Two Sample Comparison based on Semi-Competing Risks Data

Jin-Jian Hsieh*, Weijing Wang* & A. Adam Ding†
u9226803@stat.nctu.edu.tw, wjwang@stat.nctu.edu.tw, ding@neu.edu
** Institute of Statistics, National Chiao-Tung University*
Hsin-Chu, Taiwan, R.O.C.
†Department of Math, Northeastern University
Boston, MA 02115, U.S.A.

Summary

Semi-competing risks data are commonly seen in biomedical applications. In this article, we consider the problem of two-sample comparison based on a non-terminal event, say disease progression, which is subject to censoring by a terminal event such as death. Existence of possible dependent censoring complicates the analysis. The proposed methodology is developed under two types of assumptions. First, separate copula models are assumed for the two groups and then a flexible model measuring the group difference is imposed on the progression time. A competing approach for estimating the group-difference parameter is the method proposed by Lin et al. (1996) which requires making additional marginal assumption on the terminal event and implicitly assumes that the dependence structures in the two groups are the same. Although here we put an extra copula assumption on the observable region, we also propose a model checking approach to assess its validity. Via simulations, we compare the two competing methods based on their finite-sample performances and robustness properties if the imposed assumptions are violated. The proposed method is applied to a data set of bone marrow transplant provided in Klein and Moeschberger (2003).

Key words : Dependent censoring; Multiple events data; Transformation model; Copula model; Model selection.

1 Introduction

Many medical studies involve analysis of multiple endpoints. Those events may be classified into two types, namely terminal versus non-terminal. Death is an example of terminal events in the sense that its occurrence precludes the development of others. Examples of non-terminal events, which are subject to censoring by any terminal events, include disease progression or recurrence. If the relationship between the two events are completely unspecified, the marginal distribution of the time to a non-terminal event is not identifiable due to possible dependent censoring.

Let X be the time to the non-terminal event of major interest, usually a status of disease progression, Y be the time to death and C be the time to the external censoring event. Observed variables consist of $\tilde{X} = X \wedge Y \wedge C$, $\tilde{Y} = Y \wedge C$, $\delta_x = I(X \leq Y \wedge C)$ and $\delta_y = I(Y \leq C)$. Such a data structure is called semi-competing risks data by Fine et al. (2001). There has been increasing research attention in developing statistical methods for analyzing semi-competing risks data. For example, investigation of the degree of association between the two events has been pursued by Day et al. (1997) and Fine et al. (2001) which assume the Clayton model and Wang (2003) under a general copula assumption.

In the article, we consider making two-sample comparison based on progression time X. Due to dependent censoring, the marginal distribution of X is not identifiable nonparametrically. The papers by Lin et al. (1996) and Chang (2000) modeled the marginal effects on both X and Y but did not specify the joint distribution. Specifically Lin et al. (1996) considered a bivariate location-shift model and Chang (2000) assumed a bivariate accelerated failure time model. This research direction has been further extended to general regression settings in which the non-terminal event is generalized to be recurrence events (Ghosh and Lin, 2003; Lin and Ying, 2003) while death still serves as terminal event. The technique of artificial censoring is used in these papers to handle the problem of dependent censoring. Despite theoretically appealing, efficiency of the resulting estimator is affected by the degree of artificial censoring. Furthermore, these methods implicitly assume that the dependence structures for the two groups, or for all the levels of covariates, are the same. In other words, they do not account for the situation that covariates may affect the dependence structure.

In this article, we adopt a different approach to assess the group effect on progression time under dependent censoring. Without making marginal assumptions on Y, we assume that

$$h(X) = -\theta Z + \varepsilon, \tag{1}$$

where Z is the group indicator taking value of 1 and 0, $h(t)$ is a monotone function and ε is the error term. The parameter θ which measures the group difference for X is of major interest. Model (1) can be classified into two classes. One class assumes that $h(t)$ is a known monotone function but leaves the distribution of ε to be unknown. For example, when $h(t) = t$, the model becomes a location-shift model; when $h(t) = \log(t)$, the model follows an accelerated failure time model. The other class assumes that $h(t)$ is unknown but the distribution of ε is completely specified. Examples of the second class include the Cox proportional hazard model with ε being the extreme value distribution, and the proportional odds model with ε being the standard logistic distribution.

To handle the problem of non-identifiability, we assume that (X, Y) jointly follow an Archimedean copula model in the upper wedge $\mathcal{P} = \{(x, y) : 0 < x \leq y < \infty\}$. Accounting for the possibility that the dependence structures of the two groups are different, we assume separate Archimedean copula (AC) models for the two groups such that

$$\begin{aligned} F_j(x, y) &= C_{j,\alpha_j}\{F_{x,j}(x), F_{y,j}(y)\} \\ &= \phi_{j,\alpha_j}^{-1}\{\phi_{j,\alpha_j}[F_{x,j}(x)] + \phi_{j,\alpha_j}[F_{y,j}(y)]\}, \end{aligned} \tag{2}$$

where $\phi_{j,\alpha_j}(\cdot) : [0, 1] \mapsto [0, \infty]$, $F_j(x, y) = \Pr(X \geq x, Y \geq y|Z = j)$, $F_{x,j}(x) = \Pr(X \geq x|Z = j)$ and $F_{y,j}(y) = \Pr(Y \geq y|Z = j)$. Note that not only we allow different association parameters α_0 and α_1, but also we allow $\phi_{0,\alpha_0}(\cdot)$ and $\phi_{1,\alpha_1}(\cdot)$ to be of different forms. Examples of the AC models include the Clayton model (1978): $\phi_\alpha(v) = (v^{-\alpha} - 1)/\alpha$ $(\alpha > 0)$; the Frank model (1979): $\phi_\alpha(v) = \log[(1 - \alpha)/(1 - \alpha^v)]$ $(\alpha > 0)$; the Gumbel model (1960): $\phi_\alpha(v) = \{-\log(v)\}^{\alpha+1}$ $(\alpha > 0)$ and the Log-Copula model: $\phi_\alpha(v) = \{1 - \log(v)/\alpha\gamma\}^{\alpha+1} - 1$ $(\alpha, \gamma > 0)$.

The proposed inference method for estimating θ under model (1) and (2) is discussed in Section 2. In Section 3, we propose a model checking procedure to verify the copula assumption in (2), and to select among copula models. Simulation results and data analysis are presented in section 4. Section 5 contains some concluding remarks.

2 A Two-Stage Inference procedure

Let Z_i be the group indicator of the ith subject taking value of 0 or 1, $n_j = \sum_{i=1}^{n} I(Z_i = j)$ which is the sample size for group j ($j = 0, 1$) and $n = n_0 + n_1$. Let (X_i, Y_i) ($i = 1, \ldots, n$) be independent realizations of (X, Y) which follow model (2) in the upper wedge. Let C_i ($i = 1, \ldots, n$) be *iid* realizations of the external censoring variable C which is assumed to be independent of (X, Y). We observe semi-competing risks data, $\{(\tilde{X}_i, \delta_{xi}, \tilde{Y}_i, \delta_{yi}, Z_i)\ (i = 1, 2, ..., n)\}$, where $\tilde{X}_i = X_i \wedge Y_i \wedge C_i$, $\tilde{Y}_i = Y_i \wedge C_i$, $\delta_{x,i} = I(X_i \leq Y_i \wedge C_i)$ and $\delta_{y,i} = I(Y_i \leq C_i)$. The proposed inference procedure contains two steps. The parameters in model (2), namely α_j, $F_{y,j}(y)$, $F_j(x, y)$ and $F_{x,j}(x)$ ($j = 0, 1$), are estimated in the first stage. In the second stage, the proposed estimating function of θ is constructed based on the relationship between $F_{x,0}(x)$ and $F_{x,1}(x)$.

2.1 First-Stage: estimating nuisance parameters

In the first stage, estimators of $F_j(x, y)$, $F_{y,j}(y)$, $F_{x,j}(x)$, $G(y) = \Pr(C \geq y)$ and α_j, denoted as $\hat{F}_j(x, y)$, $\hat{F}_{y,j}(y)$, $\hat{F}_{x,j}(x)$, $\hat{G}(y)$ and $\hat{\alpha}_j$ respectively, can be obtained by applying existing estimators in the literature to data of the jth group.

Based on semi-competing risks data, there are two versions of the K-M estimator for $G(y) = \Pr(C \geq y)$. The censoring variable C can be viewed as being subject to censoring either by Y or $X \wedge Y$, which yield two versions of Kaplan-Meier estimators: $\hat{G}_{v1}(y) = \prod_{u<y}[1 - (\sum_{i=1}^{n} I(\tilde{Y}_i = u, \delta_{yi} = 0) / \sum_{i=1}^{n} I(\tilde{Y}_i \geq u))]$ and $\hat{G}_{v2}(y) = \prod_{u<y}[1 - (\sum_{i=1}^{n} I(\tilde{X}_i = u, \delta_{0i} = 0) / \sum_{i=1}^{n} I(\tilde{X}_i \geq u))]$, where $\delta_0 = I(X \wedge Y \leq C) = \delta_x + \delta_y - \delta_x \delta_y$. For $x \leq y$, it follows that $F_j(x, y) = \Pr(\tilde{X} \geq x, \tilde{Y} \geq y | Z = j)/G(y)$. Hence using the plugged-in approach, $F_j(x, y)$ can be estimated by

$$\hat{F}_j(x, y) = \hat{\Pr}\{\tilde{X} \geq x, \tilde{Y} \geq y | Z = j\}/\hat{G}(y) = \frac{\sum_{i=1}^{n} I(\tilde{X}_i \geq x, \tilde{Y}_i \geq y, Z_i = j)}{n_j \hat{G}(y)}, \quad (3)$$

where $\hat{G}(y)$ can be $\hat{G}_{v1}(y)$ or $\hat{G}_{v2}(y)$. This estimator is based on the assumption that covariates Z do not affect the distribution of censoring variable C. In the situation that the distribution of C depends on discrete covariate Z, $\hat{G}(y)$ can be modified by the corresponding K-M estimator $\hat{G}_j(y)$ which uses only those data points with $Z_i = j$.

Then the estimator of $F_{y,j}(y)$ is given by

$$\hat{F}_{y,j}(y) = \frac{\sum_{i=1}^{n} I(\tilde{Y}_i \geq y, Z_i = j)}{n_j \hat{G}(y)} \ (j = 0, 1). \tag{4}$$

There exist several estimators of α_j based on semi-competing risks data. Assuming the Clayton model in the upper wedge, Day et al. (1997) used the property of constant odds ratio and Fine et al. (2001) used the relationship between α and the concordance probability to construct their estimating equations respectively. Wang (2003) generalized the method by Day et al. (1997) and proposed two estimating functions of α for general AC models. Here we adopt Wang's method to estimate α_j by using only data points with $Z_i = j$. Then based on (2), one can derive $F_{x,j}(x)$ in terms of $\phi_{j,\alpha_j}(\cdot)$, $F_j(x, y)$ and $F_{y,j}(y)$. Straightforward calculation gives

$$F_{x,j}(x) = \phi_{j,\alpha_j}^{-1}\{\phi_{j,\alpha_j}[F_j(x, y)] - \phi_{j,\alpha_j}[F_{y,j}(y)]\}.$$

The right-hand side expression above involves both x and y but gives the same value for all values of y. Fine et al. (2001) suggested to consider the relationship on the diagonal line $y = x$ and proposed to estimate $F_{x,j}(x)$ by

$$\hat{F}_{x,j}(x) = \phi_{j,\hat{\alpha}_j}^{-1}\{\phi_{j,\hat{\alpha}_j}[\hat{F}_j(x, x)] - \phi_{j,\hat{\alpha}_j}[\hat{F}_{y,j}(x)]\}. \tag{5}$$

Here we adopt their proposal.

2.2 Second-Stage: estimating the group-difference parameter

The proposed estimating equation of θ is motivated by the following test statistic for testing equivalence between the two groups. Specifically to test $F_{x,0}(t) = F_{x,1}(t)$ for every t within the range of the data, one can use the test statistic

$$U_T = \sqrt{\frac{n_0 n_1}{n}} \int W(x) \left\{ \hat{F}_{x,0}(x) - \hat{F}_{x,1}(x) \right\} dx, \tag{6}$$

where $W(x)$ is a weight function. Let $t_{(1)} \leq \ldots \leq t_{(n)}$ be the observed ordered times of $\tilde{X}$ in the pooled sample and set $t_{(0)} = 0$. One can re-write U_T as

$$U_T = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} W(t_{(i)}) \left\{ t_{(i)} - t_{(i-1)} \right\} \left\{ \hat{F}_{x,0}(t_{(i)}) - \hat{F}_{x,1}(t_{(i)}) \right\}. \tag{7}$$

As for choosing the weight function, Klein and Moeschberger (p.216) suggest to use $W(x) = \frac{n\hat{G}_0(x)\hat{G}_1(x)}{n_0\hat{G}_0(x)+n_1\hat{G}_1(x)}$ where $\hat{G}_j(x)$ is the Kaplan-Meier estimator of $\Pr(C \geq x|Z = j)$. If the censoring distributions are assumed to be same in the two groups, one may use the whole sample to estimate $\Pr(C \geq x)$.

Now we modify the test statistic U_T in (6) to construct an estimating equation for θ. Our idea is to find a transformation $\xi_\theta(\cdot)$ such that $\xi_{\theta_0}(F_{x,0}(t)) = F_{x,1}(t)$ where θ_0 is the true value of θ. Therefore $g(t,\theta_0) = \xi_{\theta_0}(F_{x,0}(t)) - F_{x,1}(t) = 0$ for all t. It follows that $\sqrt{\frac{n_0n_1}{n}}\int W(x)g(x,\theta_0)dx = 0$. We can then estimate θ by solving the corresponding estimating equation $U(\theta) = \sqrt{\frac{n_0n_1}{n}}\int W(x)\hat{g}(x,\theta)dx = 0$, where $\hat{g}(t,\theta) = \xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)$. The way of finding the appropriate transformation $\xi_\theta(\cdot)$ depends on which part in model (1) is specified.

Consider model (1) where the form of $h(\cdot)$ is specified but the distribution of ε is unknown. It follows that $F_{x,1}(x) = F_{x,0}\left[h^{-1}\{h(x)+\theta_0\}\right]$. Hence $g(x,\theta) = F_{x,0}\left[h^{-1}\{h(x)+\theta\}\right] - F_{x,1}(x)$. We can construct the following estimating function:

$$U(\theta) = \sqrt{\frac{n_0n_1}{n}}\sum_{i=1}^{n} W(t_{(i)})(t_{(i)} - t_{(i-1)})\hat{g}(t_{(i)},\theta) \quad (8)$$

$$= \sqrt{\frac{n_0n_1}{n}}\sum_{i=1}^{n} W(t_{(i)})(t_{(i)} - t_{(i-1)})\left\{\hat{F}_{x,0}\left[h^{-1}\{h(t_{(i)})+\theta\}\right] - \hat{F}_{x,1}(t_{(i)})\right\}.$$

Consider the second situation when the form of $h(\cdot)$ is unknown but the distribution of ε is known. Let $F_\varepsilon(t) = \Pr(\varepsilon \geq t)$ denote the survival function of ε. Then $F_{x,1}(x) = F_\varepsilon\left[F_\varepsilon^{-1}\{F_{x,0}(x)\} + \theta_0\right]$. Hence $g(x,\theta) = F_\varepsilon\left[F_\varepsilon^{-1}\{F_{x,0}(x)\} + \theta_0\right] - F_{x,1}(x)$. The corresponding estimating function is

$$U(\theta) = \sqrt{\frac{n_0n_1}{n}}\sum_{i=1}^{n} W(t_{(i)})(t_{(i)} - t_{(i-1)})\hat{g}(t_{(i)},\theta) \quad (9)$$

$$= \sqrt{\frac{n_0n_1}{n}}\sum_{i=1}^{n} W(t_{(i)})(t_{(i)} - t_{(i-1)})\left\{F_\varepsilon\left[F_\varepsilon^{-1}\{\hat{F}_{x,0}(x)\} + \theta_0\right] - \hat{F}_{x,1}(x)\right\}.$$

Under both situations, the proposed estimator of θ is the solution to $U(\theta) = 0$, denoted as $\hat{\theta}$. Consistency and asymptotic normality of $\hat{\theta}$ are derived in Appendix 1 and 2.

We now discuss the following two special cases or illustration.

Example 1: Cox PH model

When ε has the extreme value distribution, model (1) becomes the Cox proportional hazard model. Then $F_\varepsilon(t) = \exp\{-\exp(t)\}$. When θ equals its true value θ_0, it follows that

$$F_{x,1}(x) = \{F_{x,0}(x)\}^{\exp(\theta_0)}\,.$$

Therefore $g(x,\theta) = F_{x,0}(x)^{\exp(\theta)} - F_{x,1}(x)$. The estimating equation becomes

$$U(\theta) = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} W(t_{(i)})\{t_{(i)} - t_{(i-1)}\} \left\{ \hat{F}_{x,0}(t_{(i)})^{\exp(\theta)} - \hat{F}_{x,1}(t_{(i)}) \right\} = 0.$$

Example 2: The proportional odds model

When ε is the standard logistic distribution, model (1) becomes the proportional odds model, where $F_\varepsilon(t) = \frac{1}{1+\exp(t)}$. When θ equals its true value θ_0, it follows that

$$F_{x,1}(x) = \frac{F_{x,0}(x)}{\exp(\theta_0) - F_{x,0}(x)\exp(\theta_0) + F_{x,0}(x)},$$

and

$$g(x,\theta) = \frac{F_{x,0}(x)}{\exp(\theta) - F_{x,0}(x)\exp(\theta) + F_{x,0}(x)} - F_{x,1}(x).$$

The resulting estimating equation becomes

$$U(\theta) = \sqrt{\frac{n_0 n_1}{n}} \sum_{i=1}^{n} W(t_{(i)})\{t_{(i)} - t_{(i-1)}\}\left[\frac{\hat{F}_{x,0}(t_{(i)})}{\exp(\theta) - \hat{F}_{x,0}(t_{(i)})\exp(\theta) + \hat{F}_{x,0}(t_{(i)})} - \hat{F}_{x,1}(t_{(i)})\right] = 0.$$

In the Appendix, we show that by choosing reasonable estimators for the nuisance parameters, $\hat{\theta}$ is a consistent estimator and $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotic normal with mean-zero, where θ_0 is the true value. Since the asymptotic variance of $\hat{\theta}$ is difficult to derive analytically, we suggest to use the bootstrap or jackknife resampling method to estimate its value.

3 Selection of a Copula Model

By specifying the dependence structure between X and Y for each value of Z, we can avoid making unnecessary assumption about the covariate effect on Y. In this section, we demonstrate how to select an appropriate model of $\phi_{j,\alpha_j}(\cdot)$ for $j = 0, 1$ based on the data at hand. Without loss of generality, the discussion here is based on a homogeneous sample $\{(\tilde{X}_i, \delta_{xi}, \tilde{Y}_i, \delta_{yi})\ (i = 1, 2, ..., n)\}$.

We briefly summarize the ideas. First, we select some functions which are identifiable based on semi-competing risks data. Then, for a chosen function, we compare its nonparametric estimator and the corresponding model-based estimator based on some distance measure. The most plausible model is the one that yields the smallest distance. One can further construct a goodness-of-fit test if the distribution of the distance measure under the null hypothesis can be derived. When analytic derivations are difficult, we suggest to use the bootstrap re-sampling method to obtain the cut-off value in the test.

Two identifiable functions are considered, namely $F_X^{11}(t) = \Pr(X \geq t | \delta_x = 1, \delta_y = 1)$ and $F_Y^{11}(t) = \Pr(Y \geq t | \delta_x = 1, \delta_y = 1)$. The nonpametric estimator, denoted as $\hat{F}_X^{11}(t)$, is given by $\sum_{i=1}^n I(\tilde{X}_i \geq t, \delta_{xi} = 1, \delta_{yi} = 1) / \sum_{i=1}^n I(\delta_{xi} = 1, \delta_{yi} = 1)$. Assuming model k, the model-based estimator of $F_X^{11}(t)$, denoted as $\tilde{F}_{X,k}^{11}(t)$, can be computed as follows:

$$\tilde{F}_{X,k}^{11}(t) = \frac{\int_c \hat{\Pr}(t \leq X \leq Y \leq c)\hat{G}(dc)}{\int_c \hat{\Pr}(X \leq Y \leq c)\hat{G}(dc)},$$

where $\hat{G}(dc) = \hat{G}(c) - \hat{G}(c + \triangle)$; $\hat{\Pr}(X \leq Y \leq c) = \int_0^c \int_0^y \tilde{F}(dx, dy)$ and $\hat{\Pr}(t \leq X \leq Y \leq c) = \int_t^c \int_t^y \tilde{F}(dx, dy)$, $\tilde{F}(dx, dy) = \tilde{F}(x, y) - \tilde{F}(x + \triangle, y) - \tilde{F}(x, y + \triangle) + \tilde{F}(x + \triangle, y + \triangle)$, $\tilde{F}(x, y) = \phi_{\hat{\alpha}}^{-1}\{\phi_{\hat{\alpha}}[\hat{F}_x(x)] + \phi_{\hat{\alpha}}[\hat{F}_y(y)]\}$, $\hat{F}_x(x) = \phi_{\hat{\alpha}}^{-1}\{\phi_{\hat{\alpha}}[\hat{F}(x, x)] - \phi_{\hat{\alpha}}[\hat{F}_y(x)]\}$ and $\phi_\alpha(.)$ is the copula model k. $\hat{F}_Y^{11}(t)$ and $\tilde{F}_{Y,k}^{11}(t)$ are similar.

A formal testing procedure can be performed as follows. Consider testing $H_0 : \phi_\alpha$ has the form of model k versus $H_a : \phi_\alpha$ does not have the form of model k. Define $D_X^k = \int_0^a \left| \hat{F}_X^{11}(t) - \tilde{F}_{X,k}^{11}(t) \right| dt$ and $D_Y^k = \int_0^b \left| \hat{F}_Y^{11}(t) - \tilde{F}_{Y,k}^{11}(t) \right| dt$, where $a = max\{\tilde{X}_i | \delta_{xi} = 1, \delta_{yi} = 1\ (i = 1, ..., n)\}$ and $b = max\{\tilde{Y}_i | \delta_{xi} = 1, \delta_{yi} = 1\ (i = 1, ..., n)\}$. We can reject H_0 if $D_X^k > c_x$, where c_x is the critical value satisfying $\Pr(D_X^k > c_x | H_0) = \gamma$, the type-one error rate. Alternatively, we can reject H_0 if $D_Y^k > c_y$, where c_y is the critical value satisfying $\Pr(D_Y^k > c_y | H_0) = \gamma$. We can establish the following result.

Theorem 1
(a) If copula model k is the true model, $D_X^k \xrightarrow{P} 0$ *and* $D_Y^k \xrightarrow{P} 0$.
(b) If copula model j is the true model and $j \neq k$, $D_X^k \xrightarrow{P} c_{x,k}$ *and* $D_Y^k \xrightarrow{P} c_{y,k}$*, where*

$c_{x,k}$ and $c_{y,k}$ are non-zero constant.
(c) Let S be the set containing models under consideration. Assume that copula model j belongs to the set S. Under copula model j, $\lim_{n\to\infty} Pr(\min_{k\in S} D_X^k = D_X^j) = 1$ and $\lim_{n\to\infty} Pr(\min_{k\in S} D_Y^k = D_Y^j) = 1$.

The above theorem implies that if the correct model is included in the list of model alternatives, the probability that it will be selected as the fitted model is one if the sample size is very large. The proof of the theorem is given in Appendix 3. In practice, we often are not sure whether the model choices include the true one or not. Informally we can plot $\hat{F}_X^{11}(t)$ together with $\tilde{F}_{X,k}^{11}(t)$ for $k = 1, \ldots, K$. The best fitted model is the one which is the closest to $\hat{F}_X^{11}(t)$ or $F_Y^{11}(t)$.

Because the distribution of D_X^k and D_Y^k are difficult to derive analytically, we suggest to use the bootstrap re-sampling method to obtain the cut-off value, p-value and power. Here we briefly describe the procedure. A bootstrap sample under model k can be generated as follows. Recall that given the original data, we have obtained $\hat{G}(c)$, $\hat{F}_y(y)$ and $\hat{F}_x(x)$ under the assumption of model k. Then generate $(U_i^*, V_i^*) \sim$ model k, $U_i^* \sim U(0,1)$ and $V_i^* \sim U(0,1)$. Then set $X_i^* = s$ if $\hat{F}_x(s^+) < 1 - U_i^* \leq \hat{F}_x(s)$, $Y_i^* = t$ if $\hat{F}_y(t^+) < 1 - V_i^* \leq \hat{F}_y(t)$ and $C_i^* \sim \hat{G}(c)$. Given (X_i^*, Y_i^*, C_i^*) $(i = 1, ..., n)$, we can construct a bootstrap sample $\{(\tilde{X}_i^*, \delta_{xi}^*, \tilde{Y}_i^*, \delta_{yi}^*)\ (i = 1, 2, ..., n)\}$, where $\tilde{X}_i^* = X_i^* \wedge Y_i^* \wedge C_i^*$, $\tilde{Y}_i^* = Y_i^* \wedge C_i^*$, $\delta_{xi}^* = I(X_i^* \leq Y_i^* \wedge C_i^*)$ and $\delta_{yi}^* = I(Y_i^* \leq C_i^*)$. With a bootstrapped sample, we can compute the corresponding values of D_X^k and D_Y^k. Repeating the bootstrapping procedure many times, the distributions D_X^k and D_Y^k can be approximated by their empirical counterparts.

4 Numerical Analysis

4.1 Simulation Results

We design several simulation settings to examine the validity and robustness of our method and the method proposed by Lin et al. (1996). Data generation algorithms for the Clayton model and the Frank model have been given in Prentice and Cai (1992) and Genest (1987), respectively.

In the first analysis, we simulated $(\varepsilon, \xi)|Z$ to follow an AC model. In some settings, α_j or τ_j may be different for $j = 0, 1$. Then based on (ε, ξ, Z), the value of (X, Y) can be obtained from the models $h_1(X) = -\theta_0 Z + \varepsilon$ and $h_2(Y) = -\eta_0 Z + \xi$, where $\theta_0 = 0.5$, $\eta_0 = 0.5$. The sample sizes are set to be $n_0 = n_1 = 150$. The proposed estimator is denoted as $\hat{\theta}$ and the estimator by Lin et al. (1996) is denoted as $\hat{\theta}_L$. The purpose here is to examine the robustness of $\hat{\theta}_L$. In the computation of $\hat{\theta}$, we assume the correct copula models for the two groups. The average bias and the standard deviation based on 1000 simulation runs are reported in Table 4-1. In the first four cases, we set $h_1(t) = h_2(t) = t$, $-\theta_0 + \varepsilon \sim \exp(0.8)$, $-\eta_0 + \xi \sim \exp(1)$, $C|Z = 1 \sim U(0, 6)$ and $C|Z = 0 \sim U(0.5, 6.5)$ but the dependence structures for the two groups may vary.

Case 1: Both of $\phi_{\tau_j}(\cdot)$ $(j = 0, 1)$ follow the Clayton model with $\tau_1 = \tau_0 = 0.5$. This is the condition when the two estimators are both valid. We see that $\hat{\theta}_L$ slightly outperforms $\hat{\theta}$.

Case 2: Both of $\phi_{\tau_j}(\cdot)$ $(j = 0, 1)$ follow the Clayton model with $\tau_1 = 0.1$ and $\tau_0 = 0.8$. We see that the bias of $\hat{\theta}_L$ increases drastically.

Case 3: We set $\phi_{\tau_1}(\cdot)$ to follow the Clayton model with $\tau_1 = 0.5$ and $\phi_{\tau_0}(\cdot)$ to follow the Frank model with $\tau_0 = 0.5$. Again, we see that $\hat{\theta}_L$ becomes more biased but the effect due to different forms of association is less severe than the effect due to the discrepancy in the levels of association for the two groups.

Case 4: We set $\phi_{\tau_1}(\cdot)$ to follow the Clayton model with $\tau_1 = 0.1$ and $\phi_{\tau_0}(\cdot)$ to follow the Frank model with $\tau_0 = 0.8$.

In the next four cases, we set $h_1(t) = t$ but $h_2(t) = \log(t)$, $-\theta_0 + \varepsilon \sim \exp(0.8)$, $\exp(-\eta_0)\exp(\xi) \sim \exp(1)$, $C|Z = 1 \sim U(0, 6)$ and $C|Z = 0 \sim U(0.5, 6.5)$. Note that $h_2(t) \neq h_1(t)$ which is a condition that violates the assumption made by Lin et al. (1996).

Case 5: Both of $\phi_{\tau_j}(\cdot)$ $(j = 0, 1)$ follow the Clayton model with $\tau_1 = \tau_0 = 0.5$. The estimator $\hat{\theta}_L$ not only becomes more biased but also more variable.

case 6: Both of $\phi_{\tau_j}(\cdot)$ $(j = 0, 1)$ follow the Clayton model with $\tau_1 = 0.1$ and $\tau_0 = 0.8$.

case 7: We set $\phi_{\tau_1}(\cdot)$ to follow the Clayton model with $\tau_1 = 0.5$ and $\phi_{\tau_0}(\cdot)$ to follow the

Frank model with $\tau_0 = 0.5$.

Case 8: We set $\phi_{\tau_1}(\cdot)$ to follow the Clayton model with $\tau_1 = 0.1$ and $\phi_{\tau_0}(\cdot)$ to follow the Frank model with $\tau_0 = 0.8$. From the results of case 6 and case 8, we see that $\hat{\theta}_L$ is mostly affected by joint mis-specification of $h_2(t)$ and τ_j $(j = 0, 1)$.

In the second analysis, we generated $\{\tilde{X}_i, \tilde{Y}_i, \delta_{xi}, \delta_{yi}\}$ $(i = 1, \cdots, 150)$, where $X \sim \exp(0.8)$, $Y \sim \exp(1)$, $(X, Y) \sim$ Clayton model $(\tau = 0.5)$ and $C \sim U(0, 6)$. Suppose that there are two models under consideration where model $k = 1$ is the Clayton model and model $k = 2$ is the Frank model. In 1000 replications, the mean and standard deviation (in parenthesis) of D_X^1, D_X^2, D_Y^1, D_Y^2 are 0.0487 (0.0199), 0.1344 (0.0479), 0.0542 (0.0246) and 0.1537 (0.0468), respectively. The percentage of successfully selecting the Clayton model is 94.5% based on the order of D_X^j $(j = 1, 2)$ and 95.4% based on the order of D_Y^j $(j = 1, 2)$. Then we changed the above setting by letting $(X, Y) \sim$ Frank model $(\tau = 0.5)$. In 1000 replications, the mean and standard deviation (in parenthesis) of D_X^1, D_X^2, D_Y^1, D_Y^2 are 0.1038 (0.0357), 0.0456 (0.0194), 0.1679 (0.0562) and 0.0578 (0.0253), respectively. The percentage of successfully selecting the Frank model is 92.8% based on the order of D_X^j $(j = 1, 2)$ and 95.8% based on the order of D_Y^j $(j = 1, 2)$. Then we demonstrate the procedure of model selection based on one simulation run in which the true model is the Clayton model. Graphical diagnostic plots are shown in Figure 4-1 and Figure 4-2 which reveal that the Clayton model is a better fit. Formally, we may set up the following test: H_0 : the model follows the Clayton model versus H_a : the model follows the Frank model. The bootstrap procedure was applied 1000 times to determine the p-values and cut-off points. We obtained that $D_X^1 = 0.031$ with p-value=0.796, $D_Y^1 = 0.0323$ with p-value=0.82; the cut-off value $(\gamma = 0.05)$: $c_x = 0.0833$, $c_y = 0.0964$. Both test statistics show that we accept H_0 which is a correct decision. For the same data set, suppose we change the statement of hypothesis testing such that H_0 : the model follows the Frank model versus H_a : the model follows the Clayton model. We obtained that $D_X^2 = 0.1303$ with p-value=0.002, $D_Y^2 = 0.1544$ with p-value=0.001; the cut-off value $(\gamma = 0.05)$: $c_x = 0.0921$, $c_y = 0.0944$. According to the two p-values, we reject H_0 which is also a correct decision. The above results show that the proposed selection procedure performs quite well.

The purpose of the third analysis is to examine finite-sample performance of the proposed estimator, $\hat{\theta}$, and its robustness property under model mis-specification. Again two copula families are evaluated, namely the Clayton model and the Frank model. We consider four models for measuring the group effect: the location shift model (LS), the accelerated failure time model (AFT), the Cox proportional hazard model (PH) and the proportional odds model (PO). The sample sizes in both groups are 150 with $\tau_0 = 0.5$ and $\tau_1 = 0.6$ and the true value, θ_0, is 0.5. For the LS model, the marginal distributions follow $X|Z = 1 \sim \exp(0.8)$, $Y|Z = 1 \sim \exp(1)$, $C|Z = 1 \sim U(0,6)$, $X|Z = 0 \sim \exp(0.8) + \theta_0$, $Y|Z = 0 \sim \exp(1) + \theta_0$ and $C|Z = 0 \sim U(0,6) + \theta_0$. For the AFT model, the marginal distributions follow $X|Z = 0 \sim \exp(0.8)$, $Y|Z = 0 \sim \exp(1)$, $C|Z = 0 \sim U(0,6)$, $X|Z = 1 \sim \exp(-\theta_0)\exp(0.8)$, $Y|Z = 1 \sim \exp(-\theta_0)\exp(1)$ and $C|Z = 1 \sim U(0,6)$. For the PH model, the marginal distributions follow $X|Z = 0 \sim \exp(0.8)$, $Y|Z = 0 \sim \exp(1)$, $C|Z = 0 \sim U(0,6)$, $X|Z = 1 \sim \exp(-\theta_0)\exp(0.8)$, $Y|Z = 1 \sim \exp(-\theta_0)\exp(1)$ and $C|Z = 1 \sim U(0,6)$. For the PO model, the marginal distributions follow $X|Z = 0 \sim \exp(0.8)$, $Y|Z = 0 \sim \exp(1)$, $C|Z = 0 \sim U(0,6)$, $X|Z = 1 \sim 0.8\log((U - Ue^{-\theta_0} - 1)/(U - 1))$, $Y|Z = 1 \sim \log((V - Ve^{-\theta_0} - 1)/(V - 1))$ and $C|Z = 1 \sim U(0,6)$, where U, $V \sim U(0,1)$. The average bias and the standard deviation based on 1000 simulation runs are reported in Table 4-2. Since in general our estimator $\hat{\theta}$ using $\hat{G}_{v1}(y)$ is better than using $\hat{G}_{v2}(y)$, we only present the result based on $\hat{G}_{v1}(y)$. The proposed estimator performs quite well when the dependence structure is correctly assumed. When the dependence structure is mis-specified, the bias of $\hat{\theta}$ increases but the severity seems to less than $\hat{\theta}_L$ under the situation of model violation.

4.2 Data Analysis

The proposed methodology is applied to the bone marrow transplants data given in Klein and Moeschberger (2003, p.484). There were 137 leukemia patients receiving bone marrow transplants. Let X be the time to relapse of leukemia, Y be the time to death and C be the time from transplant to the end of study. Let $\delta_x = I(X \le Y \wedge C)$ be the relapse indicator and let $\delta_y = I(Y \le C)$ be the death indicator. We consider two samples with $Z = 0$ indicating the AML high-risk group and $Z = 1$ indicating the ALL group which together contain 83 observations. Model diagnostic plots are shown in Figure 4-3 and Figure 4-4. We found that the Clayton model is more appropriate for AML high-risk group than the Frank model. So we test $H_0 : \phi_\alpha \sim$ the Clayton model for $Z = 0$ versus H_a : not H_0. By bootstrapping 1000 times, the p-value of D_X^C

is 0.931 and the p-value of D_Y^C is 0.809. Hence we adopt the Clayton model for the AML high-risk group. For the ALL group, Figure 4-5 and 4-6 indicate that Clayton model is also more appropriate than the Frank model. We test $H_0 : \phi_\alpha \sim$ the Clayton model for $Z = 1$ versus H_a : not H_0. By bootstrapping 1000 times, the p-value of D_X^C is 0.963 and the p-value of D_Y^C is 0.638. Hence we also adopt the Clayton model for the ALL group. Using Day's method to estimate τ_j $(j = 0, 1)$, $\hat{\tau}_0$=0.7685(0.0872) and $\hat{\tau}_1$=0.7894(0.0853), where the number in the parenthesis is the estimated standard derivation using the jackknife method. The above analysis implies that the dependence structures in the two groups are similar and the two events are highly correlated. Then we decided a model for measuring the group effect on X. Figure 4-7 which shows the fitted log-log plot of $\hat{F}_x(x)$ for the two groups. Since the two curves appear parallel, we chose the Cox model to measure the group effect. Figure 4-8 depicts the survival curves, $\hat{F}_x(x)$, for the two groups. Using the proposed method under the Cox model assumption, $\hat{\theta}$=-0.407 (0.3074) with $\hat{G}(y) = \hat{G}_{v1}(y)$. With $\hat{G}(y) = \hat{G}_{v2}(y)$, $\hat{\theta}$=-0.3835 (0.302). This results implies that the risk of leukemia relapse seems higher for the high-risk group but, probably due to small sample size, the difference is not statistically significant.

5 Concluding Remarks

In this article, we consider two-sample comparison based on the failure time to a non-terminal event which is subject to censoring by a terminal event. To handle the problem of dependent censoring, we assume that failure times of the two events follow a copula model in the identifiable region but, in addition, also propose a method to examine the appropriateness of this model assumption. Compared to existing methods such as that proposed by Lin et al. (1996), our approach allows for different dependence structures in the two groups and avoids making additional model assumption on the terminal event. The simulation analysis confirms our conjecture that the estimator proposed by Lin et al. (1996) becomes unreliable if the dependence structures in the two group are very different. Although simulations show that our method is not robust either when the dependence structures are mis-specified, the proposed model checking approach somewhat reduces the possibility that such an undesirable situation would happen. The proposed strategy of model checking is to compare the non-parametric estimator with its model-based estimator of a chosen measure which is $F_X^{11}(t)$ or $F_Y^{11}(t)$ in our article. For possible future research, one may examine how to choose a measure or a combination of several measures that contains most of the model information characterized by $\phi(\cdot)$ so that the corresponding test procedure would give higher power. The proposed inference method for estimating the group-difference parameter is developed under a flexible model formulation. Extension of the method under a general regression setting may deserve further investigation.

ACKNOWLEDGEMENT:
The work was done wholly/partially while the second author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2005/2006. The visit was supported by the Institute.

APPENDIX

Appendix 1: *Asymptotic properties of* $\hat{\theta}$

For $U(\theta)$ defined in (8) and (9), and assuming that $U(.)$ is twice differentiable with respect to θ with bounded derivatives, by Taylor expansion techniques we get

$$\begin{aligned} U(\hat{\theta}) &= 0 \\ &= U(\theta_0) + U'(\theta_0)(\hat{\theta} - \theta_0) + U''(\check{\theta})(\hat{\theta} - \theta_0)^2/2, \end{aligned}$$

where $\check{\theta}$ is an intermediate value between θ_0 and $\hat{\theta}$. When the second derivative is multiplied by the square of the small number $\hat{\theta} - \theta_0$, it is negligible compared to the first derivative term. Hence, we get $-U'(\theta_0)(\hat{\theta} - \theta_0) \approx U(\theta_0)$. Formally we have the following expression:

$$\sqrt{n}(\hat{\theta} - \theta_0) = (-\frac{U'(\theta_0)}{\sqrt{n}})^{-1} U(\theta_0) + o_p(1). \tag{A.1}$$

The asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ is determined by properties of $-U'(\theta_0)/\sqrt{n}$ and $U(\theta_0)$. We first show that $U(\theta_0)$ is asymptotic normal with mean zero. Then we show $-U'(\theta_0)/\sqrt{n}$ converges to a non-zero constant.

We estimate $F_x(t)$ using the estimator proposed by Fine et al. (2001). By simple calculations, we get

$$n^{1/2}\{\xi_{\theta_0}(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)\}$$

$$= n^{1/2}\{\xi_{\theta_0}(\hat{F}_{x,0}(t)) - \xi_{\theta_0}(F_{x,0}(t))\} - n^{1/2}\{\hat{F}_{x,1}(t) - F_{x,1}(t)\}.$$

It can be shown that $n^{1/2}\{\hat{F}_{x,1}(t) - F_{x,1}(t)\}$ converges weakly to a Gaussian process and $\sup_{t\in[0,t_{(n)}]} \|\hat{F}_{x,1}(t) - F_{x,1}(t)\| \xrightarrow{a.s.} 0$. Also $n^{1/2}\{\hat{F}_{x,0}(t) - F_{x,0}(t)\}$ converges weakly to a Gaussian process. Applying Taylor series expansions, it follows that

$$\sqrt{n}\{\xi_{\theta_0}(\hat{F}_{x,0}(t)) - \xi_{\theta_0}(F_{x,0}(t))\} \stackrel{a}{=} \xi'_{\theta_0}(F_{x,0}(t))\sqrt{n}\{\hat{F}_{x,0}(t) - F_{x,0}(t)\},$$

which converges weakly to a Gaussian process. Recall that the expression of $U(\theta)$ involves $\sqrt{n_0 n_1/n}$. Assume that as $n \longrightarrow \infty$, $n_0 = Cn_1$, where $C > 0$ is constant. It follows that $\sqrt{n_0 n_1}/n$ converges to $\sqrt{C}/(C+1)$. Since the integrand of $U(\theta_0)$ converges to a mean-zero Gaussian process, $U(\theta_0)$ converges to a normal random variable with mean-zero. We can write

$$\frac{-U'(\theta_0)}{\sqrt{n}} = -\frac{\sqrt{n_0 n_1}}{n}\frac{\partial}{\partial\theta}\int_0^{t_{(n)}} W(t)[\xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)]dt\ |_{\theta=\theta_0},$$

by (A.2), which converges to

$$-\frac{\sqrt{C}}{C+1}\int_0^{t_{(n)}} W(t)\frac{\partial}{\partial\theta}\xi_\theta(F_{x,0}(t))dt\ |_{\theta=\theta_0},$$

a non-zero constant.

Based on the above analysis, one can write

$$\sqrt{n}(\hat{\theta} - \theta_0) = (-\frac{U'(\theta_0)}{\sqrt{n}})^{-1}U(\theta_0) + o_p(1),$$

which is asymptotic normal. Consistency of $\hat{\theta}$ can also be established.

Appendix 2: *Proof of (A.2):*

In Appendix 1, the following result is used to establish large-sample sample properties of $\hat{\theta}$:

$$\frac{\partial}{\partial\theta}\int_0^{t_{(n)}} W(t)\hat{g}(t,\theta)dt = \frac{\partial}{\partial\theta}\int_0^{t_{(n)}} W(t)[\xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)]dt \xrightarrow{P}$$

$$\frac{\partial}{\partial\theta}\int_0^{t_{(n)}} W(t)g(t,\theta)dt = \frac{\partial}{\partial\theta}\int_0^{t_{(n)}} W(t)[\xi_\theta(F_{x,0}(t)) - F_{x,1}(t)]dt \qquad \text{(A.2)}$$

Define $\hat{\eta}_n(\theta) = \int_0^{t_{(n)}} W(t)[\xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)]dt$ and $\eta(\theta) = \int_0^{t_{(n)}} W(t)[\xi_\theta(F_{x,0}(t)) - F_{x,1}(t)]dt$. Following Fine et al. (2001), we can show the following two conditions. First, $\xi_\theta(\hat{F}_{x,0}(t))$ converges to $\xi_\theta(F_{x,0}(t))$ uniformly and strongly. Secondly $\hat{F}_{x,1}(t)$ converges to $F_{x,1}(t)$ uniformly and strongly. Therefore

$$W(t)[\xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)] \xrightarrow{a.s.}$$

$$W(t)[\xi_\theta(F_{x,0}(t)) - F_{x,1}(t)] \quad \text{for } t \in [0, t_{(n)}]$$

and $|W(t)[\xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)]| \leq 1$ almost everywhere. It follows that

$$\int_0^{t_{(n)}} W(t)[\xi_\theta(\hat{F}_{x,0}(t)) - \hat{F}_{x,1}(t)]dt \xrightarrow{P}$$

$$\int_0^{t_{(n)}} W(t)[\xi_\theta(F_{x,0}(t)) - F_{x,1}(t)]dt.$$

which implies that $\hat{\eta}_n(\theta) \xrightarrow{P} \eta(\theta)$.

Next we show that $\frac{\partial}{\partial\theta}\hat{\eta}_n(\theta) \xrightarrow{P} \frac{\partial}{\partial\theta}\eta(\theta)$. We have shown that for every $\epsilon_1 > 0$, $\lim_{n\to\infty} \Pr\{|\hat{\eta}_n(\theta) - \eta(\theta)| \geq \epsilon_1\} = 0$. Then for every $\epsilon_1 > 0$ and for every $\xi_1 > 0$, there exists $n_1 > 0$ such that for $n > n_1$, we have $\Pr\{|\hat{\eta}_n(\theta) - \eta(\theta)| \geq \epsilon_1\} \in [0, \xi_1)$. Let $\triangle$ is a positive constant. Hence, $\hat{\eta}_n(\theta + \triangle) \xrightarrow{P} \eta(\theta + \triangle)$. For every $\epsilon_2 > 0$, $\lim_{n\to\infty} \Pr\{|\hat{\eta}_n(\theta+\triangle) - \eta(\theta+\triangle)| \geq \epsilon_2\} = 0$. Then for every $\epsilon_2 > 0$ and for every $\xi_2 > 0$, there exists $n_2 > 0$ such that for $n > n_2$, we have $\Pr\{|\hat{\eta}_n(\theta+\triangle) - \eta(\theta+\triangle)| \geq \epsilon_2\} \in [0, \xi_2)$. Let $n_3 = \max\{n_1, n_2\}$. For $n > n_3$, we have $\Pr\{|\hat{\eta}_n(\theta) - \eta(\theta)| \geq \epsilon_1\} + \Pr\{|\hat{\eta}_n(\theta + \triangle) - \eta(\theta + \triangle)| \geq \epsilon_2\} \in [0, \xi_1 + \xi_2)$. Because

$$\Pr\{|\hat{\eta}_n(\theta) - \eta(\theta)| \geq \epsilon_1\} + \Pr\{|\hat{\eta}_n(\theta + \triangle) - \eta(\theta + \triangle)| \geq \epsilon_2\}$$

$$\geq \Pr\{|\hat{\eta}_n(\theta + \triangle) - \eta(\theta + \triangle)| + |\hat{\eta}_n(\theta) - \eta(\theta)| \geq \epsilon_1 + \epsilon_2\}$$

$$\geq \Pr\{|(\hat{\eta}_n(\theta + \triangle) - \eta(\theta + \triangle)) - (\hat{\eta}_n(\theta) - \eta(\theta))| \geq \epsilon_1 + \epsilon_2\}$$

$$= \Pr\{|(\hat{\eta}_n(\theta + \triangle) - \hat{\eta}_n(\theta)) - (\eta(\theta + \triangle) - \eta(\theta))| \geq \varepsilon_1 + \varepsilon_2\}$$

$$= \Pr\{|\frac{\hat{\eta}_n(\theta + \triangle) - \hat{\eta}_n(\theta)}{\triangle} - \frac{\eta(\theta + \triangle) - \eta(\theta)}{\triangle}| \geq \frac{\epsilon_1 + \epsilon_2}{\triangle}\}.$$

Let $\frac{\varepsilon_1+\varepsilon_2}{\triangle} = \epsilon$ and $\xi_1 + \xi_2 = \xi$. For every $\epsilon > 0$ and for every $\xi > 0$, for $n > n_3$, we have $\Pr\{|\frac{\hat{\eta}_n(\theta+\triangle)-\hat{\eta}_n(\theta)}{\triangle} - \frac{\eta(\theta+\triangle)-\eta(\theta)}{\triangle}| \geq \epsilon\} \in [0, \xi)$. Therefore, $\frac{\hat{\eta}_n(\theta+\triangle)-\hat{\eta}_n(\theta)}{\triangle} \xrightarrow{P} \frac{\eta(\theta+\triangle)-\eta(\theta)}{\triangle}$. Hence, as $\triangle \longrightarrow 0$, $\frac{\partial}{\partial\theta}\hat{\eta}_n(\theta) \xrightarrow{P} \frac{\partial}{\partial\theta}\eta(\theta)$.

Appendix 3: *Proof of Theorem 1:*

(a) Suppose that model k is the true one. By Glivenko-Cantelli theorem and Theorem 3.4.2 in Fleming & Harrington (1991), we can show that as $n \to \infty$, $\sup_{0\leq t\leq a} |\hat{F}_X^{11}(t) - \tilde{F}_{X,k}^{11}(t)| \xrightarrow{P} 0$ and then $D_X^k = \int_0^a |\hat{F}_X^{11}(t) - \tilde{F}_{X,k}^{11}(t)|dt \xrightarrow{P} 0$. Similarly, $D_Y^k \xrightarrow{P} 0$.

(b) Suppose that model j is the true model. By Glivenko-Cantelli theorem and Theorem 3.4.2 in Fleming & Harrington (1991), we can show that as $n \to \infty$, $\sup_{0\leq t\leq a} |\hat{F}_X^{11}(t) - F_X^{11}(t)| \xrightarrow{a.e.} 0$. Also there exists $F_k(t)$ such that as $n \to \infty$, $\sup_{0\leq t\leq a} |\tilde{F}_{X,k}^{11}(t) - F_k(t)| \xrightarrow{p} 0$. When the model k is mis-specified, we have $F_X^{11}(t) \neq F_k(t)$ for some t. Let $\sup_{0\leq t\leq a} |F_X^{11}(t) - F_k(t)| = c$ and $\int_0^a |F_X^{11}(t) - F_k(t)|dt = c_{x,k}$, both of which are non-zero constants. Hence as $n \to \infty$, $\sup_{0\leq t\leq a} |\hat{F}_X^{11}(t) - \tilde{F}_{X,k}^{11}(t)| \xrightarrow{p} c$ and $D_X^k = \int_0^a |\hat{F}_X^{11}(t) - \tilde{F}_{X,k}^{11}(t)|dt \xrightarrow{p} c_{x,k}$. Similarly, we can show that D_Y^k approaches to a non-zero constant.

(c) Assume that copula model j is the true model. From (a) and (b), we have $D_X^j \xrightarrow{p} 0$ and, for any $k \in S\backslash\{j\}$, $D_X^k \xrightarrow{p} c_{x,k} > 0$. Let $c_{\min} = \min_{k\in S\backslash\{j\}} c_{x,k}$. For every $\xi_j > 0$, there exists n_j such that $n > n_j$, we have $\Pr(|D_X^j - 0| \geq \frac{1}{2}c_{\min}) \in [0, \xi_j)$. For every $\xi_k > 0$, there exists n_k such that $n > n_k$, we have $\Pr(|D_X^k - c_{x,k}| \geq \frac{1}{2}c_{\min}) \in [0, \xi_k)$. Let $\xi = \sum_{k\in S} \xi_k$ and $n' = \max_{k\in S} n_k$. For $n > n'$, we can show that $\Pr(\min_{k\in S\backslash\{j\}} D_X^k \leq D_X^j) \in [0, \xi)$. It follows that $\lim_{n\to\infty} \Pr(\min_{k\in S} D_X^k = D_X^j) = 1$. Similarly, $\lim_{n\to\infty} \Pr(\min_{k\in S} D_Y^k = D_Y^j) = 1$.

References

Chang, S. H. (2000). A Two-Sample Comparison for Multiple Ordered Event Data. *Biometrics*, **56**, 183-189.

Day, R., Bryant, J. and Lefkopoulou, M. (1997). Adaptation of Bivariate Frailty Models for Prediction, with Application to Biological Markers as Prognostic Indicators. *Biometrika*, **84**, 45-56.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika*, **88**, 4, 907-919.

Genest, C. and Mackay, J. (1986). The Joy of Copulas: Bivariate Distributions With Uniform Marginals. *The American Statistician*, vol. **40**, No. 4.

Genest, C. (1987). Frank's Family of Bivariate Distributions. *Biometrika*, **74**, 3, 549-555.

Genest, C. and Rivest L. P. (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, vol. **88**, No. 423.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing Two Failure Time Distribution in The Presence of Dependent Censoring. *Biometrika*, **83**, 2, 381-393.

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*. Vol. **84**, No. 406.

Prentice, R. L. and Cai, J. (1992). Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika*, **79**, 3, 495-512.

Wang, W. and Wells, M. T. (2000). Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association*. Vol. **95**, No. 449.

Wang, W. (2003). Estimating the Association Parameter for Copula Models under Dependent Censoring. *Journal of the Royal Statistical Society, Series B*. **65**, 257-273.

Model	$\hat{\theta}$	$\hat{\theta}_L$
case 1:	0.0026(0.0934)	0.0025(0.0909)
case 2:	0.0013(0.1136)	0.0969(0.0849)
case 3:	0.0022(0.0950)	0.0122(0.0888)
case 4:	0.0008(0.1100)	0.0982(0.0840)
case 5:	0.0041(0.0974)	0.0890(0.1175)
case 6:	0.0067(0.1135)	0.3387(0.1127)
case 7:	0.0025(0.1156)	0.0884(0.1170)
case 8:	0.0125(0.1152)	0.3793(0.1081)

Table 4-1: Finite sample performance of two estimators evaluated under 8 situations in favor of $\hat{\theta}$. The first number is the average bias of the estimator and the number in the parenthesis is the standard deviation based on 1000 replications.

True	Imposed	LS	AFT	PH	PO
Clayton	Clayton	0.0005(0.0910)	0.0027(0.1348)	0.0037(0.1419)	0.0029(0.2346)
	Frank	0.0156(0.1019)	0.0272(0.1703)	0.0821(0.1186)	0.1378(0.1989)
Frank	Frank	0.0038(0.0908)	0.0021(0.1399)	0.0051(0.1419)	0.0024(0.2293)
	Clayton	0.0104(0.1005)	0.0274(0.1367)	0.0895(0.1998)	0.0785(0.2998)

Table 4-2: Finite sample performance of $\hat{\theta}$ under correct and wrong model assumptions. The first number is the average bias of the estimator $\hat{\theta}$ and the number in the parenthesis is the standard deviation based on 1000 replications.

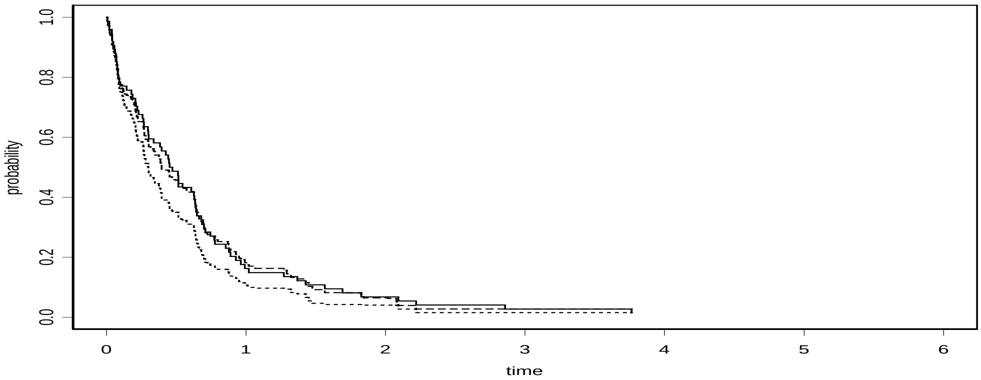

Figure 4-1: Diagnostic plot for model selection based on $F_X^{11}(t)$ when the true model is Clayton. model 1: Clayton and model 2: Frank. solid line: $\hat{F}_X^{11}(t)$; dashed line: $\tilde{F}_{X,1}^{11}(t)$; dotted line: $\tilde{F}_{X,2}^{11}(t)$.

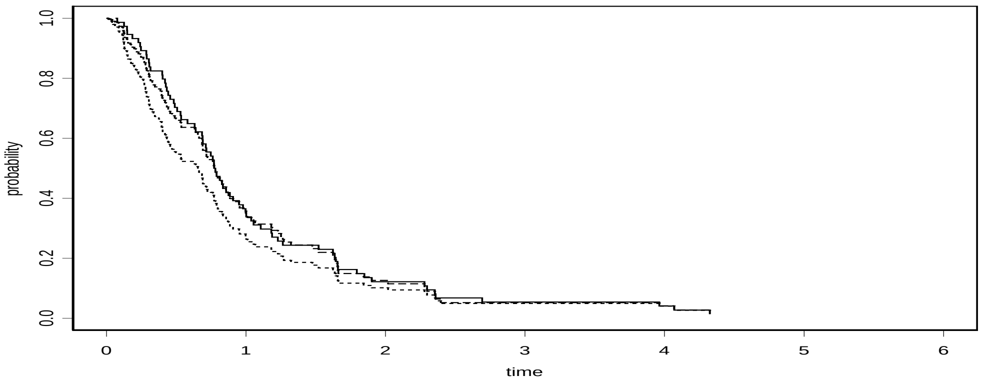

Figure 4-2: Diagnostic plot for model selection based on $F_Y^{11}(t)$ when the true model is Clayton. model 1: Clayton and model 2: Frank. solid line: $\hat{F}_Y^{11}(t)$; dashed line: $\tilde{F}_{Y,1}^{11}(t)$; dotted line: $\tilde{F}_{Y,2}^{11}(t)$.

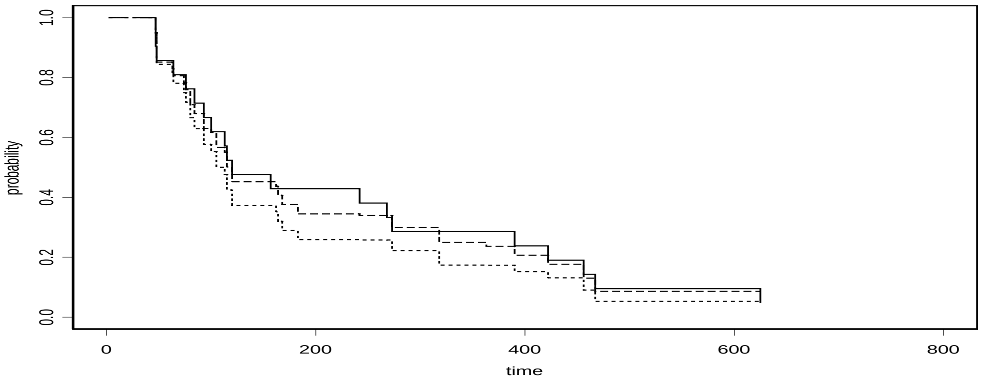

Figure 4-3: Diagnostic plot for the AML high risk group in bone marrow transplant data based on $F_X^{11}(t)$. solid line: $\hat{F}_X^{11}(t)$; dashed line: Clayton; dotted line: Frank.

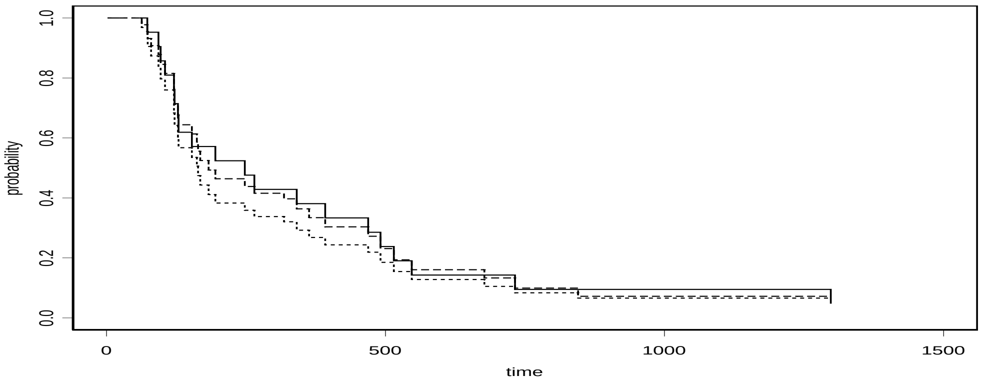

Figure 4-4: Diagnostic plot for the AML high risk group in bone marrow transplant data based on $F_Y^{11}(t)$. solid line: $\hat{F}_Y^{11}(t)$; dashed line: Clayton; dotted line: Frank.

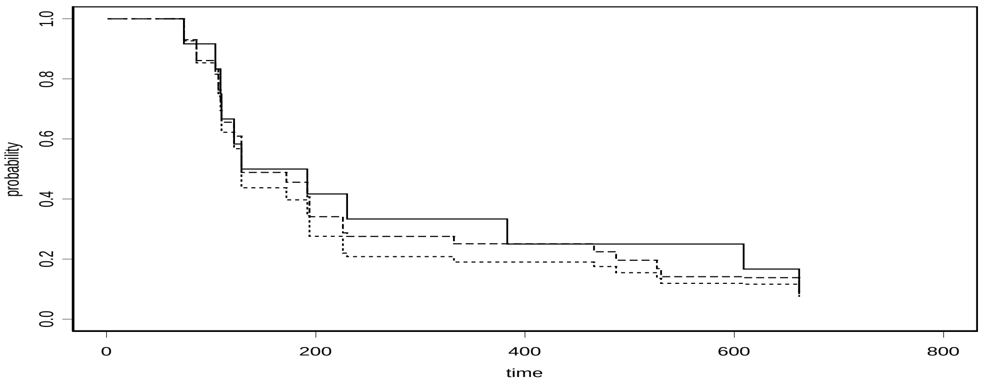

Figure 4-5: Diagnostic plot for the ALL group in bone marrow transplant data based on $F_X^{11}(t)$. solid line: $\hat{F}_X^{11}(t)$; dashed line: Clayton; dotted line: Frank.

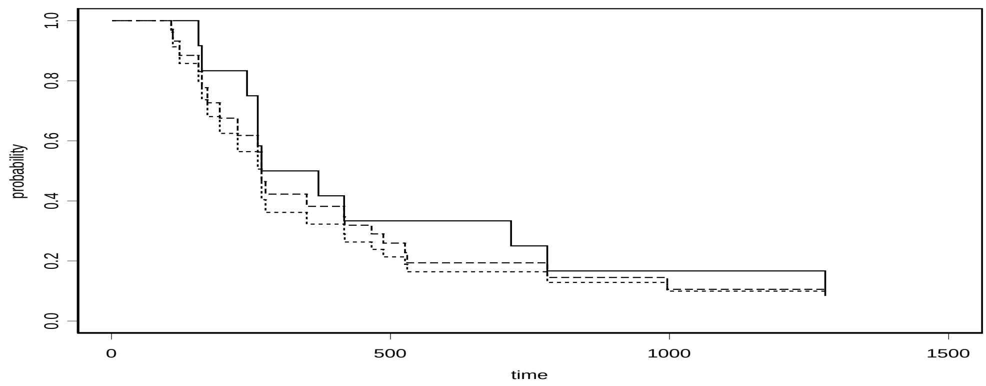

Figure 4-6: Diagnostic plot for the ALL group in bone marrow transplant data based on $F_Y^{11}(t)$. solid line: $\hat{F}_Y^{11}(t)$; dashed line: Clayton; dotted line: Frank.

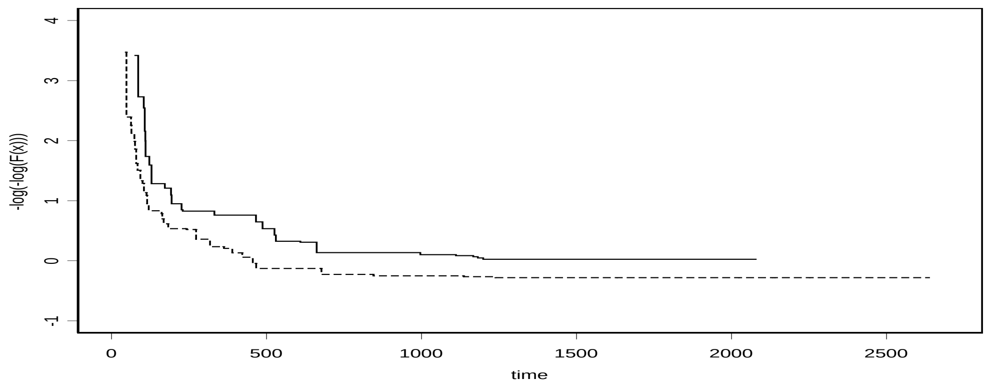

Figure 4-7: Log-Log plot for the two groups. solid line: ALL group; dashed line: AML high-risk group.

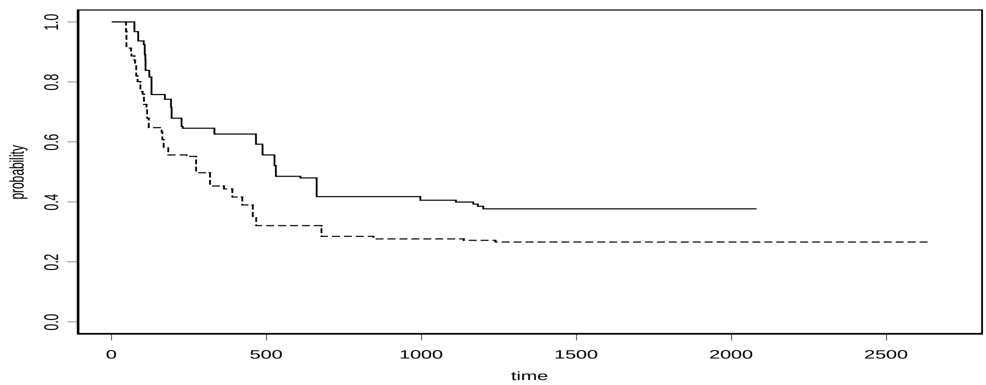

Figure 4-8: $\hat{F}_x(t)$ for the two groups. solid line: ALL group; dashed line: AML high-risk group.