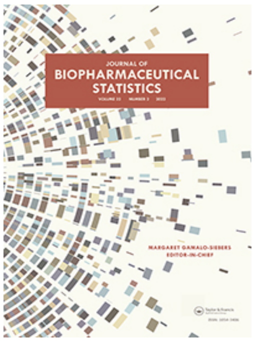

Taylor & Francis
Taylor & Francis Group

Journal of Biopharmaceutical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lbps20

Confidence intervals of mean residual life function in length-biased sampling based on modified empirical likelihood

Suthakaran Ratnasingam & Wei Ning

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
https://doi.org/10.1080/10543406.2022.2089157

Confidence intervals of mean residual life function in length-biased sampling based on modified empirical likelihood

Suthakaran Ratnasingam[a] and Wei Ning[b,c]

[a]Department of Mathematics, California State University, San Bernardino, California, USA; [b]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio, USA; [c]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China

ABSTRACT

The mean residual life (MRL) function is one of the basic parameters of interest in survival analysis. In this paper, we develop three procedures based on modified versions of empirical likelihood (EL) to construct confidence intervals of the MRL function with length-biased data. The asymptotic results corresponding to the procedures have been established. The proposed methods exhibit better finite sample performance over other existing procedures, especially in small sample sizes. Simulations are conducted to compare coverage probabilities and the mean lengths of confidence intervals under different scenarios for the proposed methods and some existing methods. Two real data applications are provided to illustrate the methods of constructing confidence intervals.

ARTICLE HISTORY
Received 25 September 2021
Accepted 19 May 2022

KEYWORDS
Mean residual life function; Length-biased data; confidence band; confidence interval; empirical likelihood

1. Introduction

There have been numerous research studies conducted to investigate the properties of the MRL function, which is widely used to model various life time data. A detailed reviews of the MRL functions can be found in Proschan and Serfling (1974), Crowley and Johnson (1982) and Guess and Proschan (1988). Many researchers have studied the theoretical properties of the MRL function. To name a few, Yang (1978) studied the MRL function on a fixed interval $0 \leq t \leq \tau < \infty$, and he showed that the estimator is strongly uniformly consistent on $[0, \tau]$. The Yang's findings were generalized to $R^+ = [0, \infty]$ with appropriate metrics by Hall and Wellner (1979). Berger et al. (1988) studied a nonprametric procedure for comparing MRL functions based on two independent samples. A semiparametric estimation of proportional mean residual life model for the censored data proposed by Chen and Cheng (2005). Further, the MRL function has many useful applications. For instance, recently, From and Ratnasingam (2021) derived new bounds for moment generating functions of various life distributions using MRL functions.

In real-world situations, we often collect a biased or weighted sample whose distribution differs from the population of interest. Length-biased sampling is a common problem in sampling design. In particular, this is a special case of left truncation with the truncation variables independent and uniformly distributed on a well-defined interval and it commonly occurs in various fields, including survival analysis, renewal processes, epidemiology, econometrics, and physics. More specifically, length-biased sampling arises when the event has already happened before the recruitment time of the study. For instance, in observational studies, the subjects who have survived at or beyond the enrollment time can be observed. In this situation, the observed samples are not randomly selected from the population of interest but with probability proportional to their length. Examples of such data can be found in Asgharian et al. (2002), Addona and Wolfson (2006) and Shen et al. (2009).

CONTACT Wei Ning wning@bgsu.edu Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403; School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081

Statistical inferences of length-biased data have drawn extensive considerations in the literature, see for example, Daniels (1942), In et al. (1969), Zelen and Feinleib (1969), Patil and Rao (1978), Simon (1980), and Wang (1996).

Much research has been conducted on the MRL function with length-biased sampling. For instance, Zhao and Qin (2006) derived the limiting distribution of the empirical likelihood (EL) ratio for MRL function for complete data. The technique was generalized by Zhao and Qin (2007) to censored survival data. Zhou and Jeong (2011) proposed EL-based ratio test to evaluate the equality of two mean residual life functions for the right censored data. A nonparametric method based on jackknife empirical likelihood through $U-$ statistic to test the equality of two mean residual functions studied by Chen et al. (2017). Fakoor (2015) proposed a nonparametric estimator of the MRL function for length-biased data. Fakoor et al. (2018) developed the MRL function inference through EL in length-biased sampling. They constructed EL-based confidence intervals and estimated coverage probabilities (CP) and the mean lengths of confidence intervals for the MRL function and compared the results with the normal approximation (NA) method. Despite of its popularity, the EL-based method suffers from two fundamental drawbacks: (1) In order to solve EL function, the convex hull must have 0 as its interior point. Owen (2001) suggested that if the convex hull does not have zero as an interior point, the EL function could be assigned a value of $-\infty$. However, this makes it difficult to find the maximum of the EL function, (2) EL method tends to suffer from an under-coverage problem, see Tsao (2013) for more details. Many methods have been proposed in the literature to mitigate these issues. For example, the adjusted empirical likelihood method (AEL) proposed by Chen et al. (2008) which affirms the existence of the solution in maximization problem and preserves the asymptotic optimality properties. To address the under-coverage problem for small sample sizes, Jing et al. (2017) proposed the transformed empirical likelihood (TEL). Recently, Stewart and Ning (2020) and Li et al. (2022) studied the transformed adjusted empirical likelihood (TAEL) which combines the advantages of AEL and TEL methods. In this paper, we propose three modified EL-based procedures to obtain confidence regions of the MRL function with length-biased data. We also show that under some regularity conditions, the limiting distribution of the modified EL-ratio for the MRL function is a standard chi-square distribution.

The rest of the paper is organized as follows. In section 2, we briefly describe fundamental properties of EL for the MRL function in length-biased sampling. The asymptotic properties of the proposed AEL, TEL, and TAEL methods are given in section 3. In section 4, we conduct simulations to study the mean lengths of the confidence intervals and coverage probabilities at various settings in the purpose of comparisons to other existing methods. In section 5, two real data applications are provided to illustrate the procedures of constructing confidence intervals. In section 6, some discussions are provided.

2. Methodology

Let X be a real-valued random variable with an unknown distribution function (d.f.) $F(\cdot)$. Let $Y_1, Y_2, \ldots, Y_n$ be a random variable denote a biased sample observed from a distribution $G(\cdot)$. Then, the version of the biased sampling of $F(\cdot)$ according to some known biasing (or weight) function $w(\cdot)$ is given as follows.

$$G(t) = \tfrac{1}{W}\int_{-\infty}^{t} w(y)dF(y), \quad t \in \mathrm{R} \tag{1}$$

where

$$W = \int w(y)dF(y). \tag{2}$$

The biased sampling problem is to estimate $F(\cdot)$ from $G(\cdot)$ on the basis of the random sample $Y_1, Y_2, \ldots, Y_n$. If $F(\cdot)$ is a d.f. on $\mathrm{R}^+ = [0, \infty)$ with a finite mean μ and let $w(x) = x, x \geq 0$, then we have

$$G(t) = \tfrac{1}{\mu}\int_{-\infty}^{t} ydF(y), \quad t \geq 0. \tag{3}$$

This is the length-biased distribution corresponding to $F(\cdot)$. For more details, refer to Fakoor et al. (2018). The corresponding mean residual life function at time t is defined as

$$M(t) = E(X - t|X > t) = \tfrac{1}{S(t)}\int_{t}^{\infty} S(x)dx, \tag{4}$$

where $S(x) = 1 - F(x)$ is a survival function of X and $t \geq 0$ which represents the probability that a subject survives beyond time x. Fakoor et al. (2018) proposed the EL-based method for the MRL function, $M(t)$, to obtain confidence interval as below. For $1 \leq i \leq n$ at a fix time t, we can define the EL at $M_0(t)$,

$$L(M_0(t)) = \left\{\sum_{i=1}^{n}\log(np_i)|p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i V_i(M_0(t)) = 0\right\}, \tag{5}$$

where $M_0(t)$ is the true value of $M(t)$ at time t, and

$$V_i(t) = \left(\tfrac{t+M_0(t)}{Y_i} - 1\right)I(Y_i \geq t). \tag{6}$$

By using the Lagrange multiplier method, $L(M_0(t))$ attains its maximum when

$$p_i = \tfrac{1}{n(1+\lambda(t)V_i(t))}, \quad i = 1, 2, \ldots, n \tag{7}$$

where $\lambda(t)$ is the Lagrange multiplier that solves the equation

$$\tfrac{1}{n}\sum_{i=1}^{n}\tfrac{V_i(t)}{n(1+\lambda(t)V_i(t))} = 0. \tag{8}$$

Note that that $\lambda(t)$ and $V_i(t)$ are functions of t. Thus, they can be evaluated at a fixed but arbitrary time, for example, t_0, such that $0 \leq t_0 < \tau$. However, for the simplicity of notations, we shall denote by $\lambda(t)$ and $V_i(t)$ instead of $\lambda(t_0)$ and $V_i(t_0)$. We should point out that the solution of this equation exists whenever zero is an interior point of the convex hull $\{\sum_{i=1}^{n} p_i V_i(M_0(t))| \sum p_i = 1, p_i \geq 0\}$. Now, the EL ratio for $M_0(t)$ is given by

$$\mathcal{R}(M_0(t)) = \prod_{i=1}^{n} np_i = \prod_{i=1}^{n}\tfrac{1}{1+\lambda(t)V_i(t)}, \tag{9}$$

and the associated empirical log-likelihood ratio is

$$l(M_0(t)) = \max\sum_{i=1}^{n}\log(1 + \lambda(t)V_i(t)). \tag{10}$$

Fakoor et al. (2018) showed that, under the condition of $E(Y^{-2}) < \infty$, and for all $t \in [0, \tau)$ the limiting distribution of $-2l(M_0(t))$ converges to χ_1^2 in distribution. However, the original EL suffers from the low coverage probability, especially for small sample sizes, see, for example, Chen et al. (2008) and Jing et al. (2017). Next, we present three modified EL-based approaches to construct confidence regions of $M(t)$ at a fixed time t for $t \in [0, \tau)$.

3. Main results

3.1. Adjusted empirical likelihood for the MRL function

We adopted the adjusted empirical likelihood (AEL) by Chen et al. (2008) for the MRL function. We define $\bar{V}_n = (1/n)\sum_{i=1}^{n} V_i(t)$ and $V_{n+1}(t) = -a_n \bar{V}_n$, where $a_n = \max\{1, \frac{1}{2}\log n\}$ according to the suggestion by Chen et al. (2008). Based on the $(n+1)$ observations, we define the adjusted empirical likelihood as

$$L^*(M_0(t)) = \max\left\{\sum_{i=1}^{n+1}\log((n+1)p_i)\middle| p_i \geq 0, \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i V_i(t) = 0\right\}. \tag{11}$$

Consequently, the adjusted empirical log-likelihood ratio is given by

$$l^*(M_0(t)) = \max \sum_{i=1}^{n+1} \log(1 + \lambda^a(t) V_i(t)). \tag{12}$$

Theorem 3.1. *Assume that $E(Y^{-2}) < \infty$. For all $t \in [0, \tau]$, let $l^*(M_0(t))$ be the adjusted log-empirical likelihood ratio function defined by (12) and $a_n = o_p(n^{2/3})$. We have*

$$-2l^*(M_0(t)) \to \chi_1^2 \tag{13}$$

in distribution.

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for $M(t)$ at a fixed time t is given as follows

$$C(t) = \{M(t) : -2l^*(M(t)) \leq \chi_{1,\alpha}^2\}, \tag{14}$$

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

Proof.

See Appendix A.

3.2. Transformed empirical likelihood for the MRL function

Besides the AEL by Chen et al. (2008), Jing et al. (2017) provided a simple transformation of the original EL (TEL) to improve the coverage probability. For a constant $\gamma \in [0, 1]$, we define

$$l_t(M_0(t), \gamma) = l(M_0(t)) \times \max\left\{1 - \frac{l(M_0(t))}{n}, 1 - \gamma\right\}, \tag{15}$$

and refer to $l_t(M_0(t), \gamma)$ as the truncated quadratic transformation of $l(M_0(t))$ defined in (10). Following Jing et al. (2017), we set $\gamma = 1/2$. Thus, the transformed empirical log-likelihood ratio can be defined as follows.

$$l_t(M_0(t)) = l_t\left(M_0(t), \gamma = \tfrac{1}{2}\right) = l(M_0(t)) \times \max\left\{1 - \frac{l(M_0(t))}{n}, \frac{1}{2}\right\}. \tag{16}$$

The corresponding transformed empirical log-likelihood ratio, denoted by $l_t(M_0(t))$, is

$$l_t(M_0(t)) = \begin{cases} l(M_0(t))\left(1 - \frac{l(M_0(t))}{n}\right) & \text{if } l(M_0(t)) \leq \frac{n}{2}, \\ l(M_0(t)) & \text{if } l(M_0(t)) > \frac{n}{2}. \end{cases} \tag{17}$$

Jing et al. (2017) pointed out that the TEL shares the same asymptotic properties with the EL. For more details readers are referred to Jing et al. (2017).

Theorem 3.2. *Assume that $E(Y^{-2}) < \infty$, for all $t \in [0, \tau]$, we have*

$$-2l_t(M_0(t)) \to \chi_1^2, \tag{18}$$

in distribution.

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for $M(t)$ at a fixed time t is given as follows

$$C(t) = \{M(t) : -2l_t(M(t)) \leq \chi_{1,\alpha}^2\}, \tag{19}$$

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

3.3. Transformed adjusted empirical likelihood for the MRL function

Transformed adjusted empirical likelihood (TAEL) is a combination of AEL and TEL methods proposed by Stewart and Ning (2020). The TAEL method comprises the advantages of AEL and TEL. For a constant $\gamma \in [0,1]$, we define

$$l_t^*(l^*(M_0(t)),\gamma) = l^*(M_0(t)) \times \max\left\{1 - \frac{l^*(M_0(t))}{n}, 1-\gamma\right\}. \tag{20}$$

where $l^*(\cdot)$ defined in (12). Thus, for $\gamma = 1/2$, the transformed empirical log-likelihood ratio $l_t^*(M_0(t))$ can be defined as,

$$l_t^*\left(l^*(M_0(t)),\tfrac{1}{2}\right) = l^*(M_0(t)) \times \max\left\{1 - \frac{l^*(M_0(t))}{n}, \frac{1}{2}\right\}. \tag{21}$$

More explicitly,

$$l_t^*(M_0(t)) = \begin{cases} l^*(M_0(t))\left(1 - \frac{l^*(M_0(t))}{n}\right) & \textit{if } l^*(M_0(t)) \leq \frac{n}{2}, \\ \frac{l^*(M_0(t))}{2} & \textit{if } l^*(M_0(t)) \geq \frac{n}{2}. \end{cases} \tag{22}$$

Theorem 3.3. *Assume that* $E(Y^{-2}) < \infty$, *for all* $t \in [0,\tau]$, *we have*

$$-2l_t^*(M_0(t)) \to \chi_1^2$$

in distribution.

Thus, an asymptotic $(1-\alpha)100\%$ confidence interval for $M(t)$ at a fixed time t is given as follows

$$C(t) = \{M(t) : -2l_t^*(M(t)) \leq \chi_{1,\alpha}^2\}, \tag{23}$$

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

Proof.

See Appendix A.

4. Simulation study

In this section, we conduct simulation studies to evaluate the performance of the proposed AEL, TEL, and TAEL-based confidence regions for the MRL function likelihood ratio in comparison with the EL and NA-based confidence regions under various sample sizes of length-biased data in terms of coverage probabilities (CP) and mean lengths (ML) of the confidence intervals. The CP is the proportion of the times that the confidence regions contain the true value of parameter among N simulation runs. In order to make fair comparisons, we adopt the same settings used in Fakoor et al. (2018). First, we obtain an i.i.d sample of $Y_1, Y_2, \ldots, Y_n$ with the proposed length-biased distributions.

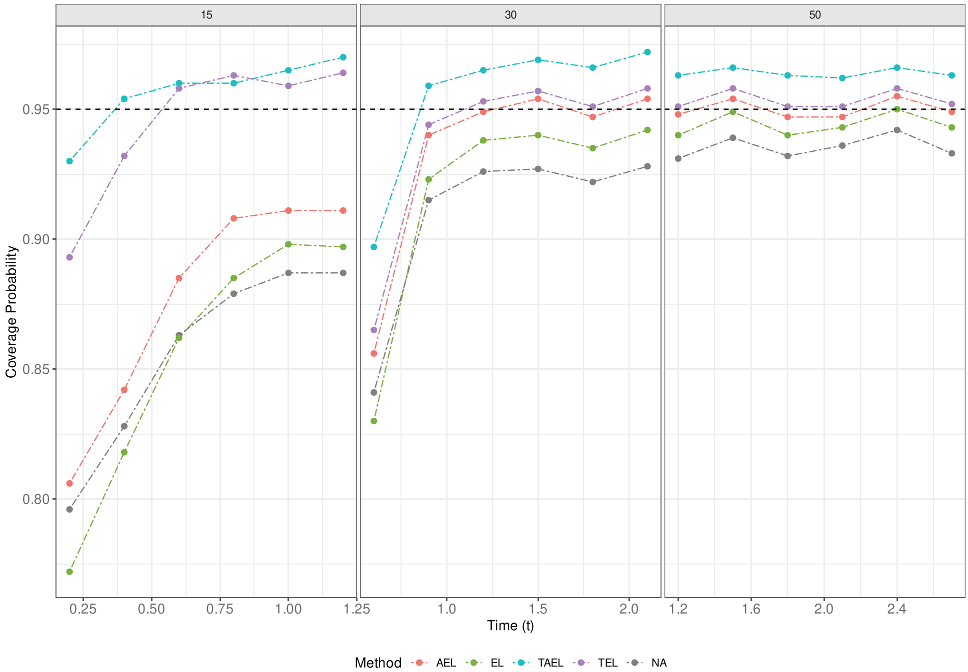

Figure 1. Coverage probabilities of the NA, EL, AEL, TEL, and TAEL methods with a range of t values for various sample sizes $n = \{15, 30, 50\}$ for the MRL of Uniform(1, 4).

We choose $n = 15, 30$, and 50 representing small and moderate sample sizes. The number of simulation replications is set to be $N = 5000$. We consider the nominal significance level $\alpha = 0.05$. The following MRL function of the respective target populations are considered: (1) Uniform(1, 4) and (2) Gamma(4, 2).

Tables 1 and 2 summarize the estimated CP of confidence intervals, and the mean lengths of the confidence intervals from different methods for the MRL of Uniform(1, 4) and Gamma(4, 2) distributions, respectively. We observe that the AEL, TEL, and TAEL-based confidence intervals have higher coverage probabilities than the EL and NA-based confidence regions. In particular, AEL, TEL, and TAEL have higher coverage accuracy for small samples, especially for $n = 15$ and 30. In particular, the coverage probability based on the TAEL method are much closer to the nominal level 0.95, and in some cases it's actually quite conservative. The confidence intervals based on NA usually perform the worst among the entire methods in general, although in some cases they are slightly better than the EL. It is also worth noting that the NA approach never achieves the nominal stated coverage. Further, it can be seen from Tables 1–2 that the TAEL-based method leads to an over-coverage problem slightly in some cases. The TEL method also provides significantly closer to nominal level, especially at small values of n. In these circumstances, the AEL or TEL methods are recommended. Our simulation results indicate that the CP tends to increase when the sample size increases, as would be expected. In addition, the mean lengths of the confidence intervals for the AEL, TEL, and TAEL are slightly longer than those corresponding to the EL and NA-based methods, but within an acceptable range. Note that as the sample size n increases, the mean length decreases for all five methods. Figures 1 and 3 demonstrate the coverage probabilities for all five methods for the MRL function based on length-biased observations of Uniform(1, 4) and Gamma(4, 2), respectively. Figures 2 and 4 illustrate the 95% pointwise confidence intervals for all five methods for the MRL function based on length-biased

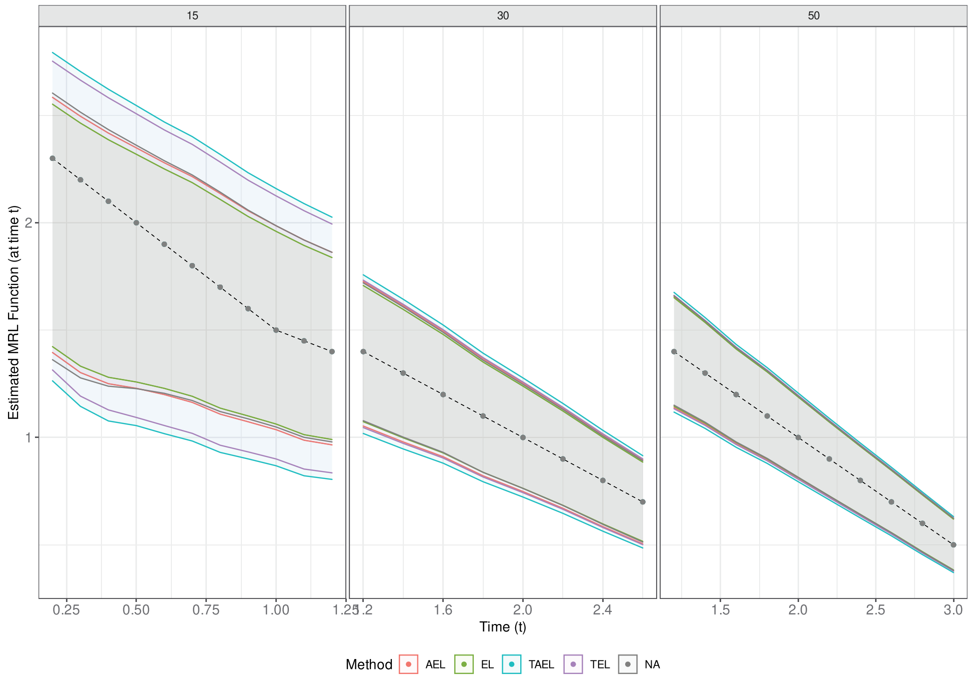

Figure 2. The point estimates (closed circles) and 95% pointwise confidence band (shaded area) of the NA, EL, AEL, TEL, and TAEL methods with a range of t values for various sample sizes $n = \{15, 30, 50\}$ for the MRL of Uniform(1, 4).

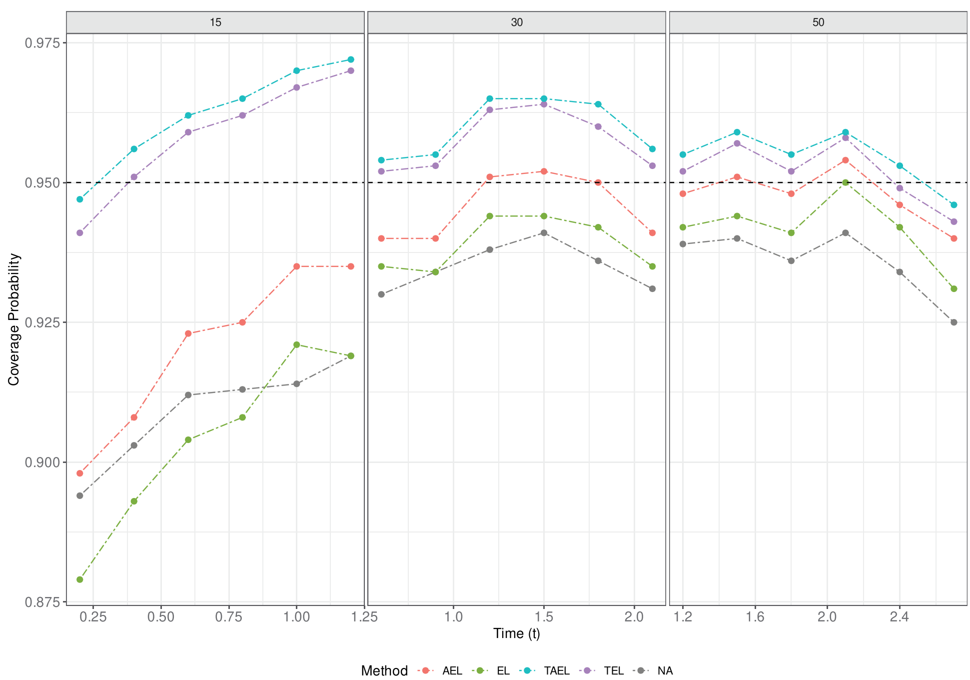

Figure 3. Coverage probabilities of the NA, EL, AEL, TEL, and TAEL methods with a range of t values for various sample sizes $n = \{15, 30, 50\}$ for the MRL of Gamma(4, 2).

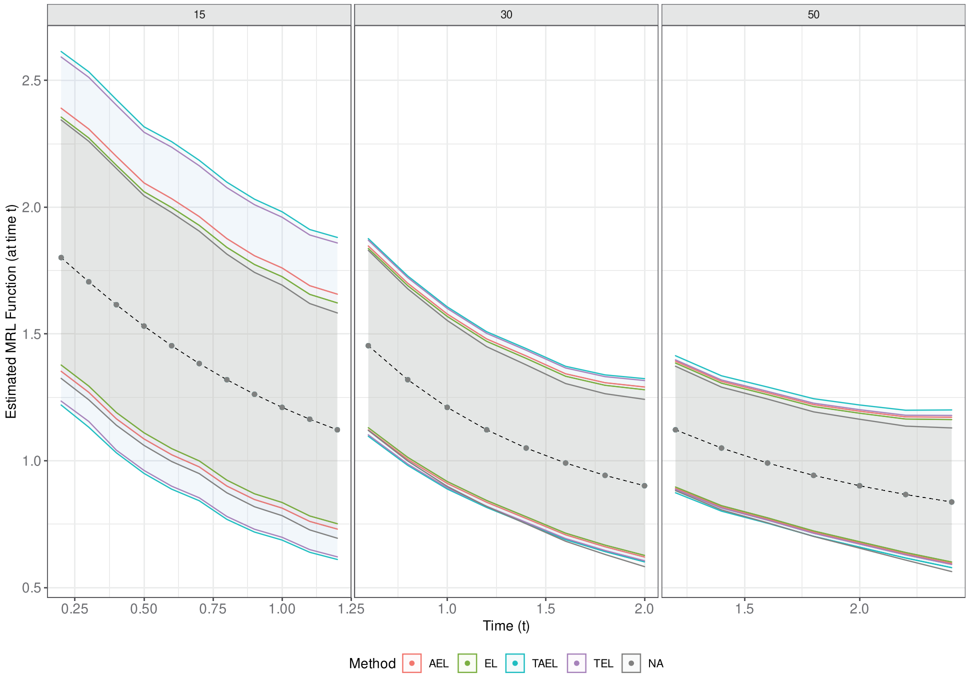

Figure 4. The point estimates (closed circles) and 95% pointwise confidence band (shaded area) of the NA, EL, AEL, TEL, and TAEL methods with a range of t values for various sample sizes $n = \{15, 30, 50\}$ for the MRL of Gamma(4, 2).

observations of Uniform(1, 4) and Gamma(4, 2), respectively. In particular, the shaded area depicts a 95% pointwise confidence band and black dashed line with closed circles in each panel represents the point estimates of the MRL at a given time point (t).

5. Applications

In this section, we apply the proposed methods to demonstrate the effectiveness of AEL, TEL, and TAEL methods by constructing confidence intervals for two real datasets as well as comparing to EL and NA methods. They are "Stanford heart transplant data" and "Medical Follow up Study data". A complete description of these data sets can be found in Crowley and Hu (1977) and Woolson (1981) respectively. In particular, length-biased sampling occurred in these datasets because only those individuals who were still alive during the study period was observed. We have also verified the stationary assumption for the sub-sample at selected ages by the method given in Asgharian et al. (2006).

5.1. Stanford heart transplant data

Our first data is survival time in days of Stanford heart transplant in Crowley and Hu (1977). The data set can be obtained from the R package "survival". We evaluate only 69 (see the appendix for instructions on how to obtain the patient's data) patients who underwent a heart transplant surgery. The aim of this application is to demonstrate how effectively the proposed methods perform when sample sizes are moderate to large. Table 3 summarizes a 95% confidence interval and the length of the confidence interval for the MRL function based on the length-biased sub-sample selected at various time t, including 0.2656, 1.7656, 3.2656, 4.7656, and 6.2656. We have also sketched the pointwise 95% EL-types confidence intervals and 95% EL-types confidence bands for the MRL function in Figures 5

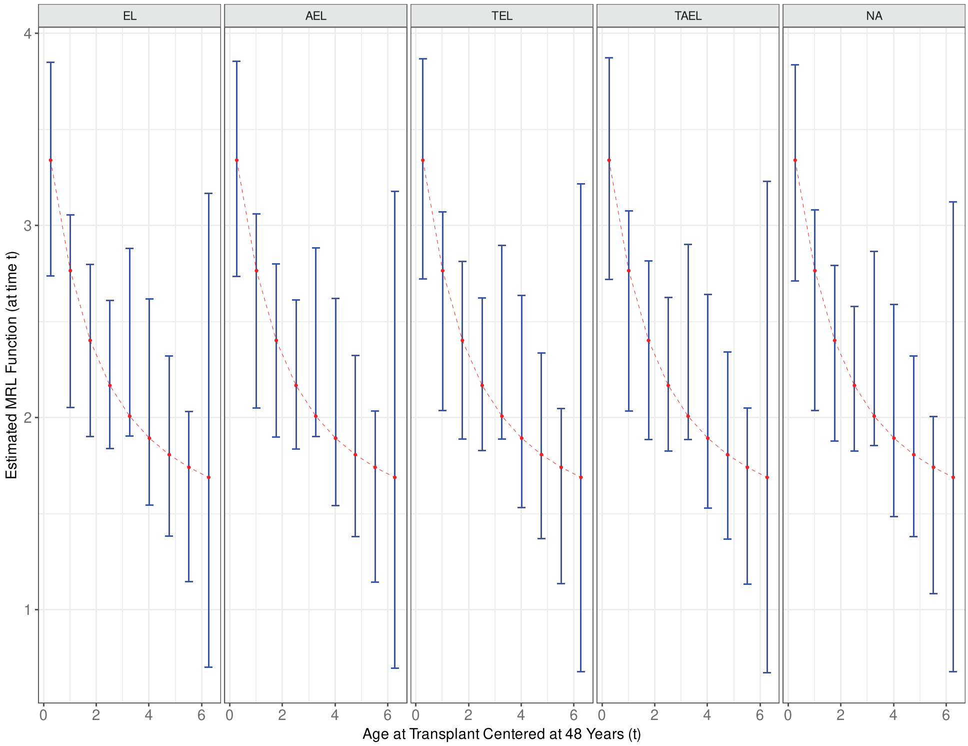

Figure 5. The point estimates (closed circles) and 95% pointwise confidence intervals of the MRL function by various methods.

and 6 respectively. The red color dashed line with closed circles in each panel represents the point estimates of the MRL function at given time t. The shaded area in Figure 6 depicts a 95% pointwise confidence band obtained from various EL-type methods and NA methods. It can be clearly seen that the MRL function steadily decreases when the time increases. We also observe that the lengths of the confidence intervals for all methods increase over time. We notice that the TAEL and TEL methods provide longer intervals than those based on NA, EL, and AEL methods. According to our simulations studies, the TAEL and TEL methods provide better CP under various settings comparing to EL and NA methods.

5.2. Medical follow up study

Our second data comes from the Medical Follow-up Study. The survival-time data were obtained from a much larger sample of 525 psychiatric inpatients who were first admitted to the University of Iowa Hospitals during the years 1935–1948. However, for the purpose of this analysis, we only considered the survival and demographic data for 26 psychiatric patients. Our goal is to show how well the proposed methods work for constructing confidence intervals for the MRL function when sample sizes are small. The data summarized in Table 1 in Woolson (1981). We have applied the proposed modified EL-based methods in order to estimate the MRL function. The 95% confidence intervals for the MRL function based on the length-biased sub-sample are calculated at selected ages 35, 37, 43, 47, 49, and 55 are presented in Table 4. Further, a 95% pointwise confidence intervals and confidence band are graphed in Figures 7 and 8 respectively. The MRL function steadily decreases when the age increases. We notice that as the sample size goes down the NA method does not present a valid result. The NA approach failed to produce the confidence interval for large values of t, such as $t = 47, 49$, and 55, although the TEL and TAEL methods did. We also observe that the EL

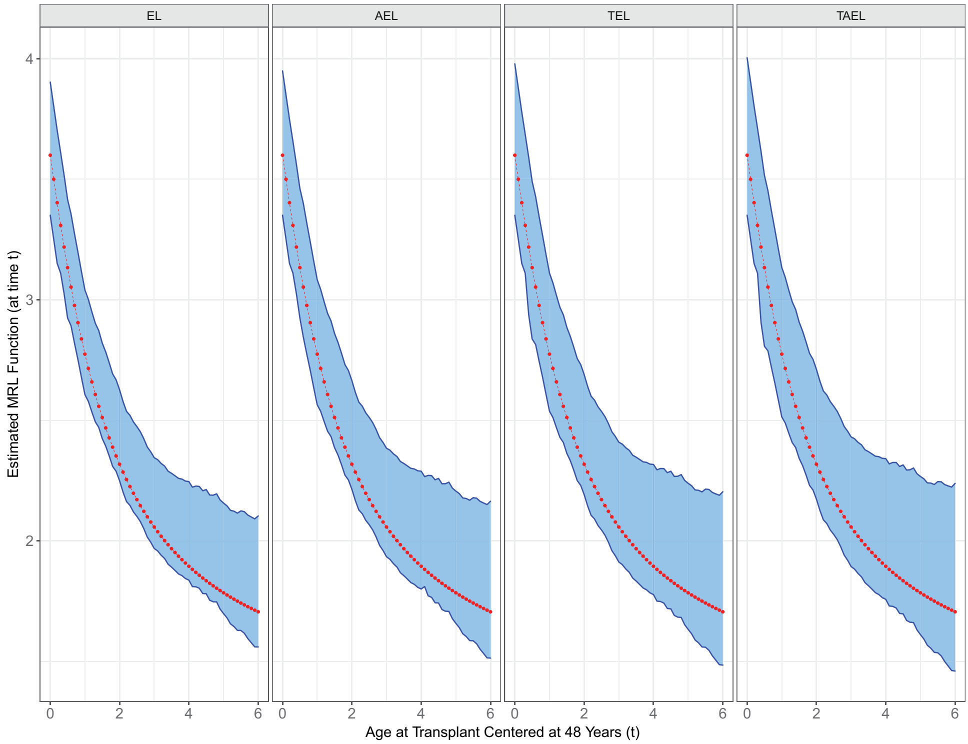

Figure 6. The point estimates (closed circles) and 95% pointwise confidence band (shaded area) of the MRL function.

Table 1. Estimated coverage probabilities (CP) of confidence regions and the mean lengths (ML) of confidence interval for the MRL of Uniform(1, 4) with the nominal significance level $\alpha = 0.05$.

		NA		EL		AEL		TEL		TAEL	
n	t	CP	ML	CP	ML	CP	ML	CP	ML	CP	ML
15	0.2	0.796	1.258	0.772	1.141	0.806	1.200	0.893	1.438	0.930	1.552
	0.4	0.828	1.174	0.818	1.088	0.842	1.148	0.932	1.427	0.954	1.515
	0.6	0.863	1.094	0.862	1.030	0.885	1.088	0.958	1.382	0.960	1.415
	0.8	0.879	1.009	0.885	0.960	0.908	1.015	0.963	1.304	0.960	1.332
	1.0	0.887	0.945	0.898	0.904	0.911	0.956	0.959	1.233	0.965	1.298
	1.2	0.887	0.881	0.897	0.846	0.911	0.896	0.964	1.157	0.970	1.219
30	6	0.841	0.826	0.830	0.809	0.856	0.858	0.865	0.875	0.897	0.940
	9	0.915	0.726	0.923	0.716	0.940	0.759	0.944	0.775	0.959	0.833
	2	0.926	0.647	0.938	0.638	0.949	0.678	0.953	0.691	0.965	0.744
	1.5	0.927	0.579	0.940	0.571	0.954	0.606	0.957	0.619	0.969	0.666
	1.8	0.922	0.517	0.935	0.510	0.947	0.541	0.951	0.552	0.966	0.594
	2.1	0.928	0.464	0.942	0.456	0.954	0.484	0.958	0.494	0.972	0.531
50	1.2	0.931	0.507	0.940	0.505	0.948	0.518	0.951	0.527	0.963	0.556
	1.5	0.939	0.454	0.949	0.451	0.954	0.463	0.958	0.471	0.966	0.498
	1.8	0.932	0.407	0.940	0.403	0.947	0.414	0.951	0.421	0.963	0.445
	2.1	0.936	0.364	0.943	0.360	0.947	0.369	0.951	0.376	0.962	0.397
	2.4	0.942	0.323	0.950	0.319	0.955	0.327	0.958	0.333	0.966	0.352
	2.7	0.933	0.283	0.943	0.278	0.949	0.285	0.952	0.290	0.963	0.306

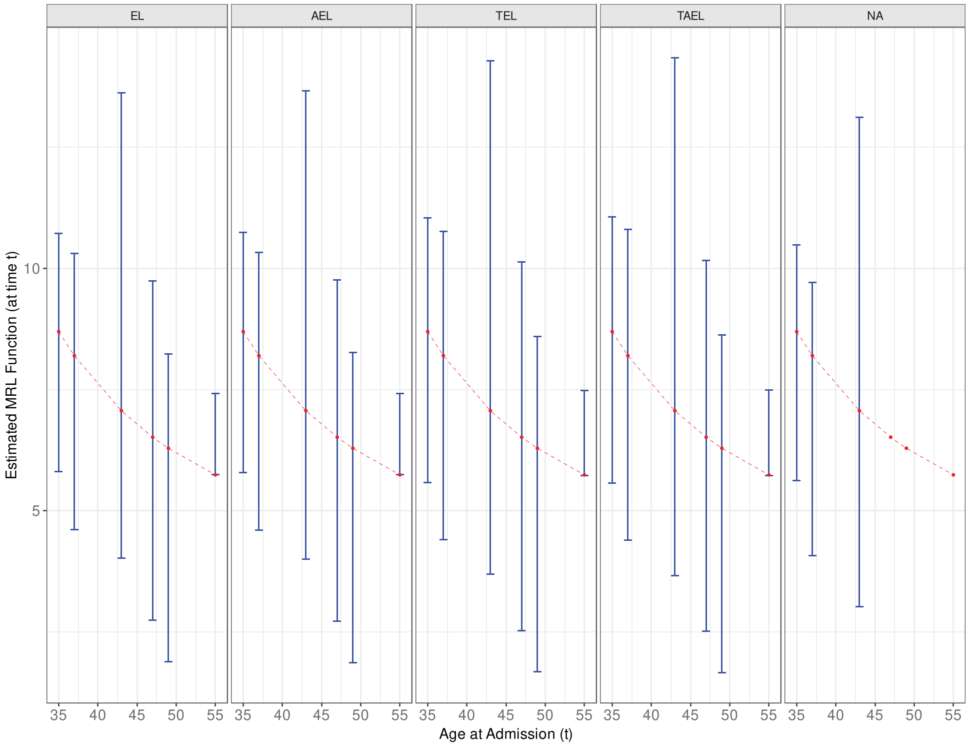

Figure 7. The point estimates (closed circles) and 95% pointwise confidence intervals of the MRL function by various methods.

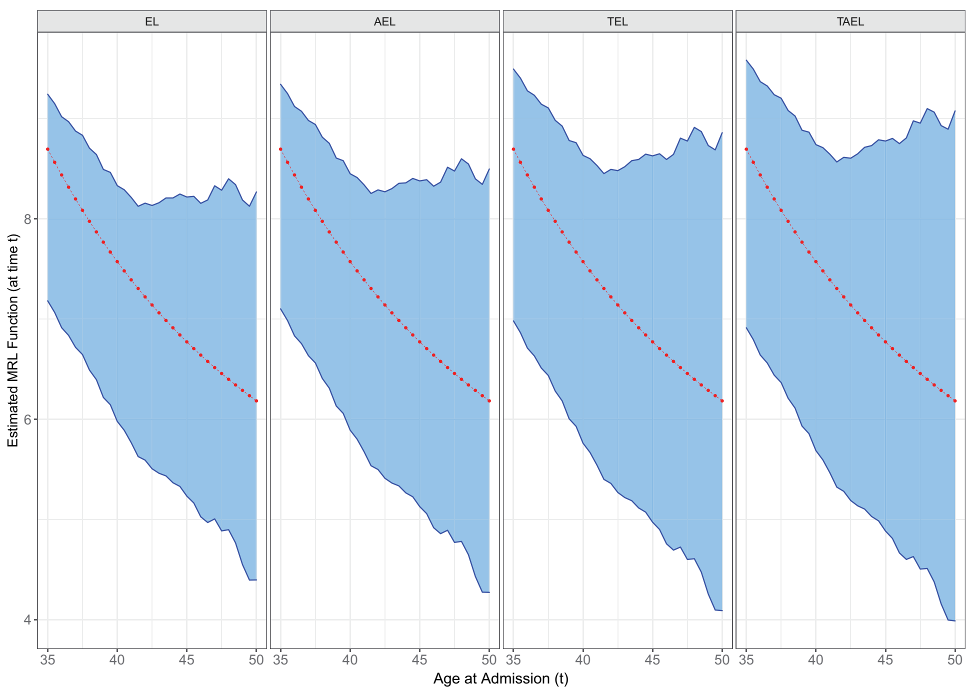

Figure 8. The point estimates (closed circles) and 95% pointwise confidence band (shaded area) of the MRL function.

Table 2. Estimated coverage probabilities (CP) of confidence regions and the mean lengths (ML) of confidence interval for the MRL of Gamma(4, 2) with the nominal significance level $\alpha = 0.05$.

		NA		EL		AEL		TEL		TAEL	
n	*t*	CP	ML	CP	ML	CP	ML	CP	ML	CP	ML
15	0.2	0.894	1.016	0.879	0.976	0.898	1.035	0.951	1.351	0.957	1.392
	0.4	0.903	1.015	0.893	0.978	0.908	1.038	0.961	1.362	0.966	1.399
	0.6	0.912	0.979	0.904	0.950	0.923	1.010	0.969	1.336	0.972	1.369
	0.8	0.913	0.943	0.908	0.920	0.925	0.977	0.972	1.298	0.975	1.331
	1	0.914	0.913	0.921	0.892	0.935	0.949	0.977	1.264	0.980	1.297
	1.2	0.919	0.890	0.919	0.871	0.935	0.926	0.980	1.237	0.982	1.269
30	0.6	0.930	0.710	0.935	0.708	0.940	0.726	0.952	0.768	0.954	0.780
	0.9	0.934	0.664	0.934	0.660	0.940	0.677	0.953	0.717	0.955	0.728
	1.2	0.938	0.635	0.944	0.631	0.951	0.647	0.963	0.685	0.965	0.696
	1.5	0.941	0.623	0.944	0.620	0.952	0.636	0.964	0.674	0.965	0.685
	1.8	0.936	0.642	0.942	0.637	0.950	0.653	0.960	0.692	0.964	0.704
	2.1	0.931	0.674	0.935	0.665	0.941	0.683	0.953	0.724	0.956	0.735
50	1.2	0.939	0.491	0.942	0.490	0.948	0.503	0.952	0.512	0.955	0.519
	1.5	0.940	0.485	0.944	0.484	0.951	0.497	0.957	0.506	0.959	0.514
	1.8	0.936	0.494	0.941	0.494	0.948	0.507	0.952	0.516	0.955	0.524
	2.1	0.941	0.521	0.950	0.519	0.954	0.533	0.958	0.543	0.959	0.551
	2.4	0.934	0.561	0.942	0.557	0.946	0.572	0.949	0.583	0.953	0.592
	2.7	0.925	0.621	0.931	0.612	0.940	0.629	0.943	0.641	0.946	0.650

Table 3. A 95% confidence intervals for MRL function for Stanford heart transplant data.

Time (*t*)	MRL	Interval	EL	AEL	TEL	TAEL	NA
0.2656	3.3401	Lower	2.7375	2.7342	2.7217	2.7181	2.7119
		Upper	3.8503	3.8541	3.8685	3.8728	3.8357
		Length	1.1128	1.1199	1.1468	1.1547	1.1238
1.7656	2.4018	Lower	1.9023	1.8996	1.8895	1.8865	1.8766
		Upper	2.7975	2.8006	2.8123	2.8157	2.7921
		Length	0.8951	0.9009	0.9228	0.9292	0.9155
3.2656	2.0073	Lower	1.9029	1.9001	1.8895	1.8864	1.8542
		Upper	2.8795	2.8831	2.8970	2.9011	2.8641
		Length	0.9766	0.9830	1.0075	1.0147	1.0099
4.7656	1.8072	Lower	1.3837	1.3811	1.3712	1.3683	1.3811
		Upper	2.3208	2.3242	2.3373	2.3411	2.3218
		Length	0.9370	0.9431	0.9661	0.9728	0.9406
6.2656	1.6890	Lower	0.7001	0.6955	0.6781	0.6731	0.6772
		Upper	3.1686	3.1789	3.2180	3.2293	3.1242
		Length	2.4684	2.4834	2.5399	2.5563	2.4470

Table 4. A 95% confidence intervals for MRL function for medical follow up study data.

Time (*t*)	MRL	Interval	EL	AEL	TEL	TAEL	NA
35	8.6958	Lower	5.8036	5.7899	5.5835	5.5642	5.6217
		Upper	10.7224	10.7418	1.0389	11.0672	10.4821
		Length	4.9188	4.9519	5.4555	5.5031	4.8604
37	8.1988	Lower	4.6116	4.5986	4.4055	4.3876	4.0723
		Upper	10.3051	0.3332	10.7663	0.8080	9.7114
		Length	5.6936	5.7345	6.3607	6.4204	5.6392
43	7.0624	Lower	4.0183	3.9970	3.6848	3.6563	3.0143
		Upper	13.6223	13.6634	14.2876	14.3466	13.1226
		Length	9.6040	9.6663	10.6028	10.6903	10.1083
47	6.5162	Lower	2.7348	2.7212	2.5267	2.5093	–
		Upper	9.7381	9.7632	10.1338	10.1678	–
		Length	7.0034	7.0420	7.6071	7.6585	–
49	6.2884	Lower	1.8795	1.8665	1.6770	1.6599	–
		Upper	8.2382	8.2607	8.5933	8.6238	–
		Length	6.3586	6.3942	6.9163	6.9639	–
55	5.7384	Lower	7.498	5.7479	5.7226	5.7205	–
		Upper	7.4176	7.4221	7.4849	7.4903	–
		Length	1.6678	1.6741	1.7623	1.7699	–

and AEL methods provide approximately the same length. The TAEL method provides the longest interval length while the EL and NA methods provide the shortest interval length. Therefore, we suggest the AEL, TEL, and TAEL methods are the better choices for the construction of confidence intervals since they produce better coverage probability at all time points (t) through the simulations.

6. Conclusion

In this article, we proposed modified empirical likelihood methods including the adjusted empirical likelihood (AEL), the transformed empirical likelihood (TEL), and the transformed adjusted empirical likelihood (TAEL) on constructing confidence intervals for the mean residual life function in length-biased sampling. The asymptotic distributions of the MRL function based on the AEL, TEL, and TAEL are derived. Simulations show that the proposed AEL and TEL methods improve the coverage probability comparing to the EL and the NA methods, especially for small sample sizes. The TAEL method provides the highest coverage probabilities, however, it occasionally suffer from an over coverage problem. Two real data applications are provided to illustrate the advantage of proposed methods. Based on the real-life applications, in general, the proposed length-biased versions of AEL, TEL, and TAEL perform better than the EL and NA due to the advantages of the nonparametric property and the existence of the solutions of the optimization procedure, especially for small sample sizes. From the simulations, we observe that, TAEL performs slightly better than AEL and TEL when the sample size is relatively small in terms of coverage probabilities while paying the price of producing slightly longer confidence intervals. As the sample size increases to moderate sample sizes, the performances of the three methods are comparable at different time points although TAEL provides slightly more conservative confidence intervals than the AEL and TEL. When in the scenarios of large sample sizes, three methods provide similar coverage probabilities and similar lengths of the confidence intervals. Therefore, in practical situations, TAEL is recommended when sample sizes are relatively small, AEL and TEL are recommended for moderate sample sizes as well as TAEL being an optional choice, and all three methods can be considered when the sample sizes are large.

Acknowledgments

The authors would like to thank two anonymous reviewers for their constructive comments and suggestions, which helped to improve this manuscript significantly.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Simons Foundation [711800].

References

Addona, V., and D. B. Wolfson. 2006. A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis* 12:267–284. doi:10.1007/s10985-006-9012-2.

Asgharian, M., C. E. M'Lan, and D. B. Wolfson. 2002. Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association* 97 (457):0 201–209. doi:10.1198/016214502753479347.

Asgharian, M., D. B. Wolfson, and X. Zhang. 2006. Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine* 25:1751–1767. doi:10.1002/sim.2326.

Berger, R. L., D. D. Boos, and F. M. Guess. 1988. Tests and confidence sets for comparing two mean residual life functions. *Biometrics* 44 (1):0 103–115. doi:10.2307/2531899.

Chen, Y. Q., and S. Cheng. 2005. Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika* 92:19–29. doi:10.1093/biomet/92.1.19.

Chen, J., A. M. Variyath, and B. Abraham. 2008. Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics* 17 (2):426–443. doi:10.1198/106186008X321068.

Chen, Y.-J., W. Ning, and A. K. Gupta. 2017. Jackknife empirical likelihood test for mean residual life functions. *Communications in Statistics-Theory and Methods* 46 (7):3111–3122. doi:10.1080/03610926.2015.1054943.

Crowley, J., and M. Hu. 1977. Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* 72:27–36. doi:10.1080/01621459.1977.10479903.

Crowley, J., and R. A. Johnson. 1982. *Survival analysis. ims. lecture notes*. Hayward, CA: Inst of Mathematical Statistic.

Daniels, H. E. 1942. A new technique for the analysis of fibre length distribution in wool. *Journal of the Textile Institute Transactions* 33:137–150. doi:10.1080/19447024208659258.

Fakoor, V. 2015. On the nonparametric mean residual life estimator in length-biased sampling. *Communications in Statistics - Theory and Methods* 44:512–519. doi:10.1080/03610926.2012.748328.

Fakoor, V., A. Shariati, and M. Sarmad. 2018. The mrl function inference through empirical likelihood in length-biased sampling. *Journal of Statistical Planning and Inference* 196:115–131. doi:10.1016/j.jspi.2017.11.001.

From, S. G., and S. Ratnasingam. 2021. Some new bounds for moment generating functions of various life distributions using mean residual life functions. *Journal of Statistical Theory and Practice* 15 (36). https://doi.org/10.1007/s42519–021–00176–1

Guess, F., and F. Proschan. 1988. Mean residual life, theory and applications. In *Handbook of statistics, 7: Reliability and quality control*, ed. P. R. Krishnaiah and C. R. rao, 215-224. New York: North-Holland.

Hall, W. J., and J. A. Wellner. Estimation of mean residual life. in technical report. *Department of Statistics*, University of Rochester, 1979.

In, D. R. C., N. L. Johnson, and H. Smith, eds. 1969. *some sampling problems in technology. new developments in survey sampling*. New York: Wiley. 506–527.

Jing, B.-Y., M. Tsao, and W. Zhou. 2017. Transforming the empirical liklihood towards better accuracy. *The Canadian Journal of Statistics* 45 (3):340–352. doi:10.1002/cjs.11328.

Li, M., S. Ratnasingam, and W. Ning. 2022. Empirical-likelihood-based confidence intervals for quantile regression models with longitudinal data. *Journal of Statistical Computation and Simulation* 1–18. In press. doi:10.1080/00949655.2022.2043322

Owen, A. B. 1990. Empirical likelihood confidence regions. *The Annals of Statistics* 18:90–120. doi:10.1214/aos/1176347494.

Owen, A. B. 2001. *Empirical likelihood*. New York: Champan Hall.

Patil, G. P., and C. R. Rao. 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* 34:179–184. doi:10.2307/2530008.

Proschan, F., and R. J. Serfling. 1974. *Reliability and biometry-statistical analysis of life length*. Philadelphia, PA: SIAM

Shen, Y., J. Ning, and J. Qin. 2009. Analyzing length-biased data with semiparametric transformation and accelerated failure time models. *Journal of the American Statistical Association* 104 (487):1192–1202. doi:10.1198/jasa.2009.tm08614.

Simon, R. 1980. Length biased sampling in etiologic studies. *American Journal of Epidemiology* 111:444–452. doi:10.1093/oxfordjournals.aje.a112920.

Stewart, P., and W. Ning. 2020. Modified empirical likelihood-based confidence intervals for data containing many zero observations. *Computational Statistics* 35 (4):2019–2042. doi:10.1007/s00180-020-00993-1.

Tsao, M. 2013. Extending the empirical likelihood by domain expansion. *Canadian Journal of Statistics* 41 (2):257–274. doi:10.1002/cjs.11175.

Wang, M. C. 1996. Hazards regression analysis for length-biased data. *Biometrika* 83 (2):343–354. doi:10.1093/biomet/83.2.343.

Woolson, R. F. 1981. Rank tests and a one-sample logrank test for comparing observed survival data to a standard population. *Biometrics* 37 (4):687–696. doi:10.2307/2530150.

Yang, G. L. 1978. Estimation of a biometric function. *The Annals of Statistics* 6:112–116. doi:10.1214/aos/1176344070.

Zelen, M., and M. Feinleib. 1969. On the theory of screening for chronic diseases. *Biometrika* 56 (3):0 601–614. doi:10.1093/biomet/56.3.601.

Zhao, Y., and G. S. Qin. 2006. Inference for the mean residual life function via empirical likelihood. *Communications in Statistics - Theory and Methods* 35 (6):1025–1036. doi:10.1080/03610920600672294.

Zhao, Y., and G. S. Qin. 2007. Empirical likelihood inference for the mean residual life under random censorship. *Statistics & Probability Letters* 77:549–557. doi:10.1016/j.spl.2006.09.018.

Zhou, M., and J. H. Jeong. 2011. Empirical likelihood ratio test for median and mean residual lifetime. *Statistics in Medicine* 30:152–159. doi:10.1002/sim.4110.

Appendix A. Proofs of Theorems

Proof. **Theorem 3.1**

The proof of this theorem is similar to Theorem 1 given in Chen et al. (2008). Let $\lambda(t)$ be the solution to

$$\sum_{i=1}^{n+1} \frac{V_i(t)}{1+\lambda^a(t)V_i(t)} = 0. \tag{24}$$

We first show that $\lambda^a(t) = O_p(n^{-1/2})$. According to Lemma 1 given Fakoor et al. (2018), since $E(V_1^2(t)) < \infty$, by using Lemma 3 of Owen (1990), we have that $V^* = \max_{1\le i\le n}\|V_i\| = o_p(n^{1/2})$ and $\bar{V}_n = O_p(n^{-1/2})$. Let $\rho = \|\lambda^a(t)\|$ and $\hat{\lambda}^a(t) = \lambda^a(t)/\rho$. Multiplying $\hat{\lambda}^a(t)/n$ to both sides gives,

$$\begin{aligned} 0 &= \frac{\hat{\lambda}^a(t)}{n} \sum_{i=1}^{n+1} \frac{V_i(t)}{(1+\lambda^a(t)V_i(t))} \\ &= \frac{\hat{\lambda}^a(t)}{n} \sum_{i=1}^{n+1} V_i(t) - \rho \sum_{i=1}^{n+1} \frac{(\hat{\lambda}^a(t)V_i(t))^2}{(1+\rho\hat{\lambda}^a(t)V_i(t))} \\ &\le \hat{\lambda}^a(t)\bar{V}_n(1-a_n/n) - \frac{\rho}{n(1+\rho V^*(t))} \sum_{i=1}^{n} (\hat{\lambda}^a(t)V_i(t))^2 \\ &= \hat{\lambda}^a(t)\bar{V}_n - \frac{\rho}{n(1+\rho V^*(t))} \sum_{i=1}^{n} (\hat{\lambda}^a(t)V_i(t))^2 + O_p(n^{-3/2}a_n). \end{aligned} \tag{25}$$

The inequality above is valid because the $(n+1)$th term in the second summation is non-negative. Following Chen et al. (2008), for any given $\varepsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} (\lambda^a(t)V_i(t))^2 \ge 1 - . \in \tag{26}$$

Therefore, as long as $a_n = o_p(n)$, so (25) implies that,

$$\frac{\rho}{(1+\rho V^*(t))} \le \hat{\lambda}^a(t)\frac{\bar{V}_n(t)}{(1-\in)} = O_p(n^{-1/2}). \tag{27}$$

Thus, we get $\rho = O_p(n^{-1/2})$ and hence $\lambda^a(t) = O_p(n^{-1/2})$. Now consider,

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n+1} \frac{V_i(t)}{1+\lambda^a(t)V_i(t)} \\ &= \bar{V}_n(t) - \lambda^a(t)\hat{V}_n(t) + o_p(n^{-1/2}). \end{aligned} \tag{28}$$

where $\hat{V}_n = (1/n)\sum_{i=1}^n V_i(t)^2$. Hence, when $n \to \infty, \lambda^a(t) = \hat{V}_n^{-1}\bar{V}_n + o_p(n^{-1/2})$. Now, we expand $l^*(M_0(t))$ as follows

$$\begin{aligned} l^*(M_0(t)) &= \sum_{i=1}^{n+1} \log(1+\lambda^a(t)V_i(t)) \\ &= \sum_{i=1}^{n+1} \left\{ \lambda^a(t)V_i(t) - \frac{(\lambda^a(t)V_i(t))^2}{2} \right\} + o_p(1). \end{aligned} \tag{29}$$

Substituting the expansion of λ^a, we get that

$$\begin{aligned} -2l^*(M_0(t)) &= n\hat{V}_n^{-1}\bar{V}_n^2 + o_p(1) \\ &\xrightarrow{d} \chi_1^2. \end{aligned} \tag{30}$$

This completes the proof.

Proof. **Theorem 3.2**

We consider the same arguments used in Jing et al. (2017). We will look at four criteria separately.

- (C_1) $0 \le l_t(M_0(t)) \le l(M_0(t))$;
- (C_2) $l_t(M_0(t))$ is a monotonically increasing function of $l(M_0(t))$;
- (C_3) $l_t(M_0(t_0)) = l(M_0(t_0)) + o_p(1)$;
- (C_4) For any $\tau_1 \in [0, +\infty)$ the level-τ_1 contour of $l_t(M_0(t))$, $\{M_0(t) : l_t(M_0(t)) = \tau_1\}$ is the same in shape as some level-τ_2 contour of $\{M_0(t) : l_t(M_0(t)) = \tau_2\}$; and $l_t(M_0(\tilde{t})) \le l_t(M_0(t))$ for $t \ne \tilde{t}$.

We evaluate criteria (C_1) through (C_4) given below.

- (C_1) We can easily see that from the original empirical log-likelihood $l(M_0(t))(\ge 0)$. This implies that

$$0 < \max\{1 - l(M_0(t))/n, 1/2\} \leq 1. \tag{31}$$

Hence, $0 \leq l_t(M_0(t)) \leq l(M_0(t))$.

• (C_2) For $l(M_0(t)) \in [0, n/2]$, we have $l_t(M_0(t)) = l(M_0(t)) \times \max\{1 - l(M_0(t))/n, 1/2\}$. Specifically, $l_t(M_0(t))$ is a strictly monotonically increasing function of $l(M_0(t))$ over the interval $[0, n/2]$. Thus, for $l(M_0(t)) > n/2$, we have $l_t(M_0(t)) = l(M_0(t))/2$. This is also a strictly monotonically increasing function of $l(M_0(t))$. Therefore, $l_t(M_0(t))$ is non-negative, continuous, and strictly monotonically increasing over $l(M_0(t)) \in [0, +\infty]$.

• (C_3) Fakoor et al. (2018) showed that the limiting distribution of $-2l(M_0(t_0))$ is $\chi^2(1)$, distribution, we have that $l(M_0(t_0)) = O_p(1)$. Consequently, we have $l(M_0(t_0)) \leq n/2$ with probability approaching to 1. Thus, $l_t(M_0(t_0)) = l(M_0(t_0)) \times \max\{1 - l(M_0(t_0))/n, 1/2\}$. Using this fact and that $l(M_0(t_0)) = O_p(1)$ give us (C_3).

• (C_4) For a level-τ_1 contour of the transformed empirical log-likelihood ratio $\{M_0(t) : l_t(M_0(t)) = \tau_1\}$, as $l_t(M_0(t))$ is a strictly monotonically increasing function of $l(M_0(t))$, let $\tau_2 = l_t^{-1}(\tau_1)$, then $\{M_0(t) : l_t(M_0(t)) = \tau_1\} = \{M_0(t) : l(M_0(t)) = \tau_2\}$. Further, as $l(M_0(t))$ typically has a unique minimum at $\tilde{M}_0(t)$, the second part of (C_4) also follows from the monotonicity of $l_t(M_0(t))$.

This completes the proof.

Proof. **Theorem 3.3**

In order to prove Theorem 3.3 we will follow the same strategy used in Theorem 3.2. Thus, details are omitted to conserve space.

Appendix B. Applications

```
library(survival)
data(heart, package = “survival”)
heart_yes = heart[heart$transplant = = 1,]
head(heart_yes)
n = length(heart_yes$age)
n
>[1] 69
```