STATISTICS IN MEDICINE
*Statist. Med.* (in press)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2327

# Semiparametric inferences for association with semi-competing risks data

Debashis Ghosh*,†

*Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48105, U.S.A.*

SUMMARY

In many biomedical studies, it is of interest to assess dependence between bivariate failure time data. We focus here on a special type of such data, referred to as semi-competing risks data. In this article, we develop methods for making inferences regarding dependence of semi-competing risks data across strata of a discrete covariate $Z$. A class of rank statistics for testing constancy of association across strata are proposed; its asymptotic properties are also derived. We develop a novel resampling-based technique for calculating the variances of the proposed test statistics. In addition, we develop methods for combining test statistics for assessing marginal effects of $Z$ on the dependent censoring variable as well as its effects on association. The finite-sample properties of the proposed methodology are assessed using simulation studies, and they are applied to data from a leukaemia transplantation study. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: copula; correlated failure time; dependent censoring; Kendall's tau; multivariate survival analysis

## 1. INTRODUCTION

Recently, there has been much attention in the medical and scientific literature devoted to the analysis of multiple event data. Such data are generally of two forms: recurrent failure time data, in which subjects experience repeated episodes of the same events during the course of the study, and ordered failure time data, in which subjects experience a progression of events of different types that signify a deterioration in health status. Our focus will be on consideration of the latter data.

We consider a study, reported in Reference [1], of leukaemia patients receiving bone marrow transplants. As described in Reference [2], the outcome of such a procedure is quite

*Correspondence to: Debashis Ghosh, Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Room M4057, Ann Arbor, MI 48109-2029, U.S.A.
†E-mail: ghoshd@umich.edu

Contract/grant sponsor: National Institutes of Health; contract/grant number: 5 P30 CA46592

*Received 29 October 2004*
*Accepted 6 June 2005*

complex. If the transplantation is successful and there are no harmful side effects, then the patient will recover. If it is not, then the patient will relapse. However, potential intermediate events that occur after transplantation, such as development of acute or chronic graft *versus* host disease, development of infections, and return of the platelet count to normal levels, can alter the risk of survival. In Reference [1], one of the associations studied by the authors was between time to platelet recovery (platelet count $\geqslant 40 \times 10^9/l$) and survival. This association was studied in the entire study population as well as in a subgroup of patients with low risk AML. While hazard ratios were given for these two populations, it is of interest to the investigators in Reference [1] to see if the association between platelet recovery and survival might differ by disease type (low risk AML, high risk AML and ALL).

If we let $T$ and $D$ denote the times to platelet recovery and death, then it is obvious that $D$ can potentially censor $T$ but not *vice versa*. After transplantation, patients can potentially have platelet recovery and not die, have platelet recovery and then die, or die without normal platelet recovery. This type of data structure has been termed semi-competing risks data [3]. This is contrast to classical competing risks data, in which the minimum of $T$ and $D$ is observed [4]. There exist several works on the analysis of semi-competing risks data. Methods for estimation of treatment effect in such a setup have been proposed [5, 6]; they require that the distribution of $(T, D)$, conditional on covariates, satisfies a bivariate location-shift or accelerated failure time model [7]. Other authors [3, 8, 9] develop testing and estimation procedures for the dependence between $T$ and $D$. As discussed by these authors, because of the data structure, the bivariate distribution of $(T, D)$ is identifiable on the region where $T \leqslant D$. By contrast, the bivariate distribution is not identifiable with competing risks data [4]. It is thus possible to assess dependence with semi-competing risks data.

An area discussed in Reference [9] that has not been currently addressed in the literature is incorporation of covariates into consideration of dependence with semi-competing risks data. For example, in References [3, 9], the dependence between relapse and survival is assessed in the entire study population. However, no adjustment for covariates such as sex or treatment group is made in the analysis. From the transplantation example, we care about the effect of $T$ on $D$ in the overall study as well as in the subpopulations defined by leukaemia type. In this paper, we develop methods for inference regarding dependence between bivariate failure times. Our focus is primarily on testing. We develop a class of score statistics for testing the null hypothesis of constant dependence across levels of a discrete-valued covariate. It is based on the Clayton–Oakes frailty model considered for semi-competing risks data by previous authors [3, 8, 9]. In Section 2, we define the data structures and models used to derive the test statistics. The proposed statistics and their asymptotic properties are studied in Section 3 with a review of existing methods in Section 3.1. The use of novel resampling techniques for variance estimation is described here as well. We also discuss the issue of combining statistics in order to gain power. The finite-sample behaviour of the proposed methods are assessed by simulation studies and applications to real data in Section 4. Finally, we conclude with some discussion in Section 5.

## 2. DATA AND MODELS

We start by making the following definitions. Let $a \wedge b$ denote the minimum of two numbers $a$ and $b$. Define $I(A)$ to be the indicator function for the event $A$. Let $T$ be the failure time of interest, $D$ time to dependent censoring and $C$ time to independent censoring. Let

 

$Z$ be a covariate that takes discrete values $(1,2,\ldots,K)$, where $K>1$. We assume that $(T,D)$ is independent of $C$ given $Z$; however, $T$ and $D$ may be dependent, as specified below. We observe the data $(X_i,\delta_i^X,Y_i,\delta_i^Y,Z_i)$, $i=1,\ldots,n$, $n$ independent and identically distributed observations from $(X,\delta^X,Y,\delta^Y,Z)$, where $X=T\wedge D\wedge C$, $\delta^X=I(T\leqslant D\wedge C)$, $Y=D\wedge C$ and $\delta^Y=I(D\leqslant C)$.

In the setup we are considering, $T$ cannot be greater than $D$; this implies that the joint distribution of $(T,D)$ on the region where $T>D$ is not identifiable based on the observed data. However, it is possible to identify the joint distribution on the wedge $(T\leqslant D)$ [3].

We formulate a constant cross-ratio [10] function model on the upper wedge $T\leqslant D$. The cross-ratio function is defined by

$$\theta(s,t)=\frac{\lambda_T(s|D=t)}{\lambda_T(s|D\geqslant t)} \tag{1}$$

where $\lambda_T(t|A)=\lim_{\Delta t\to 0}\,\mathrm{d}/\mathrm{d}t\ \Pr(T<t+\Delta t|T\geqslant t,D\in A)$, and $A$ is a subset of the interval $(0,\infty)$. If $\theta(s,t)$ in (1) is constant for all $s$ and $t$, i.e.

$$\theta(s,t)=\theta \tag{2}$$

This assumption is identical to the cross-ratio function induced by the gamma frailty model of Clayton and Oakes [11–13]. If $\theta(s,t)=1$, then this implies independence of $T$ and $D$ on the upper wedge. Note that independence here really refers to quasi-independence [14]. What this means is that while the joint distribution is defined on $T\leqslant D$, if we were to consider any subrectangular portion within the region $T\leqslant D$, then $T$ and $D$ are independent on that subrectangle.

Observe that the cross-ratio function model is being formulated only for the upper wedge of the joint distribution of $(T,D)$. This is due to the fact that the bivariate distribution of $(T,D)$ is not identified on the lower wedge. With semi-competing risks data, the cross-ratio function has been referred to as the predictive hazard ratio [8]. They considered this model and proposed a test of the independence of $T$ and $D$. In Reference [3], a closed form estimator of $\theta$ using modified weighted concordance estimating functions [12, 13] along with an asymptotic variance estimator was developed.

A more flexible formulation for the constant cross-ratio model occurs through the use of copulas [15]. The copula corresponding to the Clayton–Oakes model allows for negative correlation, although the distribution is only absolutely continuous (i.e. has a density) when $\theta>-0.5$. The copula approach was adopted in Reference [9], where two estimating procedures, one based on estimating functions and the other based on a two-stage likelihood approach, were proposed.

In the presence of the discrete covariate $Z$, the natural extension of the model described previously is $\theta(s,t|Z=z)=\theta_z$, where $\theta(s,t|Z=z)=\lambda_T(s|D=t,Z=z)/\lambda_T(s|D\geqslant t,Z=z)$ and

$$\lambda_T(t|A,Z=z)=\lim_{\Delta t\to 0}\frac{\mathrm{d}}{\mathrm{d}t}\Pr(T<t+\Delta t|T\geqslant t,D\in A,Z=z) \tag{3}$$

with $A$ an interval in $(0,\infty)$. Thus, a Clayton–Oakes frailty model is assumed for each stratum defined by $Z$. Our interest is in testing the null hypothesis that the predictive hazard ratio does not depend on $Z$, i.e. $H_0:\theta_z=\theta$. This hypothesis corresponds to no interaction effect between $Z$ and the association parameter for the joint distribution of $(T,D)$ on the upper wedge.

 

## 3. TESTING AND ESTIMATION PROCEDURES

### 3.1. Association between $T$ and $D$ for $K=1$

To motivate the proposed testing procedures, let us first study the case where $K=1$. For $i=1,\ldots,n$ and $j=1,\ldots,n$, define $\tilde{X}_{ij}=X_i\wedge X_j$, $\tilde{Y}_{ij}=Y_i\wedge Y_j$, $\tilde{C}_{ij}=C_i\wedge C_j$ and $D_{ij}=I(\tilde{X}_{ij}<\tilde{Y}_{ij}<\tilde{C}_{ij})$. We start with model (2). In Reference [3], the following class of estimating functions were used to estimate $\theta$:

$$U_1(\theta)=\sum_{i<j} W(\tilde{X}_{ij},\tilde{Y}_{ij})D_{ij}\{\Delta_{ij}-\theta/(1+\theta)\} \tag{4}$$

where $W(u,v)$ is a weight function that converges uniformly to $w(u,v)$, a bounded deterministic function, and $\Delta_{ij}=I\{(X_i-X_j)(Y_i-Y_j)>0\}$, $i,j=1,\ldots,n$. Note that the estimating function being constructed is based on the indicator that the $i$th and $j$th pairs of observations are concordant; in the absence of censoring, this is the same as the estimating function used to estimate Kendall's $\tau$. If one sets $U_1(\theta)$ in (4) equal to zero, then one obtains the following closed-form estimator for $\theta$:

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

In Reference [3], it is shown that the asymptotic distribution of $n^{-3/2}U(\theta)$ is normal with mean 0 and variance

$$I(\theta)=\lim_{n\to\infty} 2n^{-3}\sum_{k<l<m} Q_{kl}Q_{km}+Q_{kl}Q_{lm}+Q_{km}Q_{lm}$$

where $Q_{ij}=W(\tilde{X}_{ij},\tilde{Y}_{ij})D_{ij}\{\Delta_{ij}-\theta/(1+\theta)\}$, $i,j=1,\ldots,n$. A simple estimator of $I\equiv I(\theta)$, $\hat{I}$, can be found by plugging in $\hat{\theta}$ into $Q_{ij}$ and taking empirical averages. A test of no association between $T$ and $D$ on the upper wedge can be constructed using $n^{-3/2}U(1)/\hat{I}^{-1/2}$. By standard Taylor series arguments, $n^{1/2}(\hat{\theta}-\theta)$ converges in distribution to a normal random variable with mean zero and variance $I/J(\theta)^2$, where

$$J(\theta)=\lim_{n\to\infty} n^{-2}\sum_{i<j} W(\tilde{X}_{ij},\tilde{Y}_{ij})D_{ij}(1+\theta)^{-2}$$

A consistent estimator for $J(\theta)$ can be found by plugging in empirical quantities and $\hat{\theta}$ for $\theta$.

Note that the estimation procedure for $\theta$ proposed here is semi-parametric in that the marginal distributions of $T$ and $D$ are not needed for specification.

### 3.2. Association between $T$ and $D$ for $K>1$

We now consider the case where $K>1$. Using the gamma frailty model, we wish to test the null hypothesis that $H_0$ : $\theta(s,t|Z)=\theta$. We propose using the following class of test

 

statistics:

$$T_1 = \sum_{k=1}^{K} \sum_{i<j} W_z(\tilde{X}_{ij}, \tilde{Y}_{ij})(D_{ijz} - D_{ij})\Delta_{ij} \tag{5}$$

where $D_{ijz} = I(\tilde{X}_{ij} < \tilde{Y}_{ij} < \tilde{C}_{ij}, Z_i = Z_j = z)$, and $W_k$ is a weight function similar to that described in Section 3.1. In Appendix A, we prove that under the null hypothesis (i.e. the association between $T$ and $D$ is constant for all values of $Z$), $n^{-3/2}T_1$ has a limiting normal distribution with mean zero. The formula for its variance can also be found there. The class of statistics $T_1$ represents differences between stratified and unstratified analyses; under the null hypothesis of no interaction between $Z$ and the association parameter, they should yield consistent results. A similar idea is exploited in the construction of $k$-sample log-rank statistics for univariate survival data [16].

As can be seen in Appendix A, the variance for the limiting distributions of these random variables is fairly complicated. In other multivariate survival contexts, various authors [9, 17] have advocated using the bootstrap and jackknife for variance estimation. Here, we will use a variation of a resampling method proposed in Reference [18] for estimating the variance of the test statistic. Note that (5) has the following form:

$$T_1 = \sum_{k=1}^{K} \sum_{i<j} T_{ijk} \tag{6}$$

where $T_{ijk} = W_k(\tilde{X}_{ij}, \tilde{Y}_{ij})(D_{ijz} - D_{ij})\Delta_{ij}$. The statistic can be approximated by a $U$-process of order two [19]. To apply the method of [18], we generate $n$ N(0, 1) random variables $(G_1, \ldots, G_n)$ and calculate perturbations of (6):

$$T_1^* = \sum_{k=1}^{K} \sum_{i<j} T_{ijk} G_i G_j \tag{7}$$

Notice that in (7), the only stochastic components are $(G_1, \ldots, G_n)$. We choose normal random variables for resampling in (7), but we can use any random sample $(G_1, \ldots, G_n)$ such that $E(G_1) = 0$ and $\mathrm{Var}(G_1) = 1$. By arguments similar to those given in Reference [18], under the null hypothesis, $n^{-3/2}T_1^*$ and $n^{-3/2}T_1$ have the same limiting distribution. This leads to the following algorithm for calculating the variance of the test statistics $T_1$:

1. Generate $n$ iid N(0, 1) random variables $(G_1, \ldots, G_n)$ and calculate $T_1^*$.
2. Repeat step 1 $M$ times.

This resampling procedure is quite fast. In practice, we usually take $M = 1000$. We can then estimate a 95 per cent confidence interval of $T_1$ in one of two ways. The first is to calculate a standard error based on the empirical distribution of $T_1^*$. A second way is to take the 2.5th and 97.5th percentiles of the empirical distribution of $T_1^*$.

The choice of weight function $W_k(u, v)$ for $T_1$ will depend on the class of local alternatives that are under consideration. In Section 4, we study the finite-sample properties of several weight functions for the class of statistics defined by $T_1$.

It should also be noted that we are performing the testing within the gamma frailty model. The alternative hypothesis is that the cross-ratios are not equal across strata defined by $Z$. In the simulation studies in Section 4.1, we study the robustness of the testing procedure when the gamma frailty model is not correct.

 

### 3.3. Combining test statistics

A unique feature of semi-competing risks data relative to classical competing risks data is that the marginal distribution of $D$ is identifiable. Assessing the effects of leukaemia type on the transplantation process can be reformulated in one manner as a multiple endpoints problem in which the effect of $Z$ can affect various aspects of the transplantation process. One could imagine that there might be covariate effects of $Z$ on $D$ as well as that on dependence between $T$ and $D$. Denote the hypotheses being tested as $H_0^D$ and $H_0^A$, respectively. One could imagine combining inference about $D$ between strata (e.g. using a $k$-sample log-rank test) with the methods proposed in this paper. A simple method for doing this is a sequential method of Holm [20]. Let $p_1$ and $p_2$ denote the $p$-values from the $k$-sample log-rank statistic and one of the tests proposed in the previous section. Let the ordered $p$-values be denoted as $p_{(1)} \leqslant p_{(2)}$ and overall type I error level be $\alpha$. The procedure is as follows:

1. If $p_{(1)} \leqslant \alpha/2$, then continue to step 2. Otherwise, fail to reject $H_0^D$ and $H_0^A$.
2. If $p_{(2)} \leqslant \alpha$, then reject $H_0^D$ and $H_0^A$. Otherwise, reject the hypothesis corresponding to $p_{(1)}$ but fail to reject the hypothesis corresponding to $p_{(2)}$.

It can be shown that the overall type I error rate for this multiple testing procedure is $\alpha$. An advantage of this testing procedure is that it allows the investigator to determine what the separate effects of $Z$ on $D$ and the association between $T$ and $D$ and their statistical significances are while at the same time allowing for a single probability statement regarding the effects of $Z$ on $D$ and on the association between $T$ and $D$.

Because the marginal distribution of $T$ is not identifiable based on the observable data, how to combine information on $T$ with the procedures here is not as clear. In the work of Reference [21], they focus on the estimation of the distribution of $T$ in this setup. It may be possible to combine ideas from that work with those proposed here.

## 4. NUMERICAL EXAMPLES

### 4.1. Simulation studies

To assess the finite-sample properties of the proposed methods, a series of simulation studies was conducted. In the first set of simulation studies, we wanted to determine the accuracy of variances using the resampling-based method relative to the model-based variance. We studied the behaviour of (5) with $W_z = 1$. We considered $K = 2$ and generated data from a bivariate Clayton model under two scenarios. In the first, $\theta_1 = \theta_2 = 2$, while in the second, $\theta_1 = 2$ and $\theta_2 = 4$. The marginal distributions for $T$ and $D$ were exponential with means 2 and 6. An independent $U(0, 3)$ random variable was generated for censoring; this yielded 10 per cent censoring for $T$ in the first scenario and 25 per cent censoring in the second. We used the model-based variance estimator derived in Appendix A and that for the variance based on the empirical distribution of $T^*$. We consider sample sizes $n = 50, 100$ and 150 and assumed equal sample size in each stratum. For each simulation setting, 2000 samples were generated, and 1000 resamplings were performed within each simulation sample. The results are presented in Table I. Based on the estimated variances, it seems as if the resampling-based variance estimator tends to overestimate the true variance in smaller samples. This bias diminishes in larger samples. The estimators also appear to be fairly concordant.

 

Table I. Comparison of resampling-based and model-based variance estimators.

| | $\theta_1 = \theta_2 = 2$ | | | $\theta_1 = 2,\ \theta_2 = 4$ | | |
|---|---|---|---|---|---|---|
| $n$ | MOD | RES | Bias | MOD | RES | Bias |
| 50 | 0.25 | 0.23 | 0.02 | 0.32 | 0.29 | 0.04 |
| 100 | 0.19 | 0.18 | 0.01 | 0.25 | 0.23 | 0.03 |
| 150 | 0.16 | 0.16 | 0.01 | 0.20 | 0.20 | 0.01 |

MOD represents the estimator of the variance of (5) using model-based estimator from Appendix A, averaged over 1000 simulations; RES represents the estimator of variance using empirical distribution of $T_1^*$, averaged over 1000 simulations. Bias is the mean absolute deviation between the two estimators, averaged over 1000 simulations.

Having determined from the previous set of studies that the resampling-based method is accurate, we next compared the testing procedures in Section 3.2 in terms of size and power. We consider three weight functions, which are extensions of those considered in Reference [9]:

$$W_{1k}(s,t) = 1$$

$$W_{2k}(s,t) = \sum_{i=1}^{n} I(X_i > s, Y_i > t, Z_i = k) \Big/ \sum_{i=1}^{n} I(X_i > s, Y_i > t)$$

and

$$W_{3k}(s,t) = \sum_{i=1}^{n} I(X_i > s, Y_i > t) \Big/ \sum_{i=1}^{n} I(X_i > s, Y_i > t, Z_i = k)$$

The resampling-based method for variance estimation was utilized in the simulation studies. Again, sample sizes $n = 50, 100$ and 150 were used. The same numbers of simulation samples and resamplings were used as in the previous set of simulations. We set $K = 2$ and generated data from two models. The first is the bivariate Clayton model. For calculations of size, we took $\theta_1 = \theta_2 = 2$, while for power, we set $\theta = 2$ and $\theta_2 = 4$. The second model is a bivariate normal distribution for $(\log D, \log T)$, given $Z = z$ with mean zero vector and covariance matrix

$$\begin{pmatrix} 1 & \rho_z \\ \rho_z & 1 \end{pmatrix}$$

For calculations of size, we set $\rho_1 = \rho_2 = 0.1$, while for power, we take $\rho_1 = 0.1$ and $\rho_2 = 0.5$. Even though the bivariate normal model does not have a constant cross-ratio function, we wanted to determine if there was any robustness of the proposed testing procedure. The simulation results are summarized in Tables II and III. We find that the tests tend to be very slightly anticonservative in smaller samples, although this behaviour diminishes in larger samples. The percentile and empirical standard errors tend to perform quite similarly here. In terms of the weight function, weight $W_1$ tends to yield higher power. The robustness of the

 

Table II. Empirical sizes of test statistics.

| | | $\theta_1 = \theta_2 = 2$ | | | $\rho_1 = \rho_2 = 0.1$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | Method | $W_1$ | $W_2$ | $W_3$ | $W_1$ | $W_2$ | $W_3$ |
| 50 | SE | 0.052 | 0.054 | 0.053 | 0.055 | 0.054 | 0.052 |
| | Percentile | 0.053 | 0.051 | 0.052 | 0.052 | 0.053 | 0.051 |
| 100 | SE | 0.049 | 0.048 | 0.050 | 0.049 | 0.048 | 0.049 |
| | Percentile | 0.048 | 0.048 | 0.048 | 0.049 | 0.049 | 0.048 |
| 150 | SE | 0.048 | 0.049 | 0.048 | 0.048 | 0.049 | 0.048 |
| | Percentile | 0.048 | 0.048 | 0.049 | 0.049 | 0.047 | 0.049 |

Table III. Empirical powers of test statistics.

| | | $\theta_1$=2, $\theta_2$=4 | | | $\rho_1 = 0.1$, $\rho_2 = 0.5$ | | |
|---|---|---|---|---|---|---|---|
| $n$ | | $W_1$ | $W_2$ | $W_3$ | $W_1$ | $W_2$ | $W_3$ |
| 50 | SE | 0.42 | 0.38 | 0.33 | 0.41 | 0.36 | 0.32 |
| | Percentile | 0.45 | 0.40 | 0.35 | 0.42 | 0.37 | 0.32 |
| 100 | SE | 0.73 | 0.65 | 0.60 | 0.69 | 0.66 | 0.63 |
| | Percentile | 0.70 | 0.62 | 0.59 | 0.67 | 0.64 | 0.60 |
| 300 | SE | 0.96 | 0.93 | 0.91 | 0.92 | 0.89 | 0.85 |
| | Percentile | 0.96 | 0.93 | 0.91 | 0.92 | 0.88 | 0.85 |

procedure seems to be adequate based on the bivariate normal distribution results for size. For power, it is more difficult to interpret results because there is a violation of the gamma frailty model.

### *4.2. Transplantation data*

We now return to the transplantation study discussed in Section 1. In this multicentre clinical trial, patients with acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL) underwent bone marrow transplantation and were followed prospectively. As noted in Klein and Moeschberger [2], the recovery from transplantation is quite complex. These data have been analysed previously in terms of joint estimation of the two failure time endpoints [3] as well as the dependence between disease-free survival and survival for the entire population using the gamma frailty approach [9]. Here, we choose to focus on the latter goal of assessing dependence between time to platelet recovery and time to death. We wish to determine if the dependence between platelet recovery and survival varies by disease type (ALL, AML low risk, AML high risk).

We first start by considering the gamma frailty model for semi-competing risks data with respect to time to platelet recovery and death. The estimate of $\theta$, ignoring disease type, is 1.23. We computed the estimator of $\theta$ in the three subgroups using an unweighted estimator; these

 

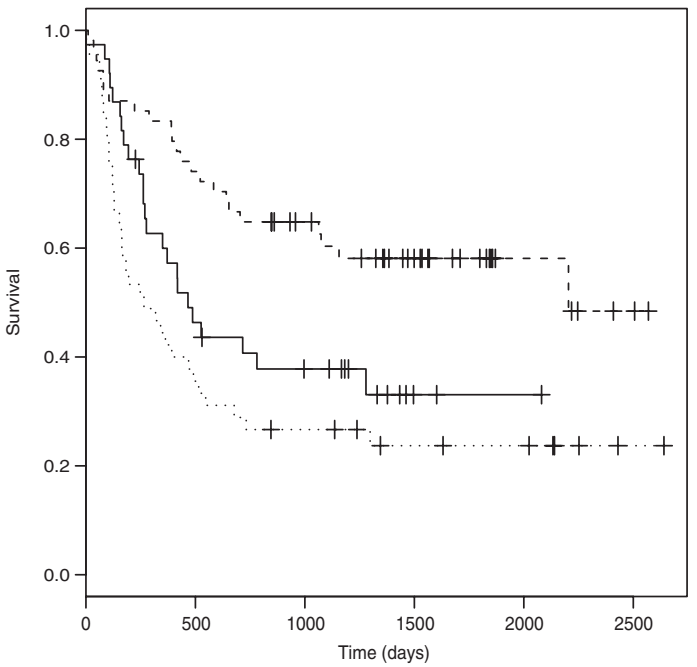


Figure 1. Overall survival by leukaemia group: solid line = ALL, dashed line = AML low risk, dotted line = AML high risk.

are 0.60, 2.37 and 1.15 for the ALL, AML low risk and AML high risk groups, respectively. Calculating (5) with the three choices of weight functions described in the previous section, we find strong evidence for an interaction between the dependence parameter and risk group. The $p$-value was less than 0.0001 for all three weight functions.

Next, we wanted to determine the effect and associated significance of leukaemia type on two aspects of the post-transplantation process; that on survival and that on the dependence between platelet recovery and survival. To adjust for the multiple testing issue, the sequential method from Section 3.3 is used. The survival curves are plotted in Figure 1. The survival distribution is significantly different between the three groups ($p$-value from log-rank test = 0.0004). Application of the sequential testing procedure yields a significant difference in both survival and association between platelet recovery and survival between the three groups.

## 5. DISCUSSION

The gamma frailty model has been proposed recently in the literature for the analysis of semi-competing risks data [3, 8, 9]. There have been two goals in the use of this model. One is the estimation of the joint distribution of the time to event and time to dependent censoring; this was studied primarily in References [3, 9]. If the dependent censoring is death, as is the case in the transplantation study, then the interpretation of the joint distribution is quite controversial. The second use of the model, addressed in Reference [8], is to provide a dependence measure

 

between the two failure times. It is this use of the model that we seek to extend in this paper. While the question of dependence can also be addressed using time-dependent covariates in a proportional hazards model, it can be problematic to interpret for practitioners. The cross-ratio interpretation of the gamma frailty model potentially has more appeal.

While there has been recent work in the area of studying dependence with semi-competing risks data, regression generalizations have not appeared yet. It seems quite plausible that the dependence between two failure times might depend on covariates or might have an interaction with a covariate. In this paper, we have proposed a testing procedure for association in semi-competing risks. It represents a generalization of the work of previous authors [3, 8, 9]. The procedures proposed in the paper test for an interaction effect between the dependence parameter with a discrete covariate.

It should be noted that the proposed sequential procedure in Section 3.3 cannot detect the situation when $H_0^D$ is accepted, but $H_0^A$ is rejected. This may happen when $Z$ shortens $T$ but prolongs $T \wedge D$ given that $T < D$. Also, more aggressive treatment may be applied to the subjects under study once their events of $T$ occur which would make the comparison based on $D$ implausible. Alternatively, one may test the null hypothesis using $T \wedge D$ first and then test $H_0^A$ next. Note that the failure time $T \wedge D$ is also an identifiable quantity.

The test statistics that have been constructed are based on the gamma frailty model and thus strictly speaking can only be used for testing the null hypothesis of constant cross-ratio across strata *versus* the alternative hypothesis of non-constant cross-ratio across strata. More formally, we could consider procedures in which we first assess goodness of fit for the Clayton–Oakes model and then perform the test proposed here. One sequential testing procedure would be the following:

1. If the goodness of fit method of Reference [3] fails to reject the null hypothesis that the Clayton–Oakes model holds, proceed to step 2. Otherwise, stop.
2. Perform the test proposed in the paper.

How to combine these testing procedures optimally remains an open question.

The issue of interaction between dependence and disease type has been considered here. However, it might also be the case that disease type is confounding the dependence between the failure times. Thus, regression modelling procedures would be required.

## APPENDIX A: ASYMPTOTIC NORMALITY OF $n^{-3/2}T_1$

Define the following estimating functions:

$$U_1(\theta) = \sum_{i<j} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}) D_{ij}\{\Delta_{ij} - \theta/(1+\theta)\}$$

and

$$U_{1k}(\theta_k) = \sum_{i<j} W_k(\tilde{X}_{ij}, \tilde{Y}_{ij}) I(Z_i = Z_j = k) D_{ij}\{\Delta_{ij} - \theta_k/(1+\theta_k)\}$$

 

$k=1,\ldots,K$. Let us consider the vector $n^{-3/2}\{U_{11}(\theta_1),\ldots,U_{1K}(\theta_K),U(\theta)\}$. By multivariate extensions of usual results from the theory of $U$-statistics [22, Chapter 12], this vector converges in distribution to a normal random vector with mean zero vector and $(K+1)\times(K+1)$ variance–covariance matrix $\Sigma$ with element

$$\sigma_{ij}=\lim_{n\to\infty} 2n^{-3}\sum_{h<l<m}\delta_{hli}\delta_{hmj}Q_{hl}Q_{hm}+\delta_{hli}\delta_{lmj}Q_{hl}Q_{lm}+\delta_{lmi}\delta_{hmj}Q_{lm}Q_{hm}$$

where $\delta_{ijk}$ is defined by

$$\delta_{ijk}=\begin{cases} I(Z_i=Z_j=k), & 1\leqslant i,j\leqslant K \\ 1, & i=K+1 \text{ or } j=K+1 \end{cases}$$

Under $H_0:\theta_1=\cdots=\theta_K=\theta$, we have by addition and subtraction that

$$\begin{aligned} T_1 &= \sum_{k=1}^{K}\sum_{i<j} W_k(\tilde{X}_{ij},\tilde{Y}_{ij})I(Z_i=Z_j=k)D_{ij}\left\{\Delta_{ij}-\frac{\theta_k}{1+\theta_k}\right\} \\ &\quad -\sum_{k=1}^{K}\sum_{i<j} W_k(\tilde{X}_{ij},\tilde{Y}_{ij})D_{ij}\left\{\Delta_{ij}-\frac{\theta}{1+\theta}\right\} \\ &= \sum_{k=1}^{K}\sum_{i<j}\psi_{ijk} \end{aligned} \tag{A1}$$

where $\psi_{ijk}=W_k(\tilde{X}_{ij},\tilde{Y}_{ij})I(Z_i=Z_j=k)D_{ij}\{\Delta_{ij}-\theta_k/(1+\theta_k)\}-W_k(\tilde{X}_{ij},\tilde{Y}_{ij})D_{ij}\{\Delta_{ij}-\theta/(1+\theta)\}$, $i,j=1,\ldots,n$, $k=1,\ldots,K$. Note that $T_1=\mathbf{A}\{U_{11}(\theta_1),\ldots,U_{1K}(\theta_K),U(\theta)\}$, where $\mathbf{A}$ is the $1\times(K+1)$ vector $(1,\ldots,1,-K)$. Under $H_0$, vector $n^{-3/2}T$ converges in distribution to a normal random vector with mean zero and variance $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}}$.

ACKNOWLEDGEMENTS

The author would like to thank Dr Wei-Yann Tsai for useful discussions. This research is supported in part by the National Institutes of Health through the University of Michigan's Cancer Center Support Grant (5 P30 CA46592).

REFERENCES

1. Copelan EA, Biggs JC, Thompson JM, Crilley P, Szer J, Klein JP, Kapoor N, Avalso BR, Cunningham I, Atkinson K, Downs K, Harmon GS, Daly MB, Brodsky I, Bulova SI, Tutschka PJ. Treatment for acute myelocytic leukemia with allogenic bone marrow transplantation following preparation with Bu/Cy2. *Blood* 1991; **78**:838–843.
2. Klein JP, Moeschberger ML. *Survival Analysis*: *Techniques for Censored and Truncated Data*. Springer: New York, 1997.
3. Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika* 2001; **88**:907–919.
4. Tsiatis AA. A nonindentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the U.S.A.* 1975; **72**:20–22.

 

5. Lin DY, Robins JM, Wei LJ. Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* 1996; **83**:381–393.
6. Chang SH. A two-sample comparison for ordered event data. *Biometrics* 2000; **56**:183–189.
7. Cox DR, Oakes D. *Analysis of Survival Data*. Chapman & Hall: London, 1984.
8. Day R, Bryant J, Lefkopolou M. Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika* 1997; **84**:45–56.
9. Wang W. Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society*, *Series B* 2003; **65**:257–273.
10. Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 1989; **84**:487–493.
11. Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; **65**:141–151.
12. Oakes D. A model for association in bivariate survival data. *Journal of the Royal Statistical Society*, *Series B* 1982; **44**:414–422.
13. Oakes D. Semiparametric inference in a model for association in bivariate survival data. *Biometrika* 1986; **73**:353–361.
14. Tsai WY. Testing the assumption of independence of truncation time and failure time. *Biometrika* 1990; **77**:169–177.
15. Genest C, Rivest LP. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 1993; **88**:1034–1043.
16. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: New York, 1993.
17. Wang W, Wells MT. Model selection and semiparametric inference for bivariate failure-time data (with discussion). *Journal of the American Statistical Association* 2003; **95**:62–76.
18. Parzen M, Wei LJ, Ying Z. A resampling method based on pivotal functions. *Biometrika* 1994; **81**:341–350.
19. Arcones M, Giné E. Limit theorems for U-processes. *Annals of Probability* 1993; **21**:1494–1542.
20. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979; **6**:65–70.
21. Jiang H, Chappell R, Fine JP. Estimating the distribution of nonterminal event time in the presence of mortality or informative dropout. *Controlled Clinical Trials* 2003; **24**:135–146.
22. Van der Vaart A. *Asymptotic Statistics*. Cambridge University Press: Cambridge, 2000.

 