Association measure of doubly interval censored data using a Kendall's τ estimator

Seo-Hyun Kang[a], Yang-Jin Kim[1,b]

[a]Ajou Medical Center, Korea; [b]Department of Statistics, Sookmyung Women's University, Korea

Abstract

In this article, our interest is to estimate the association between consecutive gap times which are subject to interval censoring. Such data are referred as doubly interval censored data (Sun, 2006). In a context of serial event, an induced dependent censoring frequently occurs, resulting in biased estimates. In this study, our goal is to propose a Kendall's τ based association measure for doubly interval censored data. For adjusting the impact of induced dependent censoring, the inverse probability censoring weighting (IPCW) technique is implemented. Furthermore, a multiple imputation technique is applied to recover unknown failure times owing to interval censoring. Simulation studies demonstrate that the suggested association estimator performs well with moderate sample sizes. The proposed method is applied to a dataset of children's dental records.

Keywords: doubly interval censored data, Kendall's τ, induced dependent censoring, IPCW, multiple imputation

1. Introduction

There are several approaches to analyze multivariate failure times where subjects experience more than one type of event or repeatedly same type of events. These can also be classified as either parallel or sequential events, respectively (Cook and Lawless, 2007). While an ordering between two events does not exist in a parallel scale context, the occurrence and timing of the initial event affect the occurrence of the sequential event in a sequential scale.

Diverse approaches have been proposed to analyze sequential events in life history processes. A multistate model has been frequently applied. For example, in an HIV-related study, the initial state is an infection-free state, the intermediate state (the initial event) is HIV infection and the terminal state (the second event) represents AIDS diagnosis, respectively. For analyzing such sequential events, two time scales such as the gap time and the total time are imposed depending on the features of process and the research purpose. For example, the gap time provides more suitable information to evaluate the treatment efficacy on the duration of the second event and a semi-Markov model is applied to investigate the effect of covariates.

Let U and T represent the times to the first and second events($U < T$), respectively. When U and T are interval censored, the available data are composed of $\{(L_U, R_U], (L_T, R_T]\}$, respectively. Then a duration time or a gap time $V = T - U$ between these events is also interval censored, $(L_V, R_V]$, where $L_V = L_T - R_U$ and $R_V = R_T - L_U$, respectively. This data structure is denoted as doubly interval censored data (Sun, 2006) and is presented at Figure 1. De Gruttola and Lagakos (1989) proposed

---

[1] Corresponding author: Department of Statistics, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yonsan-Gu, Seoul 04310, Korea. E-mail: yjin@sookmyung.ac.kr
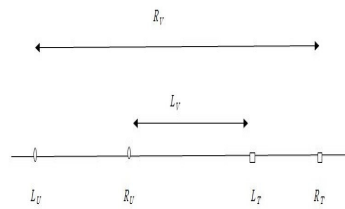
Figure 1: *Visualization of doubly interval censored data.*

a nonparametric estimator of distributions of (U, V) based on the EM algorithm at discrete time supports and Kim *et al.* (1993) extended this method to the proportional hazard model for estimating the effect of covariate on the duration time. Kim (2006) adopted a bivariate frailty effect under a piecewise constant baseline hazard model. Dejardin and Lesaffre (2013) applied a stochastic EM algorithm by extending Pan (2001)'s approach. However, there are few studies about the association measure for doubly interval censored data while some approaches were suggested for bivariate interval censored data where two events are observed on the parallel scale. For example, Betensky and Finkelstein (1999) implemented a multiple imputation technique based on the nonparametric estimator of a bivariate survival function. Lesaffre and Bogaerts (2008) suggested a two-stage procedure for estimating the association. In the first stage, the joint density of log failure time is estimated with a smoothing method based on a pair of grids and corresponding weights. At the second stage, these grids and weights are implemented to calculate the Kendall's τ estimate as a global association measure as well as the cross ratio function as a local association measure. Wang and Ding (2000) also applied a two stage estimation procedure to estimate a constant parameter α of a copula function $C_\alpha(1 - F_1, 1 - F_2)$ for doubly current status data (or case I interval censored data). They showed the estimator of α has a $n^{1/2}$ convergence rate while the nonparametric estimator $\hat{F}$ follows a $n^{1/3}$ convergence rate. Sun *et al.* (2006) extended to general bivariate interval censored data.

The Kendall's τ is the most frequently referred association measure under bivariate survival data. Oakes (1982, 2008) derived the estimator of Kendall's τ for a right censored data and discussed the consistency. Wang and Wells (2000) derived nonparametric estimators of τ under diverse censoring schemes. Lakhal-Chaieb *et al.* (2009) applied the inverse probability censoring weighted (IPCW) technique to improve Oakes' estimator and Lakhal-Chaieb *et al.* (2010) extended this approach to estimate the association between two serial gap times. Fu *et al.* (2016) proposed several Kendall's τ estimators for estimating the serial associations as the function of previous gap times in a context of recurrent event data. Our interest is to propose an association measure of doubly interval censored data by extending Lakhal-Chaieb *et al.*'s estimator. The reminder of this article is organized as follows. In Section 2, we introduce the notations and propose the nonparametric estimator of τ for doubly interval censored data. The simulation results are presented in Section 3 and the real data application appears in Section 4. Section 5 contains general remarks and topics for future research.

2. Estimation of Kendall's τ

2.1. Kendall's τ

In this subsection, we briefly review the estimation of Kendall's τ for ordinary failure times. Let (T_{1i}, T_{2i}) and (T_{1j}, T_{2j}) be independent realizations from (T_1, T_2). Kendall's τ is defined as

$$\tau = \Pr\left((T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0\right) - \Pr\left((T_{1i} - T_{1j})(T_{2i} - T_{2j}) < 0\right), \tag{2.1}$$

where $(T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0$ indicates a concordant pair and $(T_{1i} - T_{1j})(T_{2i} - T_{2j}) < 0$ indicates a discordant pair, respectively. Therefore, (2.1) can be defined as the difference between the concordance fraction and the discordance one. In addiyion, τ has a value with a range $[-1, 1]$ and becomes zero if T_1 and T_2 are independent. For the complete data case, Kendall's τ can be estimated by

$$\hat{\tau} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} a_{ij} b_{ij},$$

where $a_{ij} = 1$ if $T_{1i} > T_{1j}$; $a_{ij} = -1$ if $T_{1i} < T_{1j}$ and $b_{ij} = 1$ if $T_{2i} > T_{2j}$ and $b_{ij} = -1$ if $T_{2i} < T_{2j}$. Then the score, $a_{ij} b_{ij} = 1$ if $(T_{1i} - T_{1j})(T_{2i} - T_{2j}) > 0$ and -1 if $(T_{1i} - T_{1j})(T_{2i} - T_{2j}) < 0$.

Under a right censoring with parallel scale, the observed data on the ith subject are $(X_{1i}, X_{2i}, \delta_{1i}, \delta_{2i})$, where $X_{1i} = T_{1i} \wedge C_{1i}$ and $X_{2i} = T_{2i} \wedge C_{2i}$ with censoring times C_{1i} and C_{2i} and $\delta_{1i} = I(T_{1i} \le C_{1i})$ and $\delta_{2i} = I(T_{2i} \le C_{2i})$ are the censoring indicators, respectively. Under independent censoring, Oakes (1982) suggested the following estimator

$$\hat{\tau}_O = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} R_{ij} a_{ij} b_{ij}, \tag{2.2}$$

where R_{ij} is an indicator of the (i, j) pair being comparable, that is, $R_{ij} = I(\tilde{X}_1^{ij} < \tilde{C}_1^{ij}, \tilde{X}_2^{ij} < \tilde{C}_2^{ij})$ with $\tilde{X}_1^{ij} = \min(X_{1i}, X_{1j})$, $\tilde{X}_2^{ij} = \min(X_{2i}, X_{2j})$, $\tilde{C}_1^{ij} = \min(C_{1i}, C_{1j})$ and $\tilde{C}_2^{ij} = \min(C_{2i}, C_{2j})$. However, (2.2) results in biased estimate when τ is not zero. To adjust for bias, Lakhal-Chaieb et al. (2009) proposed an estimator based on the inverse probability of censoring weights (IPCW) as follows,

$$\tau_w = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \frac{R_{ij} a_{ij} b_{ij}}{\hat{p}_{ij}^2}, \tag{2.3}$$

where $\hat{p}_{ij}$ is the estimate of $p_{ij} = P(C_1 > X_{ij}, C_2 > Y_{ij} | X_{ij}, Y_{ij})$. They also discussed the modified versions of (2.3) under several censoring patterns of C_1 and C_2 and compared the simulation results with other estimators.

For serial event times $(T_{1i} < T_{2i})$, define C_i as a censoring time with a survival distribution G. Set $\tilde{U}_i = \tilde{T}_{1i}, \tilde{V}_i = \tilde{T}_{2i} - \tilde{T}_{1i}, \delta_{1i} = I(\tilde{U}_i < C_i)$ and $\delta_{2i} = I(\tilde{U}_i + \tilde{V}_i < C_i)$, where $\tilde{T}_{1i} = \min(T_{1i}, C_i)$ and $\tilde{T}_{2i} = \min(T_{2i}, C_i)$. When T_{2i} is censored by C_i, $\tilde{V}_i$ is associated with both $\tilde{U}_i$ and C_i, causing the late occurrence of the first event to have s greater chance for the sequential event to be censored, which is referred to as an induced dependent censoring. Therefore, the earlier the first event occurs, the greater the chance of observing the second event. Similarly, the late occurrence of the first event is inclined to be censored and the corresponding information cannot be reflected to the estimation procedure. In order to adjust such biased sampling, a following IPCW technique is applied (Lakhal-Chaieb et al., 2010).

Define $\tilde{V}_{ij} = \min(\tilde{V}_i, \tilde{V}_j)$. Then the conditional probability of a pair being comparable is

$$p_{ij} = P\left(C_i > \tilde{U}_i + \tilde{V}_{ij}, C_j > \tilde{U}_j + \tilde{V}_{ij} | \tilde{U}_i, \tilde{U}_j, \tilde{V}_{ij}\right) = G\left(\tilde{U}_i + \tilde{V}_{ij}\right) \times G\left(\tilde{U}_j + \tilde{V}_{ij}\right),$$

which is estimated by $\hat{p}_{ij} = \hat{G}(U_i + \tilde{V}_{ij})\hat{G}(U_j + \tilde{V}_{ij})$ and $\hat{G}$ is the Kaplan-Meier estimator based on $\{\tilde{U}_i + \tilde{V}_i, 1 - \delta_{1i}\delta_{2i}\}$.

$$\hat{\tau}_w = \left\{ \sum_{1 \le i < j \le n} \frac{L_{ij}}{\hat{p}_{ij}} \right\}^{-1} \sum_{1 \le i < j \le n} \frac{L_{ij} a_{ij} b_{ij}}{\hat{p}_{ij}}, \tag{2.4}$$

where $L_{ij} = I(C_i > U_{1i} + \tilde{V}_{ij}, C_j > U_{1j} + \tilde{V}_{ij})$.

2.2. Estimation with doubly interval censored data

Returning to doubly-interval censored data, instead of observing two sequential event times (U, T), observable data are composed of $\{(L_U, R_U], (L_T, R_T], \delta_U, \delta_T\}$, where $\delta_U = I(R_U < \infty)$ and $\delta_T = I(R_T < \infty)$. Thus, a gap time $V = T - U$ is also interval censored, $(L_V, R_V]$, where $L_V = L_T - R_U, R_V = R_T - L_U$. $\delta_U = 0$ implies that the first event is right censored at L_U and thus $R_U = L_T = R_T = \infty$ and $L_V = 0, R_V = \infty$. When $\delta_T = 0$, the sequential second event is censored at L_T and the gap time is also right censored with $(L_V, R_V) = (L_T - R_U, \infty)$. To estimate the association between interval censored gap times (U, V), the IPCW is applied to the simulated failure times sampled from a multiple imputation technique as follows,

(i) Estimate a nonparametric estimator of a censoring distribution, $\hat{G}$ with $(\tilde{C}_i, 1 - \delta_{yi})$, where $\tilde{C}_i = L_U$ for $\delta_U = 0$, $\tilde{C}_i = L_T$ and $\delta_U = 1, \delta_T = 0$ and $\tilde{C}_i = R_T$ for $\delta_U = \delta_T = 1$.

(ii) Estimate the nonparametric estimator of (U, V), $\hat{f}(u)$ and $\hat{w}(v)$ by applying the self-consistency algorithm (Lagakos and Gruttola, 1989, Sun, 2006) as follows,

   (ii-a) Set $u_1 < u_2 < \cdots < u_r$ for the possible mass points of the U_i's and $v_1 < v_2 < \cdots < v_s$ for the mass points of V_i's. Then define $f_j = \Pr(U = u_j), j = 1, \ldots, r$ and $w_h = \Pr(V = v_h), h = 1, \ldots, s$. Here $\sum_{j=1}^{r} f_j = 1$ and $\sum_{h=1}^{s} w_h = 1$.

   (ii-b) Define $\alpha_{ijh} = I(L_{Ui} < u_j \le R_{Ui}, L_{Ti} < u_j + v_h \le R_{Ti}), j = 1, \ldots, r, h = 1, \ldots, s$. Then the full likelihood is expressed as

   $$L_f = \prod_{i=1}^{n} \sum_{j=1}^{r} \sum_{h=1}^{s} \alpha_{ijh} f_j w_h.$$

   (ii-c) Define $I_{ijh} = E[I(U_i = u_j, T_i = u_j + v_h)|w, f]$ as the conditional expectation of the first event $U_i = u_j$ and the second event $T_i = u_j + v_h$ given w and f. However, this quantity is unavailable under interval censoring. By adopting the EM algorithm, iterate between (2.5) and (2.6) until convergence is reached.

   $$\tilde{I}_{ijh} = \frac{\alpha_{ijh} f_j w_h}{\sum_{j'=1}^{r} \sum_{h'=1}^{s} \alpha_{ij'h'} f_{j'} w_{h'}}, \tag{2.5}$$

   $$\hat{f}_j = \frac{1}{n} \sum_{i=1}^{n} \sum_{h=1}^{s} \tilde{I}_{ijh}, \quad \hat{w}_h = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{r} \tilde{I}_{ijh}. \tag{2.6}$$

   The derivation of (2.5) is an extension of the self-consistency algorithm to doubly interval censored data.

(iii) Generate M sets composing of n pairs of $(\tilde{U}_i^k, \tilde{V}_i^k), i = 1, \ldots, n; k = 1, \ldots, M$ from the estimated distributions $\hat{f}, \hat{w}$ obtained at (ii). In more detail, for the subject with $\delta_{Ui} = 0$, $\tilde{U}_i^k$ is set to L_{Ui}. In addition, when $(\delta_{Ui}, \delta_{Ti}) = (1, 0)$, $\tilde{V}_i^k$ is set to $L_{Ti} - R_{Ui}$.

(iv) For $k = 1, \ldots, M$, we obtain $\hat{\tau}_w^k$ by applying (2.4) with $(U_i^k, V_i^k)_{i=1}^n$. The variance of the estimates, $\mathrm{Var}(\hat{\tau}_w^k)$ is obtained with a bootstrap procedure with $B = 20$

(v) With $\hat{\tau}_w^k, \mathrm{Var}(\hat{\tau}_w^k), (k = 1, \ldots, M)$, the final estimates are obtained by

$$\bar{\tau}_w = \frac{1}{M} \sum_{k=1}^{M} \hat{\tau}_w^k, \quad \mathrm{Var}(\bar{\tau}_w) = \sum_{k=1}^{M} \frac{1}{M} \mathrm{Var}(\hat{\tau}_w^k) + \frac{1}{M-1} \sum_{k=1}^{M} (\hat{\tau}_w^j - \bar{\tau}_w)^2 \left(1 + \frac{1}{M}\right). \tag{2.7}$$

Table 1: Simulation result of τ_w and τ_0 with $(cf_1, cf_2) = (5\%, 10\%)$

n	Imputation M	τ	$\bar{\tau}_w$	SEE	SDE	CP	$\bar{\tau}_0$
		0.200	0.200	0.098	0.097	0.947	0.157
		0.500	0.482	0.082	0.082	0.947	0.404
	5	0.666	0.653	0.063	0.061	0.956	0.563
		0.800	0.775	0.049	0.042	0.975	0.683
50		0.200	0.189	0.105	0.109	0.920	0.152
		0.500	0.486	0.087	0.081	0.960	0.407
	10	0.666	0.648	0.069	0.065	0.953	0.553
		0.800	0.773	0.050	0.044	0.980	0.682
		0.200	0.192	0.069	0.068	0.940	0.168
		0.500	0.491	0.057	0.054	0.950	0.417
	5	0.666	0.657	0.044	0.041	0.953	0.560
		0.800	0.779	0.033	0.032	0.950	0.703
100		0.200	0.192	0.073	0.069	0.940	0.156
		0.500	0.493	0.059	0.055	0.966	0.413
	10	0.666	0.652	0.047	0.047	0.965	0.559
		0.800	0.782	0.032	0.030	0.980	0.695
		0.200	0.198	0.048	0.044	0.970	0.161
		0.500	0.495	0.039	0.036	0.966	0.420
	5	0.666	0.658	0.031	0.030	0.956	0.540
		0.800	0.782	0.020	0.020	0.950	0.697
200		0.200	0.198	0.050	0.050	0.940	0.161
		0.500	0.493	0.041	0.040	0.953	0.418
	10	0.666	0.656	0.032	0.031	0.975	0.558
		0.800	0.782	0.020	0.020	0.960	0.697

3. Simulation

In this section, we perform a simulation study to evaluate the proposed method and compare the naive estimator ($\bar{\tau}_0$) which ignores the induced dependent censoring. three hundred datasets are generated with sample sizes $n = 50, 100$ and 200 and four values $\tau = 0.2, 0.5, 0.66, 0.80$. For generating correlated pairs (U, V), a Clayton copula model is applied. The censoring variable C is generated from an exponential distribution with the parameter controlling the censoring fractions $cf_1 = P(C < U)$ and $cf_2 = P(C < U + V)$. Tables 1–3 show the simulation results for three censoring cases; the estimates, SEE(standard error based on (6)), SDE(standard deviation of estimates), CP(95 empirical coverage rate of $\bar{\tau}_w$) and the estimates of $\bar{\tau}_0$ at $(cf_1, cf_2) = (5\%, 10\%), (10\%, 25\%)$, and $(25\%, 45\%)$. As expected, $\bar{\tau}_w$ outperforms $\bar{\tau}_0$ which is underestimated under all the conditions. The bootstrap estimator for the standard error of ($\bar{\tau}_w$) provides a suitable result by comparing the empirical one and satisfying 95% coverage rates. The influence of the imputation numbers appears to be negligible by showing similar results. However, the effect of censoring rate seems to be significant. As the censoring rate (Table 3) increases, the suggested estimates show underestimated values and unstable empirical coverage rates. Furthermore, $\bar{\tau}_0$ indicates more seriously biased results in the heavy censoring case.

4. Real data analysis

In this section, we analyzed dental data from 500 children's teeth eruption time and caries time (Bogaerts *et al.* 2017). Annual dental examinations were performed by dentists and teeth conditions such as plaque status, eruption times and caries time were recorded. In this analysis, our main interest was to check the association between the eruption time (U) and the duration time to cariese ($T - U$) of four molars (16, 26, 36, 46th tooth). However, instead of the exact timings, the eruption time was either

Table 2: Simulation result of τ_w and τ_0 with $(cf_1, cf_2) = (10\%, 25\%)$

n	Imputation M	τ	$\bar{\tau}_w$	SEE	SDE	CP	$\bar{\tau}_0$
50	5	0.200	0.190	0.112	0.111	0.933	0.129
		0.500	0.485	0.093	0.092	0.970	0.371
		0.666	0.641	0.074	0.068	0.970	0.502
		0.800	0.756	0.059	0.053	0.980	0.597
	10	0.200	0.200	0.111	0.113	0.933	0.150
		0.500	0.478	0.094	0.088	0.970	0.370
		0.666	0.647	0.073	0.067	0.966	0.576
		0.800	0.764	0.051	0.048	0.970	0.609
100	5	0.200	0.192	0.076	0.074	0.950	0.123
		0.500	0.489	0.064	0.061	0.950	0.373
		0.666	0.645	0.050	0.046	0.960	0.503
		0.800	0.781	0.032	0.031	0.940	0.697
	10	0.200	0.191	0.079	0.071	0.953	0.137
		0.500	0.483	0.065	0.063	0.953	0.365
		0.666	0.648	0.050	0.047	0.965	0.501
		0.800	0.769	0.037	0.035	0.930	0.617
200	5	0.200	0.197	0.054	0.053	0.943	0.137
		0.500	0.483	0.045	0.045	0.960	0.367
		0.666	0.655	0.033	0.036	0.940	0.503
		0.800	0.784	0.022	0.020	0.955	0.696
	10	0.200	0.192	0.053	0.052	0.953	0.136
		0.500	0.484	0.044	0.042	0.953	0.365
		0.666	0.648	0.033	0.031	0.965	0.501
		0.800	0.774	0.025	0.022	0.920	0.614

Table 3: Simulation result of τ_w and τ_0 with $(cf_1, cf_2) = (25\%, 45\%)$

| n	Imputation M	τ	$\bar{\tau}_w$	SEE	SDE	CP	$\bar{\tau}_0$
50	5	0.200	0.163	0.130	0.106	0.970	0.109
		0.500	0.448	0.102	0.100	0.965	0.301
		0.666	0.604	0.094	0.081	0.960	0.441
		0.800	0.751	0.073	0.052	0.980	0.542
	10	0.200	0.162	0.134	0.111	0.990	0.098
		0.500	0.469	0.114	0.100	0.970	0.327
		0.666	0.633	0.092	0.072	0.980	0.447
		0.800	0.742	0.074	0.053	0.980	0.551
100	5	0.200	0.181	0.099	0.094	0.970	0.115
		0.500	0.442	0.079	0.070	0.970	0.295
		0.666	0.615	0.062	0.053	0.980	0.439
		0.800	0.747	0.050	0.039	0.890	0.544
	10	0.200	0.160	0.093	0.094	0.930	0.090
		0.500	0.453	0.072	0.078	0.950	0.313
		0.666	0.620	0.058	0.061	0.965	0.446
		0.800	0.746	0.037	0.048	0.920	0.552
200	5	0.200	0.174	0.060	0.050	0.970	0.110
		0.500	0.458	0.052	0.043	0.980	0.312
		0.666	0.621	0.046	0.038	0.970	0.437
		0.800	0.785	0.037	0.027	0.980	0.548
	10	0.200	0.176	0.068	0.059	0.970	0.110
		0.500	0.460	0.062	0.056	0.990	0.313
		0.666	0.617	0.050	0.045	0.960	0.434
		0.800	0.758	0.032	0.027	0.980	0.548
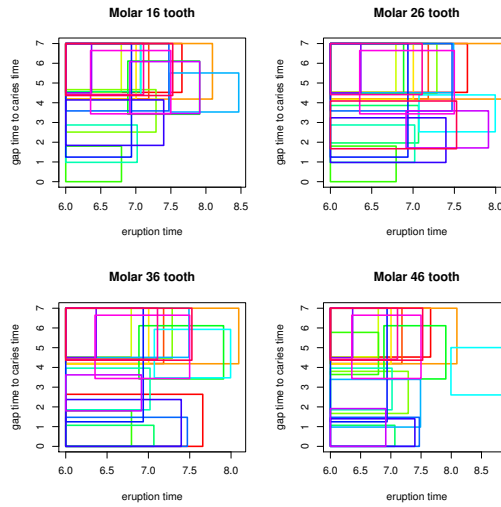
Figure 2: *Rectangle plots for eruption time and gap time to caries of molar 16,26,36, and 46 (Selected samples).*
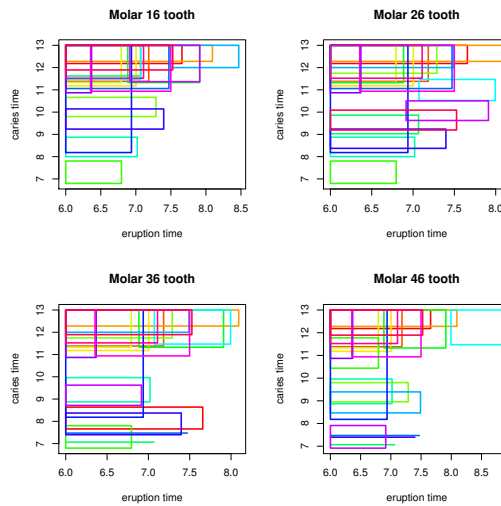

Figure 3: *Rectangle plots for eruption time and caries time of molar 16, 26, 36, and 46 (Selected samples).*

left censored $(0, L_U)$ or interval censored (L_U, R_U) and caries time is either interval censored (L_T, R_T) or right censored (R_T, ∞). According to the American Dental Association, permanent tooth eruption occurs at around 6−7 years. Therefore, the left bound of the left censored eruption time is replaced with 6.0 year. Figure 2 shows the rectangular plot of the interval censored eruption time and gap time to caries for the randomly selected twenty children's four teeth and Figure 3 shows the eruption time and time to caries for the same children in Figure 1. To check the variability of estimates under different imputation numbers, three numbers (M=10,20,40)were applied. Table 4 shows the estimates, standard errors and p-values for $H_0 : \tau = 0$ based on the unweighted τ_0 (without IPCW) and weighted τ_w (IPCW). According to the results, all estimates are negative which means that the children with

Table 4: Dental data: τ between U and V

Estimator	tooth 16 Est(SE, p-value)	tooth 26 Est(SE, p-value)	tooth 36 Est(SE, p-value)	tooth46 Est(SE, p-value)
$\tau_0(M = 10)$	−0.062(0.046, 0.177)	−0.078(0.041, 0.057)	−0.046(0.048, 0.264)	−0.097(0.044, 0.027)
$\tau_0(M = 20)$	−0.068(0.044, 0.122)	−0.080(0.046, 0.082)	−0.045(0.047, 0.276)	−0.093(0.043, 0.030)
$\tau_0(M = 40)$	−0.063(0.044, 0.152)	−0.076(0.044, 0.084)	−0.048(0.043 ,0.274)	−0.091(0.042, 0.031)
$\tau_w(m = 10)$	−0.231(0.276, 0.402)	−0.215(0.080, 0.007)	−0.200(0.194, 0.303)	−0.201(0.129, 0.119)
$\tau_w(m = 20)$	−0.203(0.234, 0.385)	−0.198(0.111, 0.074)	−0.161(0.175, 0.356)	−0.216(0.123, 0.079)
$\tau_w(m = 40)$	−0.156(0.293, 0.594)	−0.200(0.117, 0.087)	−0.168(0.201, 0.403)	−0.221(0.131, 0.091)

early eruption time tend to take a longer time to caries. $\bar{\tau}_w$ are insignificant for 16 and 36 molars but significant at 26 and 46 molars at 10% significance level. However, $\bar{\tau}_w$ and $\bar{\tau}_0$ showed somewhat different values at four molars. While the unweighted one have smaller estimates and standard errors but shows stable values irrespective of M values, $\bar{\tau}_W$s gives larger values and standard errors.

5. Discussion

In this paper, a natural extension of Kendall's τ statistic was proposed to estimate the association of doubly interval censored data. In parallel time scales, there are several methods for estimating the association of correlated failures. However, in many studies of life history processes, interest lies in the correlation between two consecutive events. For example, the occurrence of disease and death are intermediate and terminal events where the gap time between these two events is dealt using a semi-Markov model. However, it may not be plausible to assume that all subjects have the same chance to observe even when the length of follow up is independent of the event process. According to the simulation result, the ignorance of dependent censoring owing to the second event brings in biased result even under light right censoring case. The proposed estimate based on the IPCW shows good performance under light censoring rate but is underestimated at heavy censoring.

If a covariate is available, it may be of scientific interest to understand how the association changes with covariates. Furthermore, when such an association varies over time, a suitable measure such as the cross ratio function can be applied. Recently, Hu *et al.* (2011, 2019) considered a proportional cross ratio model to reflect both time dependent association and covariate effect simultaneously and analyzed the association between age at menopause and and age at menarche which are sequential events controlled by right censoring. We have extended their approach to doubly interval censored data as a related study (Lee and Kim, 2020). As another future study, the reviewer suggested a Frank copula reflecting the negative association which is commonly occurred. We have also considered this issue by incorporating with real examples and diverse simulation studies in another project.

Acknowledgements

References

Betensky R and Finkelstein DF (1999). An extension of Kendall's coefficient of concordance to bivariate interval censored data, *Statistics in Medicine*, **18**, 3101–3109.

Bogaets K, Komarek A, and Lesaffre E (2017). Survival Analysis with Interval Censored Data: A Practical Approach with Examples in R, SAS, and BUGS, CRC Press.

Bogaerts K and Lesaffre E (2008). Estimating local and global measures of association for bivariate interval censored data with a smooth estimate of density, *Statistics in Medicine*, **27**, 5941–5955.

Cook RJ and Lawless J (2007). *The Statistical Analysis of Recurrent Events*, Springer, New York.

De Gruttola V and Lagakos SW (1989). Analysis of doubly interval censored survival data with application to AIDS, *Biometrics*, **45**, 1–11.

Dejardin D and Lesaffre E (2013). Stochastic EM algorithm for doubly interval-censored data, *Biostatistics*, **14**, 766–778.

Fu TC, Su DH, and Chang SH (2016). Serial association analyses of recurrent gap time data via Kendall's tau, *Biostatistics*, **17**, 188–202.

Hu T, Lin X, and Nan B (2019). Proportional cross-hazard model, *Lifetime Data Analysis*, **25**, 480–506.

Hu T and Nan B (2011). Time-dependent cross ratio estimation for bivariate failure times, *Biometrika*, **98**, 341–354.

Kim MY, De Gruttola V, and Lagakos SW (1993). Analyzing doubly censored data with covariates, with application to AIDS, *Biometrics*, **49**, 13–22.

Kim YJ (2006). Regression analysis of doubly censored data with Frailty, *Biometrics*, **62**, 458–464.

Lakhal-Chaiaeb L, Cook RJ, and Lin X (2010). Inverse probability of censoring weighted estimated of Kendall's tau for gap time analyses, *Biometrics*, **66**, 1145–1152.

Lakhal-Chaieb L, Rivest LP, and Beaudoin D (2009). IPCW estimator for Kendall's tau under bivariate censoring, *The International Journal of Biostatistics*, **5**, 8.

Lee NR and Kim YJ (2020). Estimation of cross ratio function for doubly interval censored data, Submitted.

Oakes D (1989). A concordance test for independence in the presence of censoring, *Biometrics*, **38**, 451–455.

Oakes D (2008). On consistency of Kendall'tau under censoring, *Biometrika*, **95**, 997–1001.

Sun L, Wang L, and Sun J (2006). Estimation of the association for bivariate interval-censored failure time data, *Scandinavian Journal of Statistics*, **33**, 637–649.

Sun J (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*, Springer, Verlag, New-York.

Turnbull BW (1974). Nonparametric estimation of survivorship function with doubly censored data, *Journal of the American Statistical Association*, **69**, 169–173.

Wang W and Ding AA (2000). On assessing the association for bivariate current status data, *Biometrika*, **87**, 879–893.

Wang W and Wells M (2000). Estimation of Kendall's tau under censoring, *Statistica Sinica*, **10**, 1199–1215.