Statistics and Its Interface Volume 3 (2010) 455–464

Empirical likelihood confidence intervals for ratio of hazard rates under right censorship*

Shan Jiang and Dongsheng Tu†

Hazard ratio is an important measure for relative difference between treatment groups in clinical trials or other types of studies with time-to-event as an endpoint. Nonparametric confidence intervals for hazard ratio were derived in [26] based on asymptotic normality of the kernel estimate for hazard ratio. Simulation studies found that, however, the actual coverage probabilities of these confidence intervals were still below the nominal level. In this paper, empirical likelihood ratio method is used to construct confidence intervals for hazard ratio functions under right censorship. The asymptotic distribution of the empirical likelihood ratio is established and simulation studies show that empirical likelihood method improves the coverage probabilities of confidence intervals based on asymptotic normality.

AMS 2000 subject classifications: 60K35, 60K35, 60K35.
Keywords and phrases: Empirical likelihood, Hazard ratio, Kernel estimate, Undersmoothing.

1. INTRODUCTION

Hazard ratio is the most used statistical measure to assess the differences between treatments [21]. It is defined as the ratio of two hazard rate functions. For a subject in j-th group with a survival time T_j, hazard rate function at time t is defined as:

$$h_j(t) = \frac{f_j(t)}{1 - F_j(t)},$$

where $F_j(t)$ and $f_j(t)$ are respectively the distribution function and density function of T_j. The hazard ratio function at time t is, therefore, defined as:

$$\rho(t) = \frac{h_1(t)}{h_2(t)}.$$

Several procedures have been proposed in the literature to construct confidence intervals for hazard ratios based on data with potential censoring. The Cox proportional hazard model [3] has been the most widely used procedure over many years to estimate hazard ratio as well as construct its confidence interval, but the crucial assumption behind this procedure, proportional hazard assumption, may not be satisfied by data from epidemiologic studies or clinical trials, see the example provided by [25]. [26] derived two types of undersmoothed kernel confidence intervals for hazard ratio at a given time point t: one based on directly the asymptotic normality of kernel hazard ratio estimate and the other on the Fieller's transformation of hazard ratio estimator. It was found that, in terms of coverage probability, both undersmoothed confidence intervals performed reasonably well when proportional hazard assumption was violated. However, these procedures are still not very satisfactory, because when sample size is small, the true coverage probability is still far from the stated nominal level. This was not improved by linear transformation of kernel estimate. The requirement of estimating variance for a hazard ratio estimator may be the reason for the low accuracy of confidence intervals based on asymptotic normality.

In this paper, we explore the applications of an empirical likelihood method on construction of a confidence interval for hazard ratio. Empirical likelihood ratio confidence interval was first introduced by [16] for a single functional. A comprehensive introduction of empirical likelihood method can be found in [19]. Based on a data-driven likelihood ratio function expressed through constraints, empirical likelihood method does not need to estimate variance when constructing a confidence interval, which leads to very favorable small sample properties in comparison with its competitors. This method has been applied to some statistical problems with censored data, for example, construction of confidence interval for survival function [12], density and hazard function [24], difference of survival functions [14] and ratio of survival functions [23].

In this paper, empirical likelihood ratio function is defined for hazard ratio and shown to have a chi-square asymptotic distribution with one degree of freedom. The coverage probability of confidence interval based on this result is closer to nominal level in comparison to that based on normal approximation.

This paper is organized as follows: empirical likelihood ratio function and associated confidence interval for hazard ratio are defined in section 2. Section 3 presents results of simulation studies and application to data from a clinical trial. Proof of the major result is given in the Appendix.

*Supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
†Corresponding author.

2. METHODOLOGY

Denote T_{ji} and $C_{ji}(j = 1, 2; i = 1, 2, \ldots, n_j)$ the true survival and censoring times of subjects in two groups, respectively. The data we observe from a clinical trial or cohort study are the pairs (X_{j1}, δ_{j1}), (X_{j2}, δ_{j2}), $\ldots$, $(X_{jn_j}, \delta_{jn_j})$, where

$$\begin{cases} X_{ji} = \min(T_{ji}, C_{ji}) \\ \delta_{ji} = I(T_{ji} \le C_{ji}) \end{cases} \tag{1}$$

Here and thereafter, $I(A)$ stands for the indicator function of A. The total sample size from two groups is $n = n_1 + n_2$. Write $0 \le X_{j(1)} \le X_{j(2)} \le \cdots \le X_{j(n_j)} < \infty$ as the ordered statistics of sample $\{X_{ji}\}$ and $\delta_{j(i)}$ the concomitant of $X_{j(i)}$ for $i = 1, 2, \ldots, n_j$ and $j = 1, 2$. Let

$$r_{ji} = \sum_{k=1}^{n_j} I(X_{jk} \ge X_{j(i)}) = n_j - i + 1 \tag{2}$$

be the number of subjects that are still at risk before $X_{j(i)}$.

Now we will make some assumptions on the distribution of the true survival and censoring times. Suppose that $\{T_{ji} : i = 1, 2, \ldots, n_j\}$ are independently distributed with distribution function $F_j(t)$. The survival function of T_{ji} is defined as: $\bar{F}_j(t) = 1 - F_j(t)$. We also assume that $F_j(t)$ has continuous density $f_j(t)$. The hazard function of T_{ji} can be written as

$$h_j(t) = \frac{f_j(t)}{\bar{F}_j(t)}.$$

Suppose that $\{C_{ji} : i = 1, 2, \ldots, n_j\}$ are independently distributed with distribution function $G_j(t)$ and write

$$H_j(t) = 1 - (1 - F_j(t))(1 - G_j(t)),$$

then H_j is the distribution functions of $\{X_{ji} : i = 1, 2, \ldots, n_j\}$, $j = 1, 2$.

The likelihood function based on censored data (1) is defined as:

$$L(F_1, F_2) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} (F_j(X_{ji}) - F_j(X_{ji}-))^{\delta_{ji}} (1 - F_j(X_{ji}))^{(1-\delta_{ji})}.$$

From [12], this likelihood function can be rewritten as:

$$L(F_1, F_2) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} \lambda_{ji}^{\delta_{j(i)}} (1 - \lambda_{ji})^{r_{ji} - \delta_{j(i)}},$$

where

$$\lambda_{ji} = \frac{F(X_{j(i)}) - F(X_{j(i)}-)}{1 - F(X_{j(i)}-)}, \quad i = 1, 2, \ldots, n_j, \quad j = 1, 2.$$

Therefore, we may express cumulative hazard function $\Lambda_j(t) = -\ln \bar{F}_j(t)$ in terms of $\{\lambda_{ji} : i = 1, 2, \ldots, n_j, j = 1, 2\}$:

$$\Lambda_j(t) = -\sum_{i=1}^{n_j} \ln(1 - \lambda_{ji}) I(X_{j(i)} \le t).$$

Let $K_j(t)$ be a kernel function and $a_j = a(n_j)$ a bandwidth parameter. By the kernel smoothing method, an estimator of hazard function could be chosen from the following estimation family:

$$\tilde{h}_j(t) = -\sum_{i=1}^{n_j} \ln(1 - \lambda_{ji}) K_{ji}(t),$$

where

$$K_{ji}(t) = \frac{1}{a_j} K_j\left(\frac{t - X_{ji}}{a_j}\right). \tag{3}$$

Note that different $\{\lambda_{ji}\}$ will lead to a different estimate of $h_j(t)$. It is easy to show that $L(F_1, F_2)$ can be maximized by choosing:

$$\hat{\lambda}_{ji} = \frac{\delta_{j(i)}}{r_{ji}},$$

and this $\hat{\lambda}$ will give one of the estimators from the estimation family $\tilde{h}_j(t)$ defined as:

$$\hat{h}_j(t) = -\sum_{i=1}^{n_j} \ln\left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right) K_{ji}(t), \quad j = 1, 2. \tag{4}$$

Hazard ratio $\rho(t) = h_1(t)/h_2(t)$ is then estimated by:

$$\hat{\rho}(t) = \frac{\hat{h}_1(t)}{\hat{h}_2(t)}. \tag{5}$$

Under constraint $\eta = \tilde{h}_2(t)$ and $\eta\rho(t) = \tilde{h}_1(t)$, we can define the following empirical likelihood ratio for $\rho(t)$:

$$R(\rho(t), \eta, t) = \frac{\sup_{\lambda_{ji}} \left\{ L(F_1, F_2) : \rho(t)\eta - \tilde{h}_1(t) = 0, \eta - \tilde{h}_2(t) = 0 \right\}}{\sup_{\lambda_{ji}} \{L(F_1, F_2)\}}.$$

Then the log likelihood can be written as:

$$\begin{aligned} \ln(R(\rho(t), \eta, t)) &= \sup_{\lambda_{ji}} \left\{ \sum_{j=1}^{2} \sum_{i=1}^{n_j} (\delta_{j(i)} \ln \lambda_{ji} + (r_{ji} - \delta_{j(i)}) \ln(1 - \lambda_{ji})) : \right. \\ &\qquad \left. \rho(t)\eta - \tilde{h}_1(t) = 0, \eta - \tilde{h}_2(t) = 0 \right\} \\ &\quad - \sum_{j=1}^{2} \sum_{i=1}^{n_j} \delta_{j(i)} \ln\left(\frac{\delta_{j(i)}}{r_{ji}}\right) + (r_{ji} - \delta_{j(i)}) \ln\left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right). \end{aligned}$$

By Lagrange Multiplier Method, we can get:

$$\begin{aligned}&\ln(R(\rho(t),\eta,t))\\&\quad=\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{(r_{ji}-\delta_{j(i)})\ln\left(1+\frac{\mu_jK_{ji}(t)}{r_{ji}-\delta_{j(i)}}\right)\right.\\&\qquad\left.-r_{ji}\ln\left(1+\frac{\mu_jK_{ji}(t)}{r_{ji}}\right)\right\},\end{aligned}$$

where the Lagrange Multipliers μ_j, $j=1,2$, should satisfy:

$$\rho(t)\eta+\sum_{i=1}^{n_1}\ln\left(1-\frac{\delta_{1(i)}}{r_{1i}+\mu_1K_{1i}(t)}\right)K_{1i}(t)=0, \tag{6}$$

$$\eta+\sum_{i=1}^{n_2}\ln\left(1-\frac{\delta_{2(i)}}{r_{2i}+\mu_2K_{2i}(t)}\right)K_{2i}(t)=0. \tag{7}$$

We denote the left hand sides of equations (6) and (7) as $Q_{1n}(\eta,\mu_1,\mu_2,t)$ and $Q_{2n}(\eta,\mu_1,\mu_2,t)$, respectively. Define

$$L_j(t)=\max_i\left\{\frac{\delta_{j(i)}-r_{ji}}{K_{ji}(t)}\right\}.$$

Assume τ_1, τ_2 are two numbers such that

$$c_{F_1}\vee c_{F_2}<\tau_1<\tau_2<d_{H_1}\vee d_{H_2},$$

where $c_{F_j}=\inf\{x:F_j(x)>0\}$ and $d_{F_j}=\sup\{x:F_j(x)<1\}$. We restrict t in interval $[\tau_1,\tau_2]$. The reason t has to be restricted to this interval is that the law of iterated logarithm for Kaplan-Meier estimator [4] and kernel hazard estimator [27], major tool in the proof of Theorem 2.1 below, may not be valid outside this interval. It can be shown that for each $t\in[\tau_1,\tau_2]$, Q_{1n} is a strictly increasing function of μ_1 on interval $(L_1(t),\infty)$ for fixed n_1. When μ_1 approaches $L_1(t)$, we can find Q_{1n} decreasing to $-\infty$; when μ_1 approaches ∞, the limit of Q_{1n} will be η, which is positive. Therefore equation (6) has a unique root, and we can write it as $\mu_1(\eta,t)$. Similarly, we can show that equation (7) has a unique root $\mu_2(\eta,t)$. By implicit function theorem, we get:

$$\begin{aligned}&\frac{\partial\mu_1(\eta,t)}{\partial\eta}\\&\quad=-\rho(t)\left(\sum_{i=1}^{n_1}\frac{\delta_{1(i)}K_{1i}^2(t)}{(r_{1i}+\mu_1K_{1i}(t))(r_{1i}+\mu_1K_{1i}(t)-\delta_{1(i)})}\right)^{-1},\\&\frac{\partial\mu_2(\eta,t)}{\partial\eta}\\&\quad=-\left(\sum_{i=1}^{n_2}\frac{\delta_{2(i)}K_{2i}^2(t)}{(r_{2i}+\mu_2K_{2i}(t))(r_{2i}+\mu_2K_{2i}(t)-\delta_{2(i)})}\right)^{-1}.\end{aligned}$$

Therefore, the equation:

$$\begin{aligned}&\frac{\partial\ln(R(\rho(t),\eta,t))}{\partial\eta}\\&\quad=\frac{\partial\ln(R(\rho(t),\eta,t))}{\partial\mu_1}\frac{\partial\mu_1}{\partial\eta}+\frac{\partial\ln(R(\rho(t),\eta,t))}{\partial\mu_2}\frac{\partial\mu_2}{\partial\eta}\\&\quad=-\sum_{j=1}^{2}\sum_{i=1}^{n_j}\frac{\delta_{j(i)}K_{ji}^2(t)\mu_j}{(r_{ji}-\delta_{j(i)}+\mu_jK_{ji}(t))(r_{ji}+\mu_jK_{ji}(t))}\frac{\partial\mu_j}{\partial\eta}\\&\quad=0\end{aligned}$$

can be simplified into

$$\rho(t)\mu_1(\eta,t)+\mu_2(\eta,t)=0,$$

which is equivalent to

$$\frac{\rho(t)\mu_1(\eta,t)}{n_1a_1+n_2a_2}+\frac{\mu_2(\eta,t)}{n_1a_1+n_2a_2}=0 \tag{8}$$

We will show in our main theorem that unique root η_E of equation (8) can be found so that log likelihood ratio function $\ln(R(\rho(t),\eta,t))$ reaches its maximum.

Denote the left hand side of equation (8) as $Q_{3n}(\eta,\mu_1,\mu_2,t)$. Define the following conditions for kernel function, bandwidth and hazard function:

1. $K_j(t)(j=1,2)$ are bounded functions with compact support $[-c,c]$ such that:

$$\int_{-\infty}^{\infty}u^iK_j(u)du=\begin{cases}=1, & \text{if } i=0,\\=0, & \text{if } i=1,\\\neq 0, & \text{if } i=2.\end{cases}$$

   The first order derivative of $K_j(t)$ exists.
2. Assume that $h_1(t)>0$ and $h_2(t)>0$ hold for $t\in[\tau_1,\tau_2]$. The derivative $h_j'(t)$ of $h_j(t)$ exists and is continuous.
3. As $n_j\to\infty$, we have $a_j\to 0$, $n_ja_j\to\infty$, $n_ja_j^5\to 0$, $\liminf_{n\to\infty}n^{1/3}a_j>0$, $\ln a_j^{-1}/n_ja_j\to 0$, $\ln a_j^{-1}/\ln\ln n_j\to\infty$, and $n_ja_j/(n_1a_1+n_2a_2)\to\rho_j>0$, $j=1,2$.

Specifically, we have the following theorem.

Theorem 2.1. *Assuming conditions 1–3, for each $t\in[\tau_1,\tau_2]$, there exists a solution $\eta_E(t)$ to equation (8) almost surely as $n\to\infty$, such that $R(\rho(t),\eta,t)$ attains its maximum value at $\eta=\eta_E$, and we have for fixed t*

$$-2\ln R(\rho(t),\eta_E,t)\to\chi_1^2,\ \textit{in distribution}$$

Proof. In the Appendix. □

Remark 2.1. It is very important to select a bandwidth in our kernel smoothing estimate. [2] proposed an undersmoothing kernel bandwidth for construction of a confidence

Table 1. Actual coverage probability of confidence interval for hazard ratio ($h_1(t) = \lambda$)

$h_2(t)$	λ	t	True ratio	C_{el}	C_{tu}	C_{cox}
λ	0.075	6	1.000	0.941	0.925	0.953
		12	1.000	0.946	0.914	0.950
		24	1.000	0.950	0.901	0.950
	0.05	6	1.000	0.951	0.929	0.952
		12	1.000	0.952	0.932	0.951
		24	1.000	0.938	0.917	0.950
	0.025	6	1.000	0.947	0.930	0.942
		12	1.000	0.951	0.915	0.954
		24	1.000	0.955	0.908	0.944
$2\lambda^2 t$	0.075	6	1.111	0.945	0.928	0.699
		12	0.556	0.948	0.917	0.168
		24	0.278	0.948	0.907	0.000
	0.05	6	1.667	0.928	0.924	0.059
		12	0.833	0.944	0.929	0.884
		24	0.417	0.928	0.901	0.003
	0.025	6	3.333	0.856	0.916	0.002
		12	1.667	0.913	0.912	0.637
		24	0.833	0.935	0.910	0.507
$t\exp(-\lambda t)/\int_t^\infty u\exp(-\lambda u)du$	0.075	6	3.222	0.943	0.936	0.671
		12	2.111	0.945	0.922	0.825
		24	1.556	0.930	0.900	0.189
	0.05	6	4.333	0.950	0.927	0.312
		12	2.667	0.944	0.927	0.950
		24	1.833	0.925	0.907	0.417
	0.025	6	7.667	0.952	0.908	0.158
		12	4.333	0.939	0.921	0.865
		24	2.667	0.906	0.917	0.824

interval for a hazard function, which is defined as:

$$a_j = \frac{\exp\{(\lambda_{Cj} + \lambda_{Tj})t/3\}}{\{\lambda_{Tj}(\lambda_{Cj} + \lambda_{Tj})^2 n_j\}^{1/3}},$$

where

$$\lambda_{Tj} = \frac{\sum_{i=1}^{n_j} \delta_{ij}}{\sum_{i=1}^{n_j} X_{ij}}$$

and

$$\lambda_{Cj} = \frac{n_j - \sum_{i=1}^{n_j} \delta_{ij}}{\sum_{i=1}^{n_j} X_{ij}}.$$

This undersmoothing bandwidth alleviates estimation difficulties caused by bias and is shown to minimize the coverage error of a confidence interval for hazard rate function. This bandwidth satisfies the condition 3 for Theorem 2.1 and can be used in practice, although any bandwidth of order $O(n^{-1/3})$ can also be used. We used this bandwidth in our simulation studies and applications to real data from clinical trials.

From Theorem 2.1, an empirical likelihood confidence interval for hazard ratio function $\rho(t)$ at fixed $t \in [\tau_1, \tau_2]$ with asymptotical coverage accuracy $1 - \alpha$ can be defined as:

$$I_{n,\alpha}(t) = \{\rho(t) : -2\ln R(\rho(t), \eta_E, t) \le C_\alpha\}$$

where C_α satisfies:

$$\mathrm{P}(\chi_1^2 \le C_\alpha) = 1 - \alpha.$$

Remark 2.2. The confidence interval defined above is for a hazard ratio at a fixed time t. In practice, it may also be useful to have a simultaneous confidence interval over a given time interval. There is a technical difficulty to directly generalize the procedure developed in this paper to construct simultaneous confidence intervals since, as pointed out by [6], the stochastic process defined by kernel estimate of hazard rate is not tight. For the density function, Hall and Owen [8] derived empirical likelihood based simultaneous confidence intervals by following the technique used by Bickel and Rosenblatt [1]. Application of the same technique to construct simultaneous confidence intervals for a hazard ratio is an interesting problem for further investigation.

3. NUMERICAL STUDIES

Simulations are conducted following the same scenarios in [26]. Specifically, true survival times are assumed in the first group from an exponential distribution with parameter λ and in the second group from respectively, exponential distribution with parameter λ and Weibull and Gamma distributions with shape and scale parameters respectively γ and λ. The censoring distribution is assumed to be uniformly distributed over interval $[T_f, T_a + T_f]$, which corresponds to a clinical trial process with patients accrued uniformly

Table 2. Expected length of confidence interval for hazard ratio ($h_1(t) = \lambda$)

$h_2(t)$	λ	t		L_{el}	L_{tu}	L_{cox}
λ	0.075	6	1.000	1.441	1.603	0.617
		12	1.000	1.716	2.053	0.619
		24	1.000	3.549	32.083	0.618
	0.05	6	1.000	1.477	1.558	0.658
		12	1.000	1.597	1.901	0.657
		24	1.000	2.419	3.144	0.660
	0.025	6	1.000	1.853	1.692	0.788
		12	1.000	1.685	1.848	0.780
		24	1.000	2.060	2.580	0.778
$2\lambda^2 t$	0.075	6	1.111	1.455	1.715	0.558
		12	0.556	0.823	0.906	0.557
		24	0.278	109.996	12.984	0.555
	0.05	6	1.667	2.308	2.988	0.627
		12	0.833	1.135	1.325	0.625
		24	0.417	0.907	0.903	0.624
	0.025	6	3.333	5.926	12.938	0.979
		12	1.667	2.368	3.452	0.979
		24	0.833	1.681	1.611	0.979
$t\exp(-\lambda t)/\int_t^\infty u\exp(-\lambda u)du$	0.075	6	3.222	5.909	5.837	1.725
		12	2.111	3.410	3.635	1.729
		24	1.556	36.834	3.722	1.737
	0.05	6	4.333	10.583	9.742	2.101
		12	2.667	4.775	4.979	2.108
		24	1.833	4.338	3.995	2.109
	0.025	6	7.667	114.150	2421.374	3.856
		12	4.333	13.937	13.478	3.857
		24	2.667	8.472	6.379	3.824

into the study from time 0 to time T_a and all patients followed for at least T_f time unit before the end of the study. λ ranges from 0.075, 0.05, and 0.025 but γ is fixed at 2. In addition, we fix T_a and T_f respectively at 60 and 6, as λ varies from 0.075, 0.05 to 0.025, which gives us the censoring rate of respectively 14%, 23%, and 45% when the distribution of the survival time is exponential, 10%, 20%, and 48% when the distribution of the survival time is Weibull, and 34%, 50%, and 77% when the distribution of the survival time is Gamma. For each parameter configuration, 3,000 random samples of sizes $n_1 = 100$ and $n_2 = 100$ are generated. The proportion of confidence intervals covering the true hazard ratio over 3,000 samples are used to estimate the coverage probability for each confidence interval, and the average length of confidence intervals to estimate the length of the proposed confidence interval. The nominal significant level α used in all simulations is 0.05 and the following kernel function is used for all kernel estimates:

$$K(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \le 1) \tag{9}$$

The results of simulations are presented in Table 1 and 2 respectively for the true coverage probability and length of proposed confidence intervals. In these tables, C_{el}, C_{tu}, C_{cox} and l_{el}, l_{tu}, l_{cox} represent respectively the coverage probabilities and lengths of confidence intervals based on empirical likelihood method, asymptotic normality and Cox proportional hazard model. It can be seen from these tables that empirical likelihood method improves the confidence interval based on normal approximation in almost all cases and the lengths of these two intervals are also comparable at the majority of cases.

From Table 1, we can notice that when $h_2 = 2\lambda^2 t$, $\lambda = 0.025$ and $t = 6$, the coverage probability based on empirical likelihood procedure is only 0.856. This may be caused by relatively few events observed at this earlier time. [26] recommended to avoid making inference on hazard ratio at the time when there are too few events observed. The same recommendation may be made for use of the proposed confidence interval based on empirical likelihood.

We also applied the proposed empirical likelihood method to the same data set from a randomized clinical trial considered by [26]. This trial was designed to compare two chemotherapy regimens (CEF v.s. CMF) in women with early stage breast cancer. 710 pre-menopausal women with axillary node positive breast cancer were recruited in this trial with a median follow-up 8.8 years for all patients at end of trial.

Table 3 presents confidence intervals for a hazard ratio of death at respectively 2, 4, 6 and 8 years after randomiza-

Table 3. Estimate of hazard ratio and 95% CI of treatment CEF to CMF

Years from randomization	Number at risk CEF	Number at risk CMF	Hazard rate CEF	Hazard rate CMF	KHR[1]	CI_{na}[2]	CI_{el}[3]
2	323	326	0.0696	0.0897	0.78	0.44-1.11	0.46-1.07
4	284	265	0.0472	0.0782	0.60	0.28-0.92	0.39-0.99
6	251	239	0.0578	0.0437	1.32	0.55-2.09	0.80-2.12
8	216	207	0.0403	0.0537	0.75	0.28-1.21	0.45-1.52

tion based on respectively normal approximation and empirical likelihood methods. The empirical likelihood confidence interval is slightly shorter except at 8 years from randomization. Although both methods would conclude that CEF is significantly better than CMF at 4 years after randomization, the upper endpoint of the empirical likelihood confidence interval is closer than 1, which confirms the results from the simulation study that the confidence interval based on normal approximation may be more liberal than the empirical likelihood confidence interval.

APPENDIX A. PROOF OF THEOREM 2.1

In what follows, we assume the conditions of Theorem 2.1 are satisfied.

Lemma A.1.

$$\hat{h}_j(t) - h_j(t) = O\left(\sqrt{\frac{\ln n_j}{n_j a_j}}\right), \quad j = 1, 2.$$

Proof of Lemma A.1. Lemma A.1 can be proved following the same arguments in the proof of Theorem 2.3 in [27]. □

Lemma A.2. *As $n \to \infty$,*

$$\sqrt{n_j a_j}(h_j(t) - \hat{h}_j(t)) \to N(0, \sigma_j^2(t)),\ (j = 1, 2) \text{ in distribution,}$$

where

$$\sigma_j^2(t) = \frac{h_j}{\bar{H}_j}\int_{-c}^{c} K_j^2(t)dt.$$

Proof of Lemma A.2. Lemma A.2 can be proved from Theorem 4.2 in [13]. □

Lemma A.3. *Define $\varepsilon_n = n^{-s}$, with $\frac{1}{3} < s < \frac{1}{2}$. Let $\eta_0 = h_2(t)$ and assume that $t \in [\tau_1, \tau_2]$, then for any η satisfies $|\eta - \eta_0| \le a_1^{-1/2}\varepsilon_n$, the solutions $\mu_1(\eta, t)$ and $\mu_2(\eta, t)$ of equations (6) and (7), respectively, satisfy:*

$$\text{(A.1)} \quad \frac{\mu_1(\eta, t)}{n_1} = O(a_1^{\frac{1}{2}}\varepsilon_n) \text{ and } \frac{\mu_2(\eta, t)}{n_2} = O(a_2^{\frac{1}{2}}\varepsilon_n) \text{ a.s.}$$

Proof of Lemma A.3. For $j = 1, 2$, define

$$\text{(A.2)} \quad \begin{cases} \hat{\sigma}_j^2(t) = a_j n_j \sum_{i=1}^{n_j} \frac{\delta_{j(i)} K_{ji}^2}{r_{ji}(r_{ji} - \delta_{j(i)})}, \\ \tilde{\sigma}_j^2(t) = a_j n_j \sum_{i=1}^{n_j} \frac{\delta_{j(i)} K_{ji}^2}{r_{ji}^2}. \end{cases}$$

Similar to the proof of Proposition 3.3.1 of [20], we can show that

$$\hat{\sigma}_j^2(t) \to \sigma_j^2(t) \text{ a.s.}$$
$$\tilde{\sigma}_j^2(t) \to \sigma_j^2(t) \text{ a.s.}$$

Denote

$$\text{(A.3)} \quad \begin{cases} A_{1n}(\eta, t) = \hat{h}_1(t) - \eta\rho(t) \\ A_{2n}(\eta, t) = \hat{h}_2(t) - \eta \end{cases}$$

Since we have from (6) and (7)

$$\eta\rho(t) = -\sum_{i=1}^{n_1} \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right) K_{1i}(t)$$
$$\eta = -\sum_{i=1}^{n_2} \ln\left(1 - \frac{\delta_{2(i)}}{r_{2i} + \mu_2 K_{2i}(t)}\right) K_{2i}(t)$$

using inequality $|\ln(1-x) - \ln(1-y)| \ge |x - y|$ for $x, y \in (0, 1)$, we can get

$$\begin{aligned} &\mu_1 A_{1n}(\eta, t) \\ &= \mu_1\Bigg[-\sum_{i=1}^{n_1} \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i}}\right) K_{1i}(t) \\ &\quad + \sum_{i=1}^{n_1} \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right) K_{1i}(t)\Bigg] \\ &= |\mu_1| \left|\sum_{i=1}^{n_1} K_{1i}(t)\Bigg[\ln\left(1 - \frac{\delta_{1(i)}}{r_{1i}}\right) - \ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right)\Bigg]\right| \\ &\ge |\mu_1| \sum_{i=1}^{n_1} K_{1i}(t) \left|\frac{\delta_{1(i)}}{r_{1i}} - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right| \\ &= \mu_1^2 \sum_{i=1}^{n_1} K_{1i}^2(t) \frac{\delta_{1(i)}}{r_{1i}(r_{1i} + \mu_1 K_{1i}(t))} \\ &\ge \frac{\mu_1^2}{1 + \max_i\left(\frac{\mu_1 K_{1i}(t)}{r_{1i}}\right)} \sum_{i=1}^{n_1} \frac{K_{1i}^2(t)\delta_{1(i)}}{r_{1i}^2} \\ &= \frac{\mu_1^2 \tilde{\sigma}_1^2}{\left(1 + \max_i\left(\frac{\mu_1 K_{1i}(t)}{r_{1i}}\right)\right) n_1 a_1}. \end{aligned}$$

From condition 1, we have $|K_j(x)| \le M$, for an $M > 0$ and $j = 1, 2$, which leads to

$$\mu_1 A_{1n}(\eta,t) \ge \frac{\mu_1^2\tilde\sigma_1^2}{a_1 n_1 + M|\mu_1|\max_i(\frac{n_1}{r_{1i}})}.$$

Since for sufficiently large n_1 and n_2, we have almost surely ((4.6) in [24])

$$\max_i \left|\frac{n_j}{r_{ji}}\right| \le \frac{2}{\bar H_j(\tau_2)}, \quad j = 1, 2,$$

and

$$\tilde\sigma_1^2(t) \ge \frac{1}{2}\sigma_1^2(\tau_1).$$

Therefore, we have

$$|A_{1n}(\eta,t)| \ge \frac{|\mu_1|\sigma_1^2(\tau_1)}{2\left(a_1 n_1 + 2M|\mu_1|\bar H_1^{-1}(\tau_2)\right)}. \tag{A.4}$$

On the other hand, from definition $h_2(t) = \eta_0$, we have by lemma A.1

$$\begin{aligned} A_{1n}(\eta,t) &= \hat h_1(t) - \rho(t)\eta_0 + \rho(t)\eta_0 - \rho(t)\eta \\ &= \hat h_1(t) - h_1(t) + \rho(t)(\eta_0 - \eta) \\ &\le o(a_1^{-\frac{1}{2}}\varepsilon_n) + O(a_1^{-\frac{1}{2}}\varepsilon_n) \\ &= O(a_1^{-\frac{1}{2}}\varepsilon_n). \end{aligned} \tag{A.5}$$

Combining (A.4) and (A.5), we get

$$\frac{\mu_1(\eta,t)}{n_1} = O(a_1^{\frac{1}{2}}\varepsilon_n) \text{ a.s. for fixed } t \in [\tau_1,\tau_2].$$

Similarly, we can prove

$$\frac{\mu_2(\eta,t)}{n_2} = O(a_2^{\frac{1}{2}}\varepsilon_n) \text{ a.s. for fixed } t \in [\tau_1,\tau_2].$$

□

Lemma A.4. *Almost surely, for large n_1 and n_2, equation (8) has a solution $\eta_E(t)$, such that $R(\rho(t),\eta,t)$ reaches its maximum value $R(\rho(t),t)$ at $\eta = \eta_E(t)$.*

Proof of Lemma A.4. For any pair (j,i) which satisfies $X_{ji} < \tau_2$, we have almost surely for sufficiently large n

$$\frac{n_j}{r_{ji}} \le \frac{n_j}{\sum_{k=1}^{n_j}(X_{jk} \ge \tau_2)} \le \frac{2}{\bar H_j(\tau_2)}.$$

By Taylor Expansion and Lemma A.3, we get

$$\begin{aligned} &\ln\left[1 - \frac{\delta_{j(i)}}{r_{ji} + \mu_j K_{ji}(t)}\right] K_{ji}(t) \\ &= \ln\left[1 - \frac{\delta_{j(i)}}{r_{ji}}\left(1 + \frac{\mu_j K_{ji}(t)}{r_{ji}}\right)^{-1}\right] K_{ji}(t) \\ &= K_{ji}(t)\ln\left[1 - \frac{\delta_{j(i)}}{r_{ji}}\left(1 - \frac{\mu_j K_{ji}(t)}{r_{ji}} + O\left(\frac{\mu_j^2 K_{ji}^2(t)}{r_{ji}^2}\right)\right)\right] \\ &= K_{ji}(t)\ln\left[1 - \frac{\delta_{j(i)}}{r_{ji}} + \frac{\mu_j\delta_{j(i)}K_{ji}(t)}{r_{ji}^2} - \delta_{j(i)} \right. \\ &\quad \left. \times O\left(\frac{\mu_j^2 K_{ji}^2(t)}{r_{ji}^3}\right)\right] \\ &= K_{ji}(t)\ln\left\{\left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right)\left[1 + \left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right)^{-1}\delta_{j(i)} \right.\right. \\ &\quad \left.\left. \times \left(\frac{\mu_j K_{ji}(t)}{r_{ji}^2} + O\left(\frac{\mu_j^2 K_{ji}^2}{r_{ji}^3}\right)\right)\right]\right\} \\ &= K_{ji}(t)\ln\left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right) + K_{ji}(t)\ln\left[1 + \frac{\delta_{j(i)}\mu_j K_{ji}(t)}{(r_{ji} - \delta_{j(i)})r_{ji}} \right. \\ &\quad \left. + O\left(\frac{\mu_j^2 K_{ji}^2(t)}{r_{ji}^2(r_{ji} - \delta_{j(i)})}\right)\right] \\ &= K_{ji}(t)\ln\left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right) + K_{ji}^2(t)\frac{\delta_{j(i)}\mu_j}{(r_{ji} - \delta_{j(i)})r_{ji}} \\ &\quad + O\left(\frac{\mu_j^2 K_{ji}^3(t)}{r_{ji}^3}\right) \\ &= K_{ji}(t)\ln\left(1 - \frac{\delta_{j(i)}}{r_{ji}}\right) + K_{ji}^2(t)\frac{\delta_{j(i)}\mu_j}{(r_{ji} - \delta_{j(i)})r_{ji}} \\ &\quad + O\left(\frac{\varepsilon_n^2}{n a_j^2}\right). \end{aligned}$$

Therefore, from (6), (A.2) and the above equation, we have almost surely

$$\begin{aligned} \eta\rho(t) &= -\sum_{i=1}^{n_1} K_{1i}(t)\ln\left(1 - \frac{\delta_{1(i)}}{r_{1i} + \mu_1 K_{1i}(t)}\right) \\ &= -\sum_{i=1}^{n_1} K_{1i}(t)\ln\left(1 - \frac{\delta_{1(i)}}{r_{1i}}\right) \\ &\quad - \sum_{i=1}^{n_1} K_{1i}^2(t)\frac{\delta_{1(i)}\mu_1}{(r_{1i} - \delta_{1(i)})r_{1i}} + O\left(\frac{\varepsilon_n^2}{a_1^2}\right) \\ &= \hat h_1(t) - \frac{\mu_1(\eta,t)\hat\sigma_1^2}{n_1 a_1} + O\left(\frac{\varepsilon_n^2}{a_1^2}\right). \end{aligned}$$

Similarly, we can show that, almost surely

$$\eta = \hat h_2(t) - \frac{\mu_2(\eta,t)\hat\sigma_2^2}{n_2 a_2} + O\left(\frac{\varepsilon_n^2}{a_2^2}\right).$$

Hence, from (A.3), we get

$$\begin{aligned} \mu_1(\eta,t) &= \frac{n_1 a_1(\hat h_1(t) - \eta\rho(t))}{\hat\sigma_1^2} + O\left(\frac{n_1\varepsilon_n^2}{a_1}\right) \\ &= \frac{a_1 n_1 A_{1n}(\eta,t)}{\hat\sigma_1^2} + O\left(\frac{n_1\varepsilon_n^2}{a_1}\right) \text{ a.s.,} \end{aligned} \tag{A.6}$$

$$\mu_2(\eta,t) = \frac{a_2 n_2 A_{2n}(\eta,t)}{\hat\sigma_2^2} + O\left(\frac{n_2\varepsilon_n^2}{a_2}\right) \text{ a.s.}$$

From (1), Lemma A.3 and using Taylor Expansion again, we have

(A.7)

$$\begin{aligned}
&-2\ln(R(\rho(t),\eta,t)) \\
&= -2\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{(r_{ji}-\delta_{j(i)})\ln(1+\frac{\mu_j K_{ji}(t)}{r_{ji}-\delta_{j(i)}}) - r_{ji}\ln(1+\frac{\mu_j K_{ji}(t)}{r_{ji}})\right\} \\
&= -2\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{(r_{ji}-\delta_{j(i)})\left[\frac{\mu_j K_{ji}(t)}{r_{ji}-\delta_{j(i)}} - \frac{\mu_j K_{ji}(t)}{2(r_{ji}-\delta_{j(i)})^2} + O\left(\frac{\mu_j^3 K_{ji}^3(t)}{(r_{ji}-\delta_{j(i)})^3}\right)\right] - r_{ji}\left[\frac{\mu_j K_{ji}(t)}{r_{ji}} - \frac{\mu_j^2 K_{ji}^2}{2r_{ji}^2} + O\left(\frac{\mu_j^3 K_{ji}^3(t)}{r_{ji}^3}\right)\right]\right\} \\
&= 2\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{\frac{\mu_j^2 K_{ji}^2(t)}{2(r_{ji}-\delta_{j(i)})} + O\left(\frac{\mu_j^3 K_{ji}^3(t)}{(r_{ji}-\delta_{j(i)})^2}\right) - \frac{\mu_j^2 K_{ji}^2(t)}{2r_{ji}} - O\left(\frac{\mu_j^3 K_{ji}^3(t)}{r_{ji}^2}\right)\right\} \\
&= \sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{\frac{\delta_{j(i)} K_{ji}^2(t)\mu_j^2}{r_{ji}(r_{ji}-\delta_{j(i)})} + O\left(\varepsilon_n^3 n_j a_j^{-3/2}\right)\right\} \\
&= \frac{\mu_1^2(\eta,t)\hat\sigma_1^2}{n_1a_1} + \frac{\mu_2^2(\eta,t)\hat\sigma_2^2}{n_2a_2} + O\left(n\varepsilon_n^3 a_1^{-3/2}\right).
\end{aligned}$$

From (A.6) and (A.7)

$$\text{(A.8)}\quad -2\ln R(\rho(t),\eta,t) = \frac{a_1n_1A_{1n}^2(\eta,t)}{\hat\sigma_1^2} + \frac{a_2n_2A_{2n}^2(\eta,t)}{\hat\sigma_2^2} + O(n\varepsilon_n^3 a_1^{-3/2}).$$

If we write $\eta_n = \eta_0 + \Delta = h_2(t) + \Delta$, such that $\Delta \to 0$, $\Delta^2 a_1^{5/2}/\varepsilon_n^3 \to \infty$, and $\Delta^2 a_1 n_1/\ln\ln n \to \infty$, using Taylor Expansion of $A_{jn}^2(\eta_n,t)$, $j=1,2$ at η_0, we have almost surely

$$\text{(A.9)}\quad -2\ln R(\rho(t),\eta_n,t) = \frac{a_1n_1}{\hat\sigma_1^2}(A_{1n}(\eta_0,t)-\rho(t)\Delta)^2 + \frac{a_2n_2}{\hat\sigma_2^2}(A_{2n}(\eta_0,t)-\Delta)^2 + O(n\varepsilon_n^3 a^{-3/2}).$$

Since from Lemma A.1, we have almost surely

$$\text{(A.10)}\quad \begin{aligned} A_{1n}(\eta_0,t) &= \hat h_1 - \eta_0\rho(t) = O(n^{-1/2}a_1^{-1/2}\sqrt{\ln n}), \\ A_{2n}(\eta_0,t) &= \hat h_2 - \eta_0 = O(n^{-1/2}a_1^{-1/2}\sqrt{\ln n}), \end{aligned}$$

from (A.9) and (A.10), we have almost surely

$$-2\ln R(\rho(t),\eta_n,t) = O(na\Delta^2).$$

On the other hand, for sufficiently large n_1 and n_2, we have from (A.8) almost surely

$$\begin{aligned} -2\ln R(\rho(t),\eta_0,t) &= O(\ln\ln n) + O(n\varepsilon_n^3 a^{-3/2}) \\ &= o(na\Delta^2) \text{ (from assumption on } \Delta.) \end{aligned}$$

Therefore, for sufficiently large n_1 and n_2, we have

$$\text{(A.11)}\quad -2\ln R(\rho(t),\eta_0+\Delta,t) > -2\ln R(\rho(t),\eta_0,t), \text{ a.s.}$$

Similarly, we can obtain

$$\text{(A.12)}\quad -2\ln R(\rho(t),\eta_0-\Delta,t) > -2\ln R(\rho(t),\eta_0,t), \text{ a.s.}$$

Combining (A.11) and (A.12), we know that $-2\ln R(\rho(t),\eta,t)$ attains its minimum in the region $(\eta_0-\Delta,\eta_0+\Delta)$, say at η_E. □

Proof of Theorem 2.1. Denote $\nu_1 = \mu_1(\eta,t)/(n_1a_1)$, $\nu_2 = \mu_2(\eta,t)/(n_2a_2)$, $\nu_{1E} = \mu_1(\eta_E,t)/(n_1a_1)$ and $\nu_{2E} = \mu_2(\eta_E,t)/(n_2a_2)$. From (6)–(8), together with $n_1a_1/(n_1a_1+n_2a_2) \to p_1$ and $n_2a_2/n_1a_1+n_2a_2 \to p_2$, we have

$$S_n(\eta,\ t) = \left.\frac{\partial(Q_{1n},Q_{2n},Q_{3n})}{\partial(\eta,\nu_1,\nu_2)}\right|_{(\eta,\nu_1,\nu_2,t)=(\eta,0,0,t)} = \begin{pmatrix} \rho(t) & \hat\sigma_1^2(t) & 0 \\ 1 & 0 & \hat\sigma_2^2(t) \\ 0 & \frac{n_1a_1\rho(t)}{n_1a_1+n_2a_2} & \frac{n_2a_2}{n_1a_1+n_2a_2} \end{pmatrix},$$

and

$$S_n(\eta,\ t) \to S(\eta,t) := \begin{pmatrix} \rho(t) & \sigma_1^2(t) & 0 \\ 1 & 0 & \sigma_2^2(t) \\ 0 & p_1\rho(t) & p_2 \end{pmatrix}$$

in probability. By Taylor expansion we get

$$\begin{pmatrix}0\\0\\0\end{pmatrix} = \begin{pmatrix} Q_{1n}(\eta_E,\nu_{1E},\nu_{2E},t) \\ Q_{2n}(\eta_E,\nu_{1E},\nu_{2E},t) \\ Q_{3n}(\eta_E,\nu_{1E},\nu_{2E},t) \end{pmatrix} = \begin{pmatrix} Q_{1n}(\eta_0,0,0,t) \\ Q_{2n}(\eta_0,0,0,t) \\ 0 \end{pmatrix} + S_n(\eta_0,\ t)\begin{pmatrix}\eta_E-\eta_0 \\ \nu_{1E} \\ \nu_{2E}\end{pmatrix} + o(1).$$

Therefore,

$$\begin{pmatrix} \eta_E - \eta_0 \\ \nu_{1E} \\ \nu_{2E} \end{pmatrix} \approx S_n^{-1}(\eta_0,\ t) \begin{pmatrix} \hat{h}_1(t) - h_1(t) \\ \hat{h}_2(t) - h_2(t) \\ 0 \end{pmatrix}$$
$$\approx \begin{pmatrix} -p_1\rho(t)\sigma_2^2(t) & -p_2\sigma_1^2(t) & * \\ -p_2 & p_2\rho(t) & * \\ p_1\rho(t) & -p_1\rho^2(t) & * \end{pmatrix} \times \begin{pmatrix} \hat{h}_1(t) - h_1(t) \\ \hat{h}_2(t) - h_2(t) \\ 0 \end{pmatrix} \frac{1}{\det(S(\eta_0,\ t))},$$

where

$$\det(S_n(\eta_0,\ t)) = -p_1\rho^2(t)\sigma_2^2(t) - p_2\sigma_1^2(t).$$

This leads to

$$\begin{aligned}\nu_{1E} &\approx \left(-p_1\rho^2(t)\sigma_2^2(t) - p_2\sigma_1^2(t)\right)^{-1} \\ &\quad \times \Big(p_2(h_1(t) - \hat{h}_1(t)) - p_2\rho(t)(h_2(t) - \hat{h}_2(t))\Big) \\ &= \frac{p_2}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} \\ &\quad \times \Big(\rho(t)(h_2(t) - \hat{h}_2(t)) - (h_1(t) - \hat{h}_1(t))\Big) \\ &= \frac{p_2}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} W(t),\end{aligned}$$

where

$$W(t) = \rho(t)(h_2(t) - \hat{h}_2(t)) - (h_1(t) - \hat{h}_1(t)).$$

From Lemma A.2, we have $\sqrt{n_1a_1}\ W(t)$ is asymptotically normal distributed with mean 0 and variance

$$\text{(A.13)} \qquad \operatorname{var}(\sqrt{n_1a_1}\ W(t)) = \frac{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)}{p_2}.$$

On the other hand, from (A.7) and (8), we can get

$$\begin{aligned} &-2\ln(R(\rho(t), \eta_E, t)) \\ &\approx \frac{\mu_1^2(\eta_E, t)\sigma_1^2}{n_1a_1} + \frac{\mu_2^2(\eta_E, t)\sigma_2^2}{n_2a_2} \\ &= \frac{\mu_1^2(\eta_E, t)\sigma_1^2}{n_1a_1} + \frac{\rho^2(t)\mu_1^2(\eta_E, t)\sigma_2^2(t)}{n_2a_2} \\ &= \nu_{1E}n_1a_1\sigma_1^2(t) + \nu_{2E}n_2a_2\sigma_2^2(t) \\ &\approx \frac{n_1a_1p_2^2\sigma_1^2(t)}{(p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t))^2} W^2(t) \\ &\quad + \frac{\rho^2(t)n_1a_1p_1p_2\sigma_2^2(t)}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} W^2(t) \\ &= \frac{p_2}{p_1\rho^2(t)\sigma_2^2(t) + p_2\sigma_1^2(t)} (\sqrt{n_1a_1}W(t))^2.\end{aligned}$$

Combining the above with (A.13), we have

$$-2\ln R(\rho,\ \eta_E,\ t) \to \chi_1^2, \text{ in distribution.}$$

□

Received 14 December 2009

REFERENCES

[1] Bickel, P. and Rosenblatt, M. (1993). On some global estimate of the deviations of density estimation. *Annals of Statistics* **1** 1071–1095. MR0348906

[2] Cheng, M., Hall, P. and Tu, D. (2006). Confidence bands for hazard rates under random censorship. *Biometrika* **93** 357–366. MR2278089

[3] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)* **34** 187–220. MR0341758

[4] Csörgö, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Probability Theory and Related Fields* **62** 411–426. MR0688647

[5] Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis.* John Wiley and Sons, New York. MR1100924

[6] Gilbert, P. B., Wei, L. J., Kosorok, M. R. and Clemens, J. D. (2002). Simultaneous inferences on the contrast of two hazard functions with censored observations. *Biometrics* **58** 773–780. MR1945014

[7] Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics* **20** 675–694. MR1165587

[8] Hall, P. and Owen, A. B. (1993). Empirical likelihood confidence bands in density estimation. *Journal of Computational and Graphical Statistics* **2** 273–289. MR1272395

[9] Hall, W. J. and Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67** 133–143. MR0570515

[10] Howe, G. R. (1983). Confidence Interval Estimation for the Ratio of Simple and Standardized Rates in Cohort Studies. *Biometrics* **39** 325–331.

[11] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistic Association* **53** 457–481. MR0093867

[12] Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statistics and Probability Letters* **25** 95–104. MR1365025

[13] Lo, S. H., Mack, Y. P. and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probability Theory and Related Fields* **80** 461–473. MR0976536

[14] McKeaguea, I. W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statistics and Probability Letters* **60** 405–415. MR1947180

[15] Müller, H. and Wang J. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* **50** 61–76. MR1279435

[16] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049

[17] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annal of Statistics* **18** 90–120. MR1041387

[18] Owen, A. B. (1991). Empirical likelihood for linear models. *Annal of Statistics* **19** 1725–1747. MR1135146

[19] Owen, A. B. (2001). *Empirical likelihood* Chapman and Hall, London.

[20] Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Annals of Statistics* **11** 453–466. MR0696058

[21] Scott, W. F. (2000). The estimation of the hazard ratio in clinical trials and in meta-analysis. *The Mathematical Scientist* **25** 54–58. MR1771178

[22] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap* Springer-Verlag, New York. MR1351010

[23] Shen, J. and He, S. (2006). Empirical likelihood for the difference of two survival functions under right censorship. *Statistics and Probability Letters* **76** 169–181. MR2233389

[24] Shen, J. and He, S. (2008). Empirical likelihood confidence intervals for hazard and density functions under right censorship. *Annals of the Institute of Statistical Mathematics* **60** 575–589. MR2434412

[25] Stablein, D. M., Carter, W. H. and Novak, J. W. (1981). Analysis of survival data with non-proportional hazard functions. *Controlled Clinical Trials* **2**, 149–159.

[26] Tu, D. (2007). Under-smoothed kernel confidence intervals for the hazard ratio based on censored data. *Biometrical Journal* **49** 474–483. MR2380526

[27] Xiang, X. (1994). Law of the logarithm for density and hazard rate estimation for censored data. *Journal of Multivariate Analysis* **49** 278–286. MR1276440

Shan Jiang
Department of Math & Stats
Jeffery Hall, University Ave.
Kingston ON K7L 3N6
Canada
E-mail address: sjiang@mast.queensu.ca

Dongsheng Tu
Cancer Clinical Trials Division
Cancer Research Institute, Queen's University
10 Stuart Street
Kingston ON K7L 3N6
Canada
E-mail address: Dtu@ctg.queensu.ca