



Contents lists available at ScienceDirect

# Statistics and Probability Letters

journal homepage: [www.elsevier.com/locate/stapro](http://www.elsevier.com/locate/stapro)

## Empirical likelihood inference for the odds ratio of two survival functions under right censoring

Meng Zhao<sup>a</sup>, Yichuan Zhao<sup>b,\*</sup>, Ian W. McKeague<sup>c</sup><sup>a</sup> Data Science R & D, CareerBuilder.com, Norcross, GA 30092, USA<sup>b</sup> Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA<sup>c</sup> Department of Biostatistics, Columbia University, New York, NY 10032, USA

### ARTICLE INFO

#### Article history:

Received 26 November 2014

Received in revised form 5 September 2015

Accepted 5 September 2015

Available online 14 September 2015

#### Keywords:

Case-control

Confidence interval

Right censored data

Wald-type

### ABSTRACT

This paper develops a new method for estimating survival odds ratios under right censoring using the censored empirical likelihood. Results of a simulation study show that under small sample sizes the proposed method outperforms the corresponding Wald-type CI and the log-transformed Wald-type CI.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

The odds ratio is widely used in epidemiological studies, especially in case-control designs from which the risk ratio becomes unidentifiable (see, e.g. Cui et al., 2003; Pepe et al., 2004; Zhang et al., 2001). To sum up briefly, there are three major advantages in the applications of odds ratios (Bland and Altman, 2000): the estimation of binary relationships based on retrospective case-control data, the approximation of the risk ratio when the risks in the two groups being compared are small, as well as easier adjustment for covariate effects in logistic regression models.

In this paper, we study the problem of estimating odds ratios for comparing two survival functions with right-censored data. Right-censoring in survival data is a well-known feature of cohort studies, but it is also a common feature of retrospective case-control studies in which odds ratios play the central role. In retrospective studies, right-censoring is relevant to recall bias resulting from subjects being unable to remember the precise timing of life-course events (when used as the primary endpoint). For example, a subject suffering from early-stage Alzheimer's disease may be unable to remember events signaling the onset of the disease. There is an extensive literature on the issue of recall bias in epidemiological studies that use odds ratios to quantify associations (see, e.g. Hassan, 2006; McPhail and Haines, 2010; Parr et al., 2009; Vrijheid et al., 2009). Nonetheless, as far as we know, there have been no attempts to address the problem from the perspective of right-censoring (apart from simply discarding the right-censored observations).

We develop a nonparametric confidence interval (CI) for the odds ratio defined by two survival functions at age  $t$ , as  $t$  varies over the follow-up period of the study. Data are given in the form of two independent right-censored random samples from the survival distributions of the case and control populations. The proposed CI is developed using the empirical likelihood (EL) approach described in Owen (1988, 1990). Although Wald-type CIs for odds ratios can also be used for the

\* Corresponding author.

E-mail address: [dz2007@gmail.com](mailto:dz2007@gmail.com) (Y. Zhao).

same purpose, as discussed in [Dong and Matthews \(2012\)](#), they have some known limitations: the need for sophisticated variance estimation, restricted (symmetric) shape of CIs, and a tendency to less-than-nominal coverage in small samples. In contrast, EL based CIs have several advantages: no need for variance estimation, natural adaptation to skewness in the data, and greater accuracy in small sample settings. EL methods in survival analysis have been extensively studied, originating with [Thomas and Grunkemeier \(1975\)](#). [Li \(1995\)](#) investigated the limiting distribution of the empirical likelihood ratio for survival functions with right censoring, see [Li et al. \(2005\)](#) for a detailed review of the field. For two sample comparisons, [McKeague and Zhao \(2002\)](#) constructed a simultaneous confidence band for the ratio of two survival functions based on right censored data. [McKeague and Zhao \(2005\)](#) later extended this EL approach to the difference and ratio of distribution functions. [Shen and He \(2006\)](#) derived EL based confidence bands for the difference of two survival functions.

The paper is organized as follows. In Section 2, the approach of [Shen and He \(2006\)](#) is extended to estimate odds ratios of survival functions, and asymptotic properties of the resulting EL likelihood ratio statistic are established. Section 3 provides the results of a simulation study comparing the proposed method with the Wald-type method and the log-transformed Wald-type method. An additional illustration is presented in Section 4 using real data. Proofs are collected in the [Appendix](#).

**2. Main results**

Consider two independent samples with independent right censoring. With  $i = 1, 2$  indexing the two samples, let  $\{Z_{ij}, j = 1, 2, \dots, n_i\}$  be i.i.d. failure times from the distribution  $F_i = 1 - S_i$ . Let  $\{C_{ij}, j = 1, 2, \dots, n_i\}$  be non-negative i.i.d. censoring times from the distribution  $G_i$ , independent of  $Z_{ij}$ . The right-censored observations for each sample are then given by the pairs

$$(X_{ij}, \delta_{ij}), \quad j = 1, 2, \dots, n_i, \tag{2.1}$$

where  $X_{ij} = \min(Z_{ij}, C_{ij})$ , and  $\delta_{ij} = I(Z_{ij} \leq C_{ij})$  is the censoring indicator. Here and throughout the paper, we use  $I(A)$  to denote the indicator function of a set  $A$ .

To support major results in sequel, we herein adopt the same notations as in [Shen and He \(2006\)](#). For any cumulative distribution function  $F$ , let  $\bar{F} = 1 - F$  and define the range  $(a_F, b_F)$  as

$$a_F = \inf\{x : F(x) > 0\} \quad \text{and} \quad b_F = \sup\{x : F(x) < 1\}.$$

Denote  $H_i = 1 - \bar{F}_i \bar{G}_i$ , and given  $[t_1, t_2]$  such that  $a_{F_1} \vee a_{F_2} < t_1 < t_2 < b_{F_1} \wedge b_{F_2}$ , write the odds ratio of two survival functions as

$$\theta(t) = \frac{1 - S_1(t)}{S_1(t)} \bigg/ \frac{1 - S_2(t)}{S_2(t)}, \quad t \in [t_1, t_2].$$

Our objective is to obtain a CI for  $\theta_0(t)$ , the true value of  $\theta(t)$  at  $t = t_0 \in [t_1, t_2]$ .

Let  $0 \leq T_{i1} \leq \dots \leq T_{in_i} < \infty$  denote the ordered uncensored survival times of sample  $i$ , and write

$$r_{ij} = \sum_{k=1}^{n_i} I(X_{ik} \geq T_{ij}), \quad d_{ij} = \sum_{k=1}^{n_i} I(X_{ik} = T_{ij}, \delta_{ik} = 1),$$

as the number of subjects “at risk” prior to time  $T_{ij}$ , and the number of failures at time  $T_{ij}$ , respectively. In addition, let  $K_i(t) = \#\{j : T_{ij} \leq t\}$  and  $\Gamma$  be the space of all survival functions defined on  $[0, \infty)$ . For any  $S_1, S_2 \in \Gamma$ , define the empirical likelihood function as

$$L(S_1, S_2) = \prod_{i=1}^2 \prod_{j=1}^{n_i} [S_i(X_{ij-}) - S_i(X_{ij})]^{\delta_{ij}} [S_i(X_{ij})]^{1-\delta_{ij}}, \quad S_1, S_2 \in \Gamma.$$

At a fixed  $t$ , the nonparametric likelihood ratio for testing the null hypothesis that  $\theta(t) = \theta_0(t)$  (versus the two-sided alternative) is

$$\mathcal{R}(\theta_0, \eta, t) = \frac{\sup\{L(S_1, S_2) : S_1(t) = \eta / [\eta + \theta_0(t) - \theta_0(t)\eta], S_2(t) = \eta, S_1, S_2 \in \Gamma\}}{\sup\{L(S_1, S_2) : S_1, S_2 \in \Gamma\}}.$$

Here  $\eta$  is a nuisance parameter that separates out  $S_1$  and  $S_2$  in a convenient way. Referring to [Li \(1995\)](#), write

$$\log \mathcal{R}(\theta_0, \eta, t) = \sum_{i=1}^2 \sum_{j=1}^{K_i(t)} \left\{ (r_{ij} - d_{ij}) \log \left( 1 + \frac{\lambda_j}{r_{ij} - d_{ij}} \right) - r_{ij} \log \left( 1 + \frac{\lambda_j}{r_{ij}} \right) \right\}, \tag{2.2}$$

where the Lagrange multipliers  $\lambda_1 = \lambda_1(\eta)$  and  $\lambda_2 = \lambda_2(\eta)$  are the unique solution to the equations

$$\sum_{j=1}^{K_1(t)} \log \left( 1 - \frac{d_{1j}}{r_{1j} + \lambda_1} \right) - \log \left( \frac{\eta}{\eta + \theta_0(t) - \theta_0(t)\eta} \right) = 0, \tag{2.3}$$

$$\sum_{j=1}^{K_2(t)} \log \left( 1 - \frac{d_{2j}}{r_{2j} + \lambda_2} \right) - \log(\eta) = 0. \tag{2.4}$$

The idea is to profile out the nuisance parameter  $\eta$ . Using the implicit function theorem (Rudin, 1976), it can be shown that any value of  $\eta = \eta(t)$  that maximizes  $\mathcal{R}(\theta_0, \cdot, t)$  satisfies

$$\frac{\lambda_1(\eta)}{n} \left( \frac{\theta_0}{\eta + \theta_0 - \eta\theta_0} \right) + \frac{\lambda_2(\eta)}{n} = 0. \quad (2.5)$$

See the Appendix for the proof of (2.5). Also, let

$$\sigma_i^2(t) = \int_0^t \frac{dF_i(s)}{S_i(s-)H_i(s)}, \quad i = 1, 2, \quad (2.6)$$

where  $F_i$ ,  $G_i$  and  $H_i$  are introduced at the beginning of this section. The following result provides a way of constructing a CI for  $\theta_0(t)$ . The proof is based on arguments of Shen and He (2006) and is given in the Appendix.

**Theorem 1.** Assume that  $n_j/n \rightarrow p_j \in (0, 1)$ , as  $n \rightarrow \infty$ , where  $n = n_1 + n_2$ , and let  $a_{F_1} \vee a_{F_2} < t_1 < t_2 < b_{H_1} \wedge b_{H_2}$ . There exists a unique solution  $\eta^*(t)$  for Eq. (2.5) such that  $\eta^*(t) \in \arg \max_{\eta} \mathcal{R}(\theta_0, \eta, t)$  for all  $t \in [t_1, t_2]$  almost surely, and

$$-2 \log \mathcal{R}(\theta_0, \eta^*(t), t) \xrightarrow{d} \frac{M^2(t)}{\sigma^2(t)} \quad (2.7)$$

as a process in  $D[t_1, t_2]$ , where

$$\sigma^2(t) = \frac{(1 - S_2(t))^2 \sigma_1^2(t)}{p_1} + \frac{(1 - S_1(t))^2 \sigma_2^2(t)}{p_2}, \quad (2.8)$$

$$M(t) = \frac{\{1 - S_2(t)\}M_1(\sigma_1^2(t))}{\sqrt{p_1}} + \frac{\{1 - S_1(t)\}M_2(\sigma_2^2(t))}{\sqrt{p_2}}, \quad (2.9)$$

and  $M_1, M_2$  are independent standard Brownian motions.

Under the conditions of the theorem, for fixed  $t$ ,

$$-2 \log \mathcal{R}(\theta_0, \eta^*, t) \xrightarrow{d} \chi_1^2, \quad (2.10)$$

so a pointwise asymptotic  $100(1 - \alpha)\%$  CI for  $\theta_0(t)$  is given by

$$I_{n,\alpha}(t) = \{\theta : -2 \log \mathcal{R}(\theta, \eta^*, t) \leq \chi_{1,\alpha}^2\},$$

where  $\chi_{1,\alpha}^2$  is the upper  $\alpha$ -quantile of  $\chi_1^2$ .

### 3. Simulation study

In this section we report the results of a simulation study evaluating the performance of the proposed EL approach compared to Wald-type CIs (based on normal approximation). This is done in terms of both coverage probability and average interval length.

First we describe two competing Wald-type CIs for  $\theta_0(t)$ . Write

$$\hat{\theta}_n(t) = \frac{1 - \hat{S}_1(t)}{\hat{S}_1(t)} \bigg/ \frac{1 - \hat{S}_2(t)}{\hat{S}_2(t)}, \quad t \in [t_1, t_2],$$

where  $\hat{S}_1(t)$  and  $\hat{S}_2(t)$  are the Kaplan–Meier estimators of  $S_1(t)$  and  $S_2(t)$ , respectively. From standard results in Andersen et al. (1993), for fixed  $t \in [t_1, t_2]$

$$n_i^{1/2} \{\hat{S}_i(t) - S_i(t)\} \xrightarrow{d} S_i(t)M_i(\sigma_i^2(t)), \quad i = 1, 2,$$

so by the delta method it follows that

$$\sqrt{n}(\hat{\theta}_n(t) - \theta_0(t)) \xrightarrow{d} M^*(t), \quad (3.1)$$

where

$$M^*(t) = \frac{S_2(t)}{S_1(t)\{1 - S_2(t)\}^2} \cdot \left[ \frac{\{1 - S_2(t)\}M_1(\sigma_1^2(t))}{\sqrt{p_1}} + \frac{\{1 - S_1(t)\}M_2(\sigma_2^2(t))}{\sqrt{p_2}} \right].$$

Therefore, an asymptotic  $100(1 - \alpha)\%$  Wald-type CI for  $\theta_0(t)$  is given by

$$I_{n,\alpha}^*(t) = \hat{\theta}_n(t) \pm z_{\alpha/2} \frac{\hat{S}_2(t)}{\hat{S}_1(t)\{1 - \hat{S}_2(t)\}^2} \frac{\hat{\sigma}^*(t)}{\sqrt{n}},$$

**Table 1**  
Coverage probability and average length (in parentheses) of CIs for  $\theta_0(t)$ ,  $t = 1$ .

CR	$n_1$	$\alpha = 0.05$			$\alpha = 0.10$		
		EL	Wald-type	log (Wald-type)	EL	Wald-type	log (Wald-type)
0.10	30	0.947 (2.696)	0.917 (2.378)	0.967 (2.878)	0.894 (2.717)	0.880 (2.017)	0.922 (2.313)
	50	0.949 (2.695)	0.886 (1.667)	0.955 (1.877)	0.902 (1.859)	0.849 (1.401)	0.897 (1.524)
	80	0.947 (1.179)	0.885 (1.262)	0.947 (1.360)	0.898 (1.001)	0.868 (1.058)	0.889 (1.115)
	120	0.947 (0.919)	0.887 (1.025)	0.956 (1.077)	0.899 (0.784)	0.869 (0.858)	0.893 (0.889)
0.30	30	0.913 (3.562)	0.945 (3.272)	0.972 (4.268)	0.887 (3.232)	0.926 (2.772)	0.933 (3.353)
	50	0.927 (3.994)	0.957 (2.350)	0.972 (2.740)	0.889 (2.136)	0.937 (1.951)	0.927 (2.174)
	80	0.946 (2.355)	0.967 (1.662)	0.976 (1.825)	0.896 (1.653)	0.944 (1.376)	0.944 (1.470)
	120	0.948 (2.052)	0.972 (1.352)	0.974 (1.439)	0.897 (1.241)	0.949 (1.129)	0.933 (1.180)

where

$$\hat{\sigma}^{*2}(t) = \frac{\{1 - \hat{S}_2(t)\}^2 \hat{\sigma}_1^2(t)}{\hat{p}_1} + \frac{\{1 - \hat{S}_1(t)\}^2 \hat{\sigma}_2^2(t)}{\hat{p}_2},$$

$\hat{\sigma}_j^2(t)$  is given by (2.11) of Shen and He (2006), and  $\hat{p}_j = n_j/n$ .

Similarly, using the delta method it can be shown that

$$\sqrt{n} \left[ \log(\hat{\theta}_n(t)) - \log(\theta_0(t)) \right] \xrightarrow{d} \frac{M^*(t)}{\theta_0(t)}, \tag{3.2}$$

after which the exponential function can be applied to transform the CI back to the original scale. Therefore, for fixed  $t$ , an asymptotic  $100(1 - \alpha)\%$  Wald-type (log-transformed) CI for  $\theta_0(t)$  is given by

$$I_{n,\alpha}^{**}(t) = \exp \left[ \log(\hat{\theta}_n(t)) \pm z_{\alpha/2} \frac{\hat{\sigma}^*(t)}{\sqrt{n}(1 - \hat{S}_1(t))(1 - \hat{S}_2(t))} \right].$$

In the following simulations, both survival time and censoring time are generated independently from Weibull distributions. Four different sample sizes (30, 50, 80 and 120) are selected for the case group and matching sample sizes are assigned to the control group. In addition, two different censoring rates (CR), 10% and 30%, are considered. Under each setting, 10,000 repetitions are conducted to assess coverage probabilities as well as average CI length. More specifically, we set  $F_1 = \text{Weibull}(3, 1)$ ,  $F_2 = \text{Weibull}(6, 1)$  and generate the censoring times using the distributions  $G_1 = \text{Weibull}(3, 2)$ ,  $G_2 = \text{Weibull}(6, 1.3)$  that produce 10% CR, and  $G_1 = \text{Weibull}(3, 1.2)$ ,  $G_2 = \text{Weibull}(6, 1.1)$  that produce 30% CR.

At fixed  $t = 1$ , the proposed empirical likelihood CI,  $I_{n,\alpha}(t)$ , the Wald-type CI,  $I_{n,\alpha}^*(t)$ , and the log-transformed Wald-type CI,  $I_{n,\alpha}^{**}(t)$ , are evaluated under nominal levels  $\alpha = 0.90$  and  $0.95$ . The `nleqslv` package (Hasselmann, 2015), which allows ill-conditioned Jacobian matrices, is used to solve the system of four nonlinear equations (2.2)–(2.5). Initial values of  $\lambda_1$  and  $\lambda_2$  are set as  $n_1^{1/2}$  and  $-n_1^{1/2}$ , respectively, for the lower bound, and reversed for the upper bound. We compare the proposed method with both types of Wald CI. Coverage probability and average interval length (shown in parentheses) are reported in Table 1.

It can be seen from Table 1 that, when the sample size is small, the EL-type CI is more accurate yet slightly wider on average than the Wald-type CI. The EL-type CI has comparable average length to the log-transformed Wald-type CI in most cases. The log-transformed Wald-type CI is conservative when the censoring rate is high (30%). The EL band has the best overall performance. The advantage of the EL procedure naturally diminishes as the sample size grows.

#### 4. Application to real data

As a real data example, we present the analysis of the primary biliary cirrhosis (PBC) data (Fleming and Harrington, 1991), following McKeague and Zhao (2002, 2005, 2006). The various types of pointwise CIs we have developed are compared for the survival odds ratio between treatment (D-penicillamine) and placebo. There are 312 patients recruited in the randomized clinical trial; 158 patients received the treatment and 154 received placebo. The data are heavily right censored (187 out of 312 patients). Fig. 1 provides EL and Wald-type CIs for the odds ratio of survival between treatment and control. The estimated odds ratio based on Kaplan–Meier estimators is also displayed. The Wald-type CIs are symmetric around the estimated odds ratio. The EL-type CIs and log-transformed Wald-type CIs are asymmetric around the estimated odds ratio. The EL-type CIs are substantially narrower than the log-transformed Wald-type CIs at the start of follow-up, which is their major advantage. As can be seen from the plot, our method agrees with that of McKeague and Zhao (2002, 2005, 2006) for diagnostic indications in the sense that the drug treatment does not show significant improvement in terms of survival.

In the two-sample comparison, the survival time may depend on various covariates, i.e., risk factors. When the covariates are high dimensional, the EL method would need to be modified based on a semiparametric model such as a stratified Cox model.

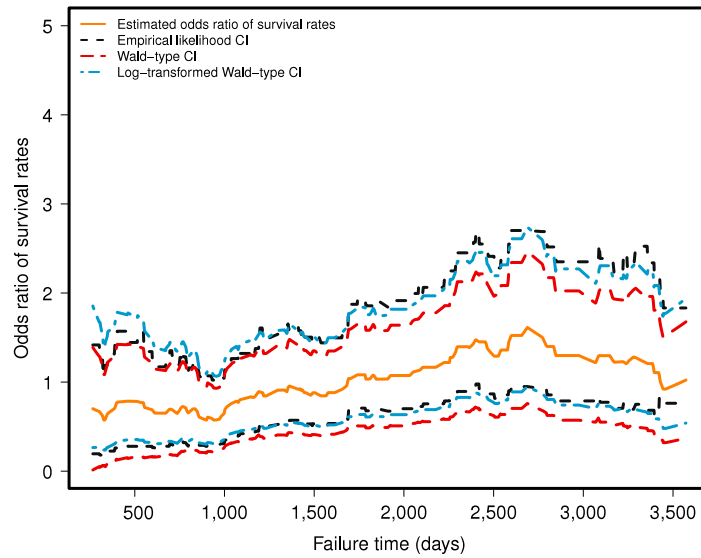


Fig. 1. 95% pointwise CIs for the odds ratio of survival functions based on the Mayo PBC data.

**Acknowledgments**

The authors thank a referee and the Co-Editor-in-Chief, Professor Davy Paindaveine, for many insightful comments that led to substantial improvements in the manuscript. The research of Ian McKeague was partially supported by NSF Grant DMS-1307838 and NIH Grant R01 GM095722. Yichuan Zhao was partially supported by NSA grant H98230-12-1-0209, NSF grant DMS-1406163, and a RIG grant from Georgia State University.

**Appendix. Proofs**

First, (2.5) follows from similar steps in Shen and He (2006, p. 171). Write  $L_i(t) = \max_{j:T_{ij} \leq t} (d_{ij} - r_{ij})$ . Since the left-hand-sides of (2.3) and (2.4) are strictly increasing functions of  $\lambda_i$  on  $(L_i(t), \infty)$ , these equations have unique roots. Hereafter we denote the roots as  $\lambda_i = \lambda_i(\eta, t)$ ,  $i = 1, 2$ . For fixed  $t$ , we can obtain an estimate for the nuisance parameter  $\eta_0 = S_2(t)$  by maximizing  $\log \mathcal{R}(\theta_0, \eta, t)$ . By the implicit function theorem (Rudin, 1976), taking derivatives w.r.t.  $\eta$  on both sides of (2.3) and (2.4), we have

$$\frac{\partial \lambda_1(\eta, t)}{\partial \eta} = \left[ \frac{(\eta + \theta_0 - \theta_0 \eta) \theta_0}{\theta_0} \sum_{j=1}^{K_1(t)} \frac{d_{1j}}{(r_{1j} + \lambda_1)(r_{1j} + \lambda_1 - d_{1j})} \right]^{-1},$$

$$\frac{\partial \lambda_2(\eta, t)}{\partial \eta} = \left[ \eta \sum_{j=1}^{K_2(t)} \frac{d_{2j}}{(r_{2j} + \lambda_2)(r_{2j} + \lambda_2 - d_{2j})} \right]^{-1}.$$

Thus, we have

$$\begin{aligned} \frac{\partial \log \mathcal{R}(\theta_0, \eta, t)}{\partial \eta} &= \frac{\partial \log \mathcal{R}(\theta_0, \eta, t)}{\partial \lambda_1} \frac{\partial \lambda_1(\eta, t)}{\partial \eta} + \frac{\partial \log \mathcal{R}(\theta_0, \eta, t)}{\partial \lambda_2} \frac{\partial \lambda_2(\eta, t)}{\partial \eta} \\ &= - \sum_{j=1}^{K_1(t)} \frac{\lambda_1(\eta, t) d_{1j}}{(r_{1j} + \lambda_1)(r_{1j} + \lambda_1 - d_{1j})} \frac{\partial \lambda_1(\eta, t)}{\partial \eta} - \sum_{j=1}^{K_2(t)} \frac{\lambda_2(\eta, t) d_{2j}}{(r_{2j} + \lambda_2)(r_{2j} + \lambda_2 - d_{2j})} \frac{\partial \lambda_2(\eta, t)}{\partial \eta} \\ &= 0 \end{aligned}$$

which simplifies to (2.5).

Adapting Shen and He (2006) to the present context, write  $\varepsilon_n = n^{-s}$ , where  $s \in (1/3, 1/2)$ , and  $n = n_1 + n_2$ . Note that  $\eta_0 = S_2(t)$ .

**Lemma 1.** Under the conditions of Theorem 1, for sufficiently large  $n$ , there exists a solution to Eq. (2.5), denoted as  $\eta^*(t)$ , such that  $\eta^*(t) = \arg \sup_{\eta} \mathcal{R}(\theta_0, \eta, t)$  almost surely, and

$$\eta^*(t) \rightarrow \eta_0 = S_2(t), \tag{A.1}$$

as  $n \rightarrow \infty$  uniformly for all  $t \in [t_1, t_2]$  almost surely.

**Proof.** As in Shen and He (2006), write

$$\begin{aligned} A_1(\eta(t), t) &= \log \left( \frac{\eta(t)}{\eta(t) + \theta_0(t) - \theta_0(t)\eta(t)} \right) - \log \hat{S}_1(t), \\ A_2(\eta(t), t) &= \log(\eta(t)) - \log \hat{S}_2(t). \end{aligned} \tag{A.2}$$

By the functional delta method (see, e.g. van der Vaart, 2000), we know

$$\log S_j(t) - \log \hat{S}_j(t) = 1/S_j(t)(S_j(t) - \hat{S}_j(t)) + o_p \left( (S_j(t) - \hat{S}_j(t)) \right) \quad j = 1, 2.$$

For  $t \in [t_1, t_2]$ ,  $S_j(t)$  is bounded, thus  $\log S_j(t) - \log \hat{S}_j(t) = O_p(S_j(t) - \hat{S}_j(t))$ . Now from Csörgö and Horváth (1983), we know  $(\log S_j(t) - \log \hat{S}_j(t))$  is also  $O((n_j / \log \log n_j)^{-1/2})$ . In addition, under the condition of Theorem 1, it is easy to see that given  $n$  large enough,  $|\eta(t) - \eta_0(t)| \leq \epsilon_n$  uniformly in  $[t_1, t_2]$ , where  $\epsilon_n$  is defined just before the statement of Lemma 1. Then, by the Taylor expansion there exists  $\eta_0(t) < \eta^*(t) < \eta(t)$  such that

$$\begin{aligned} A_1(\eta(t), t) &= \log(\theta_0(t) + \eta(t)) - \log(\theta_0(t) + \eta_0(t)) + \log(\theta_0(t) + \eta_0(t)) - \log(\hat{S}_1(t)) \\ &= \frac{\theta_0(t)(\eta(t) - \eta_0(t))}{\eta(t)^*\{\eta^*(t) + \theta_0(t) - \theta_0(t)\eta^*(t)\}} + \log S_1(t) - \log \hat{S}_1(t) \\ &= O(\epsilon_n) + O((n_1 / \log n_1)^{-1/2}) \\ &= O(\epsilon_n), \end{aligned} \tag{A.3}$$

uniformly over  $t \in [t_1, t_2]$  for  $n$  sufficiently large a.s., where the second term of (A.3) is by Theorem 2 of Csörgö and Horváth (1983). Moreover, Theorem 2 of Csörgö and Horváth (1983) implies the LIL of the Kaplan–Meier estimator, from which we know

$$\sup_{t \in [t_1, t_2]} |\hat{S}_j(t) - S_j(t)| = O((n_j / \log \log n_j)^{-1/2}) = o(\epsilon_n) \quad \text{a.s.}$$

Thus,  $A_j(\eta_0(t), t) = o(\epsilon_n)$  also hold uniformly for  $t \in [t_1, t_2]$ . Then, by (4.3) of Shen and He (2006), we have

$$\begin{aligned} \lambda_1 A_1(\eta(t), t) &= |\lambda_1| \sum_{j=1}^{K_1(t)} \left| \log \left( 1 - \frac{d_{1j}}{r_{1j} + \lambda_1} \right) - \log \left( 1 - \frac{d_{1j}}{r_{1j}} \right) \right| \\ &\geq \frac{\lambda_1^2}{n_1 + |\lambda_1| \max_{j: T_{1j} \leq t} \{ |n_1 / r_{1j}| \}} \tilde{\sigma}_1^2(t), \end{aligned} \tag{A.5}$$

where

$$\tilde{\sigma}_1^2(t) = n_1 \sum_{j=1}^{K_1(t)} \frac{d_{1j}}{r_{1j}^2}. \tag{A.6}$$

Shen and He (2006) have proved that almost surely for  $t \in [t_1, t_2]$ ,

$$\tilde{\sigma}_1^2(t) = \sigma_1^2(t) + o(1) \geq \sigma_1^2(t_1)/2, \tag{A.7}$$

and

$$\max_{j: T_{1j} \leq t} \left| \frac{n_1}{r_{1j}} \right| = \frac{1}{\bar{H}_1(t)} \leq \frac{2}{\bar{H}_1(t_2)}. \tag{A.8}$$

Thus, combining (A.4)–(A.8), for  $t \in [t_1, t_2]$ , we have almost surely that for large  $n$ ,

$$|A_1(\eta(t), t)| \geq \frac{|\lambda_1|}{n_1 + 2|\lambda_1|/\bar{H}_1(t_2)} \cdot \frac{\sigma_1^2(t_1)}{2}. \tag{A.9}$$

Plugging (A.9) into (A.4), it is easy to see that

$$\frac{\lambda_1}{n_1} = O(\varepsilon_n) \quad (\text{A.10})$$

uniformly over  $t \in [t_1, t_2]$  for  $n$  sufficiently large a.s.

Similarly, it can also be shown that

$$\frac{\lambda_2}{n_2} = O(\varepsilon_n) \quad (\text{A.11})$$

uniformly over  $t \in [t_1, t_2]$  for  $n$  sufficiently large a.s. Similar to (A.4), by (4.10) of Shen and He (2006), we show uniformly for  $t \in [t_1, t_2]$  that

$$\begin{aligned} \log \left( 1 - \frac{d_{ji}}{r_{ji} + \lambda_j(\eta_n, t)} \right) &= \log \left\{ 1 - \frac{d_{ji}}{r_{ji}} \left( 1 + \frac{\lambda_j(\eta_n, t)}{r_{ji}} \right)^{-1} \right\} \\ &= \log \left( 1 - \frac{d_{ji}}{r_{ji}} \right) + \log \left\{ 1 + \frac{d_{ji}}{r_{ji}(r_{ji} - d_{ji})} \lambda_j(\eta_n, t) + \left( 1 - \frac{d_{ji}}{r_{ji}} \right)^{-1} d_{ji} \cdot O \left( \frac{\lambda_j(\eta_n, t)}{r_{ji}^3} \right) \right\} \\ &= \log \left( 1 - \frac{d_{ji}}{r_{ji}} \right) + \frac{d_{ji}}{r_{ji}(r_{ji} - d_{ji})} \lambda_j(\eta_n, t) + O \left( \frac{\varepsilon_n^2}{n} \right). \end{aligned} \quad (\text{A.12})$$

Note that the first two terms of (A.12) are based on the Taylor expansion, and additionally

$$\begin{aligned} \left( 1 - \frac{d_{ji}}{r_{ji}} \right)^{-1} d_{ji} \cdot O \left( \frac{\lambda_j(\eta_n, t)}{r_{ji}^3} \right) &= O \left( \frac{d_{ji} \lambda_j^2(\eta_n, t)}{(r_{ji} - d_{ji}) r_{ji}^2} \right) \\ &= O \left( \frac{\lambda_j^2(\eta_n, t)}{r_{ji}^3} \right) = O \left( \frac{\lambda_j^2(\eta_n, t)}{n_j^3} \right). \end{aligned} \quad (\text{A.13})$$

Since  $\lambda_j/n_j = O(\varepsilon_n)$  uniformly for  $t \in [t_1, t_2]$ , we know (A.13) holds uniformly. This proves that (A.12) holds uniformly. Thus, using (2.11), (2.12) and (4.4) of Shen and He (2006) and combining (2.3) and (2.4) with (A.12), it is easy to see that

$$\begin{aligned} \log \left( \frac{\eta}{\eta + \theta_0 - \theta_0 \eta} \right) &= \log \hat{S}_1(t) + \frac{\lambda_1(\eta_n, t) \hat{\sigma}_1^2(t)}{n_1} + O(\varepsilon_n^2) \quad \text{a.s.}, \\ \log(\eta) &= \log \hat{S}_2(t) + \frac{\lambda_2(\eta_n, t) \hat{\sigma}_2^2(t)}{n_2} + O(\varepsilon_n^2) \quad \text{a.s.}, \end{aligned} \quad (\text{A.14})$$

by using (A.10) and (A.11).

Recasting (A.14), by (A.4) we know

$$\lambda_j(\eta, t) = \frac{n_j}{\hat{\delta}_j(t)} A_j(\eta, t) + O(n_j \varepsilon_n^2) \quad (\text{A.15})$$

holds uniformly. Then, based on the proof of (4.12) of Shen and He (2006), by (2.2), (A.1), (A.2) and (A.14), it can be shown in the same sense that

$$\begin{aligned} -2 \log \mathcal{R}(\theta_0, \eta_0, t) &= \sum_{i=1}^2 \frac{n_i}{\hat{\sigma}_i^2(t)} \{A_i(\eta_0, t)\}^2 + O(n \varepsilon_n^3) \\ &= o(n \varepsilon_n^2) \end{aligned} \quad (\text{A.16})$$

holds uniformly over  $t \in [t_1, t_2]$  for  $n$  sufficiently large a.s. This implies that the inequality,

$$-2 \log \mathcal{R}(\theta_0, \eta_0 + \varepsilon_n, t) > -2 \log \mathcal{R}(\theta_0, \eta_0, t)$$

holds uniformly over  $[t_1, t_2]$  for  $n$  sufficiently large a.s. The reverse inequality can be proved similarly. This means  $-2 \log \mathcal{R}(\theta_0, \eta, t)$  attains its minimum in  $(\eta_0 - \varepsilon_n, \eta_0 + \varepsilon_n)$  uniformly for  $t \in [t_1, t_2]$ . Using the unique solution condition for a stationary point (see, e.g. Rudin, 1976), we know there exists  $\eta^* = \arg \max_{\eta} \mathcal{R}(\theta_0, \eta)$  in  $(\eta_0 - \varepsilon_n, \eta_0 + \varepsilon_n)$  that satisfies (A.1) uniformly for  $t \in [t_1, t_2]$ .  $\square$

**Proof of Theorem 1.** The proof is similar to that of Theorem 2.1 of Shen and He (2006), so we adopt the same notations herein. Let  $\lambda_1 = n_i \gamma_i$ ,  $i = 1, 2$ , and plug these into the left-hand sides of (2.3)–(2.5). Thereafter, denoting the left-hand sides

of these three equations by  $E_i(\eta, \gamma_1, \gamma_2, t)$ ,  $i = 1, 2, 3$ , respectively, and similar to Shen and He (2006), the Jacobian matrix is calculated as

$$\hat{J}(\eta, t) = \frac{\partial(E_1, E_2, E_3)}{\partial(\eta, \gamma_1, \gamma_2)} \Big|_{(\eta, \gamma_1, \gamma_2, t) = (\eta, 0, 0, t)}$$

$$= \begin{pmatrix} -\frac{\theta_0(t)}{\eta\{\eta + \theta_0(t) - \theta_0(t)\eta\}} & \hat{\sigma}_1^2(t) & 0 \\ -\frac{1}{\eta} & 0 & \hat{\sigma}_2^2(t) \\ 0 & \hat{p}_1 \frac{\theta_0(t)}{\eta + \theta_0(t) - \theta_0(t)\eta} & \hat{p}_2 \end{pmatrix},$$

where  $\hat{p}_i = n_i/n$ ,  $i = 1, 2$ .

Therefore, denoting  $\gamma_i^* = \gamma_i(\eta^*, t)$ ,  $i = 1, 2$ , by the Taylor expansion it is easy to see that

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} E_1(\eta_0, 0, 0, t) \\ E_2(\eta_0, 0, 0, t) \\ E_3(\eta_0, 0, 0, t) \end{pmatrix} + \hat{J}(\eta_0, t) \begin{pmatrix} \eta^* - \eta_0 \\ \gamma_1^* \\ \gamma_2^* \end{pmatrix} + O_p(\varepsilon_n^2).$$

Here we replace  $D_0(t) + \eta$  in (4.13) of Shen and He (2006) by  $\frac{\eta\{\eta + \theta_0(t) - \theta_0(t)\eta\}}{\theta_0(t)}$ , then using (2.11), we can show that

$$\hat{J}(\eta, t) \xrightarrow{p} J(\eta, t) = \begin{pmatrix} -\frac{\theta_0(t)}{\eta\{\eta + \theta_0(t) - \theta_0(t)\eta\}} & \sigma_1^2(t) & 0 \\ -\frac{1}{\eta} & 0 & \sigma_2^2(t) \\ 0 & p_1 \frac{\theta_0(t)}{\eta + \theta_0(t) - \theta_0(t)\eta} & p_2 \end{pmatrix}, \tag{A.17}$$

as  $n \rightarrow \infty$ .

By Lemma 1, similar to the derivation of (4.15) of Shen and He (2006), we get

$$\frac{\lambda_1^2(\eta^*, t)}{n_1} = \{\sqrt{n_1}\gamma_1(\eta^*, t)\}^2 \xrightarrow{d} \frac{p_1 p_2^2}{\det(J(\eta_0, t))^2 \eta_0^2 (1 - \eta_0)^2} \left[ \frac{\{1 - S_2(t)\}M_1(\sigma_1^2(t))}{\sqrt{p_1}} + \frac{\{1 - S_1(t)\}M_2(\sigma_2^2(t))}{\sqrt{p_2}} \right]^2. \tag{A.18}$$

From (2.5),

$$\frac{\lambda_1 \theta_0(t)}{\eta^* \{\eta^* + \theta_0(t) - \eta^* \theta_0(t)\}} + \frac{\lambda_2}{\eta^*} = 0.$$

Note that  $A_j(\eta_0, t) = \log S_j(t) - \log \hat{S}_j(t) = o(\varepsilon_n)$  (Csörgő and Horváth, 1983). Because (A.16) holds uniformly, we can show

$$\begin{aligned} -2 \log \mathcal{R}(\theta_0(t), \eta^*, t) &= \sum_{j=1}^2 \frac{\lambda_j^2(\eta^*, t)}{n_j} \hat{\sigma}_j^2(t) + o_p(1) \\ &= \frac{\lambda_1(\eta^*, t)}{n_1} \left[ \hat{\sigma}_1^2(t) + \frac{\hat{p}_1}{\hat{p}_2} \frac{\theta_0^2(t)}{\{\eta^* + \theta_0(t) - \eta^* \theta_0(t)\}^2} \hat{\sigma}_2^2(t) \right] + o_p(1) \\ &= \frac{\lambda_1^2(\eta^*, t)}{n_1} \frac{\det(\hat{J}(\eta^*, t))(\eta^*)^2}{\hat{p}_2} + o_p(1). \end{aligned}$$

From Lemma 1, and (4.14) and (4.15) of Shen and He (2006), it is easy to see that

$$\begin{aligned} -2 \log \mathcal{R}(\theta_0, \eta^*, t) &\xrightarrow{d} \frac{p_1 p_2}{\det(J(\eta_0, t)) \eta_0 (1 - \eta_0)^2} M^2(\sigma^2(t)) \\ &= \frac{1}{\sigma^2(t)} \left[ \frac{\{1 - S_2(t)\}M_1(\sigma_1^2(t))}{\sqrt{p_1}} + \frac{\{1 - S_1(t)\}M_2(\sigma_2^2(t))}{\sqrt{p_2}} \right]^2. \quad \square \end{aligned}$$

### References

Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N., 1993. *Statistical Models Based on Counting Processes*. Springer, New York.  
 Bland, J., Altman, D., 2000. The odds ratio. *Br. Med. J.* 320, 1468.  
 Csörgő, S., Horváth, P., 1983. The rate of strong uniform consistency for the product-limit estimator. *Probab. Theory Related Fields* 62, 411–426.



- Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D., Kafonek, D., Brandenburg, S., Wu, Y., He, X., Powe, N., Feinberg, A., 2003. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 299, 1753–1755.
- Dong, B., Matthews, D.E., 2012. Empirical likelihood for cumulative hazard ratio estimation with covariate adjustment. *Biometrics* 68 (2), 408–418.
- Fleming, T.R., Harrington, D.P., 1991. *Counting Processes and Survival Analysis*. Wiley, New York.
- Hassan, E., 2006. Recall bias can be a threat to retrospective and prospective research designs. *Int. J. Epidemiol.* 3 (2).
- Hasselman, B., 2015. Solve systems of nonlinear equations. <http://cran.r-project.org/web/packages/nleqslv/nleqslv.pdf>.
- Li, G., 1995. On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. Probab. Lett.* 25, 95–104.
- Li, G., Li, R., Zhou, M., 2005. Empirical likelihood in survival analysis. In: Fan, J., Li, G. (Eds.), *Contemporary Multivariate Analysis and Design of Experiments*. World Scientific, Singapore.
- McKeague, I.W., Zhao, Y., 2002. Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statist. Probab. Lett.* 60, 405–415.
- McKeague, I.W., Zhao, Y., 2005. Comparing distribution functions via empirical likelihood. *Int. Biostat.* 1–20, article 5.
- McKeague, I.W., Zhao, Y., 2006. Width-scaled confidence bands for survival functions. *Statist. Probab. Lett.* 76, 327–339.
- McPhail, S., Haines, T., 2010. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. *Health. Qual. Life. Outcomes.* 8, 65.
- Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.
- Owen, A.B., 1990. Empirical likelihood and confidence regions. *Ann. Statist.* 18, 90–120.
- Parr, C.L., Hjartaker, A., Laake, P., Lund, E., Veierød, M.B., 2009. Recall bias in melanoma risk factors and measurement error effects: a nested case-control study within the Norwegian Women and Cancer Study. *Am. J. Epidemiol.* 169, 257–266.
- Pepe, M., Janes, H., Longton, G., Leisenring, W., Newcomb, P., 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am. J. Epidemiol.* 159, 882–890.
- Rudin, W., 1976. *Principles of Mathematical Analysis*, third ed.. McGraw-Hill, Asia.
- Shen, J., He, S., 2006. Empirical likelihood for the difference of two survival functions under right censorship. *Statist. Probab. Lett.* 76, 169–181.
- Thomas, D.R., Grunkemeier, G.L., 1975. CI estimation for survival probabilities for censored data. *J. Amer. Statist. Assoc.* 70, 865–871.
- van der Vaart, A.W., 2000. *Asymptotic Statistics*. Cambridge University Press.
- Vrijheid, M., Armstrong, B.K., Bédard, D., Brown, J., Deltour, I., Iavarone, I., Krewski, D., Lagorio, S., Moore, S., Richardson, L., Giles, G.G., McBride, M., Parent, M.E., Siemiatycki, J., Cardis, E., 2009. Recall bias in the assessment of exposure to mobile phones. *J. Expo. Sci. Environ. Epidemiol.* 19, 369–381.
- Zhang, R., Brennan, M.-L., Fu, X., Aviles, R., Pearce, G., Penn, M., Topol, E., Sprecher, D., Hazen, S., 2001. Association between myeloperoxidase levels and risk of coronary artery disease. *J. Amer. Med. Assoc.* 286, 2136–2412.