Generalized Log-Rank Tests for Partly Interval-Censored Failure Time Data

Xingqiu Zhao[*,1], **Qiang Zhao**[2], **Jainguo Sun**[3], and **Jong S. Kim**[4]

[1] Department of Mathematics and Statistics, McMaster University, 1280 Main Street West Hamilton, Ontario L8S 4K1, Canada
[2] Department of Mathematics, Texas State University, 601 University Drive San Marcos, TX 78666, USA
[3] Department of Statistics, University of Missouri, Columbia, MO 65211, USA
[4] Department of Mathematics and Statistics, Portland State University, P.O. Box 751 Portland Oregon 97207, USA

Received 30 July 2007, revised 24 December 2007, accepted 29 January 2008

Summary

In this paper, we consider incomplete survival data: partly interval-censored failure time data where observed data include both exact and interval-censored observations on the survival time of interest. We present a class of generalized log-rank tests for this type of survival data and establish their asymptotic properties. The method is evaluated using simulation studies and illustrated by a set of real data from a diabetes study.

Key words: Asymptotic distribution; Empirical process; Log-rank test; Partial interval-censoring; Survival comparison.

1 Introduction

This paper discusses nonparametric comparison of survival functions based on incomplete survival data: partly interval-censored failure time data (Peto and Peto, 1972; Huang, 1999; Kim, 2003). By partly interval-censored data, we mean that for some subjects, the exact failure times are observed, but for the remaining subjects, the survival time of interest is observed only to belong to an interval instead of being exactly known or right-censored as usually assumed (Li, 2003). Partly interval-censored data often occur in medical follow-up studies; see the Framingham Heart Disease Study (Odell, Anderson and D'Agostino, 1992) and the Danish Diabetes Study (Ramlau-Hansen, Jespersen and Andersen, 1987).

Survival comparison is usually one of main goals in survival studies. For the case of right-censored failure time data, there are a number of well-established methods (e.g., Fleming and Harrington, 1991; Kalbfleisch and Prentice, 2002). For the case of interval-censored failure time data, several authors have discussed the problem; see Peto and Peto (1972), Finkelstein (1986), Sun (1996, 1998), and Sun, Zhao and Zhao (2005). In contrast, there exists limited research for the analysis of partly interval-censored data. Peto and Peto (1972) discussed partly interval-censored data, treating an exact observation as an interval-censored observation with a very short interval. Turnbull (1976) described a general scheme of incomplete failure time data and derived self-consistency equations for computing the maximum likelihood estimator of the survival functions. Huang (1999) studied asymptotic properties of the nonparametric maximum likelihood estimator of a distribution function based on partly interval-

[*] Corresponding author: e-mail: zhaox7@math.mcmaster.ca, Phone: +001 905 525 9140, Fax: +001 905 522 0935

censored data. Kim (2003) used the proportional hazards model for regression analysis of partly interval-censored data.

In this paper, we develop nonparametric test procedures for partly interval-censored data using the idea behind the generalized log-rank tests for interval-censored data (Sun et al., 2005). The test procedures are presented in Section 2. In Section 3, the asymptotic distributions of the proposed test statistics are derived and Section 4 reports some simulation results for evaluating the proposed methodology. They suggest that the approach works well for the practical situations considered. In Section 5 we apply the approach to a set of partly interval-censored data from a diabetes study and Section 6 contains some concluding remarks.

2 Generalized Log-Rank Tests

Consider a survival study that involves n independent subjects from k different populations. Let T_{li} denote the survival time of interest for subject i from the l-th population and n_l the number of subjects from population l with survival function $G_l(t)$ and distribution function $F_l(t) = 1 - G_l(t)$, $i = 1, \ldots, n_l$, $l = 1, 2, \ldots, k$, where $n_1 + \ldots + n_k = n$. Suppose that for the l-th population, we observe the exact failure times $\{T_{li}, i = 1, \ldots, n_{l1}\}$ for n_{l1} subjects and interval-censored failure times given by $\{U_{li}, V_{li}, \Delta_{li} = I(T_{li} \le U_{li}), \Gamma_{li} = I(U_{li} < T_{li} \le V_{li}), i = n_{l1} + 1, \ldots, n_l\}$ for the remaining n_{l2} $(= n_l - n_{l1})$ subjects, where U_{li} and V_{li} are non-negative random variables independent of T_{li} such that $U_{li} < V_{li}$ with probability one. For given l and i, where $l = 1, \ldots, k$, and $i = 1, \ldots, n_{l1}$, express T_{li} as $(T_{li}-, T_{li}]$, but for $i = n_{l1} + 1, \ldots, n_l$, define

$$(L_{li}, R_{li}] = \begin{cases} (0, U_{li}], & T_{li} \le U_{li}, \\ (U_{li}, V_{li}], & U_{li} < T_{li} \le V_{li}, \\ (V_{li}, \infty), & T_{li} > V_{li}. \end{cases}$$

Our goal is to test the hypothesis $H_0 : G_1(t) = \ldots = G_k(t)$.

Let $G_0(t)$ denote the common survival function under H_0 and $\hat{G}_n(t)$ its nonparametric maximum likelihood estimate (NPMLE), whose determination will be discussed below. Set $N_j = \sum_{l=1}^{k} n_{lj}, j = 1, 2$. To test H_0, motivated by Sun et al. (2005), we propose to use the test statistic $U_\xi = (U_\xi^{(1)}, \ldots, U_\xi^{(k)})^T$, where

$$U_\xi^{(l)} = \frac{N_1}{n_{l1}} \sum_{i=1}^{n_{l1}} \frac{\xi\{\hat{G}_n(T_{li}-)\} - \xi\{\hat{G}_n(T_{li})\}}{\hat{G}_n(T_{li}-) - \hat{G}_n(T_{li})} + \frac{N_2}{n_{l2}} \sum_{i=n_{l1}+1}^{n_l} \frac{\xi\{\hat{G}_n(L_{li})\} - \xi\{\hat{G}_n(R_{li})\}}{\hat{G}_n(L_{li}) - \hat{G}_n(R_{li})}$$

and ξ is a known function over $(0, 1)$. ξ will be defined more formally in the next section. Obviously, different ξ can be used and will yield different test statistics in practice. When $N_1 = 0$, that is, only interval-censored failure time data are available, the statistics U_ξ reduces to that proposed in Sun et al. (2005), which is a generalization of the log-rank test discussed in Peto and Peto (1972). When $N_2 = 0$, that is, all the failure times are exactly known, the NPMLE $\hat{G}_n(\cdot)$ is $1 - \hat{F}_n(\cdot)$, where $\hat{F}_n(\cdot)$ is the empirical distribution function. In this case, letting $\xi(x) = x \log(x)$, we have

$$\begin{aligned} \sum_{i=1}^{n_l} \frac{\xi\{\hat{G}_n(T_{li}-)\} - \xi\{\hat{G}_n(T_{li})\}}{\hat{G}_n(T_{li}-) - \hat{G}_n(T_{li})} &\approx \sum_{i=1}^{n_l} \xi'(\hat{G}_n(T_{li})) \\ &= \sum_{i=1}^{n_l} [1 + \log(\hat{G}_n(T_{li}))] \\ &= \sum_{i=1}^{n_l} [1 + \log(1 - R_{li}/n)], \end{aligned}$$

where R_{li} denotes the rank of T_{li} in the combined sample of size n. When $\Delta_{li} = \Gamma_{li} = 0$, we get right-censored data. Peto and Peto (1972) discussed the test statistic with $\xi(x) = x \log(x)$ for right-censored

data. That is, the proposed test statistics can be seen as generalizations of the log-rank test for exact failure time data and also right-censored data.

To use U_ξ, we need to determine $\hat{G}_n(t)$. The simplest method for this, which is used in simulation studies and the example below, is perhaps the direct application of the Turnbull's self-consistency algorithm (Turnbull, 1976). Some other methods can be found in Sun (2006).

3 Asymptotic Distributions

In this section, we will establish the asymptotic distribution of U_ξ. Let $\eta(x) = 1 - \xi(1-x)$ and assume that $\lim_{x\to 0}\eta(x) = \lim_{x\to 1}\eta(x) = c_0$, where c_0 is a constant. Then we can rewrite $U_\xi^{(l)}$ as

$$\begin{aligned} U_\eta^{(l)} = & \frac{N_1}{n_{l1}} \sum_{i=1}^{n_{l1}} \frac{\eta\{\hat{F}_n(T_{li})\} - \eta\{\hat{F}_n(T_{li}-)\}}{\hat{F}_n(T_{li}) - \hat{F}_n(T_{li}-)} \\ & + \frac{N_2}{n_{l2}} \sum_{i=n_{l1}+1}^{n_l} \left[\Delta_{li} \frac{\eta\{\hat{F}_n(U_{li})\} - c_0}{\hat{F}_n(U_{li})} + \Gamma_{li} \frac{\eta\{\hat{F}_n(V_{li})\} - \eta\{\hat{F}_n(U_{li})\}}{\hat{F}_n(V_{li}) - \hat{F}_n(U_{li})} \right. \\ & \left. + (1 - \Delta_{li} - \Gamma_{li}) \frac{c_0 - \eta\{\hat{F}_n(V_{li})\}}{1 - \hat{F}_n(V_{li})} \right]. \end{aligned}$$

Let $F_0(\cdot) = 1 - G_0(\cdot)$ and $H(\cdot)$ and $h(\cdot)$ denote the distribution and density functions of (U_i, V_i), respectively. In addition, Let $P_{n_{l1}}$ be the empirical measure of the exact observations of the l-th population, P_{N_1} the empirical measure of the survival time based on all exact failure time data, P_0 the probability measure of T. Let λ_2 and ν_2 denote the Lebesgue measure on R^2 and counting measure on the set $\{(0,1),(1,0),(0,0)\}$, respectively. Define

$$q_{F_0,H}(u,v,\delta,\gamma) = h(u,v)\,\{F_0(u)\}^\delta\,\{F_0(v) - F_0(u)\}^\gamma\,\{1 - F_0(v)\}^{1-\delta-\gamma}$$

with respect to $\lambda_2 \otimes \nu_2$, which is the density function of $(U_i, V_i, \Delta_i, \Gamma_i)$. Also define $\mathrm{d}Q_0 = q_{F_0,H}\,\mathrm{d}(\lambda_2 \otimes \nu_2)$ and for $l = 1, \ldots, k$,

$$Q_{n_{l2}}(u,v,\delta,\gamma) = \frac{1}{n_{l2}} \sum_{i=n_{l1}+1}^{n_l} 1_{\{(U_{li},V_{li}) \le (u,v),\,(\Delta_{li},\Gamma_{li}) = (\delta,\gamma)\}},$$

and

$$Q_{N_2}(u,v,\delta,\gamma) = \frac{1}{N_2} \sum_{l=1}^{k} \sum_{i=n_{l1}+1}^{n_l} 1_{\{(U_{li},V_{li}) \le (u,v),\,(\Delta_{li},\Gamma_{li}) = (\delta,\gamma)\}}.$$

Set

$$f_F(t) = \begin{cases} \eta'\{F(t)\}, & \text{if} \quad F(t) = F(t-), \\ \frac{\eta\{F(t)\} - \eta\{F(t-)\}}{F(t) - F(t-)}, & \text{otherwise} \end{cases}$$

and

$$K_F(u,v,\delta,\gamma) = \delta\,\frac{\eta\{F(u)\} - c_0}{F(u)} + \gamma\,\frac{\eta\{F(v)\} - \eta\{F(u)\}}{F(v) - F(u)} + (1 - \delta - \gamma)\,\frac{c_0 - \eta\{F(v)\}}{1 - F(v)}.$$

Then $U_\eta^{(l)}$ can be expressed as

$$U_\eta^{(l)} = N_1 P_{n_{l1}}(f_{\hat{F}_n}) + N_2 Q_{n_{l2}}(K_{\hat{F}_n}).$$

For $l = 1, \ldots, k$, we assume that (T_{li}, U_{li}, V_{li}) $(i = n_{l1}+1, \ldots, n_l)$ satisfy the regularity conditions given in Groeneboom and Wellner (1992, pp. 81–82). As Huang (1999, pp. 504–505) pointed out, the

uniform consistency of $\hat{F}_n = 1 - \hat{G}_n$ can be shown as $\min\{N_1, N_2\} \to \infty$ by using the method of Groeneboom and Wellner (1992). Also we assume that $F_0(t)$ has a support in $[0, M]$ with a continuous density function and that there exist $0 < \delta_0, \varepsilon_0 < M/2$ and $M_0 < M$ such that $\Pr(U < \delta_0) = 0$, $\Pr(U + \varepsilon_0 \le V \le M_0) = 1$, $0 < F_0(\delta_0) < F_0(M_0) < 1$ and $\min_{\delta_0 \le t \le M_0 - \varepsilon_0}\{F_0(t + \varepsilon_0) - F_0(t)\} \ne 0$, where M is a constant. Furthermore, suppose that $P\left\{\liminf_{n\to\infty} \min_{\substack{1\le i\le n_{l1}\\ 1\le l\le k}} T_{li} > 0\right\} = 1$. These conditions usually hold for periodic follow-up studies. The asymptotic distribution of U_ξ is given in the following theorem.

Theorem 3.1 *Suppose that the above assumptions hold and* η *has continuous, bounded third-order derivative on* $[a, 1]$ *for any finite positive number* a. *Also suppose that as* $n \to \infty$, $n_{l1}/n \to p_{l1}$, $n_{l2}/n \to p_{l2}$ *where* $0 < p_{l1} < 1$. *Let* $p_j = \sum_{l=1}^{k} p_{lj}$, $j = 1, 2$. *Then under* H_0 *and as* $n \to \infty$, $U_\eta/\sqrt{n}$ *has an asymptotic normal distribution with mean zero and covariance matrix* $\Sigma = (\sigma_{lr})_{k\times k}$, *where*

$$\sigma_{lr} = \begin{cases} p_1\left(\frac{p_1}{p_{l1}} - 1\right) P_0(f_{F_0}^2) + p_2\left(\frac{p_2}{p_{l2}} - 1\right) Q_0(K_{F_0}^2), & \text{if} \quad l = r, \\ -p_1 P_0(f_{F_0}^2) - p_2 Q_0(K_{F_0}^2), & \text{otherwise}. \end{cases}$$

The proof of the above theorem is sketched in the appendix. For estimation of Σ, a natural consistent estimate is given by $\hat{\Sigma} = (\hat{\sigma}_{lr})_{k\times k}$, where

$$\hat{\sigma}_{lr} = \begin{cases} \frac{N_1}{n}\left(\frac{N_1}{n_{l1}} - 1\right) P_{N_1}(f_{\hat{F}_n}^2) + \frac{N_2}{n}\left(\frac{N_2}{n_{l2}} - 1\right) Q_{N_2}(\hat{K}_{\hat{F}_n}^2), & \text{if} \quad l = r, \\ -\frac{N_1}{n} P_{N_1}(f_{\hat{F}_n}^2) - \frac{N_2}{n} Q_{N_2}(K_{\hat{F}_n}^2), & \text{otherwise}. \end{cases}$$

The test of the hypothesis H_0 can then be performed as follows:

(i) If $\frac{n_{l1}}{N_1} = \frac{n_{l2}}{N_2}, l = 1, \ldots, k$, then we have that $\sum_{l=1}^{k} \frac{n_{l1}}{N_1} U_\eta^{(l)} = 0$. Let U_0 denote the first $(k-1)$ components of U_η and $\hat{\Sigma}_0$ the matrix by deleting the last row and column of $\hat{\Sigma}$. Then the hypothesis H_0 can be tested by using the statistic $\chi_0^2 = U_0^t \hat{\Sigma}_0^{-1} U_0/n$, which has asymptotically the χ^2 distribution with $(k-1)$ degrees of freedom.

(ii) If the condition in (i) is not satisfied, then the hypothesis H_0 can be tested by using the statistic $\chi^2 = U_\eta^t \hat{\Sigma}^{-1} U_\eta/n$, which has asymptotically the χ^2 distribution with k degrees of freedom.

4 Simulation Study

To investigate the finite sample properties of the proposed procedure, simulation studies were conducted. In the study, we considered the two-sample comparison problem with a total sample size of $n = 200$. The survival times T_i's were generated from the exponential distribution with mean $\exp(\alpha + \beta Z_i)$, where $Z_i \sim$ Binomial $(1, 0.5)$, $\alpha = 2.0$, and $\beta = -0.8, -0.4, 0.0, 0.4$, or 0.8. Also it was supposed that there exist $p = 25\%$ or 50% exact observations in the simulated data. To form intervals for interval-censored observations, we mimicked the set-up commonly used in periodic follow-up studies. In particular, we first generated C_1 and C_2 independently from uniform distributions $U(0, \theta_1)$ and $U(0, \theta_2)$, respectively. Here θ_1 and θ_2 are constants chosen to give pre-specified percentages of left-censored, interval-censored and right-censored observations among the interval-censored observations. Then U_i and V_i were defined to be the nearest half unit to C_1 and as the maximum of the nearest half unit to $C_1 + C_2$ and $U_i + 0.5$, respectively, and $(L_i, R_i]$ are formed accordingly as discussed in Section 2. The results reported below are based on 5000 replications.

Table 1 presents the empirical sizes and powers of the proposed test based on simulated partly interval-censored data with $p = 25\%$ exact observations for different values of β. In the table, we considered four different compositions of left-censored, interval-censored and right-censored observations for interval-censored observations, which are given in the first column of the table. The second and third columns give the values of parameters used in the functions $\xi(x) = x \log(x) x^{\varrho}(1-x)^{\gamma}$. For

Table 1 Estimated powers and sizes with 25% exact observations.

Percentages of censoring	ϱ	γ	β						
			−1.0	−0.8	−0.4	0.0	0.4	0.8	1.0
	0	0	1.000	0.996	0.553	0.048	0.524	0.979	0.999
$1/3 \sim 1/3 \sim 1/3$		1	1.000	0.983	0.465	0.042	0.433	0.946	0.993
	1	0	0.984	0.901	0.328	0.051	0.336	0.889	0.980
		1	0.999	0.973	0.426	0.048	0.446	0.956	0.995
Score test			1.000	0.999	0.686	0.052	0.652	0.996	1.000
	0	0	1.000	0.992	0.539	0.051	0.534	0.988	1.000
$1/4 \sim 1/2 \sim 1/4$		1	0.998	0.972	0.438	0.044	0.432	0.958	0.996
	1	0	0.977	0.865	0.291	0.048	0.298	0.866	0.973
		1	0.997	0.955	0.394	0.049	0.438	0.962	0.996
Score test			1.000	0.998	0.677	0.053	0.664	0.996	1.000
	0	0	1.000	0.994	0.566	0.052	0.555	0.989	1.000
$1/2 \sim 1/4 \sim 1/4$		1	0.999	0.987	0.496	0.043	0.478	0.969	0.997
	1	0	0.972	0.874	0.329	0.047	0.356	0.907	0.985
		1	0.998	0.961	0.431	0.050	0.467	0.969	0.998
Score test			1.000	0.999	0.699	0.054	0.678	0.997	1.000
	0	0	1.000	0.991	0.507	0.049	0.457	0.957	0.996
$1/4 \sim 1/4 \sim 1/2$		1	0.998	0.971	0.413	0.044	0.360	0.894	0.979
	1	0	0.988	0.915	0.341	0.048	0.320	0.853	0.966
		1	0.999	0.972	0.422	0.047	0.382	0.906	0.982
Score test			1.000	0.997	0.629	0.050	0.590	0.987	0.999

comparison, we also calculated and included in the table the empirical size and power of the parametric score test for $\beta = 0$ assuming that we know the underlying distribution. It can be seen from the table that the proposed test procedures seem to have the right size (nominal level 0.05 was used) and their powers are close to those of the parametric score test for many settings, suggesting that it performs well under these situations. The results for $p = 50\%$ are given in Table 2 and similar to those given in Table 1. It can be seen that when there exist more exact observations, the power generally increases as expected.

To evaluate the χ^2 distribution approximation given in the theorem to the finite distribution of the proposed test statistic, we studied the quantile plots of the test statistic against the χ^2 variable with degrees of freedom equal to 1 and 2 under different set-ups. All of them suggest that the χ^2 approximation works well.

5 Illustrative Example

Now we applied the proposed test procedure to the Steno Memorial Hospital diabetic data from Denmark that have been discussed by Andersen et al. (1992) and Kim (2003) among others. There generally exist two main types of diabetes: Type I (insulin-dependent) and Type II (non-insulin-dependent). Type I diabetes is the most severe form and occurs mainly at young ages, whereas Type II diabetes is a milder type and develops later in life. In this study, we consider 731 patients who were younger than 31, diagnosed as Type I diabetics between 1933 and 1972, and followed until death, emigration, or 31 December 1984.

The data set includes the information on gender, date of birth and age at diagnosis of the disease in the study. The survival time of interest is time from onset of diabetes to onset of diabetic nephropathy (DN),

Table 2 Estimated powers and sizes with 50% exact observations.

Percentages of censoring	ϱ	γ	β						
			−1.0	−0.8	−0.4	0.0	0.4	0.8	1.0
	0	0	1.000	0.997	0.602	0.042	0.585	0.995	1.000
$1/3 \sim 1/3 \sim 1/3$		1	1.000	0.990	0.494	0.042	0.482	0.979	0.998
	1	0	0.980	0.878	0.325	0.050	0.327	0.884	0.979
		1	0.998	0.965	0.439	0.050	0.433	0.964	0.996
Score test			1.000	0.999	0.736	0.048	0.724	0.998	1.000
	0	0	1.000	0.997	0.594	0.044	0.591	0.996	1.000
$1/4 \sim 1/2 \sim 1/4$		1	0.999	0.987	0.487	0.040	0.479	0.979	0.999
	1	0	0.972	0.858	0.292	0.050	0.299	0.868	0.972
		1	0.996	0.956	0.411	0.049	0.423	0.964	0.997
Score test			1.000	0.999	0.732	0.046	0.732	0.999	1.000
	0	0	1.000	0.997	0.602	0.044	0.600	0.996	1.000
$1/2 \sim 1/4 \sim 1/4$		1	1.000	0.990	0.520	0.040	0.509	0.984	0.999
	1	0	0.973	0.868	0.316	0.051	0.330	0.904	0.983
		1	0.996	0.961	0.430	0.049	0.441	0.971	0.998
Score test			1.000	0.999	0.745	0.047	0.740	0.998	1.000
	0	0	1.000	0.996	0.566	0.044	0.536	0.989	0.999
$1/4 \sim 1/4 \sim 1/2$		1	0.999	0.989	0.467	0.044	0.437	0.958	0.997
	1	0	0.982	0.889	0.325	0.049	0.310	0.858	0.968
		1	0.997	0.969	0.429	0.049	0.389	0.933	0.990
Score test			1.000	0.999	0.702	0.048	0.675	0.998	1.000

a major complication of Type I diabetes and a sign of kidney failure, which is defined to be present if at least four samples of 24 h urine at time intervals of at least one month contain more than 0.5 g protein. All 731 patients considered here had developed DN at time of admission or by the end of study, meaning that there is no right-censoring in the data. There were 595 exact and 136 interval-censored observations. Among the 731 patients, there were 277 females, 454 males, 222 aged less than 10, and 509 aged between 10 and 30. The focus here is on the effects of gender and age on the development of DN.

To test the overall effect of gender and age together, we conducted joint comparison with both gender and age together. The patients were divided into four groups: males aged less than 10, males aged between 10 and 30, females aged less than 10 and females aged between 10 and 30 as in Kim (2003). The test results are summarized in Table 3 by using different $\xi(x) = x \log(x)\, x^{\varrho}(1-x)^{\gamma}$. It can be seen from Table 3 that when $(\varrho, \gamma) = (0, 1)$, the test suggests that there is no sufficient evi-

Table 3 Analysis results of overall effect of gender and age for the diabetic data.

ϱ	γ	U'	χ^2	p-value
0	0	(98.60, 7.99, −13.09, −153.88)	22.50	0.00016
0	1	(10.71, 38.80, −0.33, −61.92)	7.39	0.11677
1	0	(87.89, −30.81, −12.76, −91.96)	47.99	0.00000
1	1	(28.76, −11.12, −5.29, −27.66)	28.42	0.00001

Table 4 Analysis results of separate effect of gender and age for the diabetic data.

Gender					Age				
ϱ	γ	U'	χ^2	p-value	ϱ	γ	U'	χ^2	p-value
0	0	(69.60, −42.93)	4.58	0.1015	0	0	(26.70, −61.82)	3.72	0.1560
0	1	(21.10, −12.99)	0.64	0.7254	0	1	(2.92, −7.09)	0.40	0.8181
1	0	(48.51, −29.94)	12.75	0.0017	1	0	(23.78, −54.73)	13.22	0.0013
1	1	(15.55, −9.90)	6.04	0.0489	1	1	(7.28, −16.89)	8.32	0.0156

dence for the overall effect of gender and age together on the development of DN. On the other hand, the tests with $(\varrho, \gamma) = (0, 0)$, $(1, 0)$ or $(1, 1)$ indicate that the overall effect is significant.

To test the separate effects of gender and age, we conducted separate comparison with respect to gender and age. For the age effect, we again divided the patients into two groups, either younger than 10 or otherwise. The test results are summarized in Table 4 by using different $\xi(x) = x \log(x)\, x^{\varrho}(1-x)^{\gamma}$. It can be seen from Table 4 that when $(\varrho, \gamma) = (0, 0)$ or $(0, 1)$, the test procedures suggest that there is no sufficient evidence for either gender or age effect on the onset of diabetic nephropathy. On the other hand, the method with $(\varrho, \gamma) = (1, 0)$ or $(1, 1)$ indicates that there exist some or significant effects of both gender and age.

Note that different $\xi(x)$ gave different conclusions. To explain the different p-values in Tables 3 and 4, one needs to study the weight function $\xi(x)$. When $(\varrho, \gamma) = (0, 0)$ or $(0, 1)$, it puts more weights to later survival differences, while $\xi(x)$ with $(\varrho, \gamma) = (1, 0)$ or $(1, 1)$ gives more weights to early survival

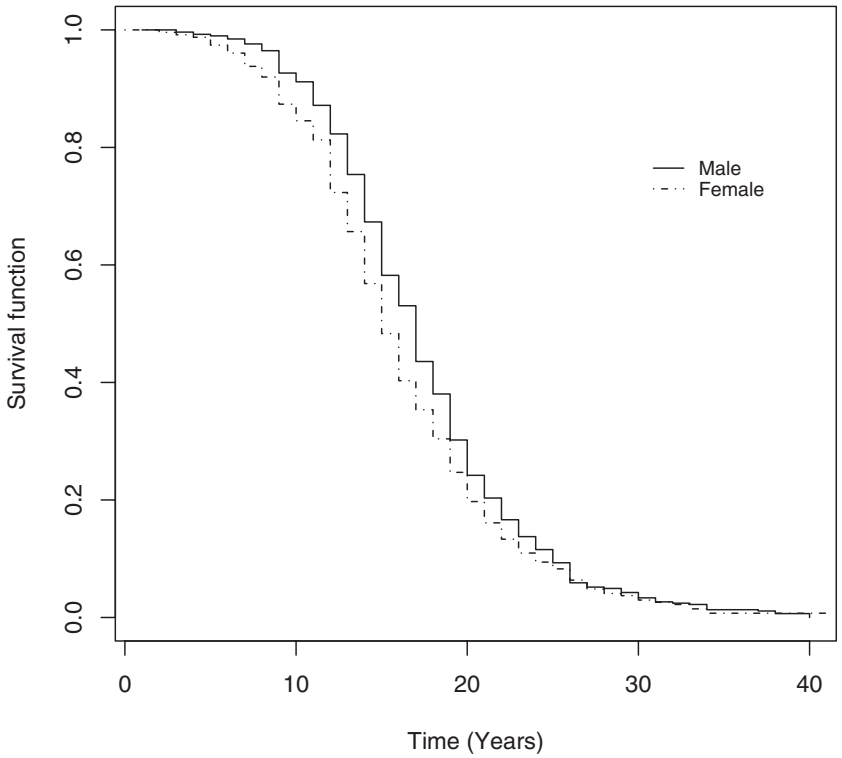

Figure 1 Estimates of the survival functions of two gender groups with diabetes data.

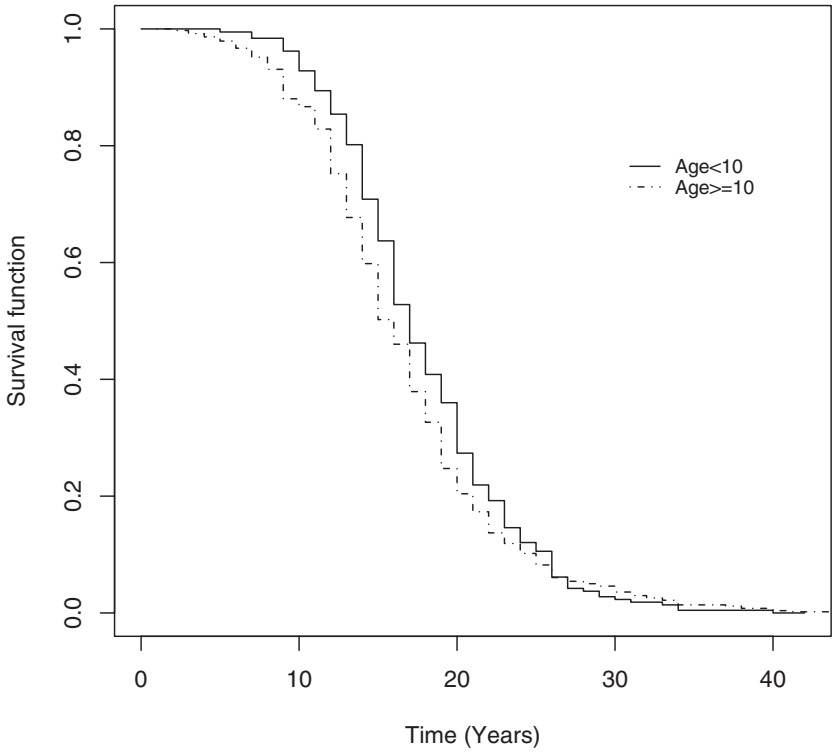

Figure 2 Estimates of the survival functions of two age groups with diabetes data.

differences. To further investigate this, we obtained the separate estimates of the survival functions with respect to gender and age and they are presented in Figures 1 and 2, respectively. They show that there seem to exist some early survival differences before 20 or 25 years between the two groups with respect to both gender and age, but no obvious differences after that time. By fitting the proportional hazards model to the data, Kim (2003) concluded that overall effect of gender and age together was not significant on the development of diabetic nephropathy.

6 Concluding Remarks

This paper discussed nonparametric comparison of survival functions when partly interval-censored failure time data are available. For the problem, a class of nonparametric tests was proposed and both finite sample and asymptotic properties of the proposed tests were established. The proposed test statistics are generalizations of the log-rank test statistic discussed in Peto and Peto (1972) and those in Sun et al. (2005) for interval-censored data. In comparison with the test procedures given in Kim (2003), in addition to the derived asymptotic distribution, the proposed procedure has the advantage that the calculation of its variance estimate is straightforward as it does not involve inverting the high-dimensional information matrix. In contrast, the variance estimation given in Kim (2003) originally involves inverting the high-dimensional information matrix, although that was resolved by his generalized missing information principle, which allows one to invert the information matrix that only involves the MLE of the regression parameter.

In comparison with right-censored failure time data, only limited research exists for partly interval-censored failure time data although they frequently occur in public health and medical studies such as clinic trials. One obstacle to this is that partial interval-censoring is much harder to deal with than

right-censoring. One consequence resulting from partial interval-censoring is that counting process and martingale theory that make the study of right-censored data relatively easy are no longer available for partly interval-censored data. Instead, the empirical process theory needs to be used to study partly interval-censored data.

A direction for future research is to study the properties of the test statistic under alternatives for selection of weight function ξ. In the cases of exact failure time data and right-censored data, the local asymptotic power of the generalized log-rank test can be discussed along the lines of Anderson et al. (1992, pp. 372–379). However, it seems to be difficult to discuss this behavior of the test for general partly interval-censored data since in this case, $\hat{G}_n$ does not have an explicit expression anymore and its property still needs to be investigated.

Appendix: Proof of Theorem 3.1

By the assumption

$$P\left\{\liminf_{n\to\infty} \min_{\substack{1\le i\le n_{l1}\\ 1\le l\le k}} T_{li} > 0\right\} = 1\,,$$

we have that for any $\varepsilon > 0$, there exists a finite positive number τ such that

$$\inf_n P\left\{\min_{\substack{1\le i\le n_{l1}\\ 1\le l\le k}} T_{li} \ge \tau\right\} \ge 1-\varepsilon\,.$$

Let $A_n = \left\{\min_{\substack{1\le i\le n_{l1}\\ 1\le l\le k}} T_{li} \ge \tau\right\}$. Then we have

$$\frac{1}{\sqrt{n}}\, U_\eta^{(l)} = \frac{N_1}{\sqrt{n}}\, P_{n_{l1}}(f_{\hat{F}_n}) + \frac{N_2}{\sqrt{n}}\, Q_{n_{l2}}(K_{\hat{F}_n})\,.$$

Note that

$$\begin{aligned} P_{n_{l1}}(f_{\hat{F}_n}) &= (P_{n_{l1}} - P_0)\left\{(f_{\hat{F}_n} - f_{F_0})\, 1_{[\tau,M]}\right\} 1_{A_n} + (P_0 - P_{N_1})\left\{(f_{\hat{F}_n} - f_{F_0})\, 1_{[\tau,M]}\right\} 1_{A_n} \\ &\quad + P_{n_{l1}}(f_{\hat{F}_n} - f_{F_0}) 1_{A_n^c} - P_{N_1}(f_{\hat{F}_n} - f_{F_0})\, 1_{A_n^c} \\ &\quad + P_{N_1}(f_{\hat{F}_n}) + P_{n_{l1}}(f_{F_0}) - P_{N_1}(f_{F_0}) \end{aligned}$$

and

$$\begin{aligned} Q_{n_{l2}}(K_{\hat{F}_n}) &= (Q_{n_{l2}} - Q_0)(K_{\hat{F}_n} - K_{F_0}) + (Q_0 - Q_{N_2})(K_{\hat{F}_n} - K_{F_0}) \\ &\quad + Q_{N_2}(K_{\hat{F}_n}) + Q_{n_{l2}}(K_{F_0}) - Q_{N_2}(K_{F_0})\,. \end{aligned}$$

Since for any $\delta' > 0$,

$$P\left\{\frac{N_1}{\sqrt{n}}\, |P_{n_{l1}}(f_{\hat{F}_n} - f_{F_0})|\, 1_{A_n^c} > \delta'\right\} \le P(A_n^c) < \varepsilon\,,$$

$$P\left\{\frac{N_1}{\sqrt{n}}\, |P_{N_1}(f_{\hat{F}_n} - f_{F_0})|\, 1_{A_n^c} > \delta'\right\} \le P(A_n^c) < \varepsilon\,,$$

and by the same arguments as those used in Sun et al. (2005), we have that

$$\sqrt{n_{l2}}\,(Q_{n_{l2}} - Q_0)\,(K_{\hat{F}_n} - K_{F_0}) \to 0$$

and

$$\sqrt{N_2}\,(Q_{N_2} - Q_0)\,(K_{\hat{F}_n} - K_{F_0}) \to 0$$

in probability as $n \to \infty$, then

$$\begin{aligned}\frac{1}{\sqrt{n}}\, U_{\eta}^{(l)} &= \frac{N_1}{\sqrt{n}}\,(P_{n_{l1}} - P_0)\,\{(f_{\hat{F}_n} - f_{F_0})\,1_{[\tau,M]}\}\,1_{A_n} + \frac{N_1}{\sqrt{n}}\,(P_0 - P_{N_1})\,\{(f_{\hat{F}_n} - f_{F_0})\,1_{[\tau,M]}\}\,1_{A_n} \\ &\quad + \left\{\frac{N_1}{\sqrt{n}}\,P_{N_1}(f_{\hat{F}_n}) + \frac{N_2}{\sqrt{n}}\,Q_{N_2}(K_{\hat{F}_n})\right\} \\ &\quad + \frac{N_1}{\sqrt{n}}\,P_{n_{l1}}(f_{F_0}) - \frac{N_1}{\sqrt{n}}\,P_{N_1}(f_{F_0}) + \frac{N_2}{\sqrt{n}}\,Q_{n_{l2}}(K_{F_0}) - \frac{N_2}{\sqrt{n}}\,Q_{N_2}(K_{F_0}) + \circ_p(1)\,.\end{aligned}$$

It is easy to see that $P_{n_{l1}}(f_{F_0})$, $P_{N_1}(f_{F_0})$, $Q_{n_{l2}}(K_{F_0})$ and $Q_{N_2}(K_{F_0})$ are U-statistics and

$$\left\{\frac{N_1}{\sqrt{n}}\,P_{n_{l1}}(f_{F_0}) - \frac{N_1}{\sqrt{n}}\,P_{N_1}(f_{F_0}) + \frac{N_2}{\sqrt{n}}\,Q_{n_{l2}}(K_{F_0}) - \frac{N_2}{\sqrt{n}}\,Q_{N_2}(K_{F_0}),\ l = 1,\ldots,k\right\}$$

has the asymptotic distribution given in the theorem. For the term $\{N_1 P_{N_1}(f_{\hat{F}_n}) + N_2 Q_{N_2}(K_{\hat{F}_n})\}$, it follows from the arguments similar to the Proposition 3.2 of Groeneboom (1996) that $\{N_1 P_{N_1}(f_{\hat{F}_n}) + N_2 Q_{N_2}(K_{\hat{F}_n})\} = 0$. Thus for the proof, it is sufficient to show that

$$\sqrt{n_{l1}}\,(P_{n_{l1}} - P_0)\,\{(f_{\hat{F}_n} - f_{F_0})\,1_{[\tau,M]}\} \to 0$$

and

$$\sqrt{N_1}\,(P_{N_1} - P_0)\,\{(f_{\hat{F}_n} - f_{F_0})\,1_{[\tau,M]}\} \to 0$$

in probability as $n \to \infty$.

From the following property of $\hat{F}_n$ used by Huang (1999)

$$\sup |\,\hat{F}_n(t) - F_0(t)| = O_p(n^{-\frac{1}{2}})\,,$$

we have that for $t \in [\tau, M]$,

$$\begin{aligned}f_{\hat{F}_n}(t) - f_{F_0}(t) &= [\eta'\{\hat{F}_n(t)\} - \eta'\{F_0(t)\}] \\ &\quad + \eta''\{\hat{F}_n(t)\}\,\{\hat{F}_n(t) - \hat{F}_n(t-)\} + O_p(n^{-1})\,.\end{aligned}$$

Define

$$\begin{aligned}\mathcal{F} &= \{F : F \text{ is a distribution function defined on } [0,M],\ F(\tau) > 0,\ F(\tau-) > 0\}\,, \\ \mathcal{G} &= \{[\eta'\{F(t)\} - \eta'\{F_0(t)\}]\,1_{[\tau,M]}(t)\ :\ F \in \mathcal{F}\}\end{aligned}$$

and

$$\mathcal{H} = \{\eta''\{F(t)\}\,\{F(t) - F(t-)\}1_{[\tau,M]}(t)\ :\ F \in \mathcal{F}\}\,.$$

Then $\mathcal{F}$ is a P-Donsker from the proof of Corollary 5.1 of Huang and Wellner (1995), and $\mathcal{G}$ and $\mathcal{H}$ are P-Donsker by using the bracket entropy theorem of van der Vaart and Wellner (1996, pp. 127–159) and the arguments similar to those used in Huang and Wellner (1995). It thus follows from the uniform asymptotic equicontinuity of the empirical process resulting from the Donsker property (van der Vaart and Wellner, 1996, pp. 168–171) that

$$\sqrt{n_{l1}}\,(P_{n_{l1}} - P_0)\,(f_{\hat{F}_n} - f_{F_0}) \to 0$$

and

$$\sqrt{N_1}\,(P_{N_1} - P_0)\,(f_{\hat{F}_n} - f_{F_0}) \to 0$$

in probability as $n \to \infty$. This completes the proof.

Acknowledgements The authors are grateful to the AE and two referees for their helpful comments and suggestions. They would like to thank Professor P. K. Andersen for supplying the data that were used in the example.

Conflict of Interests Statement
The authors have declared no conflict of interest.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1992). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Process and Survival Analysis*. John Wiley, New York.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845–854.

Groeneboom, P. (1996). Lectures on inverse problems. In *Lecture Notes in Mathematics*, 1648. Springer-Verlag, Berlin.

Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Birkhauser, Basel.

Huang, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica* **9**, 501–519.

Huang, J. and Wellner, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I. *Statistica Neerlandica*, **49**, 153–163.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data, 2nd edition.* Wiley, New York.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society, Series B* **65**, 489–502.

Li, G. (2003). Nonparametric likelihood ratio goodness-of-fit tests for survival data. *Journal Multivariate Analysis* **86**, 166–182.

Odell, P. M., Anderson, K. M., and D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics* **48**, 951–959.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, Series A* **135**, 185–207.

Ramlau-Hansen, H., Jespersen, N. C., and Andersen, P. K. (1987). Life insurance for insulin-dependent diabetics. *Scandinavian Actuarial Journal* **39**, 19–36.

Sun, J. (1996). A nonparametric test for interval-censored failure time data with application to AIDS studies. *Statistics in Medicine* **15**, 1387–1395.

Sun, J. (1998). Interval censoring. *Encyclopedia of Biostatistics*, 2090–2095. John Wiley, New York.

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

Sun, J., Zhao, Q. and Zhao, X. (2005). Generalized long-rank tests for interval-censored failure time data. *Scandinavian Journal of Statistics* **32**, 49–57.

Turnbull, B. W. (1976). The empirical distribution with arbitrarily grouped censored and truncated data. *Journal of the Royal Statistical Society, Series B* **38**, 290–295.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

