Shahid Bahonar University of Kerman
Journal of Mahani Mathematical Research
Print ISSN: 2251-7952 Online ISSN: 2645-4505

ADJUSTED EMPIRICAL LIKELIHOOD ANALYSIS OF RESTRICTED MEAN SURVIVAL TIME FOR LENGTH-BIASED DATA

Z. Mohammadian and A. Habibirad ✉

Article type: Research Article

(Received: 01 December 2023, Received in revised form 29 January 2024)

(Accepted: 28 April 2024, Published Online: 01 May 2024)

Abstract. The Restricted Mean Survival Time (RMST) serves as a valuable and extensively utilized metric in clinical trials. However, its application becomes intricate when dealing with data affected by length-biased sampling, rendering traditional inference strategies inadequate. To overcome this challenge, we advocate for the adoption of nonparametric techniques. One notably promising approach is the Empirical Likelihood (EL) method, which furnishes robust results without the need for stringent parametric assumptions. In practical scenarios, the underlying sampling distributions often remain elusive, necessitating adjustments in the case of parametric methodologies. The EL method has demonstrated its efficacy in addressing such complexities. Consequently, this paper introduces the EL method for computing RMST in situations involving both length-biased and right-censored data. Additionally, we introduce the concept of adjusted empirical likelihood (AEL) to further enhance the coverage probability, particularly when dealing with smaller sample sizes. To gauge the performance of the EL and AEL methods, we conduct simulations and rigorously compare their results. The findings unequivocally demonstrate that AEL-based confidence intervals consistently provide superior coverage probability when juxtaposed with EL-based intervals. Lastly, we substantiate the practical applicability of our proposed method by employing it in the analysis of a real dataset.

Keywords: Adjusted empirical likelihood, Empirical likelihood, Restricted mean survival time, Non-parametric, Length-biased data.
2020 MSC: 62G05, 62G20, 62N02

1. Introduction

For studies with event time endpoints, the survival function contains all information about the temporal and stochastic profile of this target variable. However, the survival probability at a particular point in time (such as t) does not transparently capture the time course from that endpoint to t. Another procedure is to summarize the profile using the RMST at time t. A useful alternative in survival analysis is the RMST an interesting and applied function.

✉ ahabibi@um.ac.ir, ORCID: 0000-0001-5402-1753
https://doi.org/10.22103/jmmr.2024.22601.1544

Publisher: Shahid Bahonar University of Kerman
How to cite: Z. Mohammadian, A. Habibirad, *Adjusted empirical likelihood analysis of restricted mean survival time for length-biased data*, J. Mahani Math. Res. 2024; 13(2): 303 - 324.

In clinical studies with time-to-event outcomes, the RMST is introduced as a useful summary measurement. When proportional hazards assumption cannot be constructed or the event rate is low, RMST has received a lot of attention due to its transparent and intuitive interpretation. RMST is a useful measure when analyzing survival data and has practical applications in assessing treatment effectiveness and making informed decisions. In a clinical trial for a new cancer treatment, researchers want to compare two different therapies for advanced-stage cancer patients. The study follows patients for a fixed period, say 5 years, and records the time until death or disease progression. By calculating the RMST for each treatment group, researchers can determine the average survival time over the 5-year period and compare which therapy provides a longer and more sustainable survival benefit.

Let X be a continuous random variable with the cumulative distribution function (cdf) F. The RMST is defined as:

$$\mu_t = \int_0^t S(x)dx, \tag{1}$$

where $S(x) = 1 - F(x)$ is the survival function at time x. The empirical estimator of μ_t is

$$\mu_n(t) = \int_0^t S_n(x)dx, \tag{2}$$

where $S_n(x) = 1 - F_n(x)$, and $F_n(\cdot)$ is an empirical distribution function based on a random sample $X_1, X_2, \ldots, X_n$ of F.

There are different methods for the analysis of RMST, including techniques based on pseudo-observations which are introduced by [4]. The area under the survival curve up to t is considered the RMST. Due to the advantages of using such a quantification, [37] proposed a curve based on the RMST over time as an alternative summary to the survival function. [35] present methods for estimating Restricted Mean Survival Time (RMST) in survival analysis, focusing on the Kaplan-Meier method, Cox Proportional Hazards (PH) model, flexible parametric model, and a pseudo-observation method. Comparisons in simulated scenarios indicate that the Kaplan-Meier method is simple and fast but lacks covariate adjustment. The unstratified Cox model suits proportional hazards, while the stratified Cox model is effective for non-proportional hazards. The flexible parametric model performs similarly to the Cox model but is more time-consuming. Pseudo-observation methods are computationally efficient but may perform worse in specific scenarios when estimating RMST differences for subjects with given characteristics. [12] presented the definition and statistical properties of the RMST, adjusted analytical methods, sample size computation, information fraction for the RMST difference, and clinical and statistical meaning and commentary. Regarding the missingness of exact observations, nonparametric analysis is associated with problems, accordingly, [36] proposed a model-free measure for the interval-censored RMST employing the linear smoothing technique. The adjusted RMST (ARMST) for covariate effects using adjusted Kaplan–Meier curves and the Kaplan–Meier Estimator was introduced by [40]. The ARMST method combines RMST and AKME concepts. Regression-based methods for covariate adjustments are compared via simulation studies. They extend RMST and ARMST to settings with competing risks, presenting estimates like restricted mean time lost (RMTL) and adjusted RMTL (ARMTL).

Survival data analysis is a vital aspect in various fields, but it becomes more complex when dealing with length-biased sampling. This phenomenon occurs when the survival times are left-truncated, and the underlying truncation time random variable follows a uniform distribution. Length-biased sampling introduces bias in the observed survival data, favoring longer survival times. This bias arises because individuals with longer survival times have a greater probability of being included in the sample compared to those with shorter survival times. As a result, the collected data may not accurately represent the true underlying distribution of survival times. Let's consider the random variable X, which comprises independent and identically distributed (*i.i.d.*) positive random variables. The cdf of these variables is denoted by F. When a nonnegative random variable Y is observed with probability proportional to its length, it has the length-biased cdf which is defined by

$$G(y) = \mu^{-1} \int_0^y x dF(x), \quad y \geq 0, \tag{3}$$

where

$$\mu = \int_0^\infty x dF(x).$$

According to (3), the distribution function F is obtained as follows:

$$F(t) = \mu \int_0^t y^{-1} dG(y), \quad t \geq 0. \tag{4}$$

[5] introduced the Non-Parametric Maximum Likelihood Estimator (NPMLE) for the length-biased survival function from right-censored data. Their study demonstrated the NPMLE's properties of uniform strong consistency, weak convergence to a Gaussian process, and asymptotic efficiency. [15] presented a valuable nonparametric estimator that effectively incorporates information about the length-biased and right-censored sampling schedule. [27] introduce expectation-maximization (EM) algorithms for estimating infinite-dimensional parameters in three scenarios for length-biased data: estimating the nonparametric distribution function, estimating the nonparametric hazard function under an increasing failure rate constraint, and jointly estimating the baseline

hazards function and the covariate coefficients under the Cox proportional hazards model. Comprehensive empirical simulation studies demonstrate the effective performance of maximum likelihood estimators, particularly with moderate sample sizes, resulting in more efficient estimates compared to estimating equation approaches. [9] made significant advancements by extending the martingale estimating equation method and the pseudo-partial likelihood approach, enabling the handling of semiparametric transformation models with length-biased and right-censored data. [31] proposed an innovative partially linear transformation model, offering a robust approach to address length-biased and right-censored data. Their method involves an iterative computational algorithm and a bootstrap resampling approach for enhanced accuracy. [29] introduced a method for quantile function estimation under length-biased and right-censored data, leading to the derivation of a nonparametric estimator. Recognizing the limitations of parametric methods for estimating the RMST under length-biased sampling, [20] introduced novel nonparametric and semiparametric regression methods explicitly tailored to address the challenges of length-biased and right-censored data. These methods have proven to be highly effective in RMST estimation in such scenarios. [14] proposed estimators for RMST under length-biased and right-censored data, employing both nonparametric and semiparametric approaches. It enhances estimation efficiency by considering the similarity in distribution between truncation time and residual time.

The EL approach is a powerful tool for constructing confidence regions in nonparametric problems, offering distinct advantages over competing methods like NA and Bootstrap. As a result, numerous researchers have extensively explored and studied the EL method. [24] introduced the empirical likelihood ratio as a compelling alternative to the bootstrap for constructing confidence regions in nonparametric issues. [43] introduced a semiparametric inference approach for comparing means and survival probabilities in the presence of right-censored data. Their work involved deriving confidence intervals based on the EL principle. [18] endeavors to demonstrate the enhancement of current estimators for error distribution in nonparametric regression models through the incorporation of additional information using the empirical likelihood method. They establish the weak convergence of the resultant improved estimator to a Gaussian process. The assessment of its performance involves a comparison of asymptotic mean squared errors and a comprehensive simulation study. [17] addressed estimation and test problems for semi-parametric two-sample density ratio models, employing the EL approach in their analysis. [11] establishes the asymptotic normality of indirect inference estimators, specifically simulation-based minimum distance estimators, within a parametric model. These estimators rely on auxiliary nonparametric maximum likelihood density estimators. Furthermore, when the parametric model is correctly specified, it is demonstrated that the asymptotic variance-covariance matrix equals the inverse of the Fisher-information matrix. In the context of length-biased sampling, [23] proposed EL-based confidence intervals and thoroughly examined their behavior for both large and small samples. [32] focused on comparing the difference of quantiles in two independent samples and constructed confidence intervals using a smoothed EL approach. Additionally, they established the limiting distribution of the empirical log-likelihood ratio as a chi-squared distribution. [28] devised a novel EM algorithm for doubly censored data and constructed confidence regions for one- or two-sample analysis of doubly censored data using the EL method. [21] developed the EL method for inferring the mean residual life (MRL) of naturally recorded items based on independent and identically distributed (iid) observations from the true distribution. However, under right censorship, the EL-based log-likelihood ratio follows a scaled chi-square distribution, leading to lower confidence interval coverage when estimating the scale parameter. To address this limitation, they devised an algorithm to directly calculate the likelihood ratio (LR) and demonstrated that the corresponding log-likelihood ratio converges to the standard chi-square distribution, resulting in improved confidence interval coverage. [3] introduced the EL for an accelerated failure time model with length-biased data. They show that the asymptotic distribution of the empirical log-likelihood ratio statistic is a weighted sum of independent chi-square distributions. They also explore the adjusted EL approach from both theoretical and practical perspectives.

In survival analysis, overlooking length-biased sampling plans can lead to significant overestimation. Conversely, previous investigations faced challenges in estimating the margin of error of commonly used summary statistics for such sampling plans. To address this gap, we adapt the EL confidence intervals to the RMST for length-biased data. While [42] explored confidence intervals for RMST and differences/ratios of two RMSTs using the EL method, the RMST inference with length-biased data has not been extensively studied through the EL method. Therefore, our aim is to obtain confidence intervals for the RMST under length-biased data, making a valuable contribution to further exploration in this crucial research area.

In situations where the dimension of the estimating function is high or the sample size is small, the EL confidence regions often exhibit coverage probabilities lower than the nominal value. To address this issue and achieve more accurate coverage probabilities, researchers have developed the AEL approach. [8] proposed an iterative set of rules for AEL that ensures convergence. They demonstrated that AEL is computationally faster than EL, and the confidence regions constructed via AEL exhibit closer-to-nominal coverage probabilities. Utilizing the AEL approach, [22] constructed confidence regions for the difference of two d-dimensional population means in the context of two-sample populations and established that the approach is Bartlett correctable. [38] applied the AEL method for estimating the cumulative baseline hazard function and constructed confidence regions for the vector of regression parameters. [7] delved into the finite-sample properties of AEL, revealing that the AEL confidence region achieves higher coverage probabilities as the level of adjustment increases. Additionally, they demonstrated that the AEL ratio function increases as the population mean deviates from the sample mean. In the analysis of right-censored data, [39] employed the influence function method via the AEL approach, demonstrating that the adjusted log-likelihood ratio follows an asymptotically Chi-squared distribution. [34] explored semi-parametric transformation models with length-biased sampling using both EL and AEL methods. [2] suggested comparing two means and constructing a confidence interval for their difference with right-censored data using the EL method. To avoid the estimation of the scale parameter in constructing confidence intervals, they considered an EL based on the iid representation of Kaplan-Meier weights involved in the empirical likelihood ratio. Additionally, they applied the AEL method and the mean empirical likelihood approach. Despite these significant advancements, the inference of the RMST with length-biased data using the AEL method remains relatively unexplored. In this study, we aim to bridge this gap by constructing a confidence interval for RMST under length-biased data through the AEL approach.

The structure of this paper is organized as follows. In Section 2, we present the EL and AEL methods for length-biased and right-censored data. Additionally, we introduce the confidence interval for the proposed methods. In Section 3, we present the numerical results obtained from our analyses. Finally, in Section 4, we provide the proof of the theorems presented in this paper.

2. Methodology and main result

In this section, we will cover the introduction of two essential methods for analyzing length-biased and right-censored data: the EL and AEL methods. We will also demonstrate the construction of confidence intervals for μ_t. Furthermore, we will explore how these methods can be applied to address the RMST difference in a two-sample problem involving length-biased and right-censored data.

2.1. **Empirical likelihood.** Suppose $Y_1, \ldots, Y_n$ be *i.i.d.* positive random variables with common cdf G, and $C_1, \ldots, C_n$ be *i.i.d.* positive random variables with a common cdf V. These two sets of random variables are assumed to be independent of each other. The observations in this random censoring model can be described as follows:

$$Z_i = \min(Y_i, C_i), \qquad \delta_i = I(Y_i \leq C_i), \qquad i = 1, \cdots, n,$$

We employed the indicator function I(A) to represent set A. The distribution H of Z satisfies $1 - H = (1 - G)(1 - V)$. Let $\tau_G = \inf\{t : G(t) = 1\}$ and $\tau_V = \inf\{t : V(t) = 1\}$. Suppose $Z_{1:n} \leq Z_{2:n} \leq \cdots \leq Z_{n:n}$ be the ordered Z-values and $\delta_{[i:n]}$ be the concomitant of the ith order statistic, that is $\delta_{[i:n]} = \delta_i$ if $Z_{i:n} = Z_i$.

Under the setting of length-biased sampling, and considering Equations (1), we have

$$\mu_t = \mu \int_0^\infty \Big(I(u \le t) + \frac{t}{u} I(u > t)\Big) dG(u),$$

where $\mu^{-1} = \int_0^\infty \frac{1}{u} dG(u)$. Therefore, we get

$$\mu_t \int_0^\infty \frac{1}{u} dG(u) = \int_0^\infty \Big(I(u \le t) + \frac{t}{u} I(u > t)\Big) dG(u). \tag{5}$$

It can be easily observed

$$E\Big(D(\mu_t)\Big) = 0, \tag{6}$$

where $D(\mu_t) = \frac{\mu_t - YI(Y \le t) - tI(Y>t)}{Y}$. In the random censorship model, it is obvious that

$$E\Big(\frac{\Big(\mu_t - ZI(Z \le t) - tI(Z > t)\Big)\delta}{Z}\Big) = 0, \tag{7}$$

(see [30]) so, the proposed estimating equation becomes

$$\frac{1}{n}\sum_{i=1}^n \left[\frac{\Big(\mu_t - Z_i I(Z_i \le t) - tI(Z_i > t)\Big)\delta_i}{Z_i\Big(1 - V(Z_i)\Big)}\right] = 0.$$

For $1 \le i \le n$, we define

$$D_i(Z, \mu_t) = \frac{\Big(\mu_t - Z_i I(Z_i \le t) - tI(Z_i > t)\Big)\delta_i}{Z_i\Big(1 - V(Z_i)\Big)}.$$

Since V is unknown, we replace it by its Kaplan-Meier estimator which is defined by

$$V_n(t) = 1 - \prod_{i=1}^n \left[\frac{n-i}{n-i+1}\right]^{I(Z_{i:n} \le t, \delta_{[i:n]}=0)},$$

so, we have

$$D_{ni}(Z, \mu_t) = \frac{\Big(\mu_t - Z_i I(Z_i \le t) - tI(Z_i > t)\Big)\delta_i}{Z_i\Big(1 - V_n(Z_i)\Big)}.$$

By replacing $(1 - V_n(Z_i))^{-1}$ in $D_i(Z, \mu_t)$ by the $i.i.d.$ representation from [13] (see Lemma 3.1) and utilizing the counting process notation,

$$\begin{aligned} W_i^{'}(\mu_t) &= \frac{D(Z_i, \mu_t)\delta_i}{1 - V(Z_i)} + \int_0^\infty \frac{\int_s^\infty D(s, \mu_t) dG(s) \overline{\delta_i}}{\overline{H}(x)} d(x) \\ &- \int_0^\infty \frac{\int_s^\infty D(s, \mu_t) dG(s) I(Z_i \ge x)}{\overline{H}(x)} d\nu(x) \end{aligned}$$

where $\overline{\delta}_i = 1 - \delta_i$, $H(x) = E\Big(I(Z_i \leq x\Big)$, $\overline{H}(x) = 1 - H(x)$, and $\nu(x) = -\log(1 - V(x))$. Estimating unknown distribution functions, we can define

$$\begin{aligned} W_{ni}^{'}(\mu_t) &= \frac{D(Z_i,\mu_t)\delta_i}{1-V_n(Z_i)} + \int_0^\infty \frac{\int_s^\infty D(s,\mu_t)dG_n(s)\overline{\delta_i}}{\overline{H}_n(x)} d(x) \\ &- \int_0^\infty \frac{\int_s^\infty D(s,\mu_t)dG_n(s)I(Z_i \geq x)}{\overline{H}_n(x)} d\nu_n(x), \end{aligned}$$

where $G_n(t) = 1 - \prod_{i=1}^n \left[\frac{n-i}{n-i+1}\right]^{I_{Z_{i:n}\leq t, \delta_{[i:n]}=1}}$, $H_n(t) = \frac{1}{n}\sum_{i=1}^n (Z_i \leq x)$ and $\nu_n = -\log(1 - V_n(x))$.

Based on W'_{ni}, we define the estimated EL ratio at the value μ_t as follows

$$L(\mu_t) = \sup\{\Pi_{i=1}^n np_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_{ni}^{'} = 0, p_i \geq 0\}. \tag{8}$$

By the Lagrange multiplier, we have

$$p_i = \frac{1}{n}\{1 + \eta W_{ni}^{'}(\mu_t)\}^{-1}, \qquad i = 1, \cdots, n, \tag{9}$$

where η is the solution of

$$\frac{1}{n}\sum_{i=1}^n \frac{W'_{ni}}{1 + \eta W'_{ni}} = 0. \tag{10}$$

Note that $\Pi_{i=1}^n p_i$, subject to $\sum_{i=1}^n p_i = 1$, attains its maximum n^{-n} at $p_i = n^{-1}$. The corresponding empirical log-likelihood ratio is defined as

$$l(\mu_t) = -2\log L(\mu_t) = 2\sum_{i=1}^n \log\{1 + \eta W'_{ni}\}. \tag{11}$$

Theorem 2.1. *Assume that the regularity conditions in the Proof hold. Then*

$$l(\mu_t) \xrightarrow{\mathcal{D}} \chi_1^2, \qquad n \longrightarrow \infty, \tag{12}$$

where χ_1^2 is a standard chi-squared random variable with one degree of freedom.

Therefore, utilizing Theorem 2.1, an asymptotic $100(1-\alpha)\%$ EL confidence interval for μ_t is

$$\mathcal{N}_1 = \{\mu_t \quad : \quad l(\mu_t) \leq \chi_{1,\alpha}^2\},$$

where $\chi_{1,\alpha}^2$ is the upper α-quantile of the distribution of χ_1^2.

2.2. **Adjusted empirical likelihood.** The AEL method, first introduced by [8], serves a critical purpose in ensuring the well-defined nature of the EL ratio even when an observation is added to the dataset. This feature is particularly advantageous in addressing the issue of under-coverage that may arise in the EL method, especially for small sample sizes. The AEL approach involves incorporating a pseudo-sample as an essential component of the data, allowing for improved accuracy and reliability in statistical inference. By utilizing

AEL, researchers can enhance the robustness of their analyses and obtain more accurate results when dealing with complex or limited datasets. We have

$$W'_{nn+1} = -\frac{a_n}{n}\sum_{i=1}^{n} W'_{ni},$$

where $a_n = \max(1, \log(n)/2)$. Based on the $n+1$ observations, the adjusted empirical log-likelihood ratio is given by

$$L^A(\mu_t) = \sup\{\Pi_{i=1}^{n+1} np_i : \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i W'_{ni}(\mu_t) = 0, p_i \geq 0\}. \tag{13}$$

Utilizing by the Lagrange multiplier, we can define that

$$l^A(\mu_t) = -2\log L^A(\mu_t) = 2\sum_{i=1}^{n+1} \log\{1 + \eta^A W'_{ni}\}, \tag{14}$$

where η^A is a solution of the equation

$$\frac{1}{n+1}\sum_{i=1}^{n+1} \frac{W'_{ni}}{1 + \eta^A W'_{ni}(\mu_t)} = 0. \tag{15}$$

Theorem 2.2. *Suppose that the regularity conditions in the Proofs hold. Therefore*

$$l^A(\mu_t) \xrightarrow{\mathcal{D}} \chi^2_1, \qquad n \longrightarrow \infty. \tag{16}$$

So, by utilizing Theorem 2.2, an asymptotic $100(1-\alpha)\%$ AEL confidence interval for μ_t is given by

$$\mathcal{A}_1 = \{\mu_t \quad : \quad l^A(\mu_t) \leq \chi^2_1(\alpha)\}.$$

3. Simulation

In this section, we present the findings of a comparative study between the EL method and the AEL method. Our investigation focuses on two essential criteria: the average length of confidence intervals (Δ) and the coverage probability (CP).

To conduct this study, we present simulations and real data. These simulations include data generated from gamma distributions, as well as real data obtained from the Channing House dataset. Throughout our analysis, we examine two significant levels, namely $\alpha = 0.10$ and $\alpha = 0.05$, with three different sample sizes: $n = 50$, $n = 100$, and $n = 150$. To evaluate the efficacy of the methods, we showcase simulations involving Gamma and Exponential distributions. The results of these simulations are presented in Tables 1 and 2, showcasing the performance of both EL and AEL methods. For the purpose of calculating average values, we perform 5000 iterations to ensure robustness in our results.

Table 1. 95% coverage probabilities and average lengths of confidence intervals for RMST of Exp distribution

			$1-\alpha=0.95$			
censored	n	Time	Δ.EL	c.p.EL	Δ.AEL	c.p.AEL
10%	50	2	0.606	0.779	0.637	0.801
		2.5	0.686	0.794	0.726	0.815
		3	0.731	0.809	0.770	0.812
		3.5	0.778	0.810	0.796	0.822
		4	0.791	0.811	0.808	0.813
	100	2	0.545	0.802	0.560	0.805
		2.5	0.605	0.806	0.612	0.816
		3	0.644	0.819	0.665	0.824
		3.5	0.680	0.821	0.697	0.834
		4	0.684	0.816	0.715	0.828
	150	2	0.511	0.813	0.523	0.816
		2.5	0.573	0.836	0.579	0.843
		3	0.611	0.838	0.620	0.841
		3.5	0.639	0.828	0.640	0.838
		4	0.638	0.824	0.658	0.831
40%	50	2	0.571	0.765	0.606	0.787
		2.5	0.626	0.778	0.641	0.795
		3	0.637	0.748	0.673	0.752
		3.5	0.654	0.737	0.679	0.748
		4	0.667	0.719	0.688	0.728
	100	2	0.511	0.791	0.531	0.804
		2.5	0.543	0.789	0.562	0.805
		3	0.564	0.781	0.583	0.788
		3.5	0.573	0.765	0.591	0.774
		4	0.589	0.755	0.596	0.762
	150	2	0.476	0.801	0.489	0.809
		2.5	0.511	0.798	0.522	0.813
		3	0.533	0.803	0.537	0.805
		3.5	0.547	0.794	0.555	0.810
		4	0.541	0.788	0.556	0.800

The dataset, consisting of n observations, denoted as $Y_1, \dots, Y_n$, is generated from the $Exp(2)$ and $Gamma(4,1)$ distributions. Additionally, we introduce censoring time observations, $C_1, \dots, C_n$, which are drawn from the $Gamma(2, \lambda_1)$ and $Gamma(5, \lambda_2)$ distribution. Here, the constants λ_1 and λ_2 play a crucial role in determining the proportion of censoring. To explore the impact of censoring, we examine two different values of λ resulting in 10% and 40% censoring proportions, respectively. The simulation results are summarized in Tables 1 and 2, 3 and 4 providing valuable insights into the performance of the analysis. In particular, we calculate the coverage probability and the average length of confidence intervals at $t = 2, 2.5, 3, 3.5, 4$, while varying the values of λ to achieve the desired 10% and 40% censoring proportions. These results shed light on the behaviour of the analysis under different scenarios, aiding in the understanding of its reliability and effectiveness.

The simulation results are summarised in Tables 1 and 2 for Exp distribution, Tables 3 and 4 for Gamma distribution. Based on the tables we can make the following conclusions:

TABLE 2. 90% coverage probabilities and average lengths of confidence intervals for RMST of Exp

			$1-\alpha=0.9$			
censored	n	Time	Δ.EL	c.p.EL	Δ.AEL	c.p.AEL
10%	50	2	0.513	0.710	0.546	0.718
		2.5	0.578	0.713	0.603	0.725
		3	0.624	0.718	0.647	0.732
		3.5	0.649	0.711	0.677	0.745
		4	0.667	0.724	0.686	0.749
	100	2	0.465	0.721	0.485	0.735
		2.5	0.517	0.728	0.525	0.739
		3	0.550	0.748	0.565	0.753
		3.5	0.574	0.739	0.589	0.758
		4	0.583	0.745	0.592	0.771
	150	2	0.437	0.725	0.449	0.738
		2.5	0.485	0.732	0.493	0.749
		3	0.518	0.765	0.529	0.771
		3.5	0.530	0.751	0.541	0.763
		4	0.546	0.755	0.552	0.781
40%	50	2	0.487	0.708	0.505	0.715
		2.5	0.525	0.696	0.546	0.704
		3	0.547	0.652	0.568	0.661
		3.5	0.553	0.638	0.572	0.667
		4	0.560	0.629	0.585	0.634
	100	2	0.429	0.718	0.446	0.729
		2.5	0.462	0.701	0.474	0.705
		3	0.477	0.682	0.499	0.704
		3.5	0.488	0.646	0.498	0.675
		4	0.489	0.654	0.508	0.669
	150	2	0.404	0.721	0.414	0.732
		2.5	0.428	0.717	0.439	0.733
		3	0.439	0.725	0.455	0.732
		3.5	0.452	0.684	0.465	0.721
		4	0.458	0.679	0.467	0.724

- It is obvious in all of tables when the sample increases in size, the CP increases. As expected, Δ of two methods decreases when the sample size grows.
- In both tables with a fixed significant level, the CP under the AEL method is far better than the EL approach.
- For all censoring rates (10%, 40%), AEL-based confidence intervals perform better than those of the EL-based confidence intervals.
- In some situations (refer to Tables 2 and 4), both EL and AEL methods exhibit low coverage probabilities. However, AEL demonstrates improvement in these coverage probabilities compared to the EL method.

4. **Real data**

To demonstrate the RMST function in the context of length-biased and right-censored data, the Channing House dataset serves as a suitable example. A complete description of this data set can be found [19]. Collected between 1964 and July 1, 1975, this dataset comprises 97 men and 365 women who

TABLE 3. 95% coverage probabilities and average lengths of confidence intervals for RMST of Gamma

			$1-\alpha=0.95$			
censored	n	Time	Δ.EL	c.p.EL	Δ.AEL	c.p.AEL
10%	50	2	0.217	0.731	0.229	0.789
		2.5	0.367	0.797	0.393	0.825
		3	0.558	0.821	0.586	0.829
		3.5	0.756	0.823	0.774	0.841
		4	0.913	0.829	0.959	0.843
	100	2	0.192	0.763	0.193	0.795
		2.5	0.339	0.825	0.348	0.837
		3	0.518	0.823	0.532	0.832
		3.5	0.672	0.830	0.690	0.844
		4	0.819	0.837	0.855	0.848
	150	2	0.183	0.777	0.181	0.801
		2.5	0.320	0.836	0.405	0.842
		3	0.482	0.844	0.494	0.848
		3.5	0.644	0.832	0.648	0.855
		4	0.784	0.848	0.793	0.851
40%	50	2	0.245	0.668	0.312	0.689
		2.5	0.309	0.711	0.408	0.722
		3	0.475	0.780	0.488	0.783
		3.5	0.662	0.799	0.701	0.808
		4	0.837	0.818	0.852	0.835
	100	2	0.148	0.672	0.166	0.698
		2.5	0.285	0.747	0.298	0.781
		3	0.437	0.789	0.453	0.813
		3.5	0.600	0.819	0.616	0.835
		4	0.774	0.831	0.781	0.843
	150	2	0.156	0.708	0.155	0.715
		2.5	0.270	0.766	0.280	0.778
		3	0.420	0.796	0.426	0.829
		3.5	0.581	0.825	0.586	0.838
		4	0.722	0.842	0.732	0.850

passed through the Channing House during the given period. Their ages at entry and exit or death were recorded, with only those who lived beyond the examination time being observed. As a result, the entry age serves as the truncation variable. Though most subjects were still alive when the data was collected, many were censored. Only 130 women and 46 men died at the Channing House during the study period, making this a left-truncated and right-censored dataset. If the truncation variable is uniformly distributed, the left-truncated dataset is length-biased. However, while the entire dataset is not length-biased, a subset comprising only those whose entry ages were above 786 months (65.5 years) is length-biased due to the uniform distribution of the truncation variable. The subset comprises 448 people, with only 14 subjects not included. Of the 173 subjects who died during the study period in the subset, the censoring ratio is approximately 61.8%. For simplicity, we use the year as the time unit.

We notice that, as age increases, the RMST increases overall. The AEL confidence intervals for the RMST at selected ages are some wider than EL confidence intervals.

TABLE 4. 90% coverage probabilities and average lengths of confidence intervals for RMST of Gamma

$1-\alpha=0.90$						
censored	n	Time	Δ.EL	c.p.EL	Δ.AEL	c.p.AEL
10%	50	2	0.187	0.667	0.193	0.701
		2.5	0.312	0.707	0.329	0.732
		3	0.484	0.733	0.500	0.757
		3.5	0.645	0.721	0.676	0.743
		4	0.792	0.711	0.816	0.739
	100	2	0.169	0.668	0.165	0.711
		2.5	0.289	0.736	0.303	0.740
		3	0.437	0.762	0.447	0.769
		3.5	0.576	0.723	0.594	0.760
		4	0.699	0.727	0.729	0.746
	150	2	0.157	0.698	0.160	0.731
		2.5	0.272	0.744	0.284	0.765
		3	0.409	0.770	0.417	0.780
		3.5	0.558	0.729	0.555	0.773
		4	0.668	0.731	0.666	0.758
40%	50	2	0.148	0.612	0.159	0.630
		2.5	0.269	0.658	0.288	0.666
		3	0.405	0.712	0.455	0.721
		3.5	0.569	0.715	0.588	0.736
		4	0.726	0.709	0.802	0.731
	100	2	0.132	0.622	0.141	0.636
		2.5	0.250	0.686	0.252	0.690
		3	0.379	0.720	0.388	0.725
		3.5	0.516	0.722	0.526	0.742
		4	0.648	0.717	0.671	0.738
	150	2	0.130	0.630	0.146	0.641
		2.5	0.233	0.670	0.244	0.699
		3	0.357	0.725	0.368	0.729
		3.5	0.494	0.726	0.511	0.765
		4	0.616	0.729	0.617	0.742

TABLE 5. 95% confidence intervals for RMST at selected ages for the Channing house data.

		T=70	T=75	T=80	T=85	T=90
EL	Lower	69.450	72.227	73.460	73.766	73.797
	Upper	69.682	72.908	74.481	74.919	74.970
AEL	Lower	69.439	72.213	73.455	73.760	73.791
	Upper	69.687	72.912	74.487	74.926	74.976

Lastly, in Table 6, we compared different confidence intervals for women against men and attained the result that men have greater RMST than women, in other words, by considering at the same age, men tend to live longer than women.

Table 6. 95% confidence intervals for RMST at selected ages for men versus women for the Channing house data.

			T=70	T=75	T=80	T=85	T=90
Women	EL	Lower	69.386	72.044	73.143	73.395	73.420
		Upper	69.657	72.815	74.262	74.644	74.690
	AEL	Lower	69.385	72.039	73.136	73.387	73.412
		Upper	69.658	72.819	74.269	74.052	74.699
Men	EL	Lower	69.540	72.405	73.987	74.425	74.465
		Upper	69.903	73.805	76.340	77.221	77.327
	AEL	Lower	69.530	72.420	73.934	74.364	74.404
		Upper	69.910	73.833	76.393	77.285	77.398

5. Conclusion

This paper is devoted to the challenging realm of computing the RMST for data affected by length-biased sampling, an issue that often perplexes traditional inference strategies. We have advocated for the utilization of nonparametric techniques, as a robust and flexible approach for addressing this complexity without the need for stringent parametric assumptions. Our exploration extended to the introduction of the AEL method, a novel concept aimed at enhancing coverage probability, especially when dealing with smaller sample sizes. Through a comprehensive set of simulations, we have provided compelling evidence that AEL-based confidence intervals consistently outperform EL-based intervals, offering researchers a more reliable tool for statistical inference in such scenarios. To solidify the practical applicability of our proposed methods, we conducted an analysis of a real dataset, demonstrating their effectiveness in a real-world context. This reaffirms the utility of the EL and AEL methods as valuable tools for researchers and practitioners in clinical trials and other fields where RMST estimation is critical. In summary, our paper has introduced and validated the use of the Empirical Likelihood and adjusted empirical likelihood methods for computing RMST in the presence of length-biased and right-censored data. We hope that these methods will find wider adoption and facilitate more accurate and robust statistical analyses in the future.

Funding

This research received no funding.

COMPLIANCE WITH ETHICAL STANDARDS

In the pursuit of advancing knowledge and understanding in the field of statistics, it is essential to uphold the highest ethical standards. This research is conducted with utmost integrity, adhering to the principles of honesty, transparency, and objectivity. All data utilized in this study is treated with the utmost respect for privacy and confidentiality, and the analysis is performed diligently and without bias. The findings presented in this article are a result of rigorous and ethical scientific inquiry, aimed at contributing meaningfully to the academic community and society at large.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Both authors contributed significantly to this work. The first author and the second author jointly conceptualized the study and designed the methodology. The first author conducted formal analysis and contributed to data curation, while the second author played a key role in visualizing the findings, managing resources, and overseeing project administration. Both authors actively participated in writing, reviewing, and editing the manuscript and have read and approved the final version of the manuscript for publication.

Data Availability Statement

The datasets generated and analyzed during the current study are called Channing House Data and also can be obtained from [16] or from the corresponding author on reasonable request. All data supporting the findings of this study are openly available in the R package boot [10]. In the spirit of open science, we encourage readers to use and build upon the data we have shared. We believe that making data openly accessible not only promotes transparency but also allows for the advancement of research through collaboration and reproducibility. If you have any inquiries regarding the data used in this study, please contact the corresponding author at ahabibi@um.ac.ir.

Appendix: Proofs of Theorems

Let us assume the following regularity conditions:
(1) G and V are distributions,
(2) $\int_0^\infty \frac{1}{t^2(1-V(t))} dG(t) < \infty$,
(3) $P(C > \tau_G) > 0$.
Assumption 1 guarantees that the time variable is continuous. Assumption 2 makes sure that the variance of $W_i^{'}$ is finite. Assumption 3 expresses that the support of C covers the support of Y. Thus, one can estimate the RMST at any point.
The following lemmas will be needed for the proofs of the theorems.

Lemma 5.1. *Let the regularity conditions holds. So, we have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}^{'}(\mu_t) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\sigma^2 = Var(W_i^{'}(\mu_t))$.

Proof. Let $H(x) = p(Z \leq x)$ and $b_H = \sup\{x : H(x) < 1\}$. We get

$$\begin{aligned}
H_n^1(x) &= \frac{1}{n}\sum_{i=1}^{n} I(Z_i \leq x, \delta_i = 1), \\
H_n^0(x) &= \frac{1}{n}\sum_{i=1}^{n} I(Z_i \leq x, \delta_i = 0), \\
H_n(x) &= H_n^0(x) + H_n^1(x) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i \leq x).
\end{aligned}$$

We can also show Kaplan-Meier estimators with the following equations

$$G_n(x) = 1 - \Pi_{s \leq x}[1 - \frac{H_n^1(s)}{\overline{H}_n(s^-)}] \quad and \quad V_n(x) = 1 - \Pi_{s \leq x}[1 - \frac{H_n^0(s)}{\overline{H}_n(s^-)}], \tag{17}$$

and also we get

$$\overline{H}_n(x) = \overline{G}_n(x)\overline{V}_n(x), \tag{18}$$

where $\overline{H}_n(x) = 1 - H_n(x)$, $\overline{G}_n(x) = 1 - G_n(x)$ and $\overline{V}_n(x) = 1 - V_n(x)$. Applying (17) and (18), we have

$$dG_n(x) = \overline{G}_n(x^-)\frac{dH_n^1(x)}{\overline{G}_n(x^-)\overline{V}_n(x^-)},$$

Therefore, we get

$$dH_n^1(x) = \overline{V}_n(x^-)dG_n(x) \quad and \quad dH_n^0(x) = \overline{G}_n(x^-)dV_n(x).$$

Therefore

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} W_{ni}^{'}(\mu_t) &= \frac{1}{n}\sum_{i=1}^{n}\{\frac{D(Z_i, \mu_t)\delta_i}{1 - V_n(Z_i)} + \int_0^{\infty} \frac{\int_s^{\infty} D(s, \mu_t)dG_n(s)\overline{\delta_i}}{\overline{H}_n(x_i)}d(x) \\
&- \int_0^{\infty} \frac{\int_s^{\infty} D(s, \mu_t)dG_n(s)I(Z_i \geq x)}{\overline{H}_n(x_i)}d\nu_n(x)\} \\
&= \int \frac{D(s, \mu_t)}{\overline{V}_n(s^-)}dH_n^1(s) + \int \frac{\int_s^{\infty} D(s, \mu_t)dG_n(s)}{\overline{H}_n(s^-)}dH_n^0(s) \\
&+ \int \int_s^{\infty} D(s, \mu_t)dG_n(s)\frac{\overline{H}_n(s^-)}{\overline{H}_n^2(s^-)}dH_n^0(s) \\
&= \int \frac{D(s, \mu_t)}{\overline{V}_n(s^-)}dH_n^1(s) = \int D(s, \mu_t)dG_n(s) \\
&= \frac{1}{n}\sum_{i=1}^{n} U(\mu_t),
\end{aligned}$$

where $U(\mu_t)$ is asymptotically equivalent to $\sum_{i=1}^{n} W_i^{'}(\mu_t)$, where $W_i^{'}(\mu_t)$ are i.i.d. random variables with mean zero for $i = 1, \ldots, n$ in the sense that

$$U(\mu_t) = \sum_{i=1}^{n} W_i^{'}(\mu_t) + o_p(n^{1/2}),$$

(see [1]). Thus by proof of Lemma 5.1 in Appendix of Liang et al (2016), Lemma 5.1 is valid. □

Lemma 5.2. *Assume that the regularity conditions holds. Then, as* $n \to \infty$, *we have*

$$\frac{1}{n}\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t))^2 \xrightarrow{\mathcal{P}} \sigma^2,$$

where $\xrightarrow{p}$ *denotes convergence in probability.*

Proof. We have to demonstrate $\left|\frac{1}{n}\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t))^2 - \frac{1}{n}\sum_{i=1}^{n}(W_i^{'}(\mu_t))^2\right| = o_p(1)$.

$$\begin{aligned} &\left|\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t))^2 - \tfrac{1}{n}\sum_{i=1}^{n}(W_i^{'}(\mu_t))^2\right| \\ &= \left|\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t) - W_i^{'}(\mu_t))(W_{ni}^{'}(\mu_t) - W_i^{'}(\mu_t) + 2W_i^{'}(\mu_t))\right| \\ &\leq \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t) - W_i^{'}(\mu_t))^2 + \left|\tfrac{2}{n}\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t) - W_i^{'}(\mu_t))W_i^{'}(\mu_t)\right| \\ &= I_1 + I_2, \end{aligned}$$

We have

$$\begin{aligned} I_1 &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{(1-V_n(Z_i))} - \frac{1}{(1-V(Z_i))}\right)^2\left(\frac{(\mu_t - Z_iI(Z_i \leq t) - tI(Z_i \geq t))\delta_i}{Z_i}\right) \\ &\leq \sup_{u \leq Z(n)}\left|\frac{V_n(u) - V(u)}{(1-V_n(u))}\right|^2 \frac{1}{n}\sum_{i=1}^{n}\frac{\left((\mu_t - Z_iI(Z_i \leq t) - tI(Z_i \geq t))^2\delta_i\right)}{(Z_i(1-V(Z_i)))^2}, \end{aligned}$$

and

$$I_2 \leq \sup_{u \leq Z(n)}\left|\frac{V_n(u) - V(u)}{(1-V_n(u))}\right|\frac{1}{n}\sum_{i=1}^{n}\frac{\Big(\mu_t - Z_iI(Z_i \leq t) - tI(Z_i \geq t)\Big)^2\delta_i}{\Big(Z_i\Big(1-V(Z_i)\Big)\Big)^2}.$$

According to our assumption

$$\int \frac{(\mu_t - yI(y \leq t) - tI(y \geq t))^2}{y^2(1-V(y))}dG(y) < \infty,$$

and the following fact due to [41]

$$\sup_{u \leq Z(n)}\left|\frac{V_n(u) - V(u)}{(1-V_n(u))}\right| = O_p(1),$$

it follows that $I_i = O_p(1)$ for $i = 1, 2$. Hence $I = I_1 + I_2 = O_p(1)$.
from $W_{ni}^{'}(\mu_t) - W_i^{'}(\mu_t) = o_p(1)$, we have

$$\frac{1}{n}\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t))^2 = \frac{1}{n}\sum_{i=1}^{n}(W_i^{'}(\mu_t))^2 + o_p(1).$$

Thus, by the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t))^2 \xrightarrow{\mathcal{P}} \sigma^2.$$

□

Proof of Theorem 2.1. Let $t \in [0, \tau)$. According to Lemmas 5.1, 5.2 and Lemma 3 of [26], we have

$$\max_{1\le i\le n} |W_{ni}^{'}(\mu_t)| = o_p(n^{1/2}), \tag{19}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left|W_{ni}^{'}(\mu_t)\right|^3 = o_p(n^{1/2}). \tag{20}$$

Therefore, considering (19), (20) and applying the same argumentations used in [25], we have

$$\left|\eta\right| = O_p(n^{-1/2}). \tag{21}$$

Utilizing the Taylor expansion, it is obvious that

$$\begin{aligned} l(\mu_t) &= 2\sum_{i=1}^{n}\log(1+\eta W_{ni}^{'}(\mu_t)) \\ &= 2\sum_{i=1}^{n}\left(\eta W_{ni}^{'}(\mu_t) - \frac{(\eta W_{ni}^{'}(\mu_t))^2}{2}\right) + R_n(t), \end{aligned} \tag{22}$$

by using Equations (20) and (21), it can be seen that

$$\begin{aligned} |R_n(t)| &\le C\sum_{i=1}^{n}|\eta W_{ni}^{'}(\mu_t)|^3 \\ &\le C|\eta|^3\sum_{i=1}^{n}|W_{ni}^{'}(\mu_t)|^3 \\ &= o_p(1). \end{aligned} \tag{23}$$

Considering (22) and (10), the following result is obtained:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \frac{W_{ni}^{'}(\mu_t)}{1+\eta W_{ni}^{'}(\mu_t)} \\
&= \sum_{i=1}^{n} W_{ni}^{'}(\mu_t)\Big[1-\eta W_{ni}^{'}(\mu_t)+\frac{(\eta W_{ni}^{'}(\mu_t))^2}{1+\eta W_{ni}^{'}(\mu_t)}\Big] \\
&= \sum_{i=1}^{n} W_{ni}^{'}(\mu_t) - \Big(\sum_{i=1}^{n}(W_{ni}^{'}(\mu_t))^2\Big)\eta + \sum_{i=1}^{n}\frac{W_{ni}^{'}(\mu_t)(\eta W_{ni}^{'}(\mu_t))^2}{1+\eta W_{ni}^{'}(\mu_t)}. \qquad (24)
\end{aligned}$$

Considering (19) and (21) as well as the use of Lemma 5.1, we conclude from the equation (24) that

$$\eta = \Big(\sum_{i=1}^{n} W_{ni}^{'}(\mu_t)^2\Big)^{-1}\sum_{i=1}^{n} W_{ni}^{'}(\mu_t) + o_p(1). \qquad (25)$$

Now by remembering (10), we obtain

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \frac{\eta W_{ni}^{'}(\mu_t)}{1+\eta W_{ni}^{'}(\mu_t)} \\
&= \sum_{i=1}^{n}(\eta W_{ni}^{'}(\mu_t)) - \sum_{i=1}^{n}(\eta W_{ni}^{'}(\mu_t))^2 + \sum_{i=1}^{n}\frac{(\eta W_{ni}^{'}(\mu_t))^3}{1+\eta W_{ni}^{'}(\mu_t)}. \qquad (26)
\end{aligned}$$

Furthermore, having (19) and (21), we get

$$\sum_{i=1}^{n}\frac{(\eta W_{ni}^{'}(\mu_t))^3}{1+\eta W_{ni}^{'}(\mu_t)} = o_p(n^{-1/2}). \qquad (27)$$

Therefore, it can be concluded from (26) and (27) that

$$\sum_{i=1}^{n}(\eta W_{ni}^{'}(\mu_t))^2 = \sum_{i=1}^{n}\eta W_{ni}^{'}(\mu_t) + o_p(1).$$

Eventually, it follows from the equations (22) and (25) and Lemmas 5.1 and 5.2 that

$$\begin{aligned}
l(\mu_t) &= \sum_{i=1}^{n}\eta W_{ni}^{'}(\mu_t) + o_p(1) \\
&= \frac{\Big(\sum_{i=1}^{n} W_{ni}^{'}(\mu_t)\Big)^2}{\sum_{i=1}^{n} W_{ni}^{'2}} + o_p(1) \\
&= [\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}^{'}(\mu_t)}{\sqrt{\sigma^2+o_p(1)}}]^2 + o_p(1) \\
&\xrightarrow{\mathcal{D}} \chi_1^2.
\end{aligned}$$

□

Proof of Theorem 2.2. By considering Lemmas 5.1 and 5.2, we have $n^{-1}\sum_{i=1}^{n} W_{ni}^{'}(\mu_t) = O_p(n^{-1/2})$ and $n^{-1}\sum_{i=1}^{n} W_{ni}^{'2}(\mu_t) = \sigma^2 + o_p(1)$. By Equation (19) and using these results, we get $|\eta^A| = O_p(n^{-1/2})$ (see [1]). Therefore, from Equation (14), $\eta^A = \frac{n^{-1}\sum_{i=1}^{n} W_{ni}^{'}(\mu_t)}{n^{-1}\sum_{i=1}^{n} W_{ni}^{'2}(\mu_t)} + o_p(n^{-1/2})$. Then,

$$\begin{aligned} l^A(\mu_t) &= 2\sum_{i=1}^{n+1}(\eta^A W_{ni}^{'}(\mu_t) - \frac{1}{2}(\eta^A)^2 W_{ni}^{'2}(\mu_t)) + o_p(1) \\ &= \frac{n^{-1}(\sum_{i=1}^{n} W_{ni}^{'}(\mu_t))^2}{n^{-1}\sum_{i=1}^{n} W_{ni}^{'2}(\mu_t)} + o_p(1) \\ &\xrightarrow{\mathcal{D}} \chi_1^2. \end{aligned}$$

□

References

[1] Alemdjrodo, K., & Zhao, Y. (2020). New empirical likelihood inference for the mean residual life with length-biased and right-censored data. Journal of Nonparametric Statistics, 32(4), 1029-1046. https://doi.org/10.1080/10485252.2020.1840568

[2] Alemdjrodo, K., & Zhao, Y. (2022). Novel empirical likelihood inference for the mean difference with right-censored data. Statistical Methods in Medical Research, 31(1), 87-104. https://doi.org/10.1177/09622802211041767

[3] Amiri, N., Fakoor, V., Sarmad, M., & Shariati, A. (2022). Empirical likelihood analysis for accelerated failure time model using length-biased data. Statistics, 56(3), 578-597. https://doi.org/10.1080/02331888.2022.2077334

[4] Andersen, P.K., Hansen, M.G., & Klein, J.P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. Lifetime data analysis, 10(4), 335-350. https://doi.org/10.1007/s10985-004-4771-0

[5] Asgharian, M., & Wolfson, D.B. (2005). Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. The Annals of Statistics, 33(5), 2109-2131. https://doi.org/10.1214/009053605000000372

[6] Bai, F., & Yang, R. (2023). Estimation of treatment effects in two sample problems under general biased sampling data. Communications in Statistics-Theory and Methods, 1-13. https://doi.org/10.1080/03610926.2023.2273208

[7] Chen, J., & Huang, Y. (2013). Finite-sample properties of the adjusted empirical likelihood. Journal of Nonparametric Statistics, 25(1), 147-159. https://doi.org/10.1080/10485252.2012.738906

[8] Chen, J., Variyath, A.M., & Abraham, B. (2008). Adjusted empirical likelihood and its properties. Journal of Computational and Graphical Statistics, 17(2), 426-443. https://doi.org/10.1198/106186008X321068

[9] Cheng, Y.J, & Huang, C.Y. (2014). Combined estimating equation approaches for semi-parametric transformation models with length-biased survival data. Biometrics, 70(3), 608-618. https://doi.org/10.1111/biom.12170

[10] Davison, A.C., & Hinkley, D.V. (1997). Bootstrap methods and their application. Number 1. Cambridge university press. https://doi.org/10.1080/00401706.2000.10486018

[11] Gach, F., & Pötscher, B.M. (2011). Nonparametric maximum likelihood density estimation and simulation-based minimum distance estimators. Mathematical Methods of Statistics, 20, 288-326. https://doi.org/10.3103/S1066530711040028

[12] Hasegawa, T., Misawa, S., Nakagawa, S., Tanaka, S., Tanase, T., Ugai, H., Wakana, A., Yodo, Y., Tsuchiya, S., & Suganami. (2020). Restricted mean survival time as a

summary measure of time-to-event outcome. Pharmaceutical statistics, 19(4), 436-453. https://doi.org/10.1002/pst.2004

[13] He, S., & Huang, X. (2003). Central limit theorem of linear regression model under right censorship. Science in China Series A: Mathematics, 46(5), 600-610. https://doi.org/10.1360/02ys0139

[14] He, Y., & Zhou, Y. (2020). Nonparametric and semiparametric estimators of restricted mean survival time under length-biased sampling. Lifetime Data Analysis, 26, 761-788. https://doi.org/10.1007/s10985-020-09498-x

[15] Huang, C.Y., & Qin, J. (2011). Nonparametric estimation for length-biased and right-censored data. Biometrika, 98(1), 177-186. https://doi.org/10.1093/biomet/asq069

[16] Hyde, J. (1977). Testing survival under right censoring and left truncation. Biometrika, 64(2), 225-230. https://doi.org/10.2307/2335688

[17] Keziou, A., & Leoni-Aubin, S. (2008). On empirical likelihood for semiparametric two-sample density ratio models. Journal of Statistical Planning and Inference, 138(4), 915-928. https://doi.org/10.1016/j.jspi.2007.02.009

[18] Kiwitt, S., Nagel, E.R., & Neumeyer, N. (2008). Empirical likelihood estimators for the error distribution in nonparametric regression models. Mathematical Methods of Statistics, 17, 241-260. https://doi.org/10.3103/S1066530708030058

[19] Klein, J.P., & Moeschberger, M.L. (2003). Survival analysis: techniques for censored and truncated data, volume 1230. Springer. https://doi.org/10.1007/b97377

[20] Lee, C.H., Ning, J., & Shen, Y. (2018). Analysis of restricted mean survival time for length-biased data. Biometrics, 74(2), 575-583. https://doi.org/10.1111/biom.12772

[21] Liang, W., Shen, J.s., & He, S.y. (2016). Likelihood ratio inference for mean residual life of length-biased random variable. Acta Mathematicae Applicatae Sinica, English Series, 32(2), 269-282. https://doi.org/10.1007/s10255-016-0562-0

[22] Liu, Y., & Yu, C.W. (2010). Bartlett correctable two-sample adjusted empirical likelihood. Journal of Multivariate Analysis, 101(7), 1701-1711. https://doi.org/10.1016/j.jmva.2010.03.008

[23] Ning, J., Qin, J., Asgharian, M., & Shen, Y. (2013). Empirical likelihood-based confidence intervals for length-biased data. Statistics in medicine, 32(13), 2278-2291. https://doi.org/10.1002/sim.5637

[24] Owen, A. (1990). Empirical likelihood ratio confidence regions. The annals of statistics, 18(1), 90-120. https://doi.org/10.1214/aos/1176347494

[25] Owen, A. (1991). Empirical likelihood for linear models. The Annals of Statistics, 1725-1747. https://doi.org/10.1214/aos/1176348368

[26] Owen, A.B. (2001). Empirical likelihood. Chapman and Hall/CRC. ISBN:9781420036152, 1420036157

[27] Qin, J., Ning, J., Liu, H., & Shen, Y. (2011). Maximum likelihood estimations and em algorithms with length-biased data. Journal of the American Statistical Association, 106(496), 1434-1449. https://doi.org/10.1198/jasa.2011.tm10156

[28] Shen, J., Yuen, K.C., & Liu, C. (2016). Empirical likelihood confidence regions for one- or two-samples with doubly censored data. Computational Statistics & Data Analysis, 93, 285-293. https://doi.org/10.1016/j.csda.2015.01.010

[29] Shi, J., Ma, H., & Zhou, Y. (2018). The nonparametric quantile estimation for length-biased and right-censored data. Statistics & Probability Letters, 134, 150-158. https://doi.org/10.1016/j.spl.2017.10.020

[30] Wang, Q. H., & Jing, B.Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. Annals of the Institute of Statistical Mathematics, 53, 517-527. https://doi.org/10.1023/A:1014617112870

[31] Wei, W., Wan, A.T., & Zhou, Y. (2018). Partially linear transformation model for length-biased and right-censored data. Journal of Nonparametric Statistics, 30(2), 332-367. https://doi.org/10.1080/10485252.2018.1424335

[32] Yang, H., Yau, C., &Zhao, Y. (2014). Smoothed empirical likelihood inference for the difference of two quantiles with right censoring. Journal of Statistical Planning and Inference, 146, 95-101. https://doi.org/10.1016/j.jspi.2013.09.010

[33] Yang, X., Du, J., & Bai, F. (2023). Semiparametric inference of treatment effects on restricted mean survival time in two sample problems from length-biased samples. Statistics & Probability Letters, 193, 109715. https://doi.org/10.1016/j.spl.2022.109715

[34] Yu, X. & Zhao, Y. (2019). Empirical likelihood inference for semi-parametric transformation models with length-biased sampling. Computational Statistics & Data Analysis, 132, 115-125. https://doi.org/10.1016/j.csda.2018.10.012

[35] Zhang, Y. (2018). A comparison of methods for estimating Restricted Mean Survival Time (Doctoral dissertation, Master's thesis, Leiden University.[100]).

[36] Zhang, C., Wu, Y., & Yin, G. (2020). Restricted mean survival time for interval-censored data. Statistics in Medicine, 39(26), 3879-3895. https://doi.org/10.1002/sim.8699

[37] Zhao, L., Claggett, B., Tian, L., Uno, H., Pfeffer, M.A., Solomon, S.D., Trippa, L., & Wei, L. (2016). On the restricted mean survival time curve in survival analysis. Biometrics, 72(1), 215-221. https://doi.org/10.1111/biom.12384

[38] Zhao, Y., & Jinnah, A. (2012). Inference for cox's regression models via adjusted empirical likelihood. Computational Statistics, 27(1),1-12. https://doi.org/10.1007/s00180-010-0225-1

[39] Zheng, J., Shen, J., & He, S. (2014). Adjusted empirical likelihood for right censored lifetime data. *Statistical Papers*, 55(3), 827-839. https://doi.org/10.1007/s00362-013-0529-7

[40] Zhong, Y., Zhao, O., Zhang, B., & Yao, B. (2022). Adjusting for covariates in analysis based on restricted mean survival times. Pharmaceutical Statistics, 21(1), 38-54. https://doi.org/10.1002/pst.2151

[41] Zhou, M. (1992). Asymptotic normality of thesynthetic data'regression estimator for censored survival data. The Annals of Statistics, 1002-1021. https://doi.org/10.1214/aos/1176348667

[42] Zhou, M. (2021). Restricted mean survival time and confidence intervals by empirical likelihood ratio. Journal of Biopharmaceutical Statistics, 31(3), 362-374. https://doi.org/10.1080/10543406.2020.1862143

[43] Zhou, Y., & Liang, H. (2005). Empirical-likelihood-based semiparametric inference for the treatment effect in the two-sample problem with censoring. Biometrika, 92(2), 271-282. https://doi.org/10.1093/biomet/92.2.271

Zahra Mohammadian
Department of Statistics
Faculty of Mathematical Sciences
Ferdowsi University of Mashhad
P.O. Box 1159 Mashhad 91775, Mashhad, Islamic Republic of Iran
Email address: zahra.mohammadian@mail.um.ac.ir

Arezoo Habibirad
Department of Statistics
Faculty of Mathematical Sciences
Ferdowsi University of Mashhad
P.O. Box 1159 Mashhad 91775, Mashhad, Islamic Republic of Iran
Email address: ahabibi@um.ac.ir