Thailand Statistician
July 2019; 17(2): 144-164
http://statassoc.or.th
Contributed paper

# Comparison of Nonparametric Survival Estimators for Interval-Censoring Mixed with Right-Censoring Type I: A Simulation Study

**Prapasiri Ratchaprapapornkul and Lily Ingsrisawang***
Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, 10900, Thailand.
*Corresponding author; e-mail: fscilli@ku.ac.th

Received**:** 18 July 2018
Revised: 15 October 2018
Accepted**:** 27 October 2018

**Abstract**

The aim of this study is to compare the performance of survival function estimators in mixed censoring between interval-censoring and right censoring type I with 20%, and 40%, via the two nonparametric methods of Kaplan-Meier and Turnbull. The Kaplan-Meier estimator is applied to the mid-point imputation. The survival function estimated by the Turnbull is a decreasing step function defined on the complement of the union of disjoint intervals, called Turnbull interval. We assume the estimate of survival function by the linear interpolation, upper bound and lower bound on each Turnbull interval. Under the conditions of our simulation study, the Kaplan-Meier estimator gives the smallest mean square error and the Turnbull estimator with linear interpolation gives a smaller mean square error than the other Turnbull estimators. Moreover, the results also show that the bias of the Turnbull estimator (mostly Turnbull estimator with linear interpolation) yield the smallest value when time between visits are 0.5 years and 1 year but the Kaplan-Meier estimator gives the smallest bias for 2 years between visits.

---

**Keywords:** Nonparametric estimator, mid-point imputation, Kaplan-Meier estimator, Turnbull estimator, linear interpolation.

## 1. Introduction

Interval-censored data arises naturally in medical or epidemiologic studies in which there is a periodic follow up. For example, De Gruttola and Lagakos (1989) analyzed the time of HIV infection among hemophiliacs from 1978 to 1987. The data blood samples were periodically collected and stored retrospective testing of stored specimens. The infection result was only known to be between the last negative and first positive test, leading to the interval-censored data.

In survival analysis, survival estimation gives information on probability of time to event. The popular survival estimator for right-censoring is the Kaplan-Meier (KM) estimate which is product-limit estimator. In case of interval censoring, the approach simply imputes a time to event and then applies the KM method to estimate survival function $(S(t))$. Typically, the midpoint imputation is used for interval-censoring (Giesecke et al. 1988, Lee et al. 1989, Schechter et al. 1989, Lui et al. 1988, Mariotto et al. 1992, and Williams et al. 2004). Law and Brookmeyer (1992) studied statistical properties of a mid-point imputation for interval censored data via simulation. The time to event was assumed to be Weibull distribution with median survival of 10 years and time between screening visits were constants as 1, 2, 4, and 8 years. They found that it was a reasonable procedure to compute the KM method when the censoring interval was no longer than two years and the bias of KM estimator increased when the interval observation was wide and varied. Lindsey and Ryan (1998) compared two parametric methods and four nonparametric methods, including Turnbull (TB), KM applied to left-point imputation $(\text{KM}_{left})$, and KM applied to right-point imputation $(\text{KM}_{right})$ methods to estimate $S(t)$ ignoring the effect of covariates in two real interval-censored data sets. The first data set contained information of breast cancer patients of total 94 observations, whereas 38 observations of the right-censored data. The time to event was time to cosmetic deterioration and its estimates were shown that the $\text{KM}_{left}$ and the $\text{KM}_{right}$ bound the TB estimator in plotting the survival functions. The second data set was the drug resistance to zidovudine in AIDS patients of total 31 interval-censored observations, in which 13 observations were right-censored data. Since there were few assessments on each patient, it was resulting in wide interval for the AIDS data set. Moreover, the length of any interval-censored data quite varied for both data sets. It was found that the AIDS data set with heavier right-censoring rate and wider interval of observation makes the TB estimator unbounded by the $\text{KM}_{left}$ and $\text{KM}_{right}$. In addition, they also claimed that incorrect imputation of interval-censoring by treating the data as right censoring may lead to invalid inference. Geskus (2001) compared 3 techniques of imputation methods: midpoint, conditional mean, and multiple imputations, and 2 likelihood maximization methods for the nonparametric estimation of the distribution of doubly-censored AIDS incubation time from the real HIV infection data and the simulation data. In the simulation study, the conditional mean imputation yielded the lowest mean squared error of the estimator of Kaplan-Meier curves when the incubation time was assumed to be Weibull distribution and time between visits were half-year. However, the result from the conditional mean imputation was very similar to the midpoint imputation when the real data was used.

Another one of the well-known methods is Turnbull estimator. Turnbull (1976) derived a self-consistency algorithm to obtain a maximum likelihood estimate of the $S(t)$ with interval-censored observations. Groeneboom and Wellner (1992) proposed iterative convex minorant algorithm to compute the nonparametric maximum likelihood estimator (NPMLE)for both general interval-censored data and current status data and studied some asymptotic properties. However, its implementation is complex algorithm compared with the self-consistency (Zhang and Jamshidian 2004). The self-consistency estimator (SCE) or the TB estimator is not necessary to be a global maximum likelihood estimator (Gu and Zhang 1993). They also proved that a SCE and a NPMLE are asymptotically equivalent. Gentleman and Geyer (1994) claimed that the TB estimator is easy to implement but slow convergence and they developed the Kuhh-Tucker conditions to verify whether the estimator is the global maximum likelihood estimator or not. Moreover, the $S(t)$ estimated by TB is undefined on the union of Turnbull intervals

$\{(q_1, p_1],\ldots,(q_m, p_m]\}$ (Fay and Shaw 2010). The class of TB estimators on these intervals can be defined in various ways under some conditions as shown in the gray rectangles in Figure 1. The values of $q_j$ and $p_j$; $j=1,2,\ldots,m,$ are obtained from the set of left and right end points of the interval-censored data $\{(L_1, R_1],\ldots,(L_n, R_n]\}$ where $m \le n$ and $0 \le q_1 \le p_1 \le q_2 \le p_2 \ldots \le q_m \le p_m$ (Turnbull 1976). On the part with undefined estimators, there are some studies assuming the curve of $S(t)$ to be a linear interpolation (Geskus 2001, and Hebeisen 2014) and some studies assuming the curve to be flat as an upper bound or a lower bound of the class of TB estimators (Hebeisen 2014). Zhang (2005) considerly separated each TB interval to be 2 equal subintervals and defined the unique survial estimator as the upper bound for the first interval and the lower bound for the second one. Hence, the TB method is not only the undefined survival estimate for some points of time to event but also is an iterative approach that does not have a close form and so, some researchers practically use the straightforward approach as KM methods (Grover et al. 2013; Harezlak and Tu 2006; Støvring and Kristiansen 2011, and Williams et al. 2004).

There are some studies compared the imputed KM estimator and the TB estimator by using real world data with unknown real survival times (Lindsey and Ryan 1998, Samuelsen and Kongerud 1994, and Sun 1995). While other few studies compared these estimators via simulation (Gorelick 2009, and Nishikawa and Tango 2004). For example, Nishikawa and Tango (2004) studied behavior of the deterministic imputation of KM estimator (i.e. left-, mid- and right-point imputation) for the interval-censored data. They gerenerated the interval censored data which the time interval between visits was nearly 3 months mixed with right-censored data type random censoring. The result showed that the mid-point imputed KM estimator $(\text{KM}_{mid})$ has smaller the mean square error (MSE) than the TB estimator. Gorelick (2009) was interested in solving the problem of crossing over between the TB and either the $\text{KM}_{left}$ or $\text{KM}_{right}$. Gorelick concluded that the probability of crossing over these estimates depends on the mean survival time, sample size, and proportion of missing vits of a patient over the course of study. In addition, the use of the imputed KM may overestimate the survival function at the late time points.

Therefore, the above situations challenges us to study the performance of the $\text{KM}_{mid}$ and the TB estimators, especially, the performance of TB estimators when assume a linear interpolation and flats as a lower bound and an upper bound of class estimators on each TB interval, namely $\text{TB}_{lin}$, $\text{TB}_{low}$, and $\text{TB}_{up}$, respectively. Moreover, The HIV patient monitoring may be every a month or every 3-6 months. In the follows up care for cancer, the patients have prescheduled visit every 3 months to 2 years which are based on their risk of recurrence, leading to a wide interval of observation. These finding are of interest to figure out how the time interval between visits $(len)$ affect to both estimators. This study is performing on 0.5, 1 and 2 years of time interval between visits. A total of 18 scenarios is evaluated, based on one sample size $(n)$ of 200, 20% and 40% of right censoring under Weibull distributions with three sizes of shape parameters; $\alpha = 0.8$, 1 and 1.5, and the three sizes of time interval between visits; $len = 0.5$, 1 and 2 years. Furthermore, the estimator that give the smaller MSE will be the more efficient estimators.

## 2. Martials and Methods

### 2.1. Imputation method and Kaplan-Meier estimator

Due to simplicity, the imputation method has been often used to deal the missing data or the interval censored data. Suppose $(L_i, R_i]$ is an interval-censored observation for the $i^{th}$ subject where $i = 1, 2, ..., n.$ In this study, we have used the midpoint imputation for each interval $(L_i, R_i].$ where $i = 1, 2, ..., n$ which can be obtained from the following equation:

$$m_i = \frac{(L_i + R_i)}{2}. \tag{1}$$

The KM estimator is a step function which depends on each time point associated with an uncensored observation. Suppose that there are $m$ interval-censored observations among the $n$ individuals where $m \le n.$ After imputation step, $m_j$ where $j = 1, 2, \ldots, m$ will be an estimator of uncensored time, called an infected time. Let $m_{(j)},\ j = 1, 2, \ldots, m$ be the ordered infected times of $n$ individuals, $n_j$ be the number of subjects at risk to be infected at time $m_{(j)},\ d_j$ be the number of infected persons at time $m_{(j)}.$ Then the KM estimator of $S(t)$ is defined as follows (Kaplan and Meier 1958) :

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < m_{(1)}, \\ \prod_{j=1}^{q} \left( \frac{n_j - d_j}{n_j} \right) & \text{if } m_{(q)} \le t < m_{(q+1)}. \end{cases} \tag{2}$$

### 2.2. Turnbull estimator and modification

Let $T_i$ be the time to event random variable which lies in the interval $(L_i, R_i]$ for the $i^{th}$ subject, where $i = 1, 2, ..., n.$ Assume that $T_1, T_2 \ldots, T_n$ are $iid$ with the probability density function $f$ and the cumulative distribution function $F$. Let $L(F) = \prod_{i=1}^{n} [F(R_i +) - F(L_i -)]$ be the likelihood function. The procedure of TB estimator has 2 steps as follows:

1. Define Turnbull intervals:

Define the innermost TB intervals to be $m$ disjoint intervals $(q_1, p_1], \ldots, (q_m, p_m]$ for $m \le n$ satisfying, for each $j = 1, 2, \ldots, m,\ q_j$ and $p_j$ lie in the set $\{L_i : 1 \le i \le n\}$ and $\{R_i : 1 \le i \le n\}$, respectively, and $0 \le q_1 \le p_1 \le q_2 \le p_2 \ldots \le q_m \le p_m$.

2. Apply the self-consistency algorithm:

Let $\bar{s} = (s_1, s_2, \ldots, s_m)$ be an m-vectors of pseudo parameters when $s_j = P(q_j < t < p_j)$ $= F(p_j +) - F(q_j -)$ for $j = 1, 2, \ldots, m.$ To estimate the class of parameters $\bar{s}$, one can maximize the following likelihood function with respect to $s_1, s_2, \ldots, s_m$ :

$$L = \prod_{i=1}^{n} [F(R_i +) - F(L_i -)] = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} F\left[ (p_j +) - F(q_j -) \right] = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} s_j, \tag{3}$$

where $\alpha_{ij} = \begin{cases} 1, & (q_j, p_j] \subseteq (L_i, R_i] \\ 0, & \text{otherwise,} \end{cases}$ and the parameters are subject to the constraint $s_j > 0$ and $\sum_{j=1}^{m} s_j = 1$ (For more details, see Turnbull 1976). The self-consistent algorithm is a special case of the EM algorithm that maximizes the likelihood function which is based on the solution of the following self-consistency equation when the initial $s_j^0 = \frac{1}{m}$ and $\sum_{j=1}^{m} s_j^0 = 1$ (Dempster et al. 1977)

$$s_j^{new} = \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_{ij} s_j^{old}}{\sum_{k=1}^{m} \alpha_{ik} s_k^{old}} \quad \text{for } j = 1, \ldots, m, \; i = 1, \ldots, n. \tag{4}$$

In other words, the self-consistency algorithm to estimate $s_j$ for all $j = 1, \ldots, m$ can be obtained by the following iterative steps:

Step1. Choose starting value $s_j^0 = \frac{1}{m}$ for all $j = 1, \ldots, m$

Step 2. Compute the $s_j^1$ for (4) for all $j = 1, \ldots, m$

Step 3. Repeat Step 2 until convergence.

Define $w_l = 1 - (\hat{s}_1 + \hat{s}_2 + \ldots + \hat{s}_l)$ for $1 \le l \le m$. Then we acquire the estimator of $S(t)$ as the following formula;

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t \le q_1 \\ w_l & \text{if } p_l \le t \le q_{l+1} \quad \text{for } 1 \le l \le m-1 \\ 0 & \text{if } t \ge p_m \end{cases} \tag{5}$$

Moreover, the TB estimator of $S(t)$ is a monotone decreasing step function obtained the NPMLE which assumed a flat only on the gaps, $[p_j, q_{j+1}]$ where $j = 1, 2, \ldots, m-1$, and also the $[0, q_1]$ and $[p_m, \infty)$ (Gómez et al. 2009). In this study, we define the estimator of $S(t)$ on each $(q_j, p_j]$ corresponding to three relationships as a linear interpolation and flats as a lower bound and an upper bound of the class estimators on each TB interval and we denote a symbolic of the estimator as $S(t)_{TB-INV}$. The example of suvival curve by three relationships can be shown in Figure 1. Define $w_0 = 1$ and $w_m = 0$ so $S(t)_{TB-INV}$ can be estimated from the (6) for $q_l < t < p_l$, where $1 \le l \le m$:

$$\hat{S}(t)_{TB-INV} = \begin{cases} w_l + (t - p_l)\dfrac{(w_{l-1} - w_l)}{(q_l - p_l)} & \text{for } \mathrm{TB}_{lin}, \\ w_l & \text{for } \mathrm{TB}_{up}, \\ w_l - 1 & \text{for } \mathrm{TB}_{low}. \end{cases} \tag{6}$$

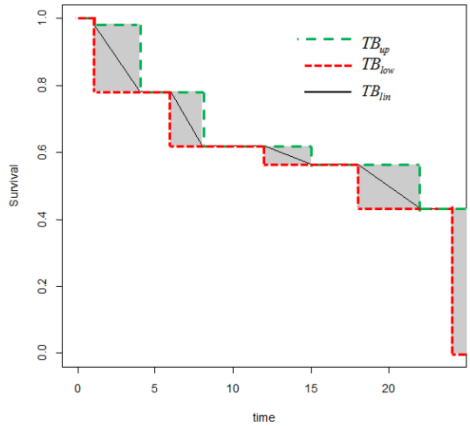


**Figure 1** The example of the survival curve that is estimated by Turnbull method

### 2.3. Simulation study

#### 2.3.1 Notation and assumptions

From the data of HIV follows up, the participants have prescheduled visits $k$ times every a specified time such as half-year and will collect the data every visit. Consider the $i^{th}$ subject, where $i=1, 2, ..., n$. The distribution of the time to event $T_i$ is assumed to be Weibull($\alpha,\gamma$) with shape parameters $\alpha = 0.8$, 1, 1.5 and scale parameters $\gamma =$ 19.382, 12.306, 6.716, respectively. The scale parameters are estimated from the real HIV data of 2-year survival probability 85% where $S(t)=\exp\left(-(t/\gamma)^{\alpha}\right)$. In this study, $len$ is considered to be constants as 0.5, 1 and 2 years except the last visit of each participant (the $len$ will be the rest of study time), no left-censored observation is concentrated here, in addition, the right censored observations are considered as type I censored. Then the censoring time is the total time in study (in years) of a patient which is estimated from percentiles of $S(t)$ with respect to levels of right censoring and levels of shape parameter (see Table 1). The number of follows up visiting $(k)$ with respect to $len$ and levels of right censoring. Let $\left(L_i, R_i\right]$ be the interval-censored observation and $v_{i1}, v_{i2}, \ldots, v_{ik}$ be visit times, starting time $(T=0)$ as $v_{i0}=0, v_{i1} \sim \text{Uniform}(0,1)$ and $v_{il}=v_{i1}+(l-1)len$ for $l>1$. In simulation, we suppose the participants may absent any appointment with probability $p=0.3$ but they will not miss any appointments at $v_{i1}$ and $v_{ik}$. To generate the interval-censored observation, if $t_i$ is lied between $v_{il}$ and $v_{i(l+1)}$, then the interval-censored data is $(L_i, R_i]=(v_{il}, v_{i(l+1)}]$ and indicator observation $\delta_i=1$. If $t_i$ is longer than the last visit time $v_{ik}$, then the observed data is right-censored as $(v_{ik}, \infty)$ and $\delta_i=0$.

The generated 200 samples data consist of mixed cases of interval-censored and right-censored observations. There are two different levels of right censoring $(r)$ consisting of 20% and 40%. A total of 18 different scenarios is set up, based on three levels of $\alpha$, two levels of $r$, one level of $n$, and three levels of $len$. The results given below are based on 500 replications.

### 2.3.2 Simulation procedure

The simulation for each scenario apply the following steps:

1. generate the real time to event $T_i$ $(i = 1,2,...,n)$ of each subject to have Weibull distributions with some specified parameters,

2. specify the time between visits $(len)$ and construct the sequence times of follows up visit according to the number of total prescheduled visit $(k)$ for each participant as $v_{i1}, v_{i2}, \ldots v_{ik}$,

3. specify the probability of missing for any visit $p = 0.3$ and generate the set of actual visit as schedules vectors $\bar{d} = \left(d_{i1}, d_{i2}, \ldots, d_{i(k-1)}\right)$ where

$$d_{ij} = \begin{cases} 1, & \text{if a patient } i^{th} \text{ meet appointment } j^{th} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

when $d_{i1} = 1,\ d_{ik} = 1$ and $d_{ij} \sim \text{Bernoulli}(1-p);$ for $j = 2, 3, \ldots, k-1$ for all $i$,

4. define the interval-censored observations $(L_i, R_i] = (v_{ij}, v_{il}]$ and the indicator observations of censoring $\delta_i = 1$ for some $0 \le j < l \le k$ where $v_{ij}$ and $v_{il}$ are the nearest visit times containing $t_i$ such that the participant meets the appointment visit $(d_{ij} = 1 = d_{il})$. Also define the right-censored observations $(R_i, \infty) = (v_{ik}, \infty)$ and $\delta_i = 0$ when $t_i$ is longer than the last $k^{th}$ visit,

5. if $\delta_i = 1$ for all $i = 1, 2, \ldots, n,$ impute the estimator of time to event by the midpoint of $(L_i, R_i]$ as (1) and estimate $S(t)$ at the $50^{th}$, $55^{th}$ ,$60^{th}$ ,$65^{th}$ ,$70^{th}$, $75^{th}$, $80^{th}$, $85^{th}$, $90^{th}$, and $95^{th}$ percentiles of time distribution from (2) to obtain $\text{KM}_{mid}$ estimator,

6. arrange Turnbull intervals from Section 2.2 and apply self-consistency estimator by iterating (4) until convergence,

7. estimate $S(t)$ at the same percentiles with Step 5 from (5) and estimate $S(t)_{TB-INV}$ from (6) to obtain $\text{TB}_{lin}$, $\text{TB}_{low}$, and $\text{TB}_{up}$ estimators,

8. repeat Step 1 to Step 7, 500 rounds,

9. calculate average survival estimates, MSE, bias and mean percentage relative change in MSE of the estimators (MPRC). The mean percentage relative change in MSE can be computed as the following equation,

$$\text{MPRC}_{\alpha,r,len}\left(M1\right) = \text{Mean}_{\alpha,r,len}\left\{\frac{\text{MSE}_{M1} - \text{MSE}_{KM}}{\text{MSE}_{KM}} \times 100\right\}. \tag{8}$$

## 3. Results and Discussion

We performed a simulation study to evaluate each scenario on the performances of the $\text{KM}_{mid}$, $\text{TB}_{lin}$, $\text{TB}_{low}$, and $\text{TB}_{up}$ estimators. For the estimates of $S(t)$ at the $50^{th}$, $55^{th}$, $60^{th}$, $65^{th}$, $70^{th}$, $75^{th}$, $80^{th}$, $85^{th}$, $90^{th}$, and $95^{th}$ percentiles, we investigated bias (BIAS), mean square error (MSE), and the mean percentage relative change in MSE (MPRC), and how these were affected by time between visits and right censoring rate when the data is generated with the different characteristics of Weibull distribution. Summaries of these result under the various scenarios are provided in Tables 2 to 5 and Figures 2 to 7.

**Table 1** Total time in study (in years) under the Weibull distribution of time to event with three sizes of shape parameters $(\alpha)$ and two levels of right censoring $(r)$

| Shape parameter | Right censoring (%) | Total time in study (year) |
|---|---|---|
| 0.8 | 20 | 35.0 |
| | 40 | 17.5 |
| 1 | 20 | 20.0 |
| | 40 | 11.0 |
| 1.5 | 20 | 9.0 |
| | 40 | 6.5 |

Note: the length of last visit time is different from another time between visits according to the total time in study.

The MPRC (see Table 2) indicated the performance of the $\text{KM}_{mid}$ which was about at least 20% better than those of all TB estimators on average. Moreover, the MPRC tended to increase as the time between visits and the shape parameter became larger.

The performance of most estimators seemed to depend on the time between visits and the shape parameter but there was no obvious difference in performances when right censoring changed from $r = 20\%$ to $r = 40\%$. So, we reported the result summary for only $r = 40\%$ (see Tables 3, 4 and 5). The result on Tables 3, 4 and 5 were shown that all four estimates achieved their percentiles almost exactly except the $75^{th}$ percentile onward for only $len = 2$ and gave the same values of MSE for the first 3 decimal digits as almost every percentile. Besides, the $\text{KM}_{mid}$ was indicated to have the smallest MSE and $\text{TB}_{lin}$ gave smaller MSE than the other two TB for most scenarios.

The Figures 2, 3 and 4 were shown that the time between visits and the shape parameter did not give obvious different results of MSE for only using the $\text{KM}_{mid}$ method. Inapposite, the $\text{TB}_{lin}$, $\text{TB}_{low}$ and $\text{TB}_{up}$ methods gave MSE slightly larger when the two factors increased for both $r = 20\%$ and $r = 40\%$. Moreover, the MSE of four estimators showed a trend to slightly decrease with increasing level of percentile especially for $len = 0.5$ and $1$ year.

However, we found that the time between visits was a factor affecting to the BIAS of four methods in which the BIAS tended to increase as the time between visits became longer specially upper or equal to $85^{th}$ percentile. Furthermore, the BIAS of $\text{KM}_{mid}$ and all TB estimators seemed to be close to zero and showed the same amount for the first 2 decimal digits as almost every percentile (see Figures 5, 6 and 7).

Although, the MSE of four estimators showed a trend to slightly decrease with increasing level of percentile, all BIAS showed a trend to slightly increase over the percentile. This implied that the estimators are more consistent over the percentiles but their accuracy is reducing. For $len = 0.5$ and $1$ years, the $\text{TB}_{lin}$ yielded the smallest BIAS for most percentiles while the $\text{KM}_{mid}$ yielded the smallest MSE for all cases. This means that the $\text{TB}_{lin}$ is more accurate but less precise than the $\text{KM}_{mid}$ when the time between visits are 0.5 and 1 year. Moreover, we also found that the $\text{KM}_{mid}$ is always overestimated $S(t)$ at upper or equal to $85^{th}$ percentile for all scenario (see Figures 5, 6 and 7).

**Table 2** The mean percentage relative change in MSE over shape parameters, levels of right censoring and time between visits

| The Mean Percentage Relative Change in MSE (MPRC) | | | | | |
|---|---|---|---|---|---|
| Shape parameters | Right censoring rate (%) | Method | Time between visits (years) | | |
| | | | $len = 0.5$ | $len = 1$ | $len = 2$ |
| 0.8 (Decreasing H.) | 20 | $TB_{up}$ | 21.980 | 44.421 | 75.133 |
| | | TB | 20.114 | 42.742 | 68.264 |
| | | $TB_{low}$ | 23.365 | 47.630 | 72.345 |
| | 40 | $TB_{up}$ | 23.734 | 44.082 | 86.716 |
| | | TB | 21.328 | 42.235 | 81.057 |
| | | $TB_{low}$ | 23.530 | 44.003 | 87.784 |
| 1 (Constant H.) | 20 | $TB_{up}$ | 27.549 | 52.328 | 105.470 |
| | | TB | 24.182 | 50.350 | 94.949 |
| | | $TB_{low}$ | 25.736 | 55.438 | 102.433 |
| | 40 | $TB_{up}$ | 25.035 | 51.274 | 111.796 |
| | | TB | 23.336 | 48.341 | 100.272 |
| | | $TB_{low}$ | 26.374 | 50.885 | 106.821 |
| 1.5 (Increasing H.) | 20 | $TB_{up}$ | 41.820 | 99.184 | 272.031 |
| | | TB | 39.285 | 96.998 | 246.146 |
| | | $TB_{low}$ | 42.479 | 103.810 | 258.490 |
| | 40 | $TB_{up}$ | 44.353 | 100.818 | 247.444 |
| | | TB | 41.938 | 96.841 | 226.443 |
| | | $TB_{low}$ | 45.170 | 102.331 | 239.233 |

Note: $TB = TB_{lin}$

**Table 3** The average of survival estimates and MSE and BIAS for 40% of right censoring and shape parameter 0.8 and sample size 200

| True value | Method | *len* 0.5 (years) | | | *len* 1 (years) | | | *len* 2 (years) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Average | MSE | BIAS | Average | MSE | BIAS | Average | MSE | BIAS |
| 0.5 | KM | 0.5009 | 0.00132 | 0.0009 | 0.5001 | 0.00114 | 0.0000 | 0.4999 | 0.00122 | -0.0001 |
| | $TB_{up}$ | 0.5025 | 0.00136 | 0.0025 | 0.5010 | 0.00134 | 0.0010 | 0.5082 | 0.00159 | 0.0082 |
| | TB | 0.5009 | 0.00137 | 0.0009 | 0.4994 | 0.00130 | -0.0006 | 0.5066 | 0.00149 | 0.0066 |
| | $TB_{low}$ | 0.4998 | 0.00136 | -0.0002 | 0.4983 | 0.00130 | -0.0017 | 0.5052 | 0.00147 | 0.0052 |
| 0.55 | KM | 0.5506 | 0.00126 | 0.0006 | 0.5504 | 0.00114 | 0.0004 | 0.5490 | 0.00120 | -0.0010 |
| | $TB_{up}$ | 0.5516 | 0.00130 | 0.0016 | 0.5513 | 0.00127 | 0.0013 | 0.5589 | 0.00160 | 0.0089 |
| | TB | 0.5502 | 0.00130 | 0.0002 | 0.5497 | 0.00124 | -0.0003 | 0.5576 | 0.00156 | 0.0076 |
| | $TB_{low}$ | 0.5488 | 0.00132 | -0.0012 | 0.5483 | 0.00126 | -0.0017 | 0.5562 | 0.00158 | 0.0062 |
| 0.6 | KM | 0.5987 | 0.00119 | -0.0013 | 0.5988 | 0.00110 | -0.0012 | 0.6014 | 0.00114 | 0.0014 |
| | $TB_{up}$ | 0.5997 | 0.00131 | -0.0003 | 0.5987 | 0.00122 | -0.0013 | 0.6060 | 0.00152 | 0.0060 |
| | TB | 0.5983 | 0.00128 | -0.0017 | 0.5976 | 0.00121 | -0.0024 | 0.6049 | 0.00152 | 0.0049 |
| | $TB_{low}$ | 0.5969 | 0.00131 | -0.0031 | 0.5966 | 0.00123 | -0.0034 | 0.6039 | 0.00152 | 0.0039 |
| 0.65 | KM | 0.6493 | 0.00109 | -0.0007 | 0.6501 | 0.00112 | 0.0001 | 0.6543 | 0.00105 | 0.0043 |
| | $TB_{up}$ | 0.6509 | 0.00115 | 0.0009 | 0.6514 | 0.00128 | 0.0014 | 0.6423 | 0.00157 | -0.0077 |
| | TB | 0.6496 | 0.00112 | -0.0004 | 0.6503 | 0.00126 | 0.0003 | 0.6405 | 0.00162 | -0.0095 |
| | $TB_{low}$ | 0.6484 | 0.00113 | -0.0016 | 0.6487 | 0.00129 | -0.0013 | 0.6389 | 0.00172 | -0.0111 |
| 0.7 | KM | 0.6989 | 0.00104 | -0.0011 | 0.6996 | 0.00097 | -0.0005 | 0.7052 | 0.00106 | 0.0052 |
| | $TB_{up}$ | 0.7000 | 0.00114 | 0.0000 | 0.6998 | 0.00118 | -0.0002 | 0.6946 | 0.00154 | -0.0054 |
| | TB | 0.6985 | 0.00114 | -0.0015 | 0.6987 | 0.00117 | -0.0013 | 0.6936 | 0.00151 | -0.0064 |
| | $TB_{low}$ | 0.6974 | 0.00118 | -0.0026 | 0.6973 | 0.00118 | -0.0027 | 0.6917 | 0.00157 | -0.0083 |
| 0.75 | KM | 0.7484 | 0.00093 | -0.0016 | 0.7521 | 0.00089 | 0.0021 | 0.7505 | 0.00097 | 0.0005 |
| | $TB_{up}$ | 0.7494 | 0.00101 | -0.0006 | 0.7516 | 0.00130 | 0.0016 | 0.7705 | 0.00211 | 0.0205 |
| | TB | 0.7480 | 0.00100 | -0.0020 | 0.7506 | 0.00128 | 0.0006 | 0.7692 | 0.00201 | 0.0192 |
| | $TB_{low}$ | 0.7465 | 0.00103 | -0.0035 | 0.7496 | 0.00127 | -0.0004 | 0.7677 | 0.00201 | 0.0177 |
| 0.8 | KM | 0.7983 | 0.00076 | -0.0017 | 0.8028 | 0.00076 | 0.0028 | 0.8170 | 0.00105 | 0.0170 |
| | $TB_{up}$ | 0.7991 | 0.00087 | -0.0009 | 0.8022 | 0.00108 | 0.0022 | 0.7810 | 0.00196 | -0.0190 |
| | TB | 0.7981 | 0.00086 | -0.0019 | 0.8012 | 0.00106 | 0.0012 | 0.7794 | 0.00203 | -0.0206 |
| | $TB_{low}$ | 0.7972 | 0.00087 | -0.0028 | 0.8002 | 0.00106 | 0.0002 | 0.7772 | 0.00213 | -0.0228 |
| 0.85 | KM | 0.8487 | 0.00061 | -0.0013 | 0.8546 | 0.00062 | 0.0046 | 0.8600 | 0.00071 | 0.0100 |
| | $TB_{up}$ | 0.8498 | 0.00078 | -0.0002 | 0.8524 | 0.00094 | 0.0024 | 0.8812 | 0.00236 | 0.0312 |
| | TB | 0.8483 | 0.00076 | -0.0017 | 0.8505 | 0.00092 | 0.0005 | 0.8745 | 0.00204 | 0.0245 |
| | $TB_{low}$ | 0.8469 | 0.00078 | -0.0031 | 0.8489 | 0.00097 | -0.0011 | 0.8744 | 0.00204 | 0.0244 |
| 0.9 | KM | 0.9024 | 0.00041 | 0.0024 | 0.9113 | 0.00051 | 0.0113 | 0.9308 | 0.00126 | 0.0308 |
| | $TB_{up}$ | 0.9011 | 0.00060 | 0.0011 | 0.9005 | 0.00084 | 0.0005 | 0.8812 | 0.00174 | -0.0188 |
| | TB | 0.8997 | 0.00059 | -0.0003 | 0.8996 | 0.00081 | -0.0004 | 0.8800 | 0.00174 | -0.0200 |
| | $TB_{low}$ | 0.8986 | 0.00061 | -0.0014 | 0.8984 | 0.00081 | -0.0016 | 0.8744 | 0.00210 | -0.0256 |
| 0.95 | KM | 0.9471 | 0.00022 | -0.0029 | 0.9545 | 0.00023 | 0.0045 | 0.9524 | 0.00021 | 0.0024 |
| | $TB_{up}$ | 0.9529 | 0.00045 | 0.0029 | 0.9545 | 0.00061 | 0.0045 | 0.9520 | 0.00065 | 0.0020 |
| | TB | 0.9515 | 0.00043 | 0.0015 | 0.9533 | 0.00062 | 0.0033 | 0.9506 | 0.00065 | 0.0006 |
| | $TB_{low}$ | 0.9500 | 0.00043 | 0.0000 | 0.9522 | 0.00063 | 0.0022 | 0.9498 | 0.00068 | -0.0002 |

Note: TB = $TB_{lin}$

**Table 4** The average of survival estimates and MSE and BIAS for 40% of right censoring and shape parameter 1 and sample size 200

| True value | Method | *len* 0.5 (years) | | | *len* 1 (years) | | | *len* 2 (years) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Average | MSE | BIAS | Average | MSE | BIAS | Average | MSE | BIAS |
| 0.5 | KM | 0.4985 | 0.00115 | -0.0015 | 0.5019 | 0.00122 | 0.0019 | 0.4969 | 0.00129 | -0.0031 |
| | $TB_{up}$ | 0.4993 | 0.00125 | -0.0007 | 0.5035 | 0.00150 | 0.0035 | 0.4996 | 0.00189 | -0.0004 |
| | TB | 0.4980 | 0.00121 | -0.0020 | 0.5018 | 0.00146 | 0.0018 | 0.4982 | 0.00185 | -0.0018 |
| | $TB_{low}$ | 0.4968 | 0.00121 | -0.0032 | 0.5003 | 0.00149 | 0.0003 | 0.4971 | 0.00187 | -0.0029 |
| 0.55 | KM | 0.5493 | 0.00115 | -0.0007 | 0.5518 | 0.00121 | 0.0018 | 0.5507 | 0.00130 | 0.0007 |
| | $TB_{up}$ | 0.5503 | 0.00121 | 0.0003 | 0.5523 | 0.00143 | 0.0023 | 0.5390 | 0.00218 | -0.0110 |
| | TB | 0.5491 | 0.00121 | -0.0009 | 0.5507 | 0.00138 | 0.0007 | 0.5380 | 0.00216 | -0.0120 |
| | $TB_{low}$ | 0.5476 | 0.00123 | -0.0024 | 0.5491 | 0.00139 | -0.0009 | 0.5362 | 0.00222 | -0.0138 |
| 0.6 | KM | 0.5989 | 0.00116 | -0.0011 | 0.6014 | 0.00122 | 0.0013 | 0.5970 | 0.00124 | -0.0030 |
| | $TB_{up}$ | 0.6000 | 0.00130 | 0.0000 | 0.6011 | 0.00161 | 0.0011 | 0.6095 | 0.00200 | 0.0095 |
| | TB | 0.5985 | 0.00128 | -0.0015 | 0.5999 | 0.00159 | -0.0001 | 0.6084 | 0.00196 | 0.0084 |
| | $TB_{low}$ | 0.5968 | 0.00130 | -0.0032 | 0.5984 | 0.00161 | -0.0016 | 0.6072 | 0.00198 | 0.0072 |
| 0.65 | KM | 0.6480 | 0.00108 | -0.0021 | 0.6505 | 0.00109 | 0.0005 | 0.6569 | 0.00118 | 0.0069 |
| | $TB_{up}$ | 0.6496 | 0.00123 | -0.0004 | 0.6506 | 0.00149 | 0.0006 | 0.6410 | 0.00211 | -0.0090 |
| | TB | 0.6480 | 0.00121 | -0.0020 | 0.6498 | 0.00147 | -0.0002 | 0.6399 | 0.00209 | -0.0101 |
| | $TB_{low}$ | 0.6466 | 0.00123 | -0.0034 | 0.6487 | 0.00148 | -0.0013 | 0.6376 | 0.00217 | -0.0124 |
| 0.7 | KM | 0.6982 | 0.00100 | -0.0018 | 0.7013 | 0.00097 | 0.0013 | 0.7034 | 0.00105 | 0.0034 |
| | $TB_{up}$ | 0.6993 | 0.00117 | -0.0007 | 0.7010 | 0.00133 | 0.0010 | 0.7088 | 0.00198 | 0.0088 |
| | TB | 0.6976 | 0.00114 | -0.0024 | 0.7000 | 0.00132 | 0.0000 | 0.7071 | 0.00189 | 0.0071 |
| | $TB_{low}$ | 0.6962 | 0.00116 | -0.0038 | 0.6990 | 0.00131 | -0.0010 | 0.7057 | 0.00187 | 0.0057 |
| 0.75 | KM | 0.7486 | 0.00095 | -0.0014 | 0.7514 | 0.00089 | 0.0014 | 0.7582 | 0.00090 | 0.0082 |
| | $TB_{up}$ | 0.7494 | 0.00115 | -0.0006 | 0.7502 | 0.00134 | 0.0002 | 0.7580 | 0.00194 | 0.0080 |
| | TB | 0.7478 | 0.00115 | -0.0022 | 0.7496 | 0.00133 | -0.0004 | 0.7554 | 0.00184 | 0.0054 |
| | $TB_{low}$ | 0.7461 | 0.00114 | -0.0039 | 0.7486 | 0.00136 | -0.0014 | 0.7532 | 0.00191 | 0.0032 |
| 0.8 | KM | 0.7983 | 0.00079 | -0.0018 | 0.8014 | 0.00071 | 0.0014 | 0.8184 | 0.00102 | 0.0184 |
| | $TB_{up}$ | 0.8003 | 0.00094 | 0.0003 | 0.8010 | 0.00111 | 0.0010 | 0.7891 | 0.00180 | -0.0109 |
| | TB | 0.7990 | 0.00095 | -0.0010 | 0.7994 | 0.00110 | -0.0006 | 0.7873 | 0.00182 | -0.0127 |
| | $TB_{low}$ | 0.7977 | 0.00097 | -0.0023 | 0.7983 | 0.00113 | -0.0017 | 0.7854 | 0.00190 | -0.0146 |
| 0.85 | KM | 0.8490 | 0.00063 | -0.0010 | 0.8542 | 0.00057 | 0.0042 | 0.8602 | 0.00066 | 0.0102 |
| | $TB_{up}$ | 0.8491 | 0.00086 | -0.0009 | 0.8513 | 0.00106 | 0.0013 | 0.8962 | 0.00364 | 0.0462 |
| | TB | 0.8471 | 0.00087 | -0.0029 | 0.8498 | 0.00104 | -0.0002 | 0.8882 | 0.00311 | 0.0382 |
| | $TB_{low}$ | 0.8458 | 0.00091 | -0.0042 | 0.8486 | 0.00107 | -0.0014 | 0.8878 | 0.00310 | 0.0378 |
| 0.9 | KM | 0.9007 | 0.00042 | 0.0007 | 0.9090 | 0.00049 | 0.0089 | 0.9308 | 0.00123 | 0.0308 |
| | $TB_{up}$ | 0.9001 | 0.00063 | 0.0001 | 0.9033 | 0.00073 | 0.0033 | 0.8962 | 0.00152 | -0.0038 |
| | TB | 0.8989 | 0.00062 | -0.0011 | 0.9018 | 0.00070 | 0.0018 | 0.8937 | 0.00147 | -0.0063 |
| | $TB_{low}$ | 0.8976 | 0.00066 | -0.0024 | 0.9005 | 0.00069 | 0.0005 | 0.8880 | 0.00181 | -0.0120 |
| 0.95 | KM | 0.9466 | 0.00027 | -0.0034 | 0.9526 | 0.00025 | 0.0026 | 0.9608 | 0.00028 | 0.0108 |
| | $TB_{up}$ | 0.9504 | 0.00045 | 0.0004 | 0.9514 | 0.00057 | 0.0014 | 0.9531 | 0.00060 | 0.0031 |
| | TB | 0.9492 | 0.00043 | -0.0008 | 0.9503 | 0.00056 | 0.0003 | 0.9524 | 0.00059 | 0.0024 |
| | $TB_{low}$ | 0.9478 | 0.00045 | -0.0022 | 0.9491 | 0.00058 | -0.0009 | 0.9514 | 0.00061 | 0.0014 |

Note: $TB = TB_{lin}$

**Table 5** The average of survival estimates and MSE and BIAS for 40% of right censoring and shape parameter 1.5 and sample size 200

| True value | Method | *len* 0.5 (years) | | | *len* 1 (years) | | | *len* 2 (years) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Average | MSE | BIAS | Average | MSE | BIAS | Average | MSE | BIAS |
| 0.5 | KM | 0.4982 | 0.00135 | -0.0018 | 0.4915 | 0.00122 | -0.0085 | 0.4913 | 0.00143 | -0.0087 |
| | $TB_{up}$ | 0.5009 | 0.00158 | 0.0009 | 0.4989 | 0.00186 | -0.0011 | 0.4802 | 0.00379 | -0.0198 |
| | TB | 0.4999 | 0.00155 | -0.0001 | 0.4980 | 0.00186 | -0.0020 | 0.4796 | 0.00378 | -0.0204 |
| | $TB_{low}$ | 0.4987 | 0.00157 | -0.0013 | 0.4970 | 0.00185 | -0.0030 | 0.4781 | 0.00390 | -0.0219 |
| 0.55 | KM | 0.5490 | 0.00131 | -0.0010 | 0.5454 | 0.00120 | -0.0046 | 0.5266 | 0.00192 | -0.0234 |
| | $TB_{up}$ | 0.5493 | 0.00173 | -0.0007 | 0.5509 | 0.00207 | 0.0009 | 0.5349 | 0.00309 | -0.0151 |
| | TB | 0.5480 | 0.00171 | -0.0020 | 0.5494 | 0.00201 | -0.0006 | 0.5336 | 0.00308 | -0.0164 |
| | $TB_{low}$ | 0.5469 | 0.00177 | -0.0031 | 0.5481 | 0.00206 | -0.0019 | 0.5322 | 0.00316 | -0.0178 |
| 0.6 | KM | 0.5992 | 0.00116 | -0.0008 | 0.5976 | 0.00108 | -0.0024 | 0.5784 | 0.00175 | -0.0216 |
| | $TB_{up}$ | 0.6008 | 0.00151 | 0.0008 | 0.6033 | 0.00191 | 0.0033 | 0.6106 | 0.00360 | 0.0106 |
| | TB | 0.5997 | 0.00148 | -0.0003 | 0.6024 | 0.00191 | 0.0024 | 0.6094 | 0.00350 | 0.0094 |
| | $TB_{low}$ | 0.5981 | 0.00152 | -0.0019 | 0.6011 | 0.00197 | 0.0011 | 0.6082 | 0.00347 | 0.0082 |
| 0.65 | KM | 0.6497 | 0.00113 | -0.0003 | 0.6474 | 0.00111 | -0.0026 | 0.6195 | 0.00209 | -0.0305 |
| | $TB_{up}$ | 0.6499 | 0.00155 | -0.0001 | 0.6482 | 0.00190 | -0.0018 | 0.6945 | 0.00575 | 0.0445 |
| | TB | 0.6489 | 0.00154 | -0.0011 | 0.6469 | 0.00187 | -0.0031 | 0.6906 | 0.00543 | 0.0406 |
| | $TB_{low}$ | 0.6481 | 0.00153 | -0.0019 | 0.6454 | 0.00191 | -0.0046 | 0.6897 | 0.00544 | 0.0397 |
| 0.7 | KM | 0.7003 | 0.00109 | 0.0003 | 0.6976 | 0.00104 | -0.0024 | 0.6977 | 0.00104 | -0.0023 |
| | $TB_{up}$ | 0.7033 | 0.00146 | 0.0033 | 0.7001 | 0.00195 | 0.0001 | 0.6945 | 0.00380 | -0.0055 |
| | TB | 0.7021 | 0.00143 | 0.0021 | 0.6985 | 0.00188 | -0.0015 | 0.6927 | 0.00375 | -0.0073 |
| | $TB_{low}$ | 0.7008 | 0.00147 | 0.0008 | 0.6974 | 0.00191 | -0.0026 | 0.6897 | 0.00397 | -0.0103 |
| 0.75 | KM | 0.7502 | 0.00092 | 0.0002 | 0.7477 | 0.00082 | -0.0023 | 0.7650 | 0.00102 | 0.0150 |
| | $TB_{up}$ | 0.7509 | 0.00136 | 0.0009 | 0.7483 | 0.00169 | -0.0017 | 0.7170 | 0.00389 | -0.0330 |
| | TB | 0.7498 | 0.00133 | -0.0002 | 0.7470 | 0.00166 | -0.0030 | 0.7160 | 0.00397 | -0.0340 |
| | $TB_{low}$ | 0.7488 | 0.00135 | -0.0012 | 0.7457 | 0.00169 | -0.0043 | 0.7139 | 0.00430 | -0.0361 |
| 0.8 | KM | 0.7995 | 0.00077 | -0.0005 | 0.7975 | 0.00076 | -0.0025 | 0.8070 | 0.00079 | 0.0070 |
| | $TB_{up}$ | 0.8031 | 0.00105 | 0.0031 | 0.8002 | 0.00161 | 0.0001 | 0.8001 | 0.00213 | 0.0001 |
| | TB | 0.8015 | 0.00102 | 0.0015 | 0.7986 | 0.00159 | -0.0014 | 0.7987 | 0.00207 | -0.0013 |
| | $TB_{low}$ | 0.8000 | 0.00106 | 0.0000 | 0.7971 | 0.00164 | -0.0029 | 0.7977 | 0.00210 | -0.0023 |
| 0.85 | KM | 0.8500 | 0.00062 | 0.0000 | 0.8502 | 0.00063 | 0.0002 | 0.8470 | 0.00064 | -0.0030 |
| | $TB_{up}$ | 0.8506 | 0.00109 | 0.0006 | 0.8499 | 0.00132 | -0.0001 | 0.9165 | 0.00650 | 0.0665 |
| | TB | 0.8487 | 0.00108 | -0.0013 | 0.8481 | 0.00127 | -0.0019 | 0.9066 | 0.00551 | 0.0566 |
| | $TB_{low}$ | 0.8468 | 0.00109 | -0.0032 | 0.8467 | 0.00133 | -0.0033 | 0.9064 | 0.00551 | 0.0564 |
| 0.9 | KM | 0.9003 | 0.00047 | 0.0003 | 0.9050 | 0.00041 | 0.0050 | 0.9165 | 0.00064 | 0.0165 |
| | $TB_{up}$ | 0.9020 | 0.00076 | 0.0020 | 0.9041 | 0.00114 | 0.0041 | 0.9165 | 0.00235 | 0.0165 |
| | TB | 0.9008 | 0.00074 | 0.0008 | 0.9024 | 0.00110 | 0.0024 | 0.9115 | 0.00213 | 0.0115 |
| | $TB_{low}$ | 0.8997 | 0.00075 | -0.0003 | 0.9014 | 0.00110 | 0.0014 | 0.9064 | 0.00237 | 0.0064 |
| 0.95 | KM | 0.9526 | 0.00025 | 0.0026 | 0.9573 | 0.00024 | 0.0073 | 0.9773 | 0.00087 | 0.0273 |
| | $TB_{up}$ | 0.9507 | 0.00044 | 0.0007 | 0.9510 | 0.00058 | 0.0010 | 0.9360 | 0.00151 | -0.0140 |
| | TB | 0.9496 | 0.00043 | -0.0004 | 0.9491 | 0.00057 | -0.0009 | 0.9348 | 0.00154 | -0.0152 |
| | $TB_{low}$ | 0.9484 | 0.00046 | -0.0016 | 0.9475 | 0.00062 | -0.0025 | 0.9329 | 0.00173 | -0.0171 |

Note: TB = $TB_{lin}$

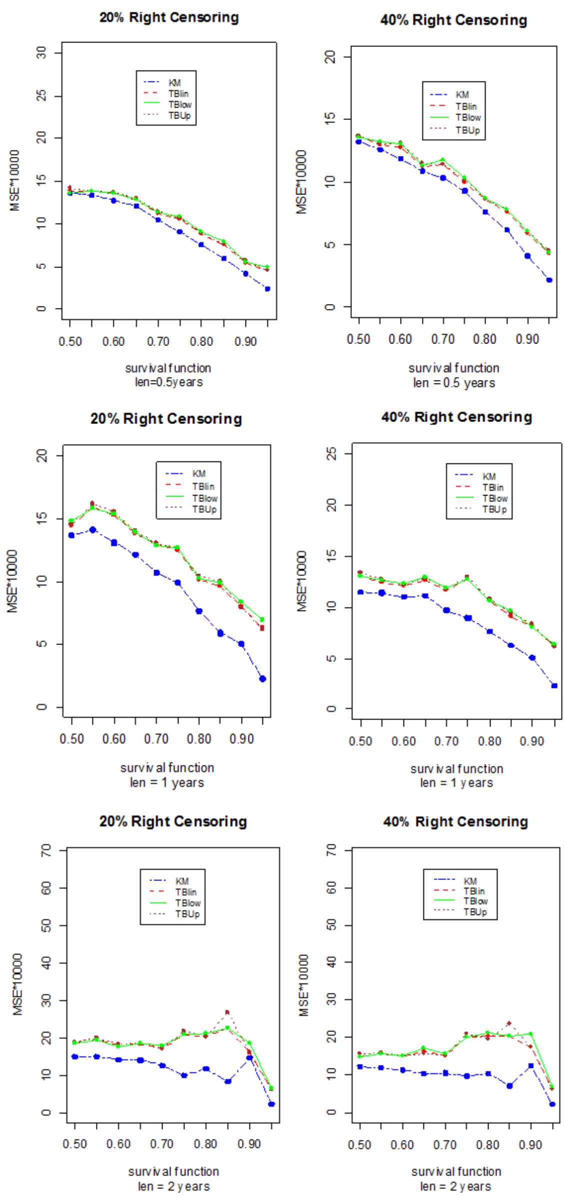


**Figure 2** The MSE×10,000 of $\hat{S}(t)$ by $TB_{lin}$, $TB_{low}$ and $TB_{up}$ for 20% and 40% of right censoring and shape parameter 0.8 over time between visit as 0.5, 1, 2 years respectively

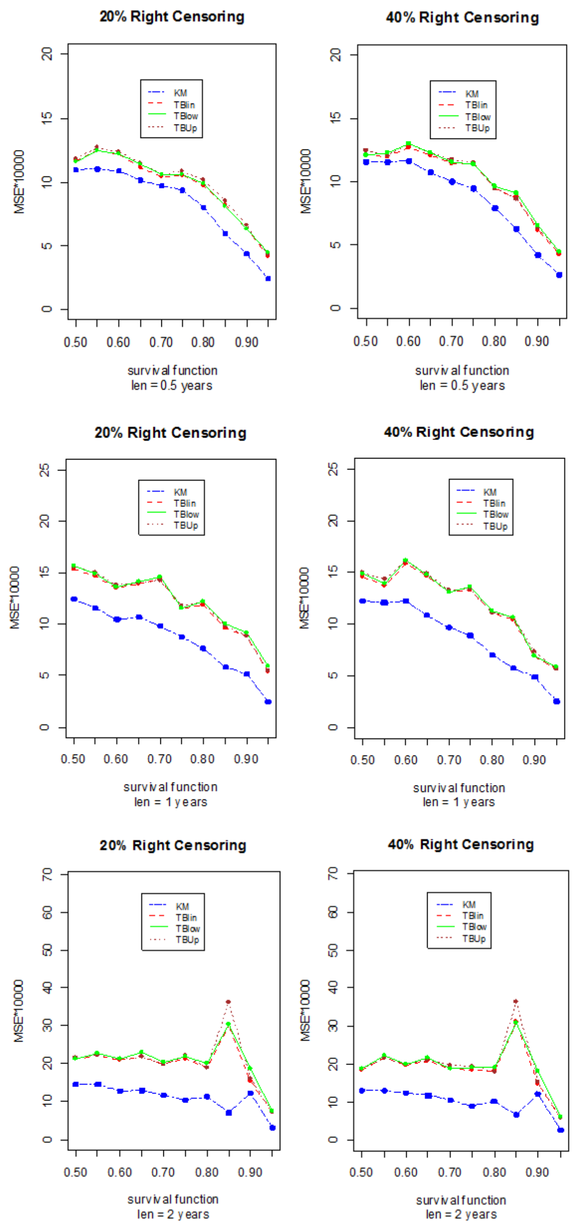


**Figure 3** The MSE×10,000 of $\hat{S}(t)$ by $TB_{lin}$, $TB_{low}$ and $TB_{up}$ for 20% and 40% of right censoring and shape parameter 1 over time between visit as 0.5, 1, 2 years respectively

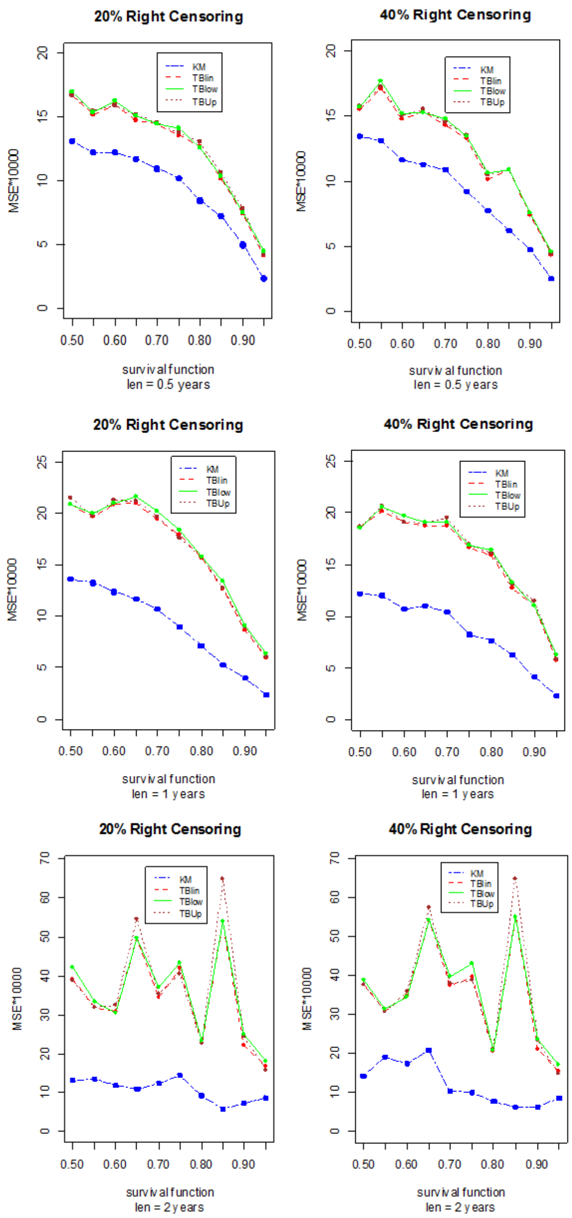


**Figure 4** The MSE×10,000 of $\hat{S}(t)$ by $TB_{lin}$, $TB_{low}$ and $TB_{up}$ for 20% and 40% of right censoring and shape parameter 1.5 over time between visit as 0.5, 1, 2 years respectively

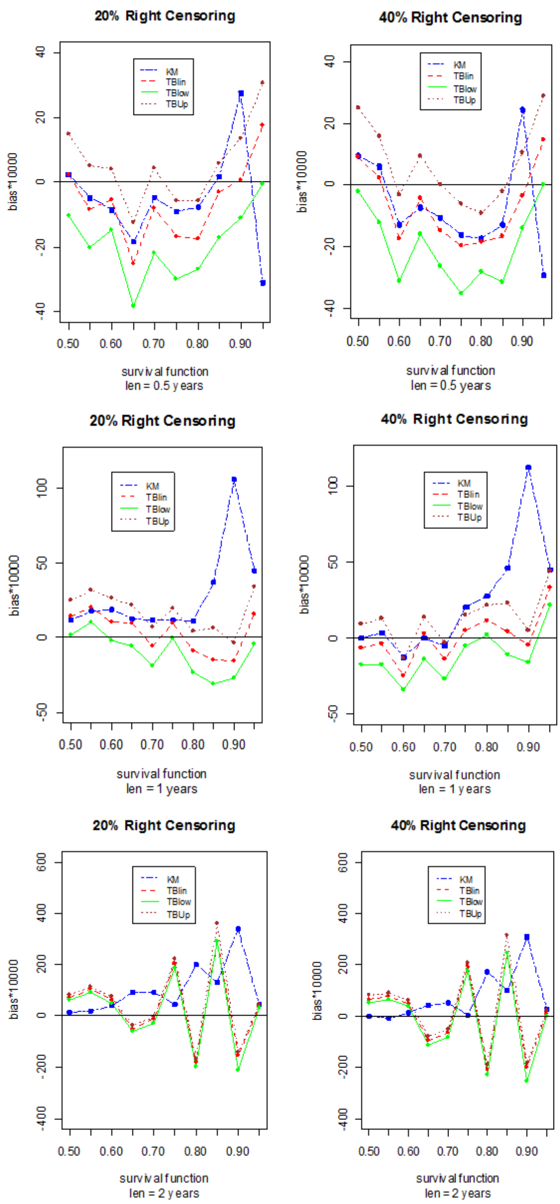


**Figure 5** The *BIAS*×10,000 of $\hat{S}(t)$ by $TB_{lin}$, $TB_{low}$ and $TB_{up}$ for 20% and 40% of right censoring and shape parameter 0.8 over time between visit as 0.5, 1, 2 years respectively

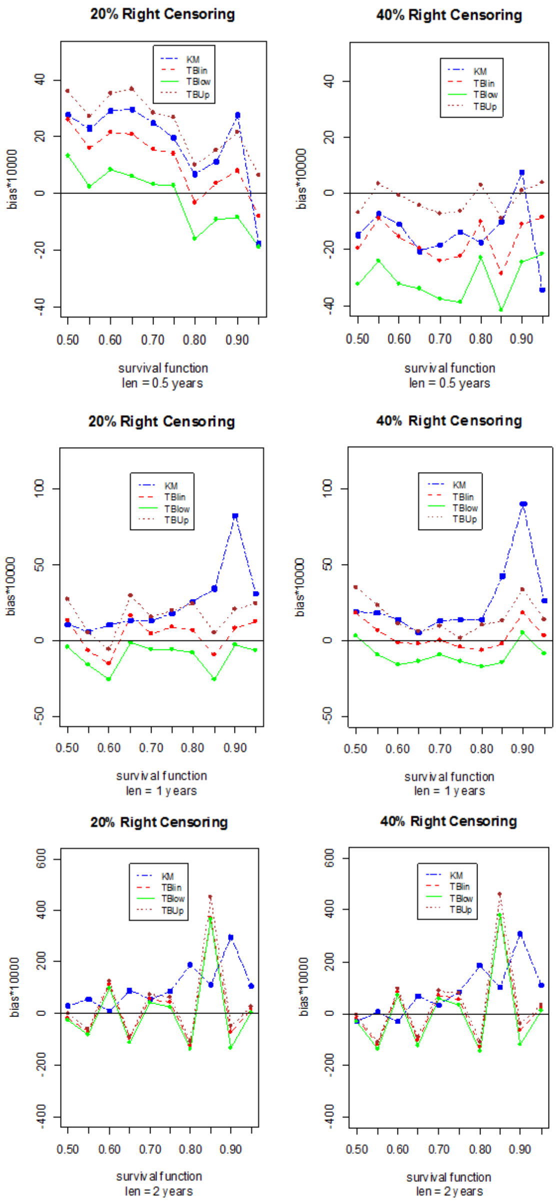


**Figure 6** The $BIAS\times10{,}000$ of $\hat{S}(t)$ by $TB_{lin}$, $TB_{low}$ and $TB_{up}$ for 20% and 40% of right censoring and shape parameter 1 over time between visit as 0.5, 1, 2 years respectively

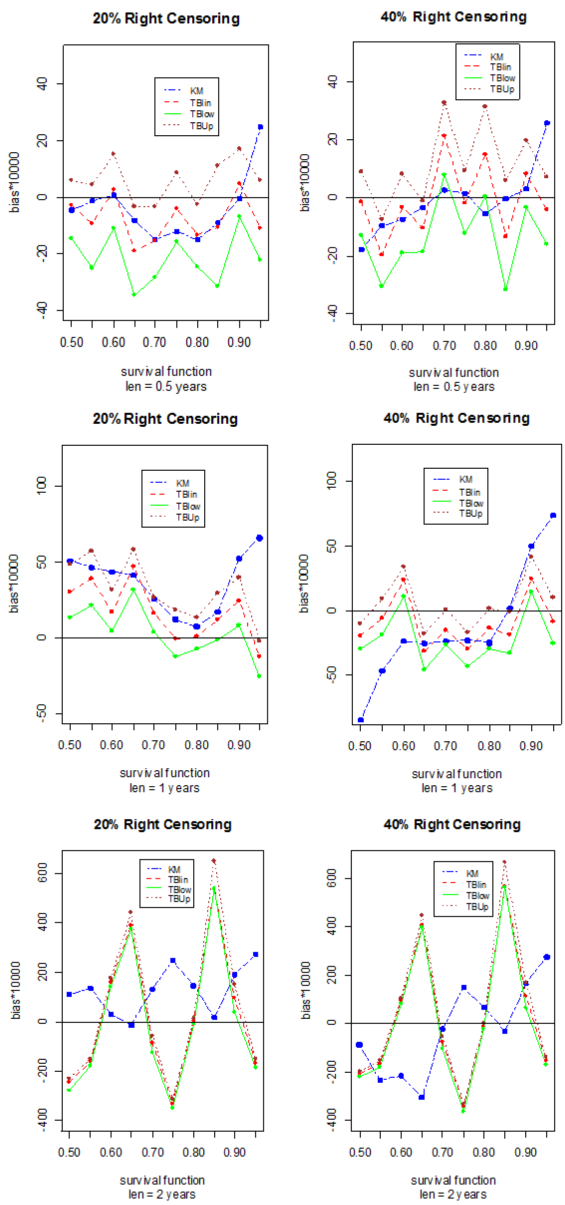


**Figure 7** The *BIAS*×10,000 of $\hat{S}(t)$ by $TB_{lin}$, $TB_{low}$ and $TB_{up}$ for 20% and 40% of right censoring and shape parameter 1.5 over time between visit as 0.5, 1, 2 years respectively

## 4. Conclusions

Although the TB estimator has been proposed over a past few decades, the KM estimator based on imputation techniques is still widely used for interval-censored data (Grover et al. 2013, Harezlak and Tu 2006, Støvring and Kristiansen 2011, Sun et al. 2013, and Williams et al. 2004).

However, the appropriateness of using the imputed KM estimator is still being argued. There are several comparisons of the performance of the TB and KM estimators with real data but there are a few studies via simulation. Typically, the simulation was set up to mimic the specific real data so that if there are several simulation studies, the useful results may be suggestion for researchers to apply the imputed KM method appropriately. In this paper we have compared the performances of the survival function estimators, KM and TB, when the time to event is subject to interval-censored data. Here, the distribution of time to event is assumed to be Weibull distribution which can also be viewed as a generalization of the exponential distribution because this is widely used in survival and reliability analysis. Also, we have compared among three shape parameters of Weibull distribution that represents the three types of hazard behavior (constant, increase, and decrease). Moreover, we compared the estimator of $S(t)$ by TB corresponding to three relationships as a linear interpolation and flats as a lower bound and an upper bound of the class estimators on each TB interval. The simulation study showed that TB with linear interpolation methods is better than the other two TB methods. Moreover, the point estimates from the TB with linear interpolation method and the KM method quite target to the true percentiles except the high percentile of time to event for the long time between visits as two years. In addition, the performance of TB estimator is close to the KM estimator, especially when the time between visits is 0.5 years. We found that all TB methods are influenced by both time between visits and shape parameter, whereas the KM method seem not to be influenced by shape parameter. In the other word, the KM method is sensitive only when time between visits is longer but all TB methods are influenced by shape parameter and time between visits. In addition, the simulation study also showed that the KM method always overestimates $S(t)$ at upper or equal to $85^{th}$ percentile and all TB methods are more varied when time between visits is two years. For the future work, the comparison in other distributions will be studied.

**References**

De Gruttola V, Lagakos SW. Analysis of doubly-censored survival data, with application to AIDS. Biometrics. 1989; 45: 1-11.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B (Methodological). 1977; 39: 1-38.

Fay MP, Shaw PA. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. J Stat Softw. 2010; 36: i02.

Gentleman R, Geyer CJ. Maximum likelihood for interval censored data: Consistency and computation. Biometrika. 1994; 81: 618-623.

Geskus RB. Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. Stat Med. 2001; 20: 795-812.

Giesecke J, Scalia-Tomba G, Berglund O, Berntorp E, Schulman S, Stigendal L. Incidence of symptoms and AIDS in 146 Swedish haemophiliacs and blood transfusion recipients infected with human immunodeficiency virus. BMJ. 1988; 297: 99-102.

Gómez G, Calle ML, Oller R, Langohr K. Tutorial on methods for interval-censored data and their implementation in R. Stat Modelling. 2009; 9(4): 259-297.

Gorelick J. Nonparametric analysis of interval-censored failure time data. PhD [dissertation]. Columbia(CO): University of Missouri-Columbia; 2009.

Groeneboom P, Wellner JA. Information bounds and nonparametric maximum likelihood estimation: Springer Science & Business Media; 1992.

Grover G, Das R, Swain PK, Deka B. On the estimation of survival of HIV/AIDS patients on antiretroviral therapy using NPMLE method: an application to interval censored data. Am J Math Stat. 2013; 3: 213-219.

Gu M, Zhang C-H. Asymptotic properties of self-consistent estimators based on doubly censored data. Ann Stat. 1993; 21: 611-624.

Harezlak J, Tu W. Estimation of survival functions in interval and right censored data using STD behavioural diaries. Stat Med. 2006; 25: 4053-4064.

Hebeisen M. Estimation from Interval-censored Time-to-event Data: Method Comparison by Simulation. [master's thesis]. Zurich(ZRH): University of Zurich; 2014.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958; 53: 457-481.

Law CG, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. Stat Med. 1992; 11: 1569-1578.

Lee CA, Phillips A, Elford J, Miller EJ, Bofill M, Griffiths PD, et al. The natural history of human immunodeficiency virus infection in a haemophilic cohort. Br J Haematol. 1989; 73: 228-234.

Lindsey JC, Ryan LM. Methods for interval-censored data. Stat Med. 1998; 17(2): 219-238.

Lui K-J, Darrow WW, Rutherford Gd. A model-based estimate of the mean incubation period for AIDS in homosexual men. Science. 1988; 240: 1333-1335.

Mariotto AB, Mariotti S, Pezzotti P, Rezza G, Verdecchia A. Estimation of the acquired immunodeficiency syndrome incubation period in intravenous drug users: a comparison with male homosexuals. Am J Epidemiol. 1992; 135: 428-437.

Nishikawa M, Tango T. Behavior of the Kaplan-Meier Estimator for Deterministic Imputations to Interval-Censored Data and the Turnbull Estimator. Japanese Journal of Biometrics. 2004; 24(2): 71-94.

Samuelsen SO, Kongerud J. Interval censoring in longitudinal data of respiratory symptoms in aluminium potroom workers: a comparison of methods. Stat Med. 1994; 13: 1771-1780.

Schechter MT, Craib KJ, Le TN, Willoughby B, Douglas B, Sestak P, et al. Progression to AIDS and predictors of AIDS in seroprevalent and seroincident cohorts of homosexual men. AIDS. 1989; 3(6): 347-353.

Støvring H, Kristiansen IS. Simple parametric survival analysis with anonymized register data: A cohort study with truncated and interval censored event and censoring times. BMC research notes. 2011; 4: 308.

Sun J. Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to AIDS studies. Biometrics. 1995; 51: 1096-1104.

Sun X, Li X, Chen C, Song Y. A review of statistical issues with progression-free survival as an interval-censored time-to-event endpoint. J Biopharm Stat. 2013; 23: 986-1003.

Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc Series B (Methodological). 1976; 38: 290-295.

Williams CF, Klinzman D, Yamashita TE, Xiang J, Polgreen PM, Rinaldo C, et al. Persistent GB virus C infection and survival in HIV-infected men. N Engl J Med. 2004; 350: 981-990.

Zhang W. Analysis of doubly censored survival data with applications to GBV-C and HIV studies. PhD [dissertation]. Iowa(IA): University of Iowa; 2005.

Zhang Y, Jamshidian M. On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. J Comput Graph Stat. 2004; 13: 123-140.