A New Flexible Dependence Measure for Semi-competing Risks

Jing Yang and **Limin Peng**

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, U.S.A

Limin Peng: lpeng@sph.emory.edu

Summary

Semi-competing risks data are often encountered in chronic disease follow-up studies that record both nonterminal events (eg. disease landmark events) and terminal events (eg. death). Studying the relationship between the nonterminal event and the terminal event can provide insightful information on disease progression. In this paper, we propose a new sensible dependence measure tailored to addressing such an interest. We develop a nonparametric estimator, which is general enough to handle both independent right censoring and left truncation. Our strategy of connecting the new dependence measure with quantile regression enables a natural extension to adjust for covariates with minor additional assumptions imposed. We establish the asymptotic properties of the proposed estimators and develop inferences accordingly. Simulation studies suggest good finite-sample performance of the proposed methods. Our proposals are illustrated via an application to Denmark diabetes registry data.

Keywords

Estimating Equation; Left truncation; Quantile; Residual lifetime; Semi-competing risks

1. Introduction

In chronic disease follow-up studies, in addition to a terminal survival outcome of primary interest, nonterminal disease landmark events are often recorded to provide extra information on disease progression. This naturally postulates a semi-competing risks setting (Fine et al., 2001), where time to a nonterminal event (T_1) can be censored by time to a terminal event (T_2) but not vice versa. The underlying disease mechanism often implicates the dependence between the nonterminal event and the terminal event. While considerably complicating the inference on T_1, such a dependency itself can offer valuable insight on disease prognosis and thus poses an important problem to study.

Statistical methods tailored to semi-competing risks data have been developed in settings without covariates (Fine et al., 2001; Wang, 2003; Peng and Fine, 2006b; Lakhal et al., 2008, among others) and with covariates (Lin et al., 1996; Peng and Fine, 2006a; Ghosh, 2006; Peng and Fine, 2007; Hsieh et al., 2008; Chen, 2012; Li and Peng, 2015, among

others). In most of existing work, the dependence between the nonterminal event and the terminal event is captured by a copula model assumed for the joint distribution of (T_1, T_2). For example, without considering covariates, Fine et al. (2001) derived a closed-form estimator for the constant cross-ratio measure under the Clayton copula model (Clayton, 1978). Wang (2003) and Lakhal et al. (2008) studied the estimation of the association parameter for the general class of Archimedean copulas. With regression models assumed for T_1 and T_2, Hsieh et al. (2008) developed an estimator for the Archimedean copula parameter when covariates are discrete. More recently, Chen (2012) studied a nonparametric maximum likelihood approach under a general specification of the copula model.

While modeling the dependence structure between T_1 and T_2 based on a copula model is intuitive and useful, such an approach can impose some implicit limitations that may often be ignored. For example, under the Clayton's copula model, the cross ratio function (Oakes, 1989) is confined to be constant over time. Peng and Fine (2007) investigated a time-dependent copula model, which allows for a time-varying association structure between T_1 and T_2. Their empirical results evidenced that a constant association between T_1 and T_2 can be inadequate in practical settings. When this occurs, a numerical example provided in Web Appendix B shows that the classical copula modeling approach assuming a constant associate parameter can lead to misleading conclusions. In addition, the interpretation of a copula parameter, constant or time-dependent, relies on the selection of the copula function. When there are covariates involved, a copula based approach is further prone to issues due to potential misspecifications of the marginal regression models for T_1 and T_2.

In this work, we propose an alternative perspective for characterizing the relation between the nonterminal event and the terminal event. The new perspective enables the accommodation of time-varying dependence without involving strong model assumptions. Our proposals are built upon a comparison of the quantile residual time to the terminal event with the occurrence of the nonterminal event versus that without the occurrence of the nonterminal event. The basic idea bears a similar flavor to that of the cross-ratio function(Oakes, 1989), which compares the hazards of the terminal event regarding the timing of the nonterminal event. In contrast to a hazard function, a quantile residual time can offer straightforward physical interpretations and usually does not require smoothing for nonparametric estimation. Quantile residual lifetime has been investigated mostly in univariate survival settings (Gelfand and Kottas, 2003; Jeong et al., 2008; Jung et al., 2009; Ma and Yin, 2010, for example); but its utility as a device to study the dependence between semi-competing risks has not been exploited.

The rest of the paper is organized as follows. In Section 2, we first introduce a new dependence measure defined on quantile residual times, well tailored to the special structure of semi-competing risks. In Section 3. We fit the new measure into a quantile regression framework, and develop a simple nonparametric estimator, which can also account for left truncation, a common complication in observational studies. We present asymptotic studies of the proposed estimator as well as inference procedures. An extension to adjust for covariates is discussed in Section 4. We conduct extensive simulation studies to evaluate the performance of our proposal. The proposed method is illustrated by an application to a

Denmark diabetic registry dataset. The numerical results are presented in Section 5, followed by a few concluding remarks in Section 6.

2. The Proposed Measure

Let $Q_\tau(Y|A) \equiv \inf\{t : \Pr(Y \leq t|A) \geq \tau\}$ denote the τ-th quantile of Y given condition A holds. For the terminal event of interest, the quantile residual time at a given time point t_0 is defined as $Q_\tau(T_2 - t_0|T_2 > t_0)$.

Our basic idea is to compare the quantile residual time to the terminal event given the nonterminal event having occurred and that without the past occurrence of the nonterminal event. That is, we consider the cross quantile residual ratio (CQRR) defined as

$$\text{CQRR}(\tau; t_0) = \frac{Q_\tau(T_2 - t_0|T_2 > t_0, T_1 > t_0)}{Q_\tau(T_2 - t_0|T_2 > t_0, T_1 \leq t_0)}, \tau \in (0, 1), t_0 > 0.$$

It is clear that a larger $CQRR(\tau, t_0)$, which reflects a larger difference in $Q_\tau(T_2 - t_0|T_2 > t_0, T_1 > t_0)$ and $Q_\tau(T_2 - t_0|T_2 > t_0, T_1 \leq t_0)$, indicates a larger impact of having $T_1 > t_0$ (versus $T_1 \leq t_0$) on the subsequent progression of T_2. Note that $CQRR(\tau, t_0)$ bears some similarity with the cross-ratio function in the semi-competing risks setting,

$$\frac{\lambda(t_2|T_1 = t_1)}{\lambda(t_2|T_1 > t_1)}, t_1 \leq t_2,$$

where $\lambda(t_2|\cdot) = \frac{d}{d\varepsilon} P(T_2 < t_2 + \varepsilon|T_2 \geq t_2, \cdot)|_{\varepsilon=0}$. Both of them assess the difference in the terminal event progression according to the timing of the nonterminating event. The distinction lies in that the cross-ratio function uses hazard functions to evaluate the progression of the terminating event, while the proposed $CQRR(\tau,; t_0)$ adopts quantile residual time, which can be directly interpreted in the time scale. Like the cross-ratio function defined above, $CQRR(\tau, t_0)$ only concerns the joint distribution of (T_1, T_2) at the upper wedge (i.e. $T_1 \leq T_2$) and hence is nonparametrically identifiable with semi-competing risks data.

We further take a log transformation on $CQRR(\tau, t_0)$. Our proposed measure for the dependence of semi-competing risks events is given by

$$\text{LCQRR}(\tau; t_0) = \log\left\{\frac{Q_\tau(T_2 - t_0|T_2 > t_0, T_1 > t_0)}{Q_\tau(T_2 - t_0|T_2 > t_0, T_1 \leq t_0)}\right\}, \tau \in (0, 1), t_0 > 0.$$

It is easy to interpret $LCQRR(\tau, t_0)$. For example, $LCQRR(\tau, t_0) > 0 (< 0)$ suggests that the nonterminating event occurring before t0 may be associated with a faster (or slower) progression to subsequent terminating event. The larger the magnitude of $LCQRR(\tau, t_0)$, the

bigger the impact of having $T_1 \leq t_0$ on the residual lifetime for T_2. When T_1 and T_2 are independent, $LCQRR(\tau, t_0) = 0$ for any $\tau \in (0, 1)$ and $t_0 > 0$. Examining $LCQRR(\tau, t_0)$ with different t_0's may help understand how the dependence between the nonterminal event and the terminal event evolves time. One may also vary the value of τ to evaluate the influence of T_1 on multiple segments of the residual time distribution of T_2.

To use $LCQRR(\tau, t_0)$ in practice, we recommend specifying τ and t_0 beforehand according to scientific interests. For example, t_0 may be chosen as time points that landmark the development of the nonterminal event. Some common choices of τ are $\tau = 0.25$, 0.5 and 0.75, which can be used to reflect below average, average, and above average progression to the terminating event. In addition to assigning discrete values to τ and t_0, one can also evaluate $LCQRR(\tau, t_0)$ over a prespecified τ-interval or t_0-interval. Doing so would permit assessing the changing pattern of the semi-competing risks dependence structure, which may shed useful scientific insight but cannot be accommodated by many traditional methods

3. The Proposed Estimation and Inference Procedures

3.1 Data and notation

We begin with a formal introduction of data and notation. Let T_1 denote time to nonterminal event, T_2 denote time to terminal event, and C denote time to censoring, which is independent of (T_1, T_2). Without considering left truncation, the observed semi-competing risks data are $X = T_1 \wedge T_2 \wedge C$, $Y = T_2 \wedge C$, $\delta = I(T_1 < Y)$ and $\eta = I(T_2 < C)$, where $\wedge$ is the minimum operator.

With truncation, the observed data consist of n independent and identically distributed replicates of $(X^*, Y^*, \delta^*, \eta^*, L^*)$, denoted by $(X_i^*, Y_i^*, \delta_i^*, \eta_i^*, L_i^*)_{i=1}^n$, where $(X^*, Y^*, \delta^*, \eta^*, L^*)$ follows the conditional distribution of (X, Y, δ, η, L) given $Y > L$. We restrict L to be always less than C, meaning that censoring only occurs after sampling time. Such assumption has been imposed in much previous work, for example, Wang (1991), Asgharian et al. (2002) and Li and Peng (2011). In addition, we assume that L is independent of (T_1, T_2) and $D = C - L$.

To simplify the presentation hereafter, we define additional notation, $A^*(t_0) = (1, I(X^* > t_0))^T$, $\tilde{A}^*(t_0) = (1, I(T_1^* > t_0))^T$, $A(t_0) = (1, I(X > t_0))^T$ and $\tilde{A}(t_0) = (1, I(T_1 > t_0))^T$. For a vector $\boldsymbol{v}$, we use $\boldsymbol{v}^{(l)}$ to denote the lth component of $\boldsymbol{v}$.

3.2 The proposed estimator

We first study the standard semi-competing risks setting without left truncation. To estimate $LCQRR(\tau, t_0)$, we consider a working quantile residual lifetime regression model, which takes the form,

$$Q_\tau(T_2 - t_0 | T_2 > t_0, I(T_1 > t_0)) = \exp\{\tilde{A}(t_0)^T \boldsymbol{\beta}_0(\tau, t_0)\}, \quad (1)$$

where $\boldsymbol{\beta}_0(\tau, t_0)$ is a 2×1 vector of unknown coefficients. In model (1), $I(T_1 > t_0)$ serves as the only covariate, which is binary. Consequently, model (1) essentially does not impose any parametric assumptions. The coefficients, $\beta_0^{(1)}(\tau, t_0)$ and $\beta_0^{(2)}(\tau, t_0)$, correspond to $\log Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 \leq t_0)$ and $\log Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 > t_0) - \log Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 \leq t_0)$ respectively. This indicates the equivalence between $LCQRR(\tau, t_0)$ and $\beta_0^{(2)}(\tau, t_0)$. Therefore, estimating $\beta_0^{(2)}(\tau, t_0)$ in the quantile regression framework leads to an estimator of $LCQRR(\tau, t_0)$.

A main challenge with fitting model (1) is that the covariate $I(T_1 > t_0)$ is not always observed because T_1 is subject to censoring by both T_2 and C. Suppose there is no independent censoring by C, and then T_2 is fully observed. In this case, we see that $I(T_1 > t_0)$ is observed and equals $I(X > t_0)$ as long as $Y > t_0$. This suggests estimating $\boldsymbol{\beta}_0(\tau, t_0)$ by a stratified quantile regression analysis, which solves the following estimating equation for $\boldsymbol{b} \in R^2$:

$$n^{-1/2}\sum_{i=1}^{n}I(Y_i>t_0)\boldsymbol{A}(t_0)\{I[\log(Y_i - t_0) \leq \boldsymbol{A}(t_0)^T\boldsymbol{b}] - \tau\}=\boldsymbol{0}. \tag{2}$$

When T_2 is subject to independent censoring by C, we still have $I(T_1 > t_0) = I(X > t_0)$ given $Y > t_0$. This nice feature allows us to adapt existing methods for quantile residual lifetime model to handle the effect of censoring. Specifically, we can use a stratified version of Ma and Yin (2010)'s estimating equation, which takes the form,

$$n^{-1/2}\sum_{i=1}^{n}\frac{I(Y_i>t_0)\eta_i}{\hat{G}_c(Y_i)}\boldsymbol{A}_i(t_0)\{I[\log(Y_i - t_0) \leq \boldsymbol{A}_i^T(t_0)\boldsymbol{b}] - \tau\}=\boldsymbol{0},$$

where $\hat{G}_c(\cdot)$ is the Kaplan-Meier estimate of the survival function of C.

When left truncation is present, we need to further modify the estimating equation (2) because $I(T_1 > t_0)$ may be missing and if observed, may not be randomly sampled. Our strategy is to weigh the observed data in an appropriate way such that the bias induced by truncation and censoring is corrected in the estimation of $\boldsymbol{\beta}_0(\tau, t_0)$. Let $D^* = C^* - L^*$. It is critical to note that under the independence between D and (T_1, T_2, L), the distributions of D and D^* are equivalent, and D^* is also independent of (T_1^*, T_2^*, L^*). This fact greatly facilitates the application of the inverse probability of censoring weighting (IPCW) in the present problem with truncated data. As elaborated in Web Appendix A, we show that $G(Y_i^* - L_i^*)$ can serve as an appropriate weight, where $G(t) = \Pr(D > t)$. More specifically, we prove that

$$E\left\{\frac{I(L^* \leq t_0)I(Y^*>t_0)\eta^*}{G(Y^* - L^*)}\boldsymbol{A}^*(t_0)\{I[\log(Y^* - t_0) \leq \boldsymbol{A}^{*T}(t_0)\boldsymbol{\beta}_0(\tau, t_0)] - \tau\}\right\}=0.$$

We propose to estimate $\boldsymbol{\beta}_0(\tau, t_0)$ by solving the following estimating equation for $\boldsymbol{b}$:

$$S_n(\boldsymbol{b}, \tau, t_0) = \boldsymbol{0}, \quad (3)$$

Where

$$S_n(\boldsymbol{b}, \tau, t_0) = n^{-1/2} \sum_{i=1}^{n} \frac{I(L_i^* \le t_0)I(Y_i^* > t_0)\eta_i^*}{\hat{G}(Y_i^* - L_i^*)} \boldsymbol{A}_i^*(t_0)\{I[\log(Y_i^* - t_0) \le \boldsymbol{A}_i^{*T}(t_0)\boldsymbol{b}] - \tau\}.$$

The resulting estimator is denoted by $\hat{\boldsymbol{\beta}}(\tau, t_0)$. Here $\hat{G}(t)$ is the Kaplan-Meier estimator of $G(t)$ obtained from $(Y_i^* - L_i^*, 1 - \eta_i^*)_{i=1}^{n}$,

$$\hat{G}(t) = \prod_{Y_i^* - L_i^* \le t} \left\{1 - \frac{\sum_{j=1}^{n} I(Y_j^* - L_j^* = Y_i^* - L_i^*, \eta_j^* = 0)}{\sum_{j=1}^{n} I(Y_i^* - L_i^* \le Y_j^* - L_j^*)}\right\}.$$

Equation (3) can be easily solved given that it is a monotone estimating equation (Fygenson and Ritov, 1994). Specifically, following similar lines of Peng and Fine (2009), we can transform the solution finding to equation (3) to locating the minimizer of the convex function $U_n(\boldsymbol{b}, \tau, t_0)$ given by

$$\begin{aligned}
&U_n(\boldsymbol{b}, \tau, t_0)\\
&= \sum_{i=1}^{n} I(L_i^* \le t_0)I(Y_i^* > t_0)\eta_i^* \left|\frac{\log(Y_i^* - t_0)}{\hat{G}(Y_i^* - L_i^*)} - \boldsymbol{b}^T \frac{\boldsymbol{A}_i^*(t_0)}{\hat{G}(Y_i^* - L_i^*)}\right| + \left|M - (2\tau\right.\\
&\quad \left. - 1)\boldsymbol{b}^T \sum_{i=1}^{n} I(L_i^* \le t_0)I(Y_i^* > t_0)\eta_i^* \frac{\boldsymbol{A}_i^*(t_0)}{\hat{G}(Y_i^* - L_i^*)}\right|
\end{aligned}$$

where M is a sufficiently large positive number that can bound

$\left|(2\tau - 1)\boldsymbol{b}^T \sum_{i=1}^{n} I(L_i^* \le t_0)I(Y_i^* > t_0)\eta_i^* \frac{\boldsymbol{A}_i^*(t_0)}{\hat{G}(Y_i^* - L_i^*)}\right|$. Minimization of the L_1-type function $U_n(\boldsymbol{b}, \tau, t_0)$ can be solved by using standard software, like the $rq()$ function in the contributed R package *quantreg*.

3.3 Asymptotic results

Given that the proposed estimator of $LCQRR(\tau, t_0)$ is the second element of $\hat{\boldsymbol{\beta}}(\tau, t_0)$, it suffices to derive the asymptotic properties of $\hat{\boldsymbol{\beta}}(\tau, t_0)$.

We assume the following regularity conditions:

C1. There exists $\nu > 0$ such that $P(D = \nu) > 0$ and $P(D > \nu) = 0$.

C2. (i) $0 < \tau_L \quad \tau_U \quad 1$; (ii) t_L and t_U are interior points of the support of X^*.

C3. (i) $\boldsymbol{\beta}_0(\tau, t_0)$ is Lipschitz continuous for $\tau \in [\tau_L, \tau_U]$ and $t_0 \in [t_L, t_U]$; (ii) $f(t|\tilde{\boldsymbol{A}}(t_0))$ is continuous and bounded above uniformly in t, t_0 and $\tilde{\boldsymbol{A}}(t_0)$, where $f(t|\tilde{\boldsymbol{A}}(t_0)) = dF(t|\tilde{\boldsymbol{A}}(t_0))/dt$ and $F(t|\tilde{\boldsymbol{A}}(t_0)) = E\{I(T_2 \leq t)|\tilde{\boldsymbol{A}}(t_0)\}$.

C4. For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{\boldsymbol{b} \in B(\rho_0), t_0 \in [t_L, t_U]} \text{eigmin} \boldsymbol{H}(\boldsymbol{b}, t_0) \geq c_0$, where $B(\rho) = \{\boldsymbol{b} \in R^2 : \inf_{\tau \in [\tau_L, \tau_U], t_0 \in [t_L, t_U]} \|\boldsymbol{b} - \boldsymbol{\beta}_0(\tau, t_0)\| \leq \rho\}$ and $\boldsymbol{H}(\boldsymbol{b}, t_0) = E[c(t_0)\tilde{\boldsymbol{A}}(t_0)^{\otimes 2} f(t_0 + \exp(\tilde{\boldsymbol{A}}^T(t_0)\boldsymbol{b})|\tilde{\boldsymbol{A}}^T(t_0))\exp(\tilde{\boldsymbol{A}}^T(t_0)\boldsymbol{b})]$. Here $\| \cdot \|$ is the Euclidean norm and $\boldsymbol{u}^{\otimes 2} = \boldsymbol{u}\boldsymbol{u}^T$ for a vector $\boldsymbol{u}$.

Define $N_i^G(t) = I(Y_i^* - L_i^* \leq t, \eta_i^* = 0)$, $Y_i(t) = I(Y_i^* - L_i^* \geq t)$, $y(t) = P(Y^* - L^* \geq t)$, $\lambda^G(t) = \lim_{\to 0} P(Y^* - L^* \in (t, t +)|Y^* - L^* \geq t)/ $, $\Lambda^G(t) = \int_0^t \lambda^G(s)ds$, and $M_i^G(t) = N_i^G(t) - \int_0^\infty Y_i(s)d\Lambda^G(s)$. Let $w(\boldsymbol{b}, \tau, t_0, t) = E\{A^*(t_0) Y(t)I(L^* \leq t_0)I(Y^* > t_0)\eta^*\{I[\log(Y^* - t_0) \leq A^{*T}(t_0)\boldsymbol{b}] - \tau\}G(Y^* - L^*)^{-1}\}$, $\boldsymbol{\zeta}_i(\tau, t_0) = \boldsymbol{\xi}_{1,i}(\tau, t_0) - \boldsymbol{\xi}_{2,i}(\tau, t_0)$, where

$$\boldsymbol{\xi}_{1,i}(\tau, t_0) = I(L_i^* \leq t_0)I(Y_i^* \leq t_0)\eta_i^* \boldsymbol{A}_i^*(t_0)\{I[\log(Y_i^* - t_0) \leq \boldsymbol{A}_i^{*T}(t_0)\boldsymbol{\beta}_0(\tau, t_0)] - \tau\}G(Y_i^* - L_i^*)^{-1}$$

and $\boldsymbol{\xi}_{2,i}(\tau, t_0) = \int_0^\infty w(\boldsymbol{\beta}_0(\tau, t_0), \tau, t_0, s)\dfrac{dM_i^G(s)}{y(s)}$, $i = 1, \ldots, n$.

We have following theorems:

Theorem 3.1: Under conditions C1–C4,

$$\lim_{n \to \infty} \sup_{\tau \in [\tau_L, \tau_U], t_0 \in [t_L, t_U]} \|\hat{\boldsymbol{\beta}}(\tau, \tau_0) - \boldsymbol{\beta}_0(\tau, \tau_0)\| \to_p 0.$$

Theorem 3.2: Under conditions C1–C4, $n\{\hat{\boldsymbol{\beta}}(\tau, t_0) - \boldsymbol{\beta}_0(\tau, t_0)\}$ weakly converge to a mean zero Gaussian process $\boldsymbol{G}(\tau, t_0)$ with covariance matrix $E\{\boldsymbol{G}(\tau', t_0')\boldsymbol{G}(\tau, t_0)^T\}$ given by

$$\boldsymbol{\Phi}(\tau', t_0', \tau, t_0) = \boldsymbol{H}\{\boldsymbol{\beta}_0(\tau', t_0'), t_0'\}^{-1} E\{\boldsymbol{\zeta}_1(\tau', t_0')\boldsymbol{\zeta}_1(\tau, t_0)^T\}[\boldsymbol{H}\{\boldsymbol{\beta}_0(\tau, t_0), t_0\}^{-1}]^T,$$

where $\tau, \tau' \in [\tau_L, \tau_U]$ and $t_0, t_0' \in [t_L, t_U]$.

Theorem 3.1 implies that the proposed estimator of $LCQRR(\tau, t_0)$ is uniformly consistent in $\tau \in [\tau_L, \tau_U]$ and $t_0 \in [t_L, t_U]$. Theorem 3.2 presents a closed form expression for the asymptotic distribution of the proposed estimator of $LCQRR(\tau, t_0)$. Detailed proofs of Theorem 3.1 and 3.2 are provided in Web Appendix A.

3.4 Inference procedures

The asymptotic covariance matrix of $n\{\hat{\boldsymbol{\beta}}(\tau, t_0) - \boldsymbol{\beta}_0(\tau, t_0)\}$ involves unknown density functions. It is straightforward to use bootstrapping procedures or adapt resampling approaches, such as Parzen, Wei, and Ying (1994) and Jin, Ying, and Wei (2001), to estimate the asymptotic covariance without requiring density estimation. Alternatively, we can also derive a consistent plug-in estimate for the covariance matrix following the lines of Peng and Fine (2009). The specific procedure follows.

1.   Calculate $\hat{\sum}(\tau, t_0, \tau, t_0) = n^{-1}\sum_{i=1}^{n}\hat{\zeta}_i(\tau, t_0)^{\otimes 2}$, where

$$\hat{\zeta}_i(\tau, t_0) = \frac{I(L_i^* \le t_0, Y_i^* > t_0)\eta_i^*}{\hat{G}(Y_i^* - L_i^*)}\boldsymbol{A}_i^*(t_0)\{I[\log(Y_i^* - t_0) \le \boldsymbol{A}_i^{*T}(t_0)\hat{\boldsymbol{\beta}}(\tau, t_0)] - \tau\}$$

$$-I(\eta_i^* = 0)\left(\sum_{j=1}^{n}\boldsymbol{A}_i^*(t_0)I(Y_j^* - L_j^* \ge Y_i^* - L_i^*)I(L_j^* \le t_0, Y_j^* > t_0)\eta_j^* \times \{I[\log(Y_j^* - t_0) \le \boldsymbol{A}_j^{*T}(t_0)\hat{\boldsymbol{\beta}}(\tau, t_0)] - \tau\}\{\hat{G}(Y_j^* - L_j^*\right.$$

2.   Use spectral decomposition to find a symmetric matrix $\boldsymbol{E}_n(\tau, t_0)$ such that

$$\hat{\sum}(\tau, t_0, \tau, t_0) = \boldsymbol{E}_n^2(\tau, t_0).$$

3.   Calculate

$\boldsymbol{D}_n(\tau, t_0) = [\boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,1}(\tau, t_0), \tau, t_0\} - \hat{\boldsymbol{\beta}}(\tau, t_0), \boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,2}(\tau, t_0), \tau, t_0\} - \hat{\boldsymbol{\beta}}(\tau, t_0)]$, where $\boldsymbol{e}_{n,j}$ is the jth column of $\boldsymbol{E}_n(\tau, t_0)$, and $\boldsymbol{S}_n^{-1}(\boldsymbol{e}, \tau, t_0)$ is defined as the solution to $\boldsymbol{S}_n(\boldsymbol{b}, \tau, t_0) - \boldsymbol{e} = 0$.

4.   A consistent estimate for the asymptotic covariance matrix of $n\{\hat{\boldsymbol{\beta}}(\tau, t_0) - \boldsymbol{\beta}_0(\tau, t_0)\}$ is given by

$$n\boldsymbol{D}_n(\tau', t_0')\boldsymbol{E}_n^{-1}(\tau', t_0')\hat{\sum}(\tau', t_0', \tau, t_0)\boldsymbol{E}_n^{-1}(\tau, t_0)\boldsymbol{D}_n^T(\tau, t_0).$$

In the special case that $\tau' = \tau$ and $t_0' = t_0$, a consistent estimate for the asymptotic variance matrix is simplified as $n\{\boldsymbol{D}_n^{\otimes 2}(\tau, t_0)\}$.

We can also develop second-stage inferences following the lines of Li and Peng (2011). For example, we can summarize $LCQRR(\tau, t_0)$ over $t_0 \in [t_L, t_U]$ by

$\Omega_\tau = \frac{1}{t_U - t_L}\int_{t_L}^{t_U}\boldsymbol{\beta}_0^{(2)}(\tau, t_0)dt_0$, which may be consistently estimated by

$\hat{\Omega}_\tau = \frac{1}{t_U - t_L}\int_{t_L}^{t_U}\hat{\boldsymbol{\beta}}_0^{(2)}(\tau, t_0)dt_0$. We can show that the limiting distribution of $n(\hat{\Omega}_\tau - \Omega_\tau)$ is a mean zero normal distribution, the variance of which may be consistently estimated by

$n\hat{\sigma}_{\Omega_\tau}^2$, where $\hat{\sigma}_{\Omega_\tau}^2$ equals the (2,2) element of

$\frac{1}{n^2}\sum_{i=1}^{n}\left\{\frac{1}{t_U - t_L}\int_{t_L}^{t_U}\sqrt{n}\boldsymbol{D}_n(\tau, t_0)\boldsymbol{E}_n^{-1}(\tau, t_0)\hat{\zeta}_i(\tau, t_0)dt_0\right\}^{\otimes 2}$. This result naturally renders a Wald-type test, $T_{\Omega_\tau} = \hat{\Omega}_\tau/\hat{\sigma}_{\Omega_\tau}$, for the null hypothesis $H_{01} : LCQRR(\tau, t_0) = 0$, $t_0 \in [t_L, t_U]$. That is, we reject H_{01} when $|T_{\Omega_\tau}| > 100(1 - a/2)$th percentile of $N(0, 1)$ distribution, where a is the desired significance level. Similar results can be obtained for the overall summary and testing of $LCQRR(\tau, t_0)$ over $\tau \in [\tau_L, \tau_U]$, corresponding to

$\Omega_{t_0} = \frac{1}{\tau_U - \tau_L}\int_{\tau_L}^{\tau_U}\boldsymbol{\beta}_0^{(2)}(\tau, t_0)d\tau$, and $H_{03} : LCQRR(\tau, t_0) = 0$, $\tau \in [\tau_L, \tau_U]$ respectively.

We can also test the constancy of $LCQRR(\tau, t_0)$ over t_0 or τ. For example, a null hypothesis of interest may take the form, $H_{02} : LCQRR(\tau, t_0) = C_\tau$, $t_0 \in [t_L, t_U]$, where C_τ is an unspecified constant and may change with τ. Let $\Xi(\tau, t_0)$ denote a known weight function satisfying $\Xi(\tau, t_0) \quad 0$ and $\int_{t_L}^{t_U}\Xi(\tau, t_0)dt_0 = 1$. If H_{02} holds, then

$\int_{t_L}^{t_U} \Xi(\tau, t_0) \boldsymbol{\beta}_0^{(2)}(\tau, t_0) dt_0 - \Omega_\tau = [\int_{t_L}^{t_U} \Xi(\tau, t_0) dt_0 - 1] C_\tau = 0$. This motivates us to construct a

test statistic for H_{02} based on $\Gamma_\tau = \sqrt{n} \{ \int_{t_L}^{t_U} \Xi(\tau, t_0) \hat{\boldsymbol{\beta}}^{(2)}(\tau, t_0) dt_0 - \hat{\Omega}_\tau \}$. Following the same

line for proving Theorem 3.2, we can show that the limiting distribution of Γ_τ under H_{02} is

normal with mean 0. A consistent variance estimate for Γ_τ may be given by $\hat{\sigma}_{\Gamma_\tau}^2$, which is the

(2,2) element of

$n^{-1} \sum_{i=1}^{n} \left[\int_{t_L}^{t_U} \{ \Xi(\tau, t_0) - \frac{1}{t_U - t_L} \} \sqrt{n} \boldsymbol{D}_n(\tau, t_0) \boldsymbol{E}_n^{-1}(\tau, t_0) \hat{\boldsymbol{\zeta}}_i(\tau, t_0) dt_0 \right]^{\otimes 2}$. A Wald-type

test for H_{02} is then given by $T_{\Gamma_\tau} = \Gamma_\tau / \hat{\sigma}_{\Gamma_\tau}$. A similar testing procedure can be developed for

testing the constancy over $t_0 \in [t_L, t_U]$, $H_{04} : LCQRR(\tau, t_0) = C_{t_0}$, $\tau \in [\tau_L, \tau_U]$.

4. An Extension to Adjusting for Covariates

Exploiting population heterogeneity in semi-competing risks dependence is often scientifically meaningful, and for example, can help uncover uncommon disease mechanisms in subgroups. To this end, we propose an extension, which adjusts for covariates (captured by $\tilde{\boldsymbol{Z}} \in R^p$) in the assessment of the dependence between the nonterminal event and the terminal event.

First, we define the covariate-adjusted log cross quantile residual ratio as

$$LCQRR(\tau, t_0 | \tilde{\boldsymbol{Z}}) = \log \left[\frac{Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 > t_0, \tilde{\boldsymbol{Z}})}{Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 \leq t_0, \tilde{\boldsymbol{Z}})} \right].$$

When all covariates of interest are discrete, one may conduct stratified analyses based on the methods in Section 3 to estimate and make inference on $LCQRR(\tau, t_0 | \tilde{\boldsymbol{Z}})$.

In many practical settings, covariates of interest can be continuous. Thus we investigate a general scenario where $\tilde{\boldsymbol{Z}}$ can include both continuous and discrete covariates. Specifically, we are interested in formulating linear covariate effects on $LCQRR$, which may be expressed as

$$LCQRR(\tau, t_0 | \tilde{\boldsymbol{Z}}) = \check{\boldsymbol{Z}}^T \boldsymbol{\alpha}_0(\tau, t_0), \quad (4)$$

where $\check{\boldsymbol{Z}} = (1, \tilde{\boldsymbol{Z}}^T)^T$. The non-intercept coefficients in $\boldsymbol{\alpha}_0(\tau, t_0)$ depict how $LCQRR$ changes per unit change in the corresponding covariate.

To address the interest in the linear effects of covariates on $LCQRR$, we consider the following quantile residual lifetime model:

$$Q_\tau(T_2 - t_0 | T_2 > t_0, I(T_1 > t_0), \tilde{\boldsymbol{Z}})$$

$$= \exp\{\boldsymbol{Z}^T(t_0)\boldsymbol{\gamma}_0(\tau, t_0)\} \equiv \exp[\boldsymbol{\gamma}_0^{(1)}(\tau, t_0) + I(T_1 > t_0)\boldsymbol{\gamma}_0^{(2)}(\tau, t_0) + \tilde{\boldsymbol{Z}}^T \boldsymbol{\gamma}_0^{3:(2+p)}(\tau, t_0) + \tilde{\boldsymbol{Z}}^T I(T_1 > t_0)\boldsymbol{\gamma}_0^{(3+p):(2+2p)}(\tau, t_0)],$$

(5)

where $\boldsymbol{Z}(t_0) = (1, I(T_1 > t_0), \tilde{\boldsymbol{Z}}^T, \tilde{\boldsymbol{Z}}^T I(T_1 > t_0))^T$, and $\boldsymbol{v}^{a:b}$ denotes the vector that includes the ath to bth components of vector $\boldsymbol{v}$. It is important to note that (5) implies

$$\mathrm{LCQRR}(\tau; t_0 | \tilde{\boldsymbol{Z}}) = \boldsymbol{\gamma}_0^{(2)}(\tau, t_0) + \tilde{\boldsymbol{Z}}^T \boldsymbol{\gamma}_0^{(3+p):(2+2p)}(\tau, t_0).$$

When there are only discrete covariates, model (5) and model (4) can be equivalent. These suggest that under slightly stronger assumptions regarding the effects of continuous covariates, model (5) defines the same linear relationship between covariates and *LCQRR* as does model (4). Compared to model (4), model (5) is more convenient to tackle. This is because model (5) takes the same form as the working quantile residual lifetime model (1) considered for the one-sample case. As shown below, this fact greatly facilitates an extension to the general case with covariates. By these considerations, we adopt model (5) as the vehicle to explore the linear covariate effects on *LCQRR*.

Suppose the observed data include n i.i.d. replicates, $(X_i^*, Y_i^*, \delta_i^*, \eta_i^*, L_i^*, \tilde{\boldsymbol{Z}}_i^*)_{i=1}^n$, where $\tilde{\boldsymbol{Z}}_i^*$ is the truncated counterpart of $\tilde{\boldsymbol{Z}}_i$ following the conditional distribution of $\tilde{\boldsymbol{Z}}$ given $Y > L$. We assume that D is independent of $(T_1, T_2, L, \tilde{\boldsymbol{Z}})$ and L is independent of T_2 given $(T_1, \tilde{\boldsymbol{Z}})$.

Define $\boldsymbol{K}^*(t_0) = (1, I(X_i^* > t_0), \tilde{\boldsymbol{Z}}_i^{*T}, \tilde{\boldsymbol{Z}}_i^{*T} I(X_i^* > t_0))^T$. Adapting the idea presented for the one-sample case, we propose to estimate $\boldsymbol{\gamma}_0(\tau, t_0)$ by solving the following estimating equation for $\boldsymbol{r} \in R^{2+2p}$:

$$\boldsymbol{S}_n(\boldsymbol{r}, \tau, t_0) = \boldsymbol{0},$$

where

$$\boldsymbol{S}_n(\boldsymbol{r}, \tau, t_0) = n^{-1/2} \sum_{i=1}^n \frac{I(L_i^* \leq t_0) I(Y_i^* > t_0) \eta_i^*}{\hat{G}(Y_i^* - L_i^*)} \boldsymbol{K}_i^*(t_0) \{I[\log(Y_i^* - t_0) \leq \boldsymbol{K}_i^{*T}(t_0)\boldsymbol{r}] - \tau\}.$$

The resulting estimator is denoted by $\hat{\boldsymbol{\gamma}}(\tau, t_0)$. It is easy to see that the subvector, $\hat{\boldsymbol{\gamma}}^{(3+p):(2+2p)}(\tau, t_0)$, can be used to describe the linear effect of $\tilde{\boldsymbol{Z}}$ on *LCQRR*. With an additional assumption that $\tilde{\boldsymbol{Z}}$ is uniformly bounded (i.e. $\sup_i \|\tilde{\boldsymbol{Z}}_i\| \leq M_1 < \infty$), we can established the same asymptotic properties and inference procedures for $\hat{\boldsymbol{\gamma}}(\tau, t_0)$ as those presented in Section 3.

5. Numerical Studies

5.1 Simulations

Simulation studies are conducted to examine the finite-sample performance of the proposed methods in the left-truncated semi-competing risks setting. Specifically, we generate (T_1, T_2) from a gamma frailty model, in which

$$P(T_1 > x, T_2 > y) = [P(T_1 > x)^{1-\theta} + P(T_2 > x)^{1-\theta} - 1]^{1/(1-\theta)},$$

with T_i following a Weibull(a_i, λ_i) distribution and $P(T_i > x) = exp(-\lambda_i x^{a_i})$, $i = 1, 2$. The truncation time L is generated from a mixture of a point mass at zero and a positive-valued random variable $\tilde{L}$. The proportion of zero truncation time is set as 20%.

The simulations are conducted under two scenarios,

Scenario 1: $T_1 \sim$ Weibull(1.4, 0.6), $T_2 \sim$ Weibull(3.5, 0.5), $\tilde{L}$ and D following uniform distributions.

Scenario 2: $T_1 \sim$ Weibull(3, 0.85), $T_2 \sim$ Weibull(3, 0.4), $\tilde{L}$ and D following Weibull distributions.

For Scenario 1, there is a low truncation level with $P(Y < L) = 0.3$, and a high dependent censoring rate with $P(\delta = 0, \eta = 1 | L \quad Y)$ close to 0.4. For Scenario 2, there is a high truncation level of 0.5 and a low dependent censoring rate around 0.15. In each scenario, we consider three different θ values, 1, 2 and 3, corresponding to independence, moderate positive association, and high positive association respectively. The detailed specifications of the marginal distributions of $\tilde{L}$ and D as well as censoring and truncation proportions are shown in Table 1.

We perform the proposed methods on 1000 simulated datasets with sample size $n = 200$ or 400 for each simulation setup, where M is set as 10^7. For Scenario 1, Figure 1 presents the empirical bias (EmpBias), empirical standard error (EmpSE) and average estimated standard error (EstSE) for the proposed estimator of $LCQRR(\tau, t_0)$ under different combinations of (θ, τ, t_0), where $\tau = 0.25, 0.5, 0.75$, $t_0 = 0.55, 0.84, 1.1$ and circles denote corresponding values. It is observed that the proposed estimator of $LCQRR(\tau, t_0)$, performs well with moderate sample size. The point estimates have small biases. The corresponding standard error estimates agree well with empirical standard errors, and the agreement generally improves as sample size increases. We have very similar observations from Figure 1 in Web Appendix B, which presents the simulation results for Scenario 2.

We also examine the proposed second-stage inferences. With fixed τ, we evaluate the average of $LCQRR$ over $t \in [t_L, t_U]$, and test whether $LCQRR(\tau, t_0)$ equals 0 for $t \in [t_L, t_U]$ and whether $LCQRR(\tau, t_0)$ is constant over $t \in [t_L, t_U]$. We consider three τ values, 0.25, 0.5, and 0.75. For Scenario 1, we set $t_L = 0.42$ and $t_U = 1.20$. For Scenario 2, we set $t_L = 0.68$ and $t_U = 1.28$. We compute integrals using left Riemann sums on intervals of equal

length 0.001 and choose the weight function $\Xi(\tau, t_0) = 2I[t_0 \quad (t_L + t_U)/2]/(t_U - t_L)$. In Table 2, we summarize the EmpBias, EmpSE and EstSE of $\hat{\Omega}_\tau$, and the empirical rejection rates (EmpRR) for the proposed Wald tests for H_{01} and H_{02}. Note that for both H_{01} and H_{02}, the EmpRR gives empirical sizes when $\theta = 1$ and empirical power when $\theta = 2, 3$. Table 2 shows that for both scenarios, the empirical biases of $\hat{\Omega}_\tau$ are small and the estimated standard errors match the empirical standard errors very well. The test for either H_{01} or H_{02} appear to have empirical sizes close to the nominal levels. The power for testing H_{01} is good, while the constancy tests appear to be conservative. The empirical power increases considerably as sample size and θ value increase for both tests.

With fixed t_0, we assess the second-stage inferences over $[\tau_L, \tau_U]$. For Scenario 1, we consider $t_0 = 0.55, 0.84, 1.10$ and set $[\tau_L, \tau_U] = [0.1, 0.87]$. For Scenario 2, we consider $t_0 = 0.85, 1.00, 1.20$ and set $[\tau_L, \tau_U] = [0.1, 0.9]$. In both scenarios, $\Xi(\tau, t_0) = 2I[\tau \quad (\tau_L + \tau_U)/2]/(\tau_U - \tau_L)$. Table 1 in Web Appendix B presents the EmpBias, EmpSE and EstSE of $\hat{\Omega}_{t_0}$ and the EmpRR for the proposed tests. Similarly, we observe small empirical biases, well-matched estimated and empirical standard errors, and pretty accurate empirical sizes. The power for the constancy tests is not high but increases as sample size increases.

5.2 Denmark Diabetes Registry Data Analysis

We apply the proposed method to a dataset from the Denmark diabetes registry study (Andersen et al., 1993). The Denmark diabetes registry study is a prospective cohort study on insulin-dependent diabetes patients referred to the Steno Memorial Hospital in Greater Copenhagen. Diabetic nephropathy (DN), an indicator of kidney failure, is a significant complication among patients with diabetes. From 1933 to 1981, 2727 patients who were diagnosed with insulin-dependent diabetes mellitus prior to age 31 and between 1933 and 1972 were accrued. At entry, patients' age at diabetes diagnosis and the presence of DN were recorded. All patients were then followed until death, emigration or December 31, 1984. In our analysis, the time origin is the age at diabetes diagnosis, with event times recorded in years since diagnosis. It is seen that time to DN and time to death naturally formed a semi-competing risks structure because death terminated the observation on time to DN, but remained observable after the occurrence of DN. Administrative left truncation on mortality was also involved. That is, patients who had died before study enrollment were excluded. Out of 2727 patients, there were 731(26.8%) experiencing DN, 718(26.3%) dead in the end and 652(24%) with diabetic onset at entry. Summary statistics for the data are presented in Table 3.

Our focus is first to quantify the relationship between DN and death by using the proposed measure $LCQRR(\tau, t_0)$. We fit model (1) to the data and adopt $M = 10^7$ as in the simulations. We restrict t0 to be within [6,40] to ensure reasonable sample sizes accumulated for strata defined by $I(X^* > t_0)$. In Figure 2, we display the results for $\tau = 0.25, 0.5, 0.75$ and t_0 values at an equally space grid on [6, 40] with step size=0.1. Estimated $LCQRR(\tau, t_0)$ are plotted in bold solid lines. The corresponding 95% pointwise confidence intervals are in dotted lines and the 95% pointwise Wald-type bootstrapping confidence intervals are in long-dashed lines. In Figure 2, we see that for all three τ values, the estimated $LCQRR(\tau, t_0)$ is generally positive; the lower bounds of confidence intervals are above 0 for t_0 less than 30,

which is roughly the third quartile of X^*. This observation is consistent with the common belief that DN is positively associated with mortality. Our formal test for H_{01} yields p-values, < 0.001, 0.002, < 0.001, respectively, for $\tau = 0.25, 0.5, 0.75$, confirming that DN is a significant prognostic factor for mortality.

We note that the confidence intervals for $LCQRR(\tau, t_0)$ with $t_0 > 30$ become wider and mostly cover 0. This may be partly due to the reduced power/efficiency as t_0 approaches the upper tail of X, resulting in smaller effective sample sizes for the proposed estimator. The insignificant difference between $LCQRR(\tau, t_0)$ and 0 with $t_0 > 30$ may also have the implication that the occurrence of DN has diminished prognostic power for mortality among patients who had lived long since diabetes diagnosis. In addition, we observe that the estimated $LCQRR(\tau, t_0)$ appears rather constant for $\tau = 0.25$ and $\tau = 0.5$, but the decreasing trend in the estimated $LCQRR(\tau, t_0)$ with $\tau = 0.75$ is quite apparent. This observation is confirmed by the constancy tests for H_{02}, which yield p-values, 0.95, 0.23, and 0.01 for $\tau = 0.25, 0.5, 0.75$ respectively. The significant changing pattern of $LCQRR(\tau, t_0)$ may second the previously conjectured inhomogeneous prognostic ability of DN on mortality.

We also choose three t_0 values, $t_0 = 15, 21, 29$, which stand for the 25th, 50th and 75th quantile of X^*, respectively, to explore the patterns of $LCQRR(\tau, t_0)$ over $\tau \in [0.1, 0.82]$. Figure 2 of Web Appendix B displays estimated $LC\hat{Q}RR(\tau, t_0)$ in bold solid lines at equally spaced τ-grids with step size 0.001, with the corresponding 95% pointwise confidence intervals in dotted lines and 95% pointwise Wald-type bootstrapping confidence intervals in long-dashed lines. We observe that $LCQRR(\tau, t_0)$ may be significantly different from 0 for all three t_0's. This is confirmed by tests for H_{03}, which give p-values, < 0.001, < 0.001, and 0.002, respectively. For $t_0 = 21$ and 29, we observe a clear decreasing trend in the estimated $LCQRR(\tau, t_0)$. Constancy tests for H_{04} yield p-values, 0.24, 0.004, 0.004, for $t_0 = 15, 21, 29$, respectively. The finding that $LCQRR(\tau, t_0)$ may decrease with τ aligns with previous results, manifesting a weak or negligible association between DN and mortality in long-term diabetes survivors.

Next, we study how diabetes onset age, a continuous covariate, affects the dependence between DN and mortality. We fit model (5) to the data and the coefficient $\gamma_0^{(4)}(\tau, t_0)$ represent the change in $LCQRR$ per one year increase in diabetes onset age. For $\tau = 0.25$, $0.5, 0.75$, we estimate $\gamma_0^{(4)}(\tau, t_0)$ at an equally spaced grid on $[8, 36]$ with step size 0.1 for t_0. In Figure 3, we display the estimates for $\gamma_0^{(4)}(\tau, t_0)$ along with their 95% pointwise confidence intervals. We see from Figure 3 that with all selected τ's, $\hat{\gamma}_0^{(4)}(\tau, t_0)$ is generally significantly positive for t_0 belong to the first half of the time interval $[8, 36]$, but loses significance from 0 for larger t_0. This suggests that for patients who were diagnosed with diabetes at older age, the occurrence of DN before t_0 may imply a bigger disadvantage in residual survival time. Such an effect of diabetes onset age may diminish for large t_0's, which point to the groups of patients who had survived for a long time since diagnosis. Tests for H_{01} over $t_0 \in [8, 22)$ confirm our observation from Figure 3, yielding three nearly zero p-values. Constancy tests for H_{02} gave p-values, $0.64, 0.11, 0.07$, respectively, for $\tau = 0.25, 0.5,$

0.75. This provides some evidence for the observed diminishing effect of diabetes onset age over t_0.

We also evaluate $\hat{\gamma}_0^{(4)}(\tau, t_0)$ over a τ-range [0.1, 0.82] for fixed t_0 values, 15, 21, 29. Results displayed in Figure 3 of Web Appendix B suggest similar findings. That is, DN may have a bigger influence on subsequent mortality for patients with later diabetes diagnosis compared to those with earlier diagnosis. Such an effect of diagnosis age may varnish when t_0 is large.

6. Remarks

In this paper, we propose a robust measure to assess the dependence of the nonterminal event and the terminal event in a semi-competing risks setting. Evaluating this measure at multiple t_0 and τ allows us to perform a comprehensive and robust evaluation of semi-competing risks dependence. It also offers the flexibility to explore the dynamic pattern of the dependence structure. The developed estimation and inference procedures well utilize the semi-competing risks structure with left truncation, and can be extended to adjust for covariates. Simulation studies show that the proposed estimation procedure performs well in finite sample cases.

Other approaches to obtaining a nonparametric estimator of $LCQRR(\tau, t_0)$ are available. For example, in the standard semi-competing risks setting without left truncation, note that $T_1 \wedge T_2$ is only subject to independent censoring by C and thus the joint survival function of (T_1, T_2) on the upper wedge can be consistently estimated by using methods, such as Lin and Ying (1993). Then we can estimate the two conditional residual quantiles in $LCQRR(\tau, t_0)$ by reversing their corresponding conditional distribution estimates. Our preference of adopting a quantile residual lifetime regression framework is primarily because of the resulting simple extension to accommodate covariates in the consideration of $LCQRR(\tau, t_0)$. Our strategy of connecting $LCQRR$ with quantile residual lifetime regression models enables a unified approach to characterizing semi-competing risks dependence with or without covariates. Existing techniques for quantile regression can readily be applied to inferences and make our work neat.

In practice, the choices of τ and t_0 mainly depend on the interest of investigators. They may be adjusted according to the empirical observations of the data. For example, the estimation efficacy may be unsatisfactory at small or large values of t_0. This is because the number of observations satisfying $X^* \leq t_0$ (or $X^* > t_0$) may be quite small when t_0 is small (or larger), making the estimate for $Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 \leq t_0)$ (or $Q_\tau(T_2 - t_0 | T_2 > t_0, T_1 > t_0)$) inaccurate or unstable. Based on our numerical experiences, we find that our method works well for estimating both $LCQRR(\tau, t_0)$ and covariance matrix when $n_{t_0,1} \wedge n_{t_0,2} > 15$, where $n_{t_0,1} = \sum_{i=1}^n I(L_i^* \leq t_0, Y_i^* > t_0, X_i^* > t_0)\eta_i^*$ and $n_{t_0,2} = \sum_{i=1}^n I(L_i^* \leq t_0, Y_i^* > t_0, X_i^* \leq t_0)\eta_i^*$. For a larger τ, we may need $n_{t_0,1}$ and $n_{t_0,2}$ to be larger. These can serve as useful empirical rules to guide the selection of τ and t_0 in real data analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, N. Statistical Models Based on Counting Processes. Springer; New York: 1993.

Asgharian M, M'Lan C, Wolfson D. Length-biased sampling with right censoring. Journal of the American Statistical Associatio. 2002; 97:201–209.

Chen YH. Maximum likelihood analysis of semicompeting risks data with semiparametric regresion models. Lifetime Data Analysis. 2012; 18

Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978; 65:141–151.

Fine JP, Jiang H, Chappell R. On semi-competing risks data. Biometrika. 2001; 88:907–919.

Fygenson M, Ritov Y. Monotone estimating equations for censored data. The Annals of Statistics. 1994; 22:732–746.

Gelfand AE, Kottas A. Bayesian semiparametric regression for median residual life. Scandinavian Journal of Statistics. 2003; 30:651–665.

Ghosh D. Semiparametric inferences for association with semi-competing risks data. Statistics in Medicine. 2006; 25:2059–2070. [PubMed: 16196081]

Hsieh JJ, Wang W, Ding AA. Regression analysis based on semicompeting risks data. Journal of the Royal Statistical Society Series B. 2008; 70:3–20.

Jeong JH, Jung SH, Costantino JP. Nonparametric inference on median residual life function. Biometrics. 2008; 64:157–163. [PubMed: 17501936]

Jin Z, Ying Z, Wei L. A simple resampling method by perturbing the minimand. Biometrika. 2001; 88:381–390.

Jung SH, Jeong JH, Bandos H. Regression on quantile residual life. Biometrics. 2009; 65:1203–1212. [PubMed: 19432781]

Lakhal L, Rivest LP, Abdous B. Estimating survival and association in a semicompeting risks model. Biometrics. 2008; 64:180–188. [PubMed: 17645782]

Li R, Peng L. Quantile regression for left-truncated semi-competing risks data. Biometrics. 2011; 67:701–710. [PubMed: 21133883]

Li R, Peng L. Quantile regression adjusting for dependent censoring from semi-competing risks. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2015; 77:107–130. [PubMed: 25574152]

Lin D, Ying Z. A simple nonparametric estimator of the bivariate survival function under univariate censoring. Biometrika. 1993; 80:573–581.

Lin DY, Robins JM, Wei LJ. Comparing two failure time distri-butions in the presence of dependent censoring. Biometrika. 1996; 83:381–393.

Ma Y, Yin G. Semiparametric median residual life model and inference. The Canadian Journal of Statistics. 2010; 34:665–679.

Oakes D. Bivariate survival models induced by frailties. Journal of the American Statistical Association. 1989; 84:414–422.

Parzen M, Wei L, Ying Z. A resampling method based on pivotal estimating functions. Biometrika. 1994; 81:341–350.

Peng L, Fine JP. Nonparametric estimation with left truncated semicompeting risks data. Biometrika. 2006a; 93:367–383.

Peng L, Fine JP. Rank estimation of accelerated lifetime models with dependent censoring. Journal of the American Statistical Association. 2006b; 101:1085–1093.

Peng L, Fine JP. Regression modeling of semi-competing risks data. Biometrics. 2007; 63:96–108. [PubMed: 17447934]

Peng L, Fine JP. Competing risks quantile regression. Journal of the American Statistical Association. 2009; 104:1140–1453.

Wang M. Nonparametric estimation from cross-sectional survival data. Journal of the American Statistical Association. 1991; 86:130–143.

Wang W. Estimating the association parameter for copula models under dependent censoring. Journal of the Royal Statistical Society Series B. 2003; 65:257–273.
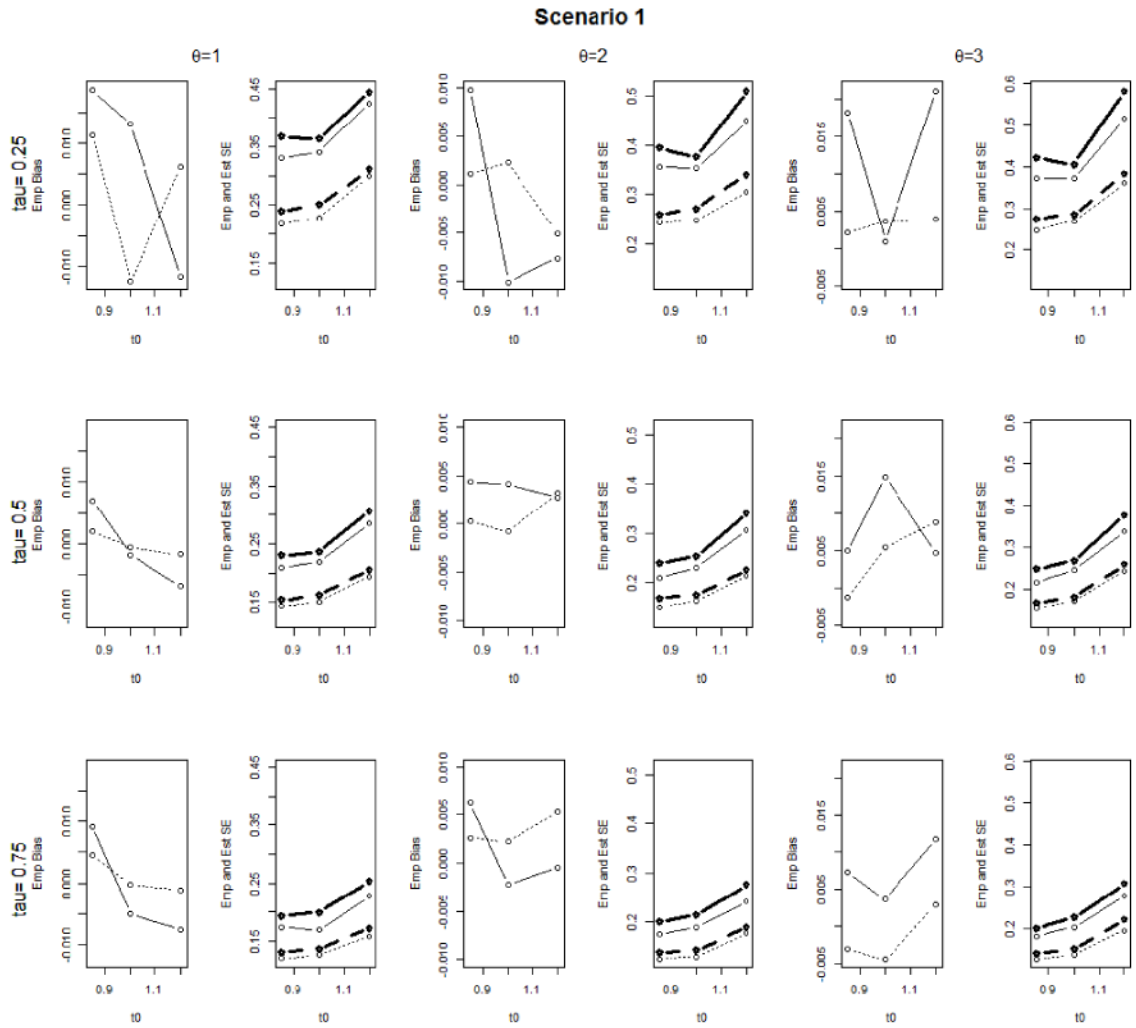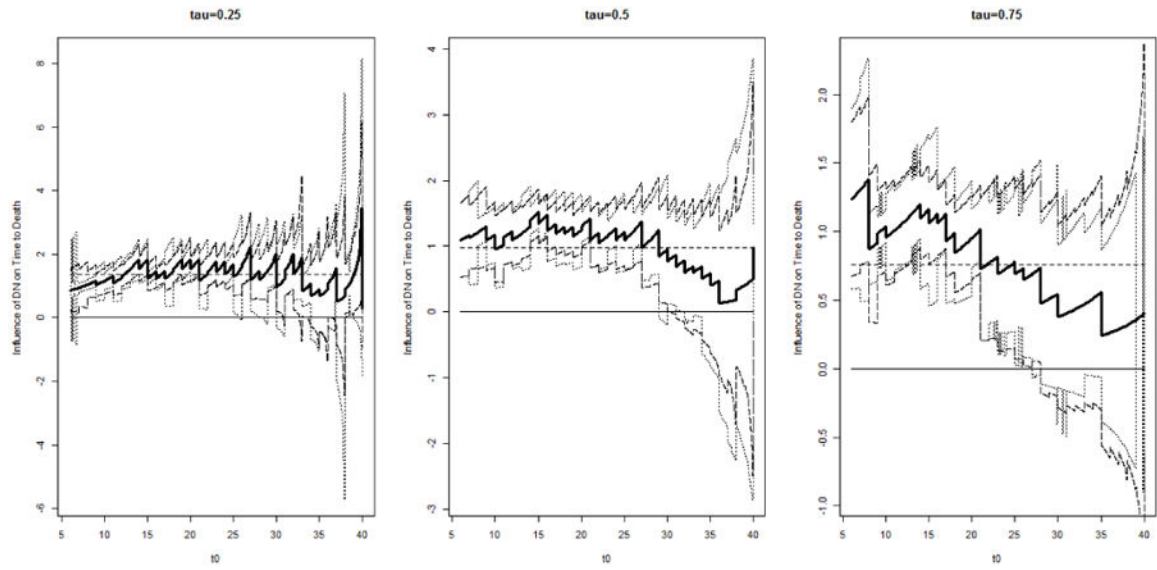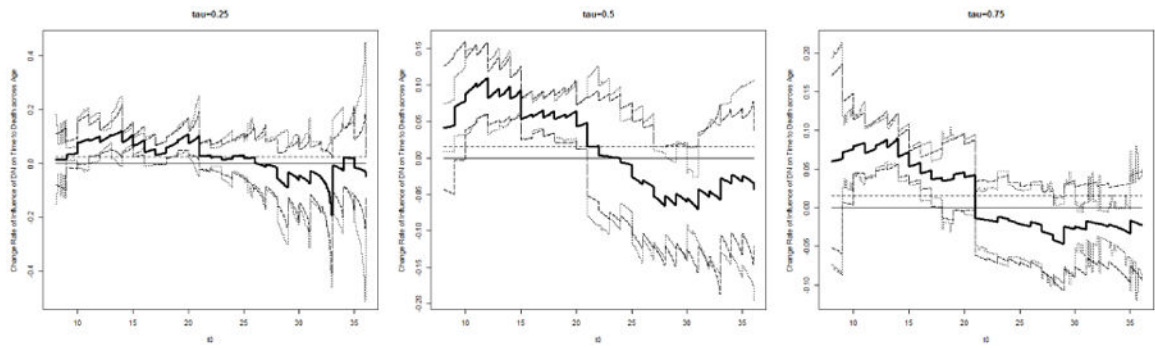
Figure 1.
Simulation results for Scenario 1: Empirical bias (EmpBias), empirical standard error (EmpSE) and average estimated standard error (EstSE) of the proposed estimator of *LCQRR*. EmpBias for $n = 200$ and EmpBias for $n = 400$ are plotted in solid lines and dotted lines respectively. EmpSE and EstSE for $n = 200$ are plotted in solid lines and bold solid lines respectively. EmpSE and EstSE for $n = 200$ are plotted in dotted lines and bold dashed lines respectively.

Figure 2.
Denmark Diabetes Registry Study: Estimated $LCQRR(\tau, t_0)$ (bold solid lines), and the corresponding 95% pointwise confidence intervals (dotted lines) and 95% pointwise Wald-type bootstrapping confidence intervals (long-dashed lines).

Figure 3.

Denmark Diabetes Registry Study: $\gamma_0^{(4)}(\tau, t_0)$ Estimated (bold solid lines), the corresponding 95% pointwise confidence intervals (dotted lines) and 95% pointwise Wald-type bootstrapping confidence intervals (long-dashed lines), and the overall influence of DN across time (horizontal dashed lines).

Table 1

Summary statistics for different simulation scenarios. All % are conditional on $L < Y$.

θ	L	D	$P(Y < L)$	$\%(\delta = 0)$	$\%(\eta = 0)$	$\%(\delta = 0, \eta = 1)$	$\%(X < L)$
				Scenario 1			
1	Un(0,1.67)	Un(0.05,3.2)	30	52	21	39	20
2	Un(0,1.67)	Un(0.17,2.6)	30	59	22	42	15
3	Un(0,1.67)	Un(0.15,2.55)	30	63	23	44	13
				Scenario 2			
1	Wei(2.6,0.35)	Wei(1.1,0.38)	50	26	20	16	40
2	Wei(1.2,0.49)	Wei(1.3,0.3)	50	27	20	15	20
3	Wei(0.5,0.55)	Wei(1.5,0.22)	50	27	20	15	9

Table 2

Empirical biases, empirical standard errors and average standard errors estimates of $\hat{\Omega}_\tau$ and empirical rejection rates for H_{01} and H_{02}

θ	τ	n	EmpBias	$\hat{\Omega}_\tau$		H_{01}	H_{02}
				EmpSE	EstSE	EmpRR	EmpRR
				Scenario 1			
				$t_0 \in [0.42, 1.20]$			
1	0.25	200	0.008	0.181	0.191	0.041	0.051
		400	0.003	0.128	0.132	0.053	0.046
	0.50	200	0.006	0.153	0.160	0.056	0.043
		400	0.005	0.112	0.110	0.052	0.046
	0.75	200	0.008	0.143	0.149	0.065	0.037
		400	0.001	0.101	0.104	0.054	0.048
2	0.25	200	0.006	0.172	0.188	0.927	0.102
		400	0.003	0.119	0.127	1.000	0.149
	0.50	200	0.007	0.140	0.153	0.928	0.142
		400	0.003	0.102	0.107	0.995	0.215
	0.75	200	0.003	0.137	0.148	0.857	0.160
		400	0.005	0.099	0.104	0.988	0.234
3	0.25	200	0.014	0.166	0.182	0.999	0.118
		400	0.001	0.112	0.122	1.000	0.213
	0.50	200	0.011	0.143	0.149	0.999	0.184
		400	0.003	0.097	0.103	1.000	0.313
	0.75	200	0.010	0.143	0.147	0.981	0.200
		400	0.005	0.097	0.103	1.000	0.347
				Scenario 2			
				$t_0 \in [0.68, 1.28]$			
1	0.25	200	-0.004	0.243	0.237	0.066	0.066
		400	0.001	0.168	0.164	0.053	0.059
	0.50	200	-0.004	0.193	0.198	0.065	0.049
		400	0.002	0.133	0.138	0.048	0.050
	0.75	200	-0.001	0.174	0.179	0.080	0.047

θ	τ	n	EmpBias	$\hat{\Omega}_{\tau}$ EmpSE	EstSE	H_{01} EmpRR	H_{02} EmpRR
		400	0.003	0.119	0.127	0.056	0.039
2	0.25	200	0.003	0.172	0.176	0.984	0.117
		400	0.001	0.124	0.120	1.000	0.211
	0.50	200	-0.004	0.140	0.141	0.984	0.128
		400	-0.001	0.098	0.097	1.000	0.194
	0.75	200	-0.001	0.128	0.133	0.965	0.096
		400	0.000	0.087	0.093	1.000	0.144
3	0.25	200	-0.007	0.133	0.137	1.000	0.161
		400	0.002	0.087	0.093	1.000	0.259
	0.50	200	-0.002	0.113	0.116	1.000	0.125
		400	0.000	0.078	0.080	1.000	0.161
	0.75	200	0.000	0.108	0.117	1.000	0.100
		400	0.001	0.075	0.081	1.000	0.118

Table 3

Summary statistics for diabetes registry data

	$n(\%)$
$(\delta, \eta) = (0, 0)$	1729(63.4%)
$(\delta, \eta) = (0, 1)$	267(9.8%)
$(\delta, \eta) = (1, 0)$	280(10.3%)
$(\delta, \eta) = (1, 1)$	451(16.5%)
$L = 0$	652(24%)
$X < L$	116(4.25%)