Published in final edited form as:

J R Stat Soc Series B Stat Methodol. 2012 March 1; 74(2): 183–202. doi:10.1111/j.
1467-9868.2011.01012.x.

Cumulative Incidence Association Models for Bivariate Competing Risks Data

Yu Cheng and

Department of Statistics and Department of Psychiatry, University of Pittsburgh Pittsburgh, PA, USA

Jason P. Fine

Department of Biostatistics and Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Summary

Association models, like frailty and copula models, are frequently used to analyze clustered survival data and evaluate within-cluster associations. The assumption of noninformative censoring is commonly applied to these models, though it may not be true in many situations. In this paper, we consider bivariate competing risk data and focus on association models specified for the bivariate cumulative incidence function (CIF), a nonparametrically identifiable quantity. Copula models are proposed which relate the bivariate CIF to its corresponding univariate CIFs, similarly to independently right censored data, and accommodate frailty models for the bivariate CIF. Two estimating equations are developed to estimate the association parameter, permitting the univariate CIFs to be estimated either parametrically or nonparametrically. Goodness-of-fit tests are presented for formally evaluating the parametric models. Both estimators perform well with moderate sample sizes in simulation studies. The practical use of the methodology is illustrated in an analysis of dementia associations.

1. Introduction

In survival analysis, the association amongst failure times is often of interest, with primary focus on pairwise associations. Bivariate survival data arise frequently in genetic family studies, demography and actuarial science. As a motivating example, we consider a study conducted in Cache County, Utah on dementia in aging population. For each pair comprising a mother and her eldest child, their onset ages of dementia, ages at death or ages at the termination of the study, whichever occurred first, were recorded along with the cause indicators. One of the goals of this study is to characterize the association in dementia among family members induced by common genetic or environmental factors. Many researchers have studied association analyses for bivariate survival data with independent right censoring; see Hougaard (2000) for an overview. However, these methods cannot be applied directly to analyze the between-family-members association in dementia onset with the presence of competing risks censoring from death that may occur prior to dementia onset.

Early work on bivariate competing risks data focused either implicitly or explicitly on association models for the “net” bivariate survival function of a particular cause of interest, in which association is interpreted in a hypothetical reality in which other causes of failure have been removed. For example, Wienke et al. (2003) studied genetic influence on death by respiratory diseases using Danish twins data, naively adopting a correlated gamma frailty model for the distribution of death from respiratory disease within families. In their analysis, failure from other causes was implicitly assumed to be independent of death due to

respiratory diseases. Tsiatis (1975) and Pruitt (1993) demonstrated theoretically that with univariate and bivariate competing risks data, the within-subject cause-specific dependence is nonparametrically nonidentifiable. Thus, the validity of the parametric analysis in Wienke et al. (2003) is unclear, owing to untestable violations of the key independence assumptions within family members for the “net” model. To avoid these practical and theoretical difficulties, recent work on association analyses for bivariate competing risks data has focused on the cause-specific hazard (CSH) functions and cumulative incidence functions (CIFs).

The CSHs and CIFs are nonparametrically estimable with competing risks and have been widely adopted in practice, as advocated in the seminal paper by Prentice et al (1978). The CSH captures the instantaneous risk for a particular cause, while the CIF quantifies the cumulative risk of an event, providing the correct probability of a cause-specific event in the current reality where all causes exist. Adapting the univariate CSH as suggested in Section 12.6 of Hougaard (2000), Bandeen-Roche and Liang (2002) and Bandeen-Roche and Ning (2008) extended Oakes (1989)’s cross hazard ratio to bivariate competing risks data. Cheng and Fine (2008) defined a CSH association measure which is the ratio of the bivariate hazard with cause-specific events from both subjects to the products of the bivariate hazards with a single event from each subject. The two association measures are equivalent and may be estimated parametrically on finite time regions.

In many applications, the cumulative risk of an event, as captured by the CIF, may be of greater interest than the instantaneous risk in the CSH. As noted by Cheng et al.(2007), the bivariate associations in the CIFs may be quite different from the bivariate associations in the CSHs and one may be misled by association analyses of the CSH. Working with CIF, Cheng et al. (2007) proposed a time-varying association measure and implemented nonparametric inferences, in which the measure is completely unspecified over time. However, the time-varying CIF association estimates were found to be unstable at time points with small numbers of events. Both semiparametric and parametric models would be practically useful, yielding estimators which are less sensitive in datasets with relatively small numbers of events. To our knowledge, such CIF association models have not been formally studied for pairwise associations, in part because of the complexity of the likelihood function for bivariate competing risks data.

Prentice et al. (1978) observed that the likelihood function for univariate competing risks data is completely specified by CSH functions and factors into a separate component for each cause. The simplicity of the resulting analyses, in which censoring from competing causes may naively be treated as independent, has contributed to the popularity of such analyses. In Section 2.1, it is shown that for bivariate competing risks data, the likelihood does not generally factor into a separate component for each bivariate CSH function. Hence, unlike the univariate case, inference about a particular bivariate CSH cannot be obtained by naively treating other causes as independent censoring events. The implication is that there are no longer computational or inferential advantages in specifying association models via bivariate CSH, as opposed to bivariate CIF.

Frailty and copula models have a long history as models for bivariate distributions (Clayton and Cuzick, 1985; Genest and Mackay, 1986a,b; Genest and Rivest, 1993; Hougaard, 1987; Oakes, 1989; Joe, 1997; Nelsen, 1999). They are, however, seldom used in competing risks settings because of the challenge imposed by dependence structure between the risk of interest and the competing risks. With univariate data, there have been attempts to specify models for the “net” cause specific survival functions. Such analyses are complicated, in part, because all model parameters must be fit jointly using the likelihood. Fine et al. (2001) considered the so-called “semi-competing risks data”, where the failure time X is subject to

censoring by the competing failure time Y , but not vice versa. They assumed that the joint distribution of (X, Y) satisfies the gamma frailty copula, with the marginal distribution of Y and subdistribution of X unspecified. A simple plug in estimator for the association parameter was presented. This copula model was extended by Wang (2003). Jiang et al. (2006) studied pseudo self-consistent estimation of the gamma frailty model for (X, Y) . Unlike this earlier work, our focus is association models for the CIF, and not the “net” failure probabilities for X and Y .

We investigate between-subject associations of the competing risks data, instead of examining the dependence structure between the latent failure times from the target and competing events. That is, the interest is in the association between (T_1, ϵ_1) and (T_2, ϵ_2) , where T_j is the time to first event, either dementia or death in our illustrative example, and ϵ_j is the cause indicator assuming dementia the cause 1 event and death the cause 2 event for the j th subject in a pair. Without loss of generality, we assume there is only a single competing event, as multiple competing risks can be grouped together without affecting the validity of our model. The CIF for individual j , cause k , is $P(T_j \leq t_j, \epsilon_j = k) = F_{j(k)}(t_j)$. Adopting a frailty modelling strategy, one might assume that $P(T_j \leq t_j, \epsilon_j = k | W_k) = \{B_{j(k)}(t_j)\}^{W_k}$, $j, k = 1, 2$, where $B_{j(k)}$ are some base cumulative incidence functions and W_k is a random effect. Then,

$$P(T_j \leq t_j, \epsilon_j = k) = F_{j(k)}(t_j) = \int \{B_{j(k)}(t_j)\}^w dF_{W_k}(w) = p_k \left[-\log \{B_{j(k)}(t_j)\} \right],$$

where F_{W_k} is the cumulative distribution function of W_k involving an unknown parameter θ_k and $p_k(u)$ is the Laplace transformation of W_k . It follows that the bivariate CIF

$$F_{kk}(t_1, t_2) = P(T_{1(k)} \leq t_1, T_{2(k)} \leq t_2, \epsilon_1 = \epsilon_2 = k) = p_k \left\{ p_k^{-1}(F_{1(k)}(t_1)) + p_k^{-1}(F_{2(k)}(t_2)) \right\}. \quad (1)$$

It is easily seen that bivariate CIFs generated by frailty models are a subclass of the Archimedean copulas studied by Genest and Mackay (1986a,b), though $F_{1(k)}$ and $F_{2(k)}$ are usually the distribution functions of some proper random variables. The model $^{***}(1)$ is a natural generalization of frailty and copula models for standard bivariate survival data without competing risks. In the sequel, we focus our discussion on cause 1 quantities with $k = 1$ in (1), although the methods are directly applicable to other causes and to cross-cause associations. For ease of notation, throughout this paper, we refer to p_1 as p .

We propose a two stage approach to estimation in (1). The univariate CIFs are first estimated separately from the association parameter by either using nonparametric or parametric methods, as discussed in Section 2.2. Next, the association parameter θ in p may be computed using either of the two estimating equations proposed in Section 2.3. The asymptotic properties of the procedure are given in Section 2.4. A simple goodness-of-fit test is given in Section 2.5. The two estimators are compared by simulation studies in Section 3. The proposed methods are applied to the Cache County Study in Section 4, with a comparison of parametric and nonparametric association models provided in Section 4.2. A discussion concludes in Section 5.

2. Modelling Association

2.1. Full likelihood method

In practice, both the target and competing events may be subject to independent censoring, e.g., by the termination of the study. Hence for j th subject in a pair, one observes Y_j which is the minimum of first event time and censoring time and η_j equals to the cause indicator if

either event occurs first or 0 if both events are right censored. A bivariate competing risks data set containing n pairs is denoted by $\{(Y_{1i}, \eta_{1i}, Y_{2i}, \eta_{2i}), i = 1, \dots, n\}$.

For the bivariate case, we construct the likelihood function based on bivariate CSH functions and a bivariate overall survival function as following. We first define bivariate CSH functions with double events from both subjects. For $k, k' = 1, 2$,

$$\Lambda_{kk'}(du, dv) = P\{T_1 \in [u, u+du), T_2 \in [v, v+dv), \epsilon_1=k, \epsilon_2=k' | (T_1, T_2) \geq (u, v)\},$$

The bivariate CSHs with a single event from one of the subjects are

$$\begin{aligned}\Lambda_{k0}(du, v-) &= P\{T_1 \in [u, u+du), \epsilon_1=k | (T_1, T_2) \geq (u, v)\}, \quad k=1, 2, \text{ and} \\ \Lambda_{0k'}(u-, dv) &= P\{T_2 \in [v, v+dv), \epsilon_2=k' | (T_1, T_2) \geq (u, v)\}, \quad k'=1, 2.\end{aligned}$$

For simplicity, we assume continuous failure times, where $\Lambda_{kk'}(du, dv) = \lambda_{kk'}(u, v)dudv$, $\Lambda_{k0}(du, v-) = \lambda_{k0}(u, v-)du$ and $\Lambda_{0k'}(u-, dv) = \lambda_{0k'}(u-, v)dv$ for $k, k' = 1, 2$. The bivariate overall survival function is $S(u, v) = P(T_1 > u, T_2 > v)$.

The likelihood function for bivariate competing risks data is

$$\begin{aligned}\mathcal{L} = \prod_{i=1}^n & \left[\{\lambda_{11}(y_{i1}, y_{i2}) S(y_{i1}-, y_{i2}-)\}^{I_{\{\eta_{i1}=1, \eta_{i2}=1\}}} \times \{\lambda_{12}(y_{i1}, y_{i2}) S(y_{i1}-, y_{i2}-)\}^{I_{\{\eta_{i1}=1, \eta_{i2}=2\}}} \right. \\ & \times \{\lambda_{21}(y_{i1}, y_{i2}) S(y_{i1}-, y_{i2}-)\}^{I_{\{\eta_{i1}=2, \eta_{i2}=1\}}} \times \{\lambda_{22}(y_{i1}, y_{i2}) S(y_{i1}-, y_{i2}-)\}^{I_{\{\eta_{i1}=2, \eta_{i2}=2\}}} \\ & \times \{\lambda_{10}(y_{i1}, y_{i2}) S(y_{i1}-, y_{i2}-)\}^{I_{\{\eta_{i1}=1, \eta_{i2}=0\}}} \times \{\lambda_{20}(y_{i1}, y_{i2}) S(y_{i1}-, y_{i2}-)\}^{I_{\{\eta_{i1}=2, \eta_{i2}=0\}}} \\ & \times \{\lambda_{01}(y_{i1}, y_{i2}) S(y_{i1}, y_{i2}-)\}^{I_{\{\eta_{i1}=0, \eta_{i2}=1\}}} \times \{\lambda_{02}(y_{i1}, y_{i2}) S(y_{i1}, y_{i2}-)\}^{I_{\{\eta_{i1}=0, \eta_{i2}=2\}}} \\ & \left. \times \{S(y_{i1}, y_{i2})\}^{I_{\{\eta_{i1}=0, \eta_{i2}=0\}}} \right].\end{aligned}\quad (2)$$

Here we adopt the Dabrowska (1988)'s representation for the bivariate survival function

$$S(y_1, y_2) = \prod_{[0, y_1]} (1 - \Lambda_{10}^*(du, 0)) \prod_{[0, y_2]} (1 - \Lambda_{01}^*(0, dv)) \prod_{[0, y]} (1 - L(du, dv)), \text{ where}$$

Π is the product integral, $[0, y] = [0, y_1] \times [0, y_2]$, and

$$L(y) = \int_{[0, y]} \frac{\Lambda_{10}^*(du, v-) \Lambda_{01}^*(u-, dv) - \Lambda_{11}^*(du, dv)}{(1 - \Lambda_{10}^*(\Delta u, v-)) \cdot (1 - \Lambda_{01}^*(u-, \Delta v))} \text{ in which } \Lambda_{10}^*, \Lambda_{01}^* \text{ and } \Lambda_{11}^* \text{ are the}$$

bivariate cumulative hazard functions corresponding to T_1 and T_2 . That is,

$$\Lambda_{10}^* = \Lambda_{10} + \Lambda_{20}, \Lambda_{01}^* = \Lambda_{01} + \Lambda_{02} \text{ and } \Lambda_{11}^* = \Lambda_{11} + \Lambda_{12} + \Lambda_{21} + \Lambda_{22}.$$

Upon inspection, one notes that $\mathcal{L}$ involves only $(\Lambda_{11}, \Lambda_{12}, \Lambda_{21}, \Lambda_{22}, \Lambda_{10}, \Lambda_{20}, \Lambda_{01}, \Lambda_{02})$. However, the likelihood does not factor into a separate component for each bivariate CSH function, as it does with univariate data. Hence, we generally cannot use the likelihood to estimate a particular CSH function while treating failures from other causes as censoring events. If the likelihood is employed, it requires specifying all cause-specific and cross-cause association models, which is much more complicated than the pairwise model (1). Rewriting $\mathcal{L}$ in terms of the bivariate CIFs also leads to complex estimation, as the likelihood does not factor across bivariate CIFs, similarly to univariate data when fitting models to CIFs.

2.2. Nonparametric and parametric estimators of univariate CIFs

The univariate CIFs with respect to cause 1 $F_{j(1)}, j = 1, 2$ can be estimated by both nonparametric and parametric methods.

If one utilizes a nonparametric estimator of the univariate CIFs in (1), a semiparametric model is obtained. We use a standard counting process estimator for univariate data (Gray, 1988). Let $N_j(t) = I(Y_j \leq t, \eta_j = 1)$, $H_j(t) = I(Y_j \geq t)$ and let PF denote the expectation of a

random function F . One can express $\widehat{F}_{j(1)} = \int \widehat{\Lambda}_{j(1)}(du) \widehat{S}_j(u-)$, where $\widehat{\Lambda}_{j(1)}(t) = \int_0^t \frac{\sum_j N_j(u)}{\sum_j H_j(u)}$ is a Nelson-Aalen estimator for the cause specific hazard for cause 1 and $\widehat{S}_j$ is the Kaplan-Meier estimator for the all cause survival function for subject $j (= 1, 2)$.

If one considers a parametric model for univariate CIFs, coupled with (1), a parametric model is obtained for the bivariate CIF. Without loss of generality, we assume $F_{j(1)}(s)$ satisfies a model depending on a finite dimensional parameter, γ_j , $j = 1, 2$. Jeong and Fine (2006) presented a parametric approach to modelling univariate CIFs using a generalized Gompertz function. For subject j , the cause 1 CIF is given below:

$$F_{j(1)}^{Gom}(s; \alpha_j, \beta_j) = 1 - \exp\left[\beta_j \left\{1 - \exp(\alpha_j s)\right\} / \alpha_j\right], \quad (3)$$

where α_j and β_j are two parameters. When $\alpha_j < 0$ and $\beta_j > 0$, it describes a CIF with the upper asymptote $1 - \exp(\beta_j / \alpha_j) < 1$. This generalized Gompertz model is flexible, however, it cannot provide an adequate fit to a CIF of sigmoidal shape, as in the data analysis in Section 4, where the nonparametric cumulative incidence estimates for dementia have two bend points. Therefore, we also explore a recently developed three-parameter logistic model (Cheng, 2009) which provides better fit to sigmoidal curves. For $j (= 1, 2)$, the model is

$$F_{j(1)}^{Log}(s; b_j, c_j, p_j) = \frac{p_j \exp\{b_j(s - c_j)\} - p_j \exp(-b_j c_j)}{1 + \exp\{b_j(s - c_j)\}}, \quad (4)$$

where the parameter p_j controls the upper asymptote, and b_j and c_j represent the dilation/contraction and shift in the time scale. The unknown parameters, γ_j , can be estimated using univariate maximum likelihood, for $j = 1, 2$ (Jeong and Fine, 2006; Cheng, 2009).

2.3. Estimating the association parameter

Two approaches are presented for estimating the association parameter, utilizing the estimates of the univariate CIFs from the previous subsection.

The basic idea of the first approach is to estimate θ by minimizing the distance between $\tilde{F}_{11}$ and a nonparametric estimator of F_{11} , denoted by $\tilde{F}_{11}(s, t)$. Nonparametric estimation of the bivariate CIF is first studied in Cheng et al. (2007) and a simple nonparametric estimator given therein is used below.

Typically, p in (1) is an explicit function of the association parameter θ , as in most frailty and copula models. Accordingly, one may obtain a model-restricted estimator of $F_{11}(s, t)$, denoted by $\tilde{F}_{11}^{(\cdot)}(s, t; \theta)$, with $F_{1(1)}$ and $F_{2(1)}$ in (1) replaced by their corresponding nonparametric or parametric estimators. The notation $\tilde{F}_{11}^{(\cdot)}$ is intended to be generic: it refers to $\tilde{F}_{11}^{NP}$, $\tilde{F}_{11}^{Gom}$ and $\tilde{F}_{11}^{Log}$ employing either the nonparametric estimators of the univariate CIFs or the corresponding parametric estimates based on $\widehat{\gamma}$, for example, either the Gompertz model or the modified logistic model. After replacing the univariate CIFs with their estimates, the only unknown quantity is θ .

The estimating function for θ is the integral of the weighted differences between the nonparametric and model-restricted estimators of the bivariate CIF. The estimator for θ , $\hat{\theta}_n^{(\cdot)}$, solves:

$$\Psi_{1n}^{(\cdot)}(\theta) = \int_0^{\tau_1} \int_0^{\tau_2} \widehat{w}(s, t) \left\{ \widehat{F}_{11}(s, t) - \tilde{F}_{11}^{(\cdot)}(s, t; \theta) \right\} ds dt = 0, \quad (5)$$

where $\tau = (\tau_1, \tau_2)$ satisfies $S(\tau_1, \tau_2) > 0$, $\widehat{w}(s, t)$ is a weight function with $\int_{[0, \tau]} \widehat{w}(s, t) ds dt = 1$, $\sup_{(s, t) \in [0, \tau]} |\widehat{w}(s, t) - w(s, t)| \xrightarrow{P} 0$ and $\sqrt{n}(\widehat{w} - w)(s, t) = O_p(1)$, where $w(s, t)$ is a bounded deterministic function. The notation $\hat{\theta}_n^{(\cdot)}$ is analogous to $\hat{F}_{11}^{(\cdot)}$, where $\cdot$ generically denotes the method of estimation for the univariate CIFs.

A second approach to estimating the association parameter θ is based on a bivariate Doob-Meyer decomposition. In this approach, a bivariate intensity function is defined with respect to a bivariate counting process, which counts the number of pairs in which both members experience the event of interest. The bivariate intensity involves both the bivariate CIF for the cause of interest and the overall bivariate survival function. After estimating the bivariate survival function nonparametrically, an estimating function is constructed for the association parameter using an integrated weighted difference, similarly to $\psi_{1n}(\theta)$.

Consider the filtration

$\mathcal{F}_{s,t}^* = \sigma \left\{ I(Y_{i1} \leq u, \eta_{i1} = k), I(Y_{i2} \leq v, \eta_{i2} = k'), I(Y_{i1} \geq u), I(Y_{i2} \geq v), u \leq s, v \leq t, i = 1, 2, \dots, n, k, k' = 1, 2 \right\}$. Define the bivariate event process and at-risk process $N_i(s, t) = I(Y_{i1} \leq s, \eta_{i1} = 1, Y_{i2} \leq t, \eta_{i2} = 1)$ and $H_i(s, t) = I(Y_{i1} \geq s, Y_{i2} \geq t)$. The bivariate decomposition

$$\frac{1}{n} \int_0^{\tau_1} \int_0^{\tau_2} \widehat{w}(s, t) \left\{ \sum_i N_i(ds, dt) - \sum_i H_i(s, t) \Lambda_{11}(ds, dt; \theta) \right\} = 0, \quad (6)$$

where $\widehat{w}(s, t)$ is defined in (5). The bivariate hazard function with respect to cause 1 satisfies $\Lambda_{11}(ds, dt; \theta) = F_{11}(ds, dt; \theta) / S(s-, t-)$. Therefore it is natural to estimate $\Lambda_{11}(ds, dt; \theta)$ in the estimating equation (6) by $\widehat{F}_{11}^{(\cdot)}(ds, dt; \theta) / \widehat{S}(s-, t-)$, where $\widehat{S}$ is the Dabrowska estimator of the bivariate survival function (Dabrowska, 1988). This yields the estimating equation:

$$\Psi_{2n}^{(\cdot)}(\theta) = \frac{1}{n} \int_0^{\tau_1} \int_0^{\tau_2} \widehat{w}(s, t) \left\{ \sum_i N_i(ds, dt) - \sum_i H_i(s, t) \tilde{F}_{11}^{(\cdot)}(ds, dt; \theta) / \widehat{S}(s-, t-) \right\} = 0. \quad (7)$$

The solution to the above equation is denoted as $\tilde{\theta}_n^{(\cdot)}$, with $\cdot$ again denoting the method of estimation for the univariate CIFs.

2.4. Large sample properties

The basic idea for proving consistency is to show that as the sample size gets large $\psi_{1n}(\theta)$ and $\psi_{2n}(\theta)$ converge to deterministic functions of θ which have unique solutions at the true value of θ . The asymptotic normality of the resulting estimators involves a careful analysis of the variability of estimators using the construct of the estimating equations. To be more specific, one may view $\hat{\theta}_n^{(\cdot)}$ as a function of the estimates of the univariate CIFs and the nonparametric estimator of F_{11} and view $\tilde{\theta}_n^{(\cdot)}$ as a function of the bivariate counting processes, the univariate CIF estimators, and the nonparametric estimator of S . The variance of the estimators of θ is a complicated function of the variances of these “inputs”.

To be more precise, let $\Psi_1(\theta) = \lim_{n \rightarrow \infty} \Psi_{1n}^{(\cdot)}(\theta)$. Suppose that $\theta_0 \in \Theta \subset \mathbb{R}$ is the true association parameter. Clearly, $\Psi_1(\theta) = \int_0^{\tau_1} \int_0^{\tau_2} w(s, t) \{F_{11}(s, t; \theta_0) - F_{11}(s, t; \theta)\} ds dt$, where w is a deterministic weight function. In Appendix A.1, we establish the following theorem regarding $\widehat{\theta}^{(\cdot)}$:

Theorem 1—Under regularity conditions listed in Appendix A.1, $\widehat{\theta}_n^{(\cdot)}$, the solution to $\Psi_{1n}^{(\cdot)}(\theta) = 0$, is consistent, and $\sqrt{n}(\widehat{\theta}_n^{(\cdot)} - \theta_0)$ converges in distribution to $-\dot{\Psi}_{1, \theta_0}^{-1}(Z_1) = 0$, where $\dot{\Psi}_{1, \theta_0} = \frac{\partial \Psi_1}{\partial \theta}|_{\theta=\theta_0} = - \int_{[0, \tau]} w(s, t) \dot{F}_{11}(s, t; \theta_0) ds dt$ with $\dot{F}_{11}(s, t; \theta_0) = \frac{\partial F_{11}(s, t; \theta)}{\partial \theta}|_{\theta=\theta_0}$ and Z_1 is a mean zero normal random variable.

The consistency result entails showing that $\psi_1(\theta)$ has a unique solution in a neighborhood of θ_0 . The variance of $\widehat{\theta}_n^{(\cdot)}$ can be obtained based on the influence functions given in Appendix A.1. However, evaluation of the influence functions is not straightforward, even for the case where the univariate CIFs are estimated parametrically. This is due to the complexity of the distribution of the nonparametric estimator of $\widehat{F}_{11}$ (Cheng et al, 2007). A general variance estimation strategy which does not require derivation of influence functions on a case by case basis for $\widehat{\theta}_n^{(\cdot)}$ is bootstrapping. The validity of such resampling is supported by Theorem 3.9.11 in van der Vaart and Wellner (2000).

In a similar manner, let $\psi_2(\theta)$ be the asymptotic limit of $\Psi_{2n}^{(\cdot)}(\theta)$. That is, $\Psi_2(\theta) = \int_0^{\tau_1} \int_0^{\tau_2} w(s, t) \{PN(ds, dt) - PH(s, t)F_{11}(ds, dt; \theta)/S(s-, t-)\}$, where P stands for the expectation of an indicator function. The large sample properties of $\widehat{\theta}_n^{(\cdot)}$ as the solution to $\Psi_{2n}^{(\cdot)}(\theta) = 0$ are given in the following theorem.

Theorem 2—Under regularity conditions in Appendix A.2, $\widehat{\theta}_n^{(\cdot)}$ is consistent, and $\sqrt{n}(\widehat{\theta}_n^{(\cdot)} - \theta_0)$ converges in distribution to $-\dot{\Psi}_{2, \theta_0}^{-1}(Z_2)$, where $\dot{\Psi}_{2, \theta_0} = \frac{\partial \Psi_2(\theta)}{\partial \theta}|_{\theta=\theta_0} = - \int_{[0, \tau]} w(s, t) PH(s, t) \frac{\dot{F}_{11}(ds, dt; \theta_0)}{S(s-, t-)}$ and Z_2 is a mean 0 normal random variable.

The proof of Theorem 2 is given in Appendix A.2. The uniqueness of the solution to ψ_2 follows similarly to that in Theorem 1. The asymptotic normality of $\widehat{\theta}_n^{(\cdot)}$ again requires a linear representation for the large sample distribution of this estimator. Since the involved influence functions are quite complicated, owing to their dependence on the Dabrowska estimator $\widehat{S}$, the variance estimator of $\widehat{\theta}_n^{(\cdot)}$ will also be obtained by bootstrapping, justified along the lines in Theorem 1.

Regarding the efficiency of the two estimators, inspecting the influence functions in the Appendix, there is no clear theoretical guidance on which estimator has smaller variance. The choice of weight function in the integrated difference estimating equations is also somewhat unclear. Using unit weights at all time points is a simple choice for the weight function. Another possibility is to downweight at time points where variability is greatest. This can be achieved by using the inverse of the estimated overall survival function.

2.5. Goodness-of-fit test of a frailty model

The specification of the bivariate CIF in terms of univariate CIFs and the association parameter necessitates the need to verify that the model provides an adequate fit to the data. Glidden (1999) proposed some goodness-of-fit procedures to check the adequacy of the gamma frailty model for multivariate failure times, under the assumption that all individuals within a cluster are subject to the same gamma frailty for all causes. We cannot apply his procedures directly here, since we only assume a pairwise dependence model for cause 1 CIFs. Furthermore, given the generality of the model (1), a more general procedure is required to check the goodness-of-fit of other association models, such as the Frank model (Frank, 1979; Genest, 1987) and the positive stable model (Hougaard, 1984, 1986a,b). Shih (1998) proposed a goodness-of-fit test for the Clayton model (Clayton, 1978; Clayton and Cuzick, 1985) by comparing the unweighted and weighted concordance estimators of the association parameter. The idea is generally applicable and may be adopted in our set-up.

If the model (1) is correctly specified, then the two estimating equations should both provide consistent estimates of θ . If the model is misspecified, then $\widehat{\theta}_n^{(\cdot)}$ and $\tilde{\theta}_n^{(\cdot)}$ should converge to different values. The proposed test exploits this fact.

We construct a goodness-of-fit test based on the difference of the two estimators. Let

$$z^{(\cdot)} = \frac{\widehat{\theta}_n^{(\cdot)} - \tilde{\theta}_n^{(\cdot)}}{\text{se}(\widehat{\theta}_n^{(\cdot)} - \tilde{\theta}_n^{(\cdot)})},$$
 where $\text{se}(\widehat{\theta}_n^{(\cdot)} - \tilde{\theta}_n^{(\cdot)})$ denotes the standard error of the difference between the two estimators. The asymptotic normality of $\widehat{\theta}_n^{(\cdot)} - \tilde{\theta}_n^{(\cdot)}$ can be established straightforwardly based on the asymptotic normality of $\widehat{\theta}_n^{(\cdot)}$ and $\tilde{\theta}_n^{(\cdot)}$. Hence, the z statistic can be approximated by a standard normal distribution. Following Theorems 1 and 2, the explicit form of $\text{se}(\widehat{\theta}_n^{(\cdot)} - \tilde{\theta}_n^{(\cdot)})$ is complicated but may be estimated using the bootstrap.

3. Simulations

We begin by simulating from the Clayton model (Clayton, 1978; Clayton and Cuzick,

1985). We assume that the frailty for cause 1, $W_1 \sim \text{gamma}\left(\frac{1}{\theta-1}, 1\right)$, $\theta > 1$. The corresponding Laplace transformation function $p(u) = (u+1)^{1/(1-\theta)}$ and its inverse function $q(u) = p^{-1}(u) = u^{1-\theta} - 1$. Following Cheng et al. (2009), we first generate potentially infinite random variables, $T_{1(1)}^*$ and $T_{2(1)}^*$ from $F_{11}(s, t) = p\{-\log(B_{1(1)}(s)) - \log(B_{2(1)}(t))\} = p\{p^{-1}(F_{1(1)}(s)) + p^{-1}(F_{2(1)}(t))\}$, where $-\log(B_{1(1)}(s)) = q(F_{1(1)}(s))$. This gives $P(T_j \leq s, \epsilon_j = 1|W = w) = \{B_{1(1)}(s)\}^w = \exp\{-w \cdot q(F_{1(1)}(s))\}$, with $y(x) = F_{1(1)}^{-1}[p(-\log(x)/w)]$. One may simulate $T_{1(1)}^*$ by $F_{1(1)}^{-1}[p(-\log(U)/w)]$, where w is drawn from the $\text{gamma}\left(\frac{1}{\theta-1}, 1\right)$ and U is drawn from the uniform (0,1) distribution. One generates $T_{2(1)}^*$ similarly using same frailty w but independent uniform(0,1).

When $T_{j(1)}^* = \infty$, $j = 1, 2$, it implies that $T_{j(2)}^* = T_j$ and $\epsilon_j = 2$. In this case, we simulate $T_{j(2)}^*$ directly from the conditional distribution of T_j given $\epsilon_j = 2$. When $T_{j(1)}^* < \infty < \infty$, then we let $T_{j(2)}^* = \infty$ and $\epsilon_j = 1$. Therefore, $T_j = T_{j(1)}^* \wedge T_{j(2)}^*$ and $\epsilon_j = I\{T_j = T_{j(1)}^*\} + 2I\{T_j = T_{j(2)}^*\}$.

Based on these data, we obtain the bivariate observation $(Y_1, \eta_1, Y_2, \eta_2)$, with $T_{j(1)}^* \wedge T_{j(2)}^* \wedge C_j$ and $\eta_j = I\{Y_j = T_{j(1)}^*\} + 2I\{Y_j = T_{j(2)}^*\}$ for $j = 1, 2$, where (C_1, C_2) are independently drawn from some bivariate distribution.

In this simulation study we let $F_{j(1)}(t) = \frac{4}{5}(1 - e^{-5t/4})$ and $F_{j(2)}(t) = \frac{1}{5}(1 - e^{-5t/4})$ for $j = 1, 2$. $N = 200$ and 400 pairs of $(T_{1(1)}^*, T_{2(1)}^*)$ are generated from the Clayton model with frailty parameter θ and equal marginals $F_{1(1)}, T_{j(2)}^*$ are simulated from the conditional distribution $P(T_j \leq t | C_j = 2) = 1 - e^{-5t/4}$ when $T_{j(1)}^* = \infty$. Otherwise $T_{j(2)}^* = \infty$. The independent censoring times C_1 and C_2 are drawn from the Uniform $(0, 5)$. The “observed” times are calculated as $Y_j = T_{j(1)}^* \wedge T_{j(2)}^* \wedge C_j$, as well as risk indicators $\eta_j = I\{Y_j = T_{j(1)}^*\} + 2I\{Y_j = T_{j(2)}^*\}$ for $j = 1, 2$.

For each data set, we estimate the univariate and bivariate CIFs nonparametrically, denoted as $\widehat{F}_{1(1)}(s)$, $\widehat{F}_{2(1)}(t)$ and $\widehat{F}_{11}(s, t)$. Under the Clayton model, the bivariate CIF can also be estimated semiparametrically for given θ as below:

$$\tilde{F}_{11}^{NP}(s, t; \theta) = \left[\left\{ \widehat{F}_{1(1)}(s) \right\}^{1-\theta} + \left\{ \widehat{F}_{2(1)}(t) \right\}^{1-\theta} - 1 \right]^{\frac{1}{1-\theta}}. \quad (8)$$

The nonparametric estimator $\widehat{F}_{11}$ and the model-restricted estimator $\tilde{F}_{11}^{NP}$ are plugged into (5) to obtain $\widehat{\theta}_n^{NP}$. On the other hand, $\tilde{F}_{11}^{NP}(s, t; \theta)$ is differentiated with respect to s and t and plugged into (7) to obtain the estimator $\tilde{\theta}_n^{NP}$.

Here we took unit weights, with $\widehat{w}(s, t) = 0$ at time points with $\widehat{S}(s, t) < 0.05$, since at those particularly large time points, the at-risk sets are small and the variations of the CIF estimators are large. As $\theta > 1$, to improve asymptotic normality, the simulations are performed in the log scale. Five hundred data sets are simulated and the average, the empirical standard errors, the bootstrap standard errors of $\widehat{\theta}_n^{NP}$ and $\tilde{\theta}_n^{NP}$ as well as the empirical coverages of the true parameter in their 95% confidence intervals are reported in the upper panel of Table 1.

We also simulate random variables from the positive stable model with varying association parameters following the same simulation strategy as above. We first obtain a pair of random variables $(V_1, V_2 = (p(-\log(U_1)/w), p(-\log(U_2)/w))$ by using the R function *rcopula*, where U_1 and U_2 are uniformly distributed random variables, p is the Laplace transformation of the positive stable distribution, w is a random variable generated from the positive stable distribution; see Yan (2007) for details. Next, the random variables $(T_{1(1)}^*, T_{2(1)}^*)$ are simulated as $(F_{1(1)}^{-1}(V_1), F_{1(1)}^{-1}(V_2))$. Note that $T_{j(2)}^*, j = 1, 2$, are simulated from the same conditional distribution as described above when $T_{j(1)}^* = \infty$. Independent censoring times are generated from the Uniform $(0, 8)$. Similar estimation procedures to those above are employed. The results are summarized in the lower panel of Table 1.

Both $\widehat{\theta}_n^{NP}$ and $\tilde{\theta}_n^{NP}$ are close to the true θ even with moderate sample sizes for both the Clayton and positive stable models. The empirical standard errors agree with the bootstrap standard errors. It is noted that $\widehat{\theta}_n^{NP}$ has smaller variance than $\tilde{\theta}_n^{NP}$. The integrand in (5) is the difference of the two estimators of the bivariate CIF, while that of (7) is the difference between the double cause 1 event process and its compensator; the latter may have larger

variation at individual time points, resulting in larger variation in $\tilde{\theta}_n^{NP}$ than in $\widehat{\theta}_n^{NP}$. However, the empirical coverages for both estimators are generally close to the nominal level 0.95. These results demonstrate the estimators' good performance when the association model is correctly specified.

4. Cache County Study

4.1. Association models for CIF

We now consider the data from the Cache County Study of Dementia in aging population. The event of interest is dementia onset, which may be unobserved due to the occurrence of death. The information on the dementia onset age or age at death was recorded for each subject along with the cause indicator. Both events, the occurrence of dementia or death, were subject to independent censoring imposed by the end of study. That is, if a subject was alive and free of dementia by the end of study, his/her age at the end of study was recorded instead. The data were collected on families, and the association of dementia among family members is of interest. Following Bandeen-Roche and Liang (2002) and Cheng et al. (2007), we analyze pairs comprising the mother and oldest child, discarding information from other children. We also exclude pairs either with missing data or whose member died or became demented prior to age 55, resulting in 3635 pairs. These pairs are summarized in the table (2).

Some commonly used copula models, such as the Clayton model (Clayton, 1978; Clayton and Cuzick, 1985), the Frank model (Frank, 1979; Genest, 1987) and the positive stable model (Hougaard, 1986a) are used to specify the bivariate CIF. The model expressions are given in Table 3. The same notation θ is used for the association parameter in each model. As pointed out by Shih and Louis (1995), different frailty distributions may induce quite different dependence structures, where the magnitude and range of θ may differ. The Clayton model characterizes positive associations between two marginals with $\theta \rightarrow 1$ corresponding to independence. Frank's distributions are positive likelihood ratio dependent (Lehmann, 1966) when $0 < \theta < 1$, are negative dependent when $\theta > 1$ and approach independence as $\theta \rightarrow 1$ (Genest, 1987). The positive stable frailty model captures positive dependence structure between two marginals when $0 < \theta < 1$ and independence structure when $\theta = 1$ (Hougaard, 2000). In Section 4.2, the models are compared using model independent association measures for bivariate CIF.

For each model, we compute the univariate CIFs either nonparametrically or parametrically assuming the modified logistic model and the Gompertz model. We then estimate the association parameter $\tilde{\theta}_n^{(\cdot)}$ based on the integrated difference of the nonparametric and the model-restricted estimators of the bivariate CIF, and $\tilde{\theta}_n^{(\cdot)}$ based on the Doob-Meyer decomposition. Uniform weights are used over the time points where the estimated overall survival function $\widehat{S}(s, t)$ is at least 0.05. The bootstrap standard errors of the estimators and the p values from the goodness-of-fit test are computed based on 2500 bootstrap samples. The results are reported in Table 3.

For $\widehat{\theta}_n^{NP}$ and $\tilde{\theta}_n^{NP}$ where the univariate CIFs are estimated nonparametrically, the Frank model and the positive stable model suggest lack of fit. The two estimates are noticeably different and the p values of the goodness-of-fit tests are less than 0.10. On the other hand, $\widehat{\theta}_n^{NP}$ and $\tilde{\theta}_n^{NP}$ from the Clayton model are in agreement and the goodness-of-fit p-value is large. We observe similar results for $\widehat{\theta}_n^{Log}$ and $\tilde{\theta}_n^{Log}$.

As demonstrated in Cheng (2009), the modified logistic model provides an adequate fit to the Cache County Study data in describing a mother and her eldest child's cumulative incidences of dementia. Consequently, the estimates and standard errors from this model agree well with those based on the nonparametric method. The Gompertz model, however, fits the data poorly (Cheng, 2009), resulting in unreliable estimates of univariate CIFs. The association estimates derived from this model are either unobtainable or rather different compared with those from the nonparametric method and the modified logistic model.

Based on the results from the nonparametric and modified logistic methods for univariate distributions, we conclude that the Clayton model is most appropriate among the three frailty models to characterize the mother-child association in cumulative probabilities of dementia and the overall strength of familial association in dementia is mild to moderately strong and statistically significant, with 0.95 confidence interval (1.06, 1.22).

As suggested by the referee, we conducted a simulation study to evaluate the accuracy of the bootstrap standard errors in this practical example. We adopted the Clayton model with the association parameter 1.14 and the modified logistic models for mother and child cumulative incidences of dementia. We assumed that the cumulative incidences of dementia and death have the following forms:

$$F_{j(1)}(s) = \frac{0.13 \exp\{0.19(s-81)\} - 0.13 \exp(-0.19 \times 81)}{1 + \exp\{0.19(s-81)\}} \text{ and} \\ F_{j(2)}(s) = \frac{0.87 \exp\{0.15(s-80)\} - 0.87 \exp(-0.15 \times 80)}{1 + \exp\{0.15(s-80)\}}, j=1, 2,$$

where the parameters were estimated from the data. A sample of 3635 pairs is repeatedly generated, with the event times rounded to integers as in the data. We then obtained the nonparametric estimator $\hat{\theta}_n^{NP}$ and its bootstrap standard error based on 250 resamples. The procedure was repeated 500 times. The empirical mean is 1.14, and the empirical standard error 0.037 is close to the mean bootstrap standard error 0.039. This suggests that the bootstrap error reported in Table 3 is accurate.

4.2. Comparison to previous analyses

The estimate of the bivariate cross hazard ratio for dementia association in Bandeen-Roche and Liang (2002) is 2.98, with standard error 0.66. This result is highly statistically significant relative to the null that there is no association in the CSHs, where the cross hazard ratio is one at all time points. In population based studies like that in Cache County, life-time risks of disease are of particular interest, so that the familial associations observed in the bivariate CSH may not be of direct interest for some researchers.

Cheng et al. (2007) proposed a time-varying association measure $\psi(s, t)$ for bivariate competing risks data which is the ratio of the bivariate CIF to the product of the marginals,

i.e., $\psi(s, t) = \frac{F_{11}(s, t)}{F_{1(1)}(s) F_{2(1)}(t)}$. The association measure was applied to the Cache County Study and estimated based on nonparametric estimation of the bivariate and univariate CIFs

as $\hat{\psi}(s, t) = \frac{\hat{F}_{11}(s, t)}{\hat{F}_{1(1)}(s) \hat{F}_{2(1)}(t)}$. The standard error at each time point was computed from 500 bootstrap samples and the corresponding pointwise confidence interval was constructed based on asymptotic normality of the estimator (Cheng et al., 2007). The estimates $\hat{\psi}(s, t)$ and the 95% pointwise confidence intervals were presented at diagonal time points in the left panel of Figure 1. The estimates $\hat{\psi}(t, t)$ have higher variabilities at early time points due

to the small number of events. In such scenarios, semiparametric or parametric modelling of the association may result in less variable estimators.

The data analysis in Section 4 indicates that the Clayton model fits the Cache County Study data well. We plug the nonparametric estimates for the univariate CIFs into the Clayton model (8) with the estimated association parameter 1.14 to obtain a semiparametric estimate of the bivariate CIF. Then $\psi(s, t)$ can be estimated semiparametrically as the ratio of the semiparametric estimate of the bivariate CIF to the product of nonparametric estimates of

the marginals. That is, $\tilde{\psi}^{NP}(s, t) = \frac{\tilde{F}^{NP}(s, t; 1.14)}{\hat{F}_{1(1)}(s) \hat{F}_{2(1)}(t)}$. The standard errors of $\tilde{\psi}^{NP}(s, t)$ are similarly computed based on 500 bootstrapped samples. The semiparametric estimates $\tilde{\psi}^{NP}(s, t)$ and the pointwise confidence intervals are also plotted at diagonal points on the right panel of Fig 1. The semiparametric estimates are much smoother than the nonparametric counterparts and the variation of the association measure from semiparametric estimation has been substantially reduced at all ages.

It is important to recognize that the parameter θ in the Clayton model corresponds to a cross hazard ratio for a certain bivariate subdistribution hazard ratio defined by the bivariate CIF, and not the usual bivariate cause specific hazard ratio (Bandein-Roche and Liang, 2002). Thus, there is no simple relationship between the estimate of θ under the Clayton model for the bivariate CIF and the bivariate CSH analyses reported above. The exact formula for θ is given by equation (4.6) in Hougaard (p. 136, 2000), with $1 - F_{11}$ in place of the bivariate all cause survival function.

One may also compute average values of $\hat{\psi}(s, t)$ using the fitted parametric and semiparametric models, as done in (Cheng et al., 2007) with the nonparametric estimates. Summarizing over all timepoint pairs gives a mean bivariate CIF association of 4.19 for the Clayton model, 2.33 for the Frank model, 2.83 for the positive stable model, and 3.03 for the nonparametric estimate. The differences in these estimates are not statistically significant. The results suggest a three to fourfold increase in risk for dementia by a given age for a child whose mother developed dementia by that age versus another child whose mother did not.

5. Discussion

In practice, cross-cause association may be of interest and can be incorporated readily by model (1) for the bivariate CIF. A strength of the pairwise model specification is that models may be fit separately to each pair of causes, greatly simplifying the computations and inferences. Efficiency might be gained by using the full likelihood described in Section 2, at the cost of simultaneously fitting models for all bivariate CIF associations. Such an approach has previously been considered for bivariate CSH analyses Bandeen-Roche and Liang (2002). Alternatively, one might consider pseudo-likelihood estimation, as in Shih and Louis (1995), using a two stage approach similar to that presented in the current paper. The difference is that one would still need to fit all pairwise models simultaneously, as with full likelihood estimation. One would first estimate all univariate CIFs. The pseudo-likelihood would then be constructed from the full likelihood with the estimated univariate CIFs replacing the theoretical quantities. The association parameters for all pairwise association models would then be estimated from the pseudo likelihood. The potential gains in efficiency would need to be weighed against the computational and inferential complexity, and potential loss of robustness. Additional work is needed in this area.

In a Bayesian analysis, the full likelihood given in Section 2.1 would be coupled with prior parameters for the cause-specific models described in the full likelihood. Such Bayesian analysis has not been widely studied under direct modelling for CIFs. Additional work is needed to determine suitable prior distributions and to study the associated computations. This Bayesian analysis would appear to be much more complicated than the pairwise method proposed in this paper, owing to the need for simultaneous inference regarding all models.

Copula models are important for analyzing independently censored bivariate survival data, where a bivariate quantity is expressed as a function of the univariate distributions and an association parameter. The model formulation is convenient, in that it permits the corresponding univariate components and the unknown association parameter to be estimated separately. The proposed models for the bivariate CIF are similar in spirit, providing a natural generalization of this modelling strategy to competing risks data.

For low incidence events, the number of events is often small at early time points, as in the Cache County Study. The comparison of nonparametric and semiparametric estimates of the time-varying association measure $\psi(s, t)$ in the data analysis section indicates that the parametric and semiparametric modelling proposed in this paper is more stable than the nonparametric counterpart. As illustrated in our numerical studies, a poor parametric fit to the univariate CIFs may lead to bias in the association estimator. We recommend that goodness-of-fit tests be conducted for parametric modelling of univariate CIFs before estimating the association parameter under those model assumptions.

One should recognize that the consistency of the goodness-of-fit test may depend on the form of the assumed association model. For the test to have good power, the two estimators should converge to different values, with larger power being achieved when the difference between these values is larger. For certain models, it is possible that the two estimators may converge to the same value, even when the model is misspecified. A more careful theoretical study of such issues is needed but is beyond the scope of the current paper.

In the semi-competing risks setting (Fine et al, 2001), the occurrence of one event may be observed subsequent to another event. For example, in the Cache County Study, death might potentially occur after dementia. All previous analyses have defined the events to be death prior to dementia and dementia prior to death, so that only the minimum of the two events is needed. It would be of interest to extend the methods in the current paper to accommodate the semi-competing risks set-up. This is a topic for future research.

Acknowledgments

The authors are grateful to the Cache County Study Steering Committee for providing the dementia dataset. We also thank the Associate Editor and referees for their constructive comments. The work was partially supported by NSF 0906449 (Cheng) and NIH 1R01 CA94893-01 (Fine).

Appendix

Appendix A:

For simplicity, in the sequel, we use generic empirical process notation. Let $X_1, \dots, X_n$ denote i.i.d. random variables and let $f(x)$ denote a function. Define $\mathbb{P}_n f = \sum_i f(X_i)$.

A.1. Asymptotic properties of $\widehat{\theta}_n^{(\cdot)}$

Recall that $\Psi_1(\theta) = \int_0^{\tau_1} \int_0^{\tau_2} w(s, t) \{F_{11}(s, t; \theta_0) - F_{11}(s, t; \theta)\} ds dt$ where $\theta_0 \in \Theta \subset \mathbb{R}$ is assumed to be the true association parameter. Suppose we have the following regularity conditions.

1. θ_0 is the unique solution to $\Psi_1(\theta) = 0$.
2. $F_{11}(s, t; \theta)$ is bounded and continuous with respect to its first two arguments for all θ in an open neighborhood of the true value θ_0 .
3. $F_{11}(s, t; \theta)$ is continuously differentiable as a function of θ , and the derivative is uniformly bounded in its first two arguments for all $\theta \in \Theta$.
4. $\sqrt{n}|\Psi_{1n}(\widehat{\theta}_n^{(\cdot)})| \xrightarrow{P} 0$, as $n \rightarrow \infty$.

We now sketch the proof of Theorem 1 which focuses on the case of nonparametric estimation of the univariate CIFs. The parametric univariate CIF case can be established similarly, assuming that the parametric model is the true model. To simplify the notation, in the sequel we omit “NP” from the superscript if there is no confusion.

Proof

We first show consistency of $\widehat{\theta}_n$. Let $A_n(s, t; \theta) = \widehat{F}_{11}(s, t) - \widetilde{F}_{11}(s, t; \theta)$ and $A(s, t; \theta) = F_{11}(s, t; \theta_0) - F_{11}(s, t; \theta)$. It can be shown that

$$\begin{aligned} & \sup_{\theta \in \Theta} \int_{[0, \tau]} |A_n(s, t; \theta) - A(s, t; \theta)| ds dt \\ & \leq \| \widehat{F}_{11}(s, t) - F_{11}(s, t; \theta_0) \|_{[0, \tau]} \cdot \tau_1 \tau_2 + \int_{[0, \tau]} \sup_{\theta \in \Theta} | \widetilde{F}_{11}(s, t; \theta) - F_{11}(s, t; \theta) | ds dt, \end{aligned}$$

where $\| \cdot \|_{[0, \tau]}$ is the uniform norm over the rectangle $[0, \tau] = [0, \tau_1] \times [0, \tau_2]$. Since $\widehat{F}_{11}(s, t)$, $\widehat{F}_{1(1)}(s)$ and $\widehat{F}_{2(1)}(t)$ are all uniformly consistent, we have

$\sup_{\theta \in \Theta} \int_{[0, \tau]} |A_n(s, t; \theta) - A(s, t; \theta)| ds dt \xrightarrow{P} 0$ as $n \rightarrow \infty$. Then

$$\sup_{\theta \in \Theta} \int_{[0, \tau]} | \widehat{w}(A_n - A)(s, t; \theta) | ds dt + \sup_{\theta \in \Theta} \int_{[0, \tau]} | (\widehat{w} - w)(s, t) A(s, t; \theta) | ds dt \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, since $\sup_{s \leq \tau_1, t \leq \tau_2} | \widehat{w}(s, t) - w(s, t) | \xrightarrow{P} 0$ and $| \widehat{w}(s, t) | \leq 1$. Given that θ_0 is the unique solution to $\Psi_1(\theta) = 0$, $\Psi_1(\theta_0) = 0$ and $\Psi_{1n}(\widehat{\theta}_n)$ is close enough to zero, we have $| \widehat{\theta}_n - \theta_0 | \xrightarrow{P} 0$, as $n \rightarrow \infty$.

To establish the asymptotic normality of $\widehat{\theta}_n$, we need the following lemma.

Lemma A.1— $\sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{I}_i(\tau_1, \tau_2; \theta_0) + o_p(1)$,
 $\widehat{I}_i = \int_{[0, \tau]} w(s, t) \left[\widehat{F}_i(s, t) - p^{(1)} \{ q(F_{1(1)}(s)) + q(F_{2(1)}(t)) \} \cdot \{ q^{(1)}(F_{1(1)}(s)) F_{1i}(s) + q^{(1)}(F_{2(1)}(t)) F_{2i}(t) \} \right] ds dt$,
 $p^{(1)}$ and $q^{(1)}$ are the first derivatives of p in (1) and the inverse $q = p^{-1}$, and $\widehat{F}_i$, F_{1i} and F_{2i} ,
 $i = 1, \dots, n$, are influence functions for the nonparametric estimators $\widehat{F}_{11}$, $\widehat{F}_{1(1)}$ and $\widehat{F}_{2(1)}$.

Proof

We examine

$$\begin{aligned}\sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0) &= \int_{[0,\tau]} \widehat{w}(s, t) \sqrt{n} A_n(s, t; \theta_0) ds dt - \int_{[0,\tau]} w(s, t) \sqrt{n} A(s, t; \theta_0) ds dt \\ &= \int_{[0,\tau]} w(s, t) \sqrt{n} [A_n(s, t; \theta_0) - A(s, t; \theta_0)] ds dt + \int_{[0,\tau]} \sqrt{n} [\widehat{w} - w](s, t) A_n(s, t; \theta_0) ds dt.\end{aligned}$$

Since $A(s, t; \theta_0) = 0$, $\|A_n(s, t; \theta_0) - A(s, t; \theta_0)\|_{[0,\tau]} \xrightarrow{P} 0$, and $\sqrt{n}(\widehat{w}(s, t) - w(s, t)) = O_p(1)$, then we have $\int_{[0,\tau]} \sqrt{n}(\widehat{w} - w)(s, t) A_n(s, t; \theta_0) ds dt = O_p(1)$. Therefore,

$$\sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0) = \int_{[0,\tau]} w(s, t) \left[\sqrt{n} \left\{ \widehat{F}_{11}(s, t) - F_{11}(s, t; \theta_0) \right\} - \sqrt{n} \left\{ \tilde{F}_{11}(s, t; \theta_0) - F_{11}(s, t; \theta) \right\} \right] ds dt + o_p(1).$$

Because the extended Nelson-Aalen estimators of the univariate cumulative CSH functions and the Kaplan-Meier estimator of univariate survival functions are asymptotically linear, by Lemma 3.9.17 of van der Vaart and Wellner (2000), there are influence functions $F_{1i}(s)$ and $F_{2i}(t)$ such that

$$\begin{aligned}\sqrt{n} \left(\widehat{F}_{1(1)}(s) - F_{1(1)}(s), \widehat{F}_{2(1)}(t) - F_{2(1)}(t) \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (F_{1i}(s), F_{2i}(t)) + o_p(1). \text{ Moreover,} \\ \text{following the arguments for the proof of Theorem 1 in Cheng et al. (2007), we can show that} \\ \sqrt{n} \left(\widehat{F}_{11}(s, t) - F_{11}(s, t; \theta_0) \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{F}_i(s, t) + o_p(1), \text{ where } \widehat{F}_i(s, t) \text{ are some influence} \\ \text{functions. By Taylor expansion,}\end{aligned}$$

$$\begin{aligned}& \sqrt{n} \left\{ \tilde{F}_{11}(s, t; \theta_0) - F_{11}(s, t; \theta_0) \right\} \\ &= p^{(1)} \{ q(F_{1(1)}(s)) + q(F_{2(1)}(t)) \} q^{(1)}(F_{1(1)}(s)) \sqrt{n} \left\{ \widehat{F}_{1(1)}(s) - F_{1(1)}(s) \right\} \\ &+ p^{(1)} \{ q(F_{1(1)}(s)) + q(F_{2(1)}(t)) \} q^{(1)}(F_{2(1)}(t)) \sqrt{n} \left\{ \widehat{F}_{2(1)}(t) - F_{2(1)}(t) \right\} + o_p(1).\end{aligned}$$

Then

$$\begin{aligned}\sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0) &= \int_{[0,\tau]} w(s, t) \left[\sqrt{n} \left\{ \widehat{F}_{11}(s, t) - F_{11}(s, t; \theta_0) \right\} - \left\{ \tilde{F}_{11}(s, t; \theta_0) - F_{11}(s, t; \theta) \right\} \right] ds dt + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{[0,\tau]} w(s, t) \left[\widehat{F}_i(s, t) - p^{(1)} \{ q(F_{1(1)}(s)) + q(F_{2(1)}(t)) \} \right. \\ &\times \left. \left\{ q^{(1)}(F_{1(1)}(s)) F_{1i}(s) + q^{(1)}(F_{2(1)}(t)) F_{2i}(t) \right\} \right] ds dt + o_p(1) \\ &\stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{I}_i(\tau_1, \tau_2; \theta_0) + o_p(1).\end{aligned}$$

Employing Lemma A.1., we now continue the proof of Theorem 1.

Proof

From the above Lemma, we have $\sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0) \rightarrow Z_1$ in distribution, where Z_1 is a mean zero normal random variable. By Taylor expansion, we have

$$|\sqrt{n}(\Psi_{1n} - \Psi_1)(\widehat{\theta}_n) - \sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0)| = \left| \frac{1}{n} \sum_{i=1}^n \dot{\widehat{I}}_i(\tau_1, \tau_2; \theta_0) \cdot (1 + o_p(1)) \sqrt{n}(\widehat{\theta}_n - \theta_0) \right| + o_p(1),$$

where $\hat{I}_i(\tau_1, \tau_2; \theta_0) = \frac{\partial \hat{I}_i(\tau_1, \tau_2; \theta)}{\partial \theta} \Big|_{\theta=\theta_0}$. Note that $E\hat{I}_i(\tau_1, \tau_2; \theta_0) = 0$, $E\hat{I}_i(\tau_1, \tau_2; \theta_0) = 0$ and $\frac{1}{n} \sum_{i=1}^n \hat{I}_i(\tau_1, \tau_2; \theta_0) = o_p(1)$. Therefore, $|\sqrt{n}(\Psi_{1n} - \Psi_1)(\hat{\theta}_n) - \sqrt{n}(\Psi_{1n} - \Psi_1)(\theta_0)| = o_p(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|)$. Let $\dot{\Psi}_{1, \theta_0} = \frac{\partial \Psi_1}{\partial \theta} \Big|_{\theta=\theta_0} = - \int_{[0, \tau]} w(s, t) \dot{F}_{11}(s, t; \theta_0) ds dt$, where $\dot{F}_{11}(s, t; \theta_0) = \frac{\partial F_{11}(s, t; \theta)}{\partial \theta} \Big|_{\theta=\theta_0}$ is uniformly bounded in its first two arguments, which results in $|\dot{\Psi}_1(\theta_0)| < \infty$. Then applying the Delta-method yields $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow -\dot{\Psi}_{1, \theta_0}^{-1}(Z_1)$ in distribution. By Theorem 3.9.11 (Delta-method for bootstrap in probability) of van der Vaart and Wellner (2000), we can estimate the variance of $\hat{\theta}_n$ by bootstrapping.

A.2. Asymptotic properties of $\tilde{\theta}_n^{(\cdot)}$

To prove Theorem 2 under nonparametric estimation of the univariate CIFs, we need the following two regularity conditions, in addition to the regularity conditions 2 and 3 given in Appendix A.1. The proof for parametric univariate CIFs is similar and is omitted.

- There is a unique $\theta_0 \in \Theta$ such that $\Psi_2(\theta_0) = 0$.
- $\sqrt{n}|\Psi_{2n}(\tilde{\theta}_n^{(\cdot)})| \xrightarrow{P} 0$, as $n \rightarrow \infty$.

Proof

Let $M_n(ds, dt) = \mathbb{P}_n N(ds, dt) - \mathbb{P}_n H(s, t) \frac{\tilde{F}_{11}(ds, dt; \theta)}{\hat{S}(s-, t-)}$ and

$M_n(ds, dt) = \mathbb{P}_n N(ds, dt) - PH(s, t) \frac{F_{11}(ds, dt; \theta)}{S(s-, t-)}$. It can be shown that

$$\begin{aligned} & \sup_{\theta \in \Theta} \int_{[0, \tau]} |M_n(ds, dt) - M(ds, dt)| \\ & \leq \|(\mathbb{P}_n - P)N\|_{[0, \tau]} + \|(\mathbb{P}_n - P)H\|_{[0, \tau]} \frac{\sup_{\theta \in \Theta} F_{11}(\tau_1, \tau_2; \theta)}{\hat{S}(\tau_1-, \tau_2-)} \\ & \quad + \frac{\|\hat{S} - S\|_{[0, \tau]} \sup_{\theta \in \Theta} F_{11}(\tau_1, \tau_2; \theta)}{\inf_{s, t} \hat{S}(s-, t-) S(s-, t-)} + \frac{\sup_{\theta \in \Theta} |\tilde{F}_{11}(\tau_1, \tau_2; \theta) - F_{11}(\tau_1, \tau_2; \theta)|}{\inf_{s, t} \hat{S}(s-, t-)} \xrightarrow{P} 0, \end{aligned}$$

as $n \rightarrow \infty$. Then,

$$\sup_{\theta \in \Theta} |\Psi_{2n}(\theta) - \Psi_2(\theta)| \leq \sup_{\theta \in \Theta} \int_{[0, \tau]} |\hat{w}(M_n - M)(ds, dt)| + \sup_{\theta \in \Theta} \int_{[0, \tau]} |(\hat{w} - w)(s, t) M(ds, dt)| \xrightarrow{P} 0$$

as $n \rightarrow \infty$, since $\sup_{s \leq \tau_1, t \leq \tau_2} |\hat{w}(s, t) - w(s, t)| \xrightarrow{P} 0$ and $|\hat{w}(s, t)| \leq 1$. Thus, $|\tilde{\theta}_n - \theta_0| \xrightarrow{P} 0$.

Next, we show that $\sqrt{n}(\Psi_{2n} - \Psi_2)(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{I}_i(\tau_1, \tau_2; \theta_0) + o_p(1)$ where $\tilde{I}_i$'s are defined below. We examine

$$\sqrt{n}(\Psi_{2n} - \Psi_2)(\theta_0) = \sqrt{n} \int_{[0, \tau]} \hat{w}(s, t) \left[\mathbb{P}_n N(ds, dt) - \mathbb{P}_n H(s, t) \tilde{F}_{11}(ds, dt; \theta_0) \left\{ \hat{S}(s-, t-)^{-1} \right\} \right].$$

Since $\|\widehat{w}(s, t) - w(s, t)\|_{[0, \tau]} \xrightarrow{P} 0$, as $n \rightarrow \infty$, without loss of generality, we can just examine

$$\begin{aligned} & \sqrt{n} \int_{[0, \tau]} w(s, t) \left\{ \mathbb{P}_n N(ds, dt) - \mathbb{P}_n H(s, t) \frac{\tilde{F}_{11}(ds, dt; \theta_0)}{\tilde{S}(s-, t-)} \right\} \\ &= \int_{[0, \tau]} w(s, t) \sqrt{n} (\mathbb{P}_n - P) N(ds, dt) - \int_{[0, \tau]} w(s, t) \sqrt{n} (\mathbb{P}_n - P) H(s, t) \frac{F_{11}(ds, dt; \theta_0)}{\tilde{S}(s-, t-)} \\ &+ \int_{[0, \tau]} \frac{w(s, t) \mathbb{P}_n H(s, t)}{\tilde{S}(s-, t-) \tilde{S}(s-, t-)} \sqrt{n} \left\{ \widehat{S}(s-, t-) - S(s-, t-) \right\} F_{11}(ds, dt; \theta_0) \\ &- \int_{[0, \tau]} \frac{w(s, t) \mathbb{P}_n H(s, t)}{\tilde{S}(s-, t-)} \sqrt{n} \left\{ \tilde{F}_{11}(ds, dt; \theta_0) - F_{11}(ds, dt; \theta_0) \right\}. \end{aligned}$$

The influence functions for the first two terms on the right hand side are straightforward. By Theorem 5.1 in Gill et al. (1995), there exists influence functions $S_i(s, t)$ such that

$\sqrt{n}(\widehat{S} - S)(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(s, t) + o_p(1)$, though the expression of S_i is rather complicated. By Lemma A.5 of Cheng et al. (2007) which is an extension of Lemma 3.9.17 of van der Vaart and Wellner (2000), we may obtain the influence functions for the third term. We now focus on the fourth term.

Suppose the third derivative of p exists and denote its j th order derivative as $p^{(j)}$, $j = 1, 2, 3$. For $q = p^{-1}$, its j th order derivative is denoted as $q^{(j)}$, $j = 1, 2, 3$. By Taylor expansion, $F_{11}(ds, dt; \theta_0) = p^{(2)}\{q(F_{1(1)}(s)) + q(F_{2(1)}(t))\}q^{(1)}\{F_{1(1)}(s)\}q^{(1)}\{F_{2(1)}(t)\}F_{1(1)}(ds)F_{2(1)}(dt)$. Let $A_1(s, t) = p^{(2)}\{q(F_{1(1)}(s)) + q(F_{2(1)}(t))\}$, $B_1(s, t) = q^{(1)}(F_{1(1)}(s))q^{(1)}(F_{2(1)}(t))$ and $C_1(ds, dt) = F_{1(1)}(ds)F_{2(1)}(dt)$, then

$$\begin{aligned} & \sqrt{n} \int_{[0, \tau]} \left\{ \tilde{F}_{11}(ds, dt; \theta_0) - F_{11}(ds, dt; \theta_0) \right\} \\ &= \int \widehat{A}_1(s, t) \widehat{B}_1(s, t) \sqrt{n} \left\{ \widehat{C}_1(ds, dt) - C_1(ds, dt) \right\} + \int \widehat{A}_1(s, t) \sqrt{n} \left\{ \widehat{B}_1(s, t) - B_1(s, t) \right\} C_1(ds, dt) \\ &+ \int \sqrt{n} \left\{ \widehat{A}_1(s, t) - A_1(s, t) \right\} B_1(s, t) C_1(ds, dt) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\int \left\{ \widehat{A}_1(s, t) B_1(s, t) \{F_{1(1)}(ds) F_{2i}(dt) + F_{1i}(ds) F_{2(1)}(dt)\} \right. \right. \\ &+ \left. \int A_1(s, t) \left[q^{(1)}(F_{2(1)}(t)) q^{(2)}(F_{1(1)}(s)) F_{1i}(s) + q^{(1)}(F_{1(1)}(s)) q^{(2)}(F_{2(1)}(t)) F_{2i}(t) \right] C_1(ds, dt) \right. \\ &+ \left. \int p^{(3)}\{q(F_{1(1)}(s)) + q(F_{2(1)}(t))\} \left\{ q^{(1)}(F_{1(1)}(s)) F_{1i}(s) + q^{(1)}(F_{2(1)}(t)) F_{2i}(t) \right\} B_1(s, t) C_1(ds, dt) \right] \\ &+ o_p(1) \\ &\stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{F}_i(s, t; \theta_0) + o_p(1), \end{aligned}$$

where $F_{1i}(s)$ and $F_{2i}(t)$ are the influence functions defined previously. Then

$$\sqrt{n}(\Psi_{2n} - \Psi_2)(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{I}_i(\tau_1, \tau_2; \theta_0) + o_p(1), \text{ where}$$

$$\begin{aligned} \tilde{I}_i(\tau_1, \tau_2; \theta_0) &= \int_{[0, \tau]} w(s, t) \{N_i(ds, dt) - PN(s, t)\} \\ &- \int_{[0, \tau]} w(s, t) \{H_i(s, t) - PH(s, t)\} \frac{F_{11}(ds, dt; \theta_0)}{\tilde{S}(s-, t-)} \\ &+ \int_{[0, \tau]} \frac{w(s, t) PH(s, t)}{\tilde{S}^2(s-, t-)} S_i(s-, t-) F_{11}(ds, dt; \theta_0) \\ &- \int_{[0, \tau]} \frac{w(s, t) PH(s, t)}{\tilde{S}(s-, t-)} \tilde{F}_i(ds, dt; \theta_0), \end{aligned}$$

where $N_i(s, t)$ is the double event process, $H_i(s, t)$ is the bivariate at risk process, and $S_i(s, t)$ and $\tilde{F}_i(s, t; \theta_0)$ are the influence function defined previously.

Since $E\tilde{I}_i(\tau_1, \tau_2; \theta_0) = 0$, $\sqrt{n}(\Psi_{2n} - \Psi_2)(\theta_0) \rightarrow Z_2$ in distribution, for a mean 0 normal random variable Z_2 . By the Taylor's expansion,

$$|\sqrt{n}(\Psi_{2n} - \Psi_2)(\tilde{\theta}_n) - \sqrt{n}(\Psi_{2n} - \Psi_2)(\theta_0)| = \left| \frac{1}{n} \sum_{i=1}^n \dot{I}_i(\tau_1, \tau_2; \theta_0) \cdot (1 + o_p(1)) \sqrt{n}(\tilde{\theta}_n - \theta_0) \right| + o_p(1),$$

where $\dot{I}_i(\tau_1, \tau_2; \theta_0) = \frac{\partial \tilde{I}_i(\tau_1, \tau_2; \theta)}{\partial \theta} \Big|_{\theta=\theta_0}$. In addition, $E \dot{I}_i(\tau_1, \tau_2; \theta_0) = 0$, then $\frac{1}{n} \sum_{i=1}^n \dot{I}_i(\tau_1, \tau_2; \theta_0)$ is $o_p(1)$, and $\sqrt{n}(\Psi_{2n} - \Psi_2)(\tilde{\theta}_n) - \sqrt{n}(\Psi_{2n} - \Psi_2)(\theta_0) = o_p(1 + \sqrt{n}|\tilde{\theta}_n - \theta_0|)$. Let

$$\dot{\Psi}_{2, \theta_0} = \frac{\partial \Psi_2}{\partial \theta} \Big|_{\theta=\theta_0} = - \int_{[0, \tau]} w(s, t) PH(s, t) \frac{\dot{F}_{11}(ds, dt; \theta_0)}{S(s-, t-)}.$$

By Condition 3 in Appendix A.1.,

$|\dot{\Psi}_{2, \theta_0}| < \infty$. Therefore, applying the delta-method, we have $\sqrt{n}(\tilde{\theta}_n - \theta_0) \rightarrow -\dot{\Psi}_{2, \theta_0}^{-1}(Z_2)$ in distribution and the validity of using the bootstrap method to estimate the variance of $\tilde{\theta}_n$.

References

- Bandeén-Roche K, Liang K. Modelling multivariate failure time associations in the presence of a competing risk. *Biometrika*. 2002; 89:299–314.
- Bandeén-Roche K, Ning J. Nonparametric estimation of bivariate failure time associations in the presence of a competing risk. *Biometrika*. 2008; 95:221–232. [PubMed: 20305739]
- Cheng Y. Modeling cumulative incidences of dementia and dementia-free death using a novel three-parameter logistic function. *The International Journal of Biostatistics*. 2009 DOI: 10.2202/1557-4679.1183.
- Cheng Y, Fine JP. Nonparametric estimation of cause-specific cross hazard ratio with bivariate competing risks data. *Biometrika*. 2008; 95:233–240.
- Cheng Y, Fine JP, Kosorok MR. Nonparametric analysis of bivariate competing risks data. *Journal of the American Statistical Association*. 2007; 102:1407–1416.
- Cheng Y, Fine JP, Kosorok MR. Nonparametric Association Analysis of Exchangeable Clustered Competing Risks Data. *Biometrics*. 2009; 65:385–393. [PubMed: 18549422]
- Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model (C/R: P108-117). *Journal of the Royal Statistical Society, Series A: General*. 1985; 148:82–108.
- 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; 65:141–152.
- Dabrowska DM. Kaplan meier estimate on the plane. *The Annals of Statistics*. 1988; 16:1475–1489.
- Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88:907–919.
- Frank MJ. On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*. 1979; 19:194–226.
- Genest C. Frank's family of bivariate distributions. *Biometrika*. 1987; 74:549–555.
- Genest C, Mackay RJ. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *The Canadian Journal of Statistics*. 1986a; 14:145–159.
- Genest C, Mackay RJ. The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*. 1986b; 40:280–283.
- Genest C, Rivest L-P. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*. 1993; 88:1034–1043.
- Gill RD, van der Laan MJ, Wellner JA. Inefficient estimators of the bivariate survival function for three models. *Annales de L'Institut Henri Poincaré Probabilités et Statistiques*. 1995; 31:545–597.
- Glidden DV. Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*. 1999; 86:381–393.
- Hougaard P. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*. 1984; 71:75–83.

- Hougaard P. A class of multivariate failure time distributions (Corr: V75 p395). *Biometrika*. 1986a; 73:671–678.
- Hougaard P. Survival models for heterogeneous populations derived from stable distributions (Corr: V75 p395). *Biometrika*. 1986b; 73:387–396.
- Hougaard P. Modelling multivariate survival. *Scandinavian Journal of Statistics*. 1987; 14:291–304.
- Hougaard, P. *Analysis of Multivariate Survival Data*. Springer; New York: 2000.
- Jeong J, Fine JP. Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society, Series C: Applied Statistics*. 2006; 55:187–200.
- Jiang H, Fine JP, Kosorok MR, Chappell R. Pseudo self-consistent estimation of a copula model with informative censoring. *Scandinavian Journal of Statistics*. 2006; 32:1–20.
- Joe, H. *Multivariate Models and Dependence Concepts*. Chapman & Hall Ltd.; Boca Raton, FL: 1997.
- Lehmann EL. Some concepts of dependence. *Annals of Mathematical Statistics*. 1966; 37:1137–1153.
- Nelsen, RB. *An Introduction to Copulas*. Springer-Verlag Inc.; New York: 1999.
- Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*. 1989; 84:487–493.
- Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure time data in the presence of competing risks. *Biometrics*. 1978; 12:737–751.
- Pruitt RC. Identifiability of bivariate survival curves from censored data. *Journal of the American Statistical Association*. 1993; 88:573–579.
- Shih JH. A goodness-of-fit test for association in a bivariate survival model. *Biometrika*. 1998; 85:189–200.
- Shih JH, Louis TA. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*. 1995; 51:1384–1399. [PubMed: 8589230]
- Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*. 1975; 72:20–22.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes*. Springer; New York: 2000.
- Wang W. Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society Series B*. 2003; 65:257–273.
- Wienke A, Holm NV, Christensen K, Skytthe A, Vaupel JW, Yashin AI. The heritability of cause-specific mortality: a correlated gamma-frailty model applied to mortality due to respiratory diseases in danish twins born 1870-1930. *Statistics in Medicine*. 2003; 22:3873–3887. [PubMed: 14673944]
- Yan J. Enjoy the joy of copulas. *Journal of Statistical Software*. 2007; 21:1–21.

Fig. 1.

Nonparametric estimate of $\psi(s, t)$ and semiparametric estimate from the Clayton model on the diagonal for dementia onset in the Cache County Study, where solid lines are point estimates and dashed lines are 95% pointwise confidence intervals.

Table 1

Simulation comparing $\widehat{\theta}_n^{NP}$ and $\widetilde{\theta}_n^{NP}$ with restriction on $\widehat{S}(s, t) \geq 0.05$; $\widehat{\theta}_n^{NP}$: solution to (5), $\widetilde{\theta}_n^{NP}$: solution to (7), Ave: average of the estimates, ESE: empirical standard error, BSE: bootstrap standard error, and Coverage: empirical coverage of the true parameter in its 95% confidence interval based on BSE.

		$\widehat{\theta}_n^{NP}$					$\widetilde{\theta}_n^{NP}$				
Clayton	log(θ)	0.41	1.10	1.39	1.61	0.41	1.10	1.39	1.61	0.41	1.39
	Ave	0.41	1.11	1.38	1.63	0.42	1.12	1.40	1.57	0.42	1.57
	ESE	0.10	0.14	0.16	0.24	0.21	0.31	0.30	0.31	0.21	0.31
	BSE	0.10	0.14	0.18	0.24	0.22	0.28	0.30	0.32	0.22	0.32
	Coverage	0.95	0.95	0.97	0.94	0.91	0.91	0.95	0.93	0.91	0.95
N = 200	Ave	0.40	1.10	1.39	1.62	0.42	1.11	1.40	1.59	0.42	1.59
	ESE	0.07	0.10	0.12	0.15	0.16	0.18	0.21	0.22	0.16	0.22
	BSE	0.07	0.10	0.13	0.18	0.15	0.16	0.22	0.22	0.15	0.22
	Coverage	0.96	0.95	0.95	0.98	0.94	0.96	0.94	0.94	0.94	0.94
Positive Stable	log(θ)	-0.22	-0.51	-0.92	-1.61	-0.22	-0.51	-0.92	-1.61	-0.22	-1.61
	Ave	-0.23	-0.52	-0.92	-1.64	-0.22	-0.50	-0.89	-1.51	-0.22	-1.51
	ESE	0.07	0.09	0.12	0.26	0.10	0.14	0.22	0.46	0.10	0.22
	BSE	0.08	0.09	0.12	0.28	0.10	0.16	0.25	0.39	0.10	0.25
	Coverage	0.95	0.96	0.97	0.95	0.91	0.97	0.93	0.84	0.91	0.93
N = 400	Ave	-0.23	-0.51	-0.92	-1.63	-0.22	-0.50	-0.91	-1.60	-0.22	-1.60
	ESE	0.06	0.07	0.08	0.17	0.08	0.10	0.18	0.38	0.08	0.18
	BSE	0.05	0.06	0.08	0.21	0.07	0.10	0.18	0.35	0.07	0.18
	Coverage	0.94	0.93	0.96	0.98	0.91	0.95	0.96	0.92	0.91	0.96

Table 2

Summary of the data from the Cache County Study

		Eldest Child		
		Dementia	Death	Censoring
Mother	Dementia	39	103	316
	Death	155	1132	1743
	Censoring	2	8	137
	Total	196	1243	2196
				3635

Table 3

Estimates of θ involved in different frailty models for CIFs; $\widehat{\theta}_n^{(\cdot)}$: solution to (5), $\widehat{\theta}_n^{(\cdot)}$: solution to (7), NP , Log and Gom stand for the nonparametric, modified logistic and Gompertz models for estimating univariate CIFs, Est: estimate of the association parameter, BSE: boot-strap standard error and pvalue; p value from the goodness-of-fit test.

Models	$C_\theta(u; v)$	Stat	$\widehat{\theta}_n^{NP}$	$\sim NP$ θ_n	$\sim Log$ θ_n	$\sim Log$ θ_n	$\sim Gom$ θ_n	$\sim Gom$ θ_n
Clayton	$(u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)},$ $\theta > 1$	Est	1.14	1.14	1.14	1.12	-	-
		BSE	0.04	0.04	0.04	0.04	-	-
		pval	0.89		0.23		-	
Frank	$\log_\theta \left\{ 1 + \frac{(\theta^u - 1)(\theta^v - 1)}{\theta - 1} \right\},$ $\theta > 0, \theta \neq 1$	Est	0.09	0.18	0.09	0.23	4.02	1.52
		BSE	0.07	0.11	0.13	0.14	0.98	1.92
		pval	0.13		0.09		0.0001	
Positive Stable	$\exp(-\{(-\log(u))^{1-\theta} + (-\log(v))^{1-\theta}\}),$ $0 < \theta \leq 1$	Est	0.77	0.81	0.77	0.83	-	0.99
		BSE	0.05	0.05	0.06	0.05	-	0.03
		pval	0.10		0.01		-	