

# A novel estimator of survival without assumption of censored survival time

Shee-Ping Chen (✉ [asdinap@yahoo.com.tw](mailto:asdinap@yahoo.com.tw))

Buddhist Tzu Chi Stem Cells Centre

---

## Method Article

**Keywords:** survival, survival time, survival curve, survival estimator, survival analysis

**DOI:** <https://doi.org/10.21203/rs.3.rs-141531/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The Kaplan-Meier estimator is commonly used to analyze time-to-event data, but it may over estimate the survival function when censored events occur. Several methods have been developed to adjust the traditional Kaplan-Meier estimator. Most of these adjusted methods are based on their assumptions for survival time of censored events. We here propose a novel estimator of survival without assumption of censored survival time, and it gives sensible estimates of survival probabilities in various censoring conditions.

## Introduction

Survival analyses are common statistical methods in medical and clinical researches. The Kaplan-Meier (KM) estimator is a popular method to analyze time-to-event data<sup>1</sup>. However, in the traditional Kaplan-Meier method, censored subjects are considered to survive to the time on the next death so that it may over estimate the survival function. It is affected by censoring and would not be reliable in the case of heavy censoring. Several other methods have been designed to adjust the traditional Kaplan-Meier estimate<sup>2-4</sup>. Previous studies have compared these adjustment methods and reported that Weighted Kaplan-Meier (WKM) and Modified Weighted Kaplan-Meier (MWKM) estimators give better estimates in the case of heavy censoring<sup>5,6</sup>. WKM and MWKM estimators are almost the same estimates except for the estimate of last censored event. We propose a novel estimator of survival without assumption of censored survival time and compare it with the KM and WKM estimates.

## Results

### Principle of the estimator

The principle of our estimator is just to exclude censored subjects from dataset when they have been censored. The estimator takes the information before subjects were censored, but abandons any information after subjects were censored. In fact, we do not get any true information from subjects after they were censored. Our proposed estimator does not make any assumption of censored survival time because we have excluded them when they are censored.

### Estimator of survival function

$$\hat{S}(t) = \frac{s_t}{N - \sum_{r=0}^t c_r} \quad (1)$$

$\hat{S}(t)$  is the estimator of survival function for any time  $t$  when last events include death events. When last events are all censored events at time  $t$ ,  $\hat{S}(t)$  is the estimator of survival function for time 0 to the time just before time  $t$ .

$s_t$  is the number of observed alive subjects at time  $t$ .

$N$  is the total number of subjects at the start ( $t=0$ ) of observation.

$c_t$  is the number of censored subjects at time  $t$ , and  $\sum_{r=0}^t c_r$  is the number of cumulative censored subjects from time 0 to time  $t$ .

## Comparison

Our proposed estimator and KM estimator will estimate the same survival function when there is no any censored subject in the dataset. We compare our proposed estimator and describe advantages of our proposed estimator at various censored conditions. The following examples used for comparisons are simple and basic censored conditions, and true survival data would be the permutations and combinations of these basic censored conditions.

### Censored and death events at different time

When censored events and death events do not occur at the same time, the KM estimate treats censored events as survival to the time on next death, but WKM estimate treats censored events as death at censored time. Our proposed estimator treats censored events neither as survival nor death, but excludes censored subjects from dataset at their censored time. We compare the three estimates of the data listed in Table 1 and illustrate the survival curves as in Supplementary Fig. 1. In the example of Table 1, the first event at time 1 is

censored event. At time 1, both KM method and our proposed estimator give 100% survival probabilities, but WKM method estimate gives an 80% survival probability. At time 2, KM and our proposed estimator give 75% survival probabilities, but WKM methods give a 60% survival probability. However, KM estimator treats the censored subject as be excluded at time 2 but our proposed estimator treats the censored subject as be excluded at time 1.

Table 1. The first event is a censored event

Time	s	d	c	$\hat{S}(t)$ KM	$\hat{S}(t)$ WKM	$\hat{S}(t)$ Proposed
0	5	0	0	$\frac{5}{5} = 100\%$	$\frac{5}{5} \times \frac{5}{5} = 100\%$	$\frac{5}{5} = 100\%$
1	4	0	1	$\frac{5}{5} = 100\%$	$\frac{4}{5} \times \frac{5}{5} = 80\%$	$\frac{4}{5-1} = 100\%$
2	3	1	0	$\left(\frac{5}{5}\right) \times \left(\frac{3}{4}\right) = 75\%$	$\left(\frac{4}{5} \times \frac{5}{5}\right) \times \left(\frac{4}{4} \times \frac{3}{4}\right) = 60\%$	$\frac{3}{5-1} = 75\%$
4	3	0	0	End of study or observation		

s: survival; d: death; c: censored event

We compare the three estimates on another data listed in Table 2 and illustrate the survival curves as in Supplementary Fig. 2. In the example of Table 2, the first event is death event, and the three estimates produce the same survival probabilities at time 1. At time 2 and time 3, the three estimates give three different survival probabilities. In the condition, the survival probability estimated by our proposed estimator is between that are estimated by KM and WKM methods.

Table 2. The first event is a death event

Time	s	d	c	$\hat{S}(t)$ KM	$\hat{S}(t)$ WKM	$\hat{S}(t)$ Proposed
0	5	0	0	$\frac{5}{5} = 100\%$	$\frac{5}{5} \times \frac{5}{5} = 100\%$	$\frac{5}{5} = 100\%$
1	4	1	0	$\frac{4}{5} = 80\%$	$\frac{5}{5} \times \frac{4}{5} = 80\%$	$\frac{4}{5} = 80\%$
2	3	0	1	$\frac{4}{5} = 80\%$	$\left(\frac{5}{5} \times \frac{4}{5}\right) \times \left(\frac{3}{4} \times \frac{4}{4}\right) = 60\%$	$\frac{3}{5-1} = 75\%$
3	2	1	0	$\left(\frac{4}{5}\right) \times \left(\frac{2}{3}\right) \approx 53.3\%$	$\left(\frac{5}{5} \times \frac{4}{5}\right) \times \left(\frac{3}{4} \times \frac{4}{4}\right) \times \left(\frac{3}{3} \times \frac{2}{3}\right) = 40\%$	$\frac{2}{5-1} = 50\%$
4	2	0	0	End of study or observation		

s: survival; d: death; c: censored event

### Censored and death events at the same time

When censored events and death events occur at the same time, the KM estimate treats censored events as survival at that time point while the WKM method down-regulates the survival probability given by the traditional KM estimate. We compare the three estimates on the data listed in Table 3 and illustrate the survival curves as in Supplementary Fig. 3.

Table 3. Censored and death events occur at the same time

Time	s	d	c	$\hat{S}(t)$ KM	$\hat{S}(t)$ WKM	$\hat{S}(t)$ Proposed
0	5	0	0	$\frac{5}{5} = 100\%$	$\frac{5}{5} \times \frac{5}{5} = 100\%$	$\frac{5}{5} = 100\%$
1	3	1	1	$\frac{4}{5} = 80\%$	$\frac{4}{5} \times \frac{4}{5} = 64\%$	$\frac{3}{5-1} = 75\%$
2	1	1	1	$\left(\frac{4}{5}\right) \times \left(\frac{2}{3}\right) \approx 53.3\%$	$\left(\frac{4}{5} \times \frac{4}{5}\right) \times \left(\frac{2}{3} \times \frac{2}{3}\right) \approx 28.4\%$	$\frac{1}{5-2} \approx 33.3\%$
4	1	0	0	End of study or observation		

s: survival; d: death; c: censored event

In some extreme cases, our proposed estimator gives minimal survival probability among the three methods. We compare the three estimates on the dataset listed in Table 4.

Table 4. Lots of censored and death events occur at the same time

Time	s	d	c	$\hat{S}(t)$ KM	$\hat{S}(t)$ WKM	$\hat{S}(t)$ Proposed
0	1000	0	0	$\frac{1000}{1000} = 100\%$	$\frac{1000}{1000} \times \frac{1000}{1000} = 100\%$	$\frac{1000}{1000} = 100\%$
1	0	600	400	$\frac{400}{1000} = 40\%$	$\frac{600}{1000} \times \frac{400}{1000} = 24\%$	$\frac{0}{1000 - 400} = 0\%$

s: survival; d: death; c: censored event

In the example of Table 4, no survival is observed at time 1. Let us consider that if the 400 censored events are random among the total 1000 subjects, then the survival probability of the censored subjects should be the same with survival probability of observed subjects. In the example, there is no survival in the 600 observed subjects and survival probability of the 600 observed subjects is zero, so the survival probability of the 400 censored subjects should be also zero. Our proposed estimator gives zero survival probability of the total

1000 subjects at time 1. For the 40% survival probability estimated by KM method, it means that the 400 censored are all alive at time 1. It is almost impossible if the 400 censored events are random among the total 1000 subjects. For the 24% survival probability estimated by WKM method, we hypothesize that there is a true 24% survival probability of the 1000 subjects at time 1 and the 400 censored events are random among the total 1000 subjects. Then, the most possible obtainable data should be that listed in Supplementary Table 1 and Table 5. The data listed in Table 5 is obviously different from Table 4, and only our proposed estimator gives exactly the true 24% survival probability of the 1000 subjects at time 1.

Table 5. Comparison of estimates for the dataset with known survival probability

Time	s	d	c	$\hat{S}(t)$ KM	$\hat{S}(t)$ WKM	$\hat{S}(t)$ Proposed
0	1000	0	0	$\frac{1000}{1000} = 100\%$	$\frac{1000}{1000} \times \frac{1000}{1000} = 100\%$	$\frac{1000}{1000} = 100\%$
1	144	456	400	$\frac{544}{1000} = 54.4\%$	$\frac{600}{1000} \times \frac{544}{1000} = 32.6\%$	$\frac{144}{1000 - 400} = 24\%$

s: survival; d: death; c: censored event

### Censored and death events occur almost simultaneously

If the time interval of observation and record is short enough in the study, censored events and death events would not occur at the exact same time point. To examine if the estimates of the three methods are affected by the time bias of observations or records, we consider an example that there are three subjects at the start, and one death event and one censored event occur at approximate the same time point. We compare the three different records of the example as listed in Table 6 and illustrate the survival curves as in Supplementary Fig. 4. Among the three methods, only our proposed method estimates similar survival curves of the three datasets with a bit of time bias on observations or records.

Table 6. Comparison of estimates for similar datasets

Time	s	d	c	$\hat{S}(t)$ KM	$\hat{S}(t)$ WKM	$\hat{S}(t)$ Proposed
0	3	0	0	$\frac{3}{3} = 100\%$	$\frac{3}{3} \times \frac{3}{3} = 100\%$	$\frac{3}{3} = 100\%$
Record 1						
1	1	1	1	$\frac{2}{3} \approx 66.7\%$	$\frac{2}{3} \times \frac{2}{3} \approx 44.4\%$	$\frac{1}{3-1} = 50\%$
Record 2						
1	2	1	0	$\frac{2}{3} \approx 66.7\%$	$\frac{3}{3} \times \frac{2}{3} \approx 66.7\%$	$\frac{2}{3} \approx 66.7\%$
1+ $\Delta t$	1	0	1	$\frac{2}{3} \approx 66.7\%$	$\left(\frac{3}{3} \times \frac{2}{3}\right) \times \left(\frac{1}{2} \times \frac{2}{2}\right) \approx 33.3\%$	$\frac{1}{3-1} = 50\%$
Record 3						
1- $\Delta t$	2	0	1	$\frac{3}{3} = 100\%$	$\frac{2}{3} \times \frac{3}{3} \approx 66.7\%$	$\frac{2}{3-1} = 100\%$
1	1	1	0	$\frac{3}{3} \times \frac{1}{2} = 50\%$	$\left(\frac{2}{3} \times \frac{3}{3}\right) \times \left(\frac{2}{2} \times \frac{1}{2}\right) \approx 33.3\%$	$\frac{1}{3-1} = 50\%$

s: survival; d: death; c: censored event;  $\Delta t$  approaches to 0.

## Discussion

In this study, we propose a novel estimator of survival without assumption of censored survival time. We think the censored subjects in survival analyses do not contribute anymore information after they are censored. Our proposed estimator excludes them from the estimates of survival probability at their censored time points, and it is not necessary to make any assumption of censored survival time. It is easy to understand that the amounts of censored subjects are subtracted from numerator and denominator of the estimate formula in our proposed estimator. The survival probabilities estimated by our proposed estimator are sensible no matter the censored events and death events occur at the same time or at different time. Further, the survival curves estimated by our proposed estimator would not be disturbed by time bias of observation or record (Table 6 and Supplementary Fig. 4.).

Our proposed estimator and WKM method have a same feature that zero survival probabilities are estimated when the last events are all censored events. Shafiq *et al.* have presented a modified WKM method to resolve the feature for WKM method<sup>4</sup>. However, we do not think this is a defect of our proposed estimator. When the last events are all censored events in a dataset, it could be regarded as the end of observation or study. It is not necessary to estimate the survival function at and after that time point. We have noted that our proposed estimator only estimates survival function from the start to the

time just before the end of observation. The end of survival curve estimated by our proposed estimator will be illustrated as that estimated by traditional KM method.

## References

1. Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457-481.
2. Jan B, et al. Weighted Kaplan Meier estimation of survival function in heavy censoring. *Pak J Stat* 2005; 21: 55-63.
3. Huang ML. A weighted estimation method for survival function. *App Math Sci* 2008; 2: 753-762.
4. Shafiq M, Shah S, Alamgir M. Modified Weighted KM Estimator. *Pak J Stat Operat Res* 2007; 3: 39-44.
5. Ramadurai M, Ponnuraja C. Non-parametric estimation of the survival probability of children affected by TB meningitis. *Int Refereed Res J* 2011; II: 216-227.
6. Zare A, et al. A comparison between KM and weighted KM methods of five-year survival estimation of patients with gastric cancer. *Acta Med Iran* 2014; 52: 764-767.

## Competing Interests

The authors declare no competing interests.