr}(X > x | Y > x)$, which is meaningful. However, the usual assumption of random censoring, that is that the estimand is $F(x, 0)$, may be invalid if X and Y are dependent on the lower wedge. On the other hand, if censoring is noninformative, then the estimators are unbiased for the marginal distribution and $\hat{F}_x$ may be less efficient. This is the price of not assuming independence.

Acknowledgement

We are grateful to the editor, the associate editor and the referee for helpful comments and to D. Oakes for the reference to Day et al. (1997). H. Jiang was supported by grants from the National Eye Institute and the National Institute of Allergies and Infectious Diseases. R. Chappell was supported by a grant from the National Eye Institute.

Appendix

Proofs

Asymptotic normality of $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$. A Taylor expansion of $U(\hat{\theta})$ in $\hat{\theta}$ around θ_0 and the consistency of $\hat{\theta}$ give

$$n^{1/2}(\hat{\theta} - \theta_0) = I^{-1}\{n^{-3/2}U(\theta_0)\} + o_p(1),$$

where I is the probability limit of $\hat{I}$. Straightforward calculations show that

$$n^{-3/2}U(\theta_0) = n^{-3/2}\sum_{i<j} Q_{ij} + o_p(1),$$

where $Q_{ij} = \tilde{W}(\tilde{S}_{ij}, \tilde{R}_{ij})D_{ij}\{\Delta_{ij} - \theta_0(1 + \theta_0)^{-1}\}$. A central limit theorem for U-statistics and Slutsky's law yield the normal distribution for $n^{1/2}(\hat{\theta} - \theta_0)$ with variance $I^{-2}J$, where J is the limit of $\hat{J}$.

Under the null hypothesis, the distributions of $\hat{\theta}_1$ and $\hat{\theta}_2$ are centred around the same θ_0. By the previous results,

$$n^{1/2}(\hat{\theta}_1 - \hat{\theta}_2) = n^{-3/2}\sum_{i<j} Q^*_{ij} + o_p(1),$$

where $Q^*_{ij} = I_1^{-1}Q_{1ij} - I_2^{-1}Q_{2ij}$, $I_k = \lim_{n\to\infty}\hat{I}_k$ and Q_{kij} is Q_{ij} with $\tilde{W}$ replaced by $\tilde{W}_k = \lim_{n\to\infty} W_k$, for $k = 1, 2$. A limit theorem for U-statistics gives asymptotic normality, with variance

$$\Gamma = \lim_{n\to\infty} 2n^{-3}\sum_{k<l<m} (Q^*_{kl}Q^*_{km} + Q^*_{kl}Q^*_{lm} + Q^*_{lm}Q^*_{km}).$$

A consistent estimator $\hat{\Gamma}$ is computed with $\hat{Q}^*_{ij}$ in place of Q^*_{ij} in Γ.

Weak convergence of $n^{\frac{1}{2}}\{\hat{F}_x(t) - F_x(t)\}$. The martingale representations for $\hat{F}_z$ and $\hat{F}_y$ (Gill, 1980, p. 37) give

$$n^{\frac{1}{2}}\{\hat{F}_z(t) - F_z(t)\} = -F_z(t)n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \pi_z(u)^{-1}\, dM_{zi}(u) + o_p(1), \tag{A1}$$

$$n^{\frac{1}{2}}\{\hat{F}_y(t)-F_y(t)\} = -F_y(t)n^{-\frac{1}{2}}\sum_{i=1}^{n}\int_0^t \pi_y(u)^{-1}\,dM_{yi}(u)+o_p(1), \tag{A2}$$

where $\pi_z(t)$ and $\pi_y(t)$ are the limits of

$$\hat{\pi}_z(t)=n^{-1}\sum_{i=1}^{n} I(S_i \geqslant t), \quad \hat{\pi}_y(t)=n^{-1}\sum_{i=1}^{n} I(R_i \geqslant t); \tag{A3}$$

$$M_{zi}(t)=I(S_i \leqslant t, \eta_{zi}=1)-\int_0^t I(S_i \geqslant u)\,d\Lambda_z(u), \quad M_{yi}(t)=I(R_i \leqslant t, \eta_{yi}=1)-\int_0^t I(R_i \geqslant u)\,d\Lambda_y(u)$$

are martingales defined with respect to the appropriate filtrations, and $\Lambda_z(u)$ and $\Lambda_y(u)$ are the cumulative hazard functions for Z and Y, respectively. Applications of the functional and finite-dimensional delta methods show that $n^{\frac{1}{2}}\{\hat{F}_x(t)-F_x(t)\}$ is asymptotically equivalent to

$$\begin{aligned} J_x(t) = {} & g_1\{F_z(t), F_y(t), \theta_0\}[n^{\frac{1}{2}}\{\hat{F}_z(t)-F_z(t)\}]+g_2\{F_z(t), F_y(t), \theta_0\}[n^{\frac{1}{2}}\{\hat{F}_y(t)-F_y(t)\}] \\ & + g_3\{F_z(t), F_y(t), \theta_0\}\{n^{\frac{1}{2}}(\hat{\theta}-\theta_0)\}, \end{aligned} \tag{A4}$$

where

$$\begin{aligned} g_1(a,b,c) &= \partial g(a,b,c)/\partial a = a^{-c}(a^{1-c}-b^{1-c}+1)^{c/(1-c)}, \\ g_2(a,b,c) &= \partial g(a,b,c)/\partial b = -b^{-c}(a^{1-c}-b^{1-c}+1)^{c/(1-c)}, \end{aligned} \tag{A5}$$

$$\begin{aligned} g_3(a,b,c) &= \frac{\partial g(a,b,c)}{\partial c} \\ &= g(a,b,c)\left\{\frac{\log(a^{1-c}-b^{1-c}+1)}{(1-c)^2}+\frac{-a^{1-c}\log(a)+b^{1-c}\log(b)}{(a^{1-c}-b^{1-c}+1)(1-c)}\right\}. \end{aligned} \tag{A6}$$

Manipulating the results for $n^{\frac{1}{2}}(\hat{\theta}-\theta_0)$ earlier in this Appendix and those for $\hat{F}_z$ and $\hat{F}_y$ in (A1) and (A2), we obtain $J_x(t)=n^{-3/2}\sum_{i<j}V_{ij}(t)+o_p(1)$, where

$$\begin{aligned} V_{ij}(t) = {} & -g_1\{F_z(t), F_y(t), \theta_0\}F_z(t)\int_0^t \pi_z(u)^{-1}\{dM_{zi}(u)+dM_{zj}(u)\} \\ & - g_2\{F_z(t), F_y(t), \theta_0\}F_y(t)\int_0^t \pi_y(u)^{-1}\{dM_{yi}(u)+dM_{yj}(u)\}+g_3\{F_x(t), F_y(t), \theta_0\}I^{-1}Q_{ij}. \end{aligned}$$

A multivariate central limit theorem for U-statistics (Wei & Johnson, 1985) gives joint asymptotic normality of $\{J_x(t_1), \ldots, J_x(t_k)\}$ for any finite collection of times $(t_1, \ldots, t_k)$. Since the random quantities in $n^{\frac{1}{2}}(\hat{\theta}-\theta_0)$ are time-independent, the third term in (A4) is naturally tight. The asymptotic equivalents for the Kaplan–Meier estimators in the first and second terms of (A4) are martingale integrals, and hence tight. Thus, $J_x(t)$ is a sum of tight processes and is also tight. Finite-dimensional convergence plus tightness give weak convergence. The covariate function,

$$\begin{aligned} \sigma(s,t) &= E\{J_x(s)J_x(t)\} \\ &= \lim_{n\to\infty} n^{-3}\sum_{k<l<m}\{V_{kl}(s)V_{km}(t)+V_{lm}(s)V_{km}(t)+V_{kl}(s)V_{lm}(t) \\ &\quad + V_{lm}(s)V_{kl}(t)+V_{km}(s)V_{kl}(t)+V_{km}(s)V_{lm}(t)\}, \end{aligned} \tag{A7}$$

is consistently estimated by replacing theoretical quantities with empirical versions.

References

Clayton, D. G. (1978). A model for association in bivariate life tables and its application to epidemiological studies of familial tendency in chronic disease epidemiology. *Biometrika* **65**, 141–51.

COPELAN, E. A., BIGGS, J. C., THOMPSON, J. M., CRILLEY, P., SZER, J., KLEIN, J. P., KAPOOR, N., AVALOS, B. R., CUNNINGHAM, I., ATKINSON, K., DOWNS, K., HARMON, G. S., DALY, M. B., BRODSKY, I., BULOVA, S. I. & TUTSCHKA, P. J. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with Bu/Cy2. *Blood* **78**, 838–43.

DAY, R., BRYANT, J. & LEFKOPOULOU, M. (1997). Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic factors. *Biometrika* **84**, 45–56.

EMOTO, S. E. & MATTHEWS, P. C. (1990). A Weibull model for dependent censoring. *Ann. Statist.* **18**, 1556–77.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

GENEST, C., GHOUDI, K. & RIVEST, L.-P. (1995). A semiparametric estimating procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–52.

GILL, R. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 121. Amsterdam: Mathematisch Centrum.

HOUGAARD, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73**, 671–8.

KLEIN, J. P. & MOESCHBERGER, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

LIN, D. Y. & YING, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573–82.

MOESCHBERGER, M. (1974). Life tests under dependent causes of failure. *Technometrics* **16**, 39–47.

OAKES, D. (1982). A model for association in bivariate survival data. *J. R. Statist. Soc.* B **44**, 414–22.

OAKES, D. (1986). Semiparametric inference in bivariate survival data. *Biometrika* **73**, 353–61.

OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.

PETERSON, A. V. (1976). Bounds for a joint distribution with subdistribution functions: application to competing risks. *Proc. Nat. Acad. Sci.* **73**, 11–3.

SHIH, J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika* **85**, 189–200.

SHIH, J. H. & LOUIS, T. A. (1995). Inferences on the association parameter in copula methods for bivariate survival data. *Biometrics* **51**, 1384–99.

SLUD, E. V. & RUBINSTEIN, L. V. (1983). Dependent competing risks and summary survival curves. *Biometrika* **70**, 643–9.

TSIATIS, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci.* **72**, 20–2.

WEI, L. J. & JOHNSON, W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika* **72**, 359–64.

ZHENG, M. & KLEIN, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–38.

[*Received February* 2000. *Revised February* 2001]