Estimation of Survival Probabilities in the Presence of Ties Short Title: Survival Probabilities in the Presence of Ties

Qamruz Zaman[1], Muhammad Atif[1], Muhammad Iqbal[1], Aisha Bibi[2], Syed Habib Shah[3], Muhammad Farooq[4], Muhammad Rafiq[4], Karl-Peter Pfeiffer[5]

[1]Department of Statistics, University of Peshawar, Pakistan
[2]Department of Statistics, Postgraduate college Mansehra, Pakistan
[3]Deparment of Management Sciences, Kohat University of Science & Technology, Pakistan
[4]Department of Statistics, University of Malakand, Pakistan
[5]Department of Medical Statistics, Informatics and Health Economics,
Medical University Innsbruck, Austria
mqamruzzamankh@hotmail.com

Abstract: In the real world we usually face the problem of discrete survival times, typically associated with the presence of ties between events and censored observations. However, the conventional Kaplan-Meier approach, as well as Greenwood´s variance estimator, do not adequately consider this fact, which leads to underestimation of true survival probabilities and variances. In this paper we therefore present a modified Kaplan-Meier approach, by explicitly considering the presence of ties. A variance estimator based on our modified Kaplan-Meier approach is developed. In absence of ties the new variance estimator equals to Greenwood variance estimator, while in censoring free data, it reduces to binomial variance estimator. A simulation study was conducted in order to compare the performance of conventional Kaplan-Meier estimator and modified Kaplan-Meier estimator on different censoring percentages. Our simulation results suggest a significant improvement in terms of bias of modified Kaplan-Meier approach in comparison to conventional Kaplan-Meier estimator. Similarly, the results of variance simulation favour the proposed, modified variance estimator. Our new approaches are illustrated on a leukaemia data set.

[Zaman Q,Atif M, Iqbal M, Aisha B, Shah S H, Farooq M,Rafiq M, Pfeiffer KP. **Estimation of Survival Probabilities in the Presence of Ties.** *Life Sci J* 2014;11(10s):155-164]. (ISSN:1097-8135). http://www.lifesciencesite.com. 25

Keywords: Kaplan-Meier Estimator, Modified Kaplan-Meier Estimator, Discrete Survival Times, Ties, Greenwood Variance Estimator.

1 Introduction

Over the last decades, statistical techniques of Survival analysis play increasingly important roles in biostatistics and modern medical research (Horton et al., 1979; Strasak et al., 2006; Emerson et al., 1983; Fleming et al., 1984). Due to considerable methodological development concerning this statistical area, today, methods of Survival analysis exist in manifold parametric, semi-parametric and non-parametric forms. However, the famous Kaplan-Meier Survival function is the most commonly used non-parametric method for estimating survival probabilities in medical research (Kaplan et al., 1958).

Survival analysis is different from conventional statistical procedures due to the concept of censoring. Let T_1, T_2, …, T_n be independently identically distributed survival times having the distribution function F(t) and let G(c) be the distribution function of independently identically distributed censoring times C_1, C_2, …, C_n. T_i and C_i are assumed to be independent. Let $X_i=\min\{T_i, C_i\}$ be the observed survival time and $\delta_i=I(T_i \le C_i)$ indicate whether the survival time is censored or event. Given this context, the problem is to estimate the survival probability $S(t)=1-F(t)$ from the ordered survival time X_i and from corresponding δ_i. Let the number of individuals who are alive just before time T_i, including those who are about to die at this time, be denoted by r_i and e_i to denote the number of persons who die at this time, the conventional Kaplan-Meier estimator is defined as

$$\hat{S}_{KM}(t)=\prod_{i=1}^{n}\left(\frac{r_i-e_i}{r_i}\right) \quad \ldots\ldots \quad (1)$$

However, the conventional Kaplan-Meier estimator considers time to be continuous, although this does not necessarily hold true for all real situations. For example, it may easily happen that a patient visits hospital after specific time periods (e.g. every week, month or quartile) and therefore visiting times are assumed to be discrete and not continuous. In the other hand, discrete survival times are typically associated with the presence of ties between censored and event time. This fact is not adequately considered by the conventional Kaplan-Meier method, as in case of ties censored observations are ignored(Collet, 1994).

Similarly, Greenwoods variance estimator, commonly used in Survival analysis and the formula of jump size do not give any attention to possible presence of ties in discrete survival times(Greenwood, 1926; Maller, 1996).

In this paper we present a modified estimator for Kaplan-Meier Survival function by explicitly considering the presence of ties. In section 2 we develop a modified approach for the Kaplan-Meier estimator and derive a variance formula for our modified survival function. A modified formula for the jump size, considering tie-cases is presented, as well. In section 3, two different simulation studies are performed in order to compare potential bias of conventional Kaplan-Meier estimator with modified Kaplan-Meier estimator and our new variance estimator with Greenwood variance estimator. In section 4, we apply our methods to the leukaemia data set of Freireich et. Al (1963). Although some previous authors already considered the idea of discrete survival times, to our knowledge up to now there is no work explicitly related to survival probabilities in the presence of ties (Hogan et al., 1998; Brookmeyer et al., 2002; Lam et al., 2003; Adebayo et al., 2005). When using the conventional Kaplan-Meier Survival function in this case, estimated survival probabilities are regularly underestimated. Our new approach leads to better results in terms of bias and variance.

2 Method

Suppose that a sample of survival data (X_i, δ_i), i=1, 2,…, n is generated from survival and censoring times. We assume the time to be discrete and the presence of ties among censoring times and event times. We further suppose that r_i denotes the number of persons at risk prior to time T_i, e_i to denote the number of events and c_i the number of censored observations at time T_i.

2.1 Modified Kaplan-Meier Survival Function

We define the modified function in two steps:

Step 1: We consider c_i and e_i to occur together at time T_i and adopt the new procedure by first subtracting c_i from r_i in order to obtain the adjusted number of persons at risk at time T_i.

$$r_i^{'} = r_i - c_i \quad \ldots\ldots \quad (2)$$

By first ignoring censoring in order to calculate censoring free survival probability at time T_i, we replace r_i by $r_i^{'}$ in the conventional Kaplan-Meier Survival function and derive

$$p_i = \left(\frac{r_i^{'} - e_i}{r_i^{'}} \right) \quad \ldots\ldots \quad (3)$$

Step 2: We now consider the second factor i.e. c_i. In order to incorporate the concept of ties into our new approach and to obtain the modified survival probability at time T_i, we use the arithmetic mean of c_i and p_i, considering p_i to be a single observation

$$p_i^* = \left(\frac{p_i + c_i}{(c_i + 1)} \right) \quad \ldots\ldots \quad (4)$$

There are mainly two reasons for choosing the arithmetic mean instead of median, although median is the most commonly used method of central tendency in survival analysis. First, we are considering all tie cases having same observed times and so there is no chance of occurrence of outliers. Second, if choosing the median instead of the arithmetic mean, at tie-times there are more chances of getting survival probabilities greater than 1, which is unrealistic and statistically not possible. By using the concept of Product-Limit probability we derive our modified Kaplan-Meier Survival function

$$S^*(t) = p_1^* * p_2^* * \ldots * p_i^* * \ldots * p_n^* \quad \ldots\ldots \quad (5)$$

By simplifying we get

$$S^*(t) = \prod_{i=1}^{n} \left(\frac{r_i^{'} - e_i + r_i^{'} * ci}{r_i^{'} * (c_i + 1)} \right) \quad \ldots\ldots \quad (6)$$

However, if no ties are present in the data $r_i^{'}$ equals to r_i and $S^*(t)$ reduces to $S(t)$ which corresponds to the conventional Kaplan-Meier estimator.

2.2 Variance Estimator Of Modified Kaplan-Meier Estimator

Greenwood variance is the most commonly used variance estimator for calculating confidence intervals in survival probabilities and is given by

$$\operatorname{var}\left(\hat{S}_{KM}(t) \right) \approx \left[\hat{S}_{KM}(t) \right]^2 \sum_{i=1}^{n} \frac{e_i}{r_i (r_i - e_i)} \quad (7)$$

where $\hat{S}_{KM}(t)$ denotes the Kaplan-Meier Survival probability. Just like the conventional Kaplan-Meier Survival function, also Greenwood's variance estimator does not consider the possibility of ties. However, in order to calculate confidence intervals and other statistics for our new approach, one also needs a modified variance estimator.

We obtain the modified variance estimator for modified Kaplan-Meier Survival function by using the delta method (Collet, 1994) . Considering equation 5, taking log and variance on both sides, we obtain

$$\operatorname{var}\left(\log\left(S^*(t) \right) \right) = \sum_i \operatorname{var}\left(\log p_i^* \right) \quad \ldots\ldots \quad (8)$$

By applying the delta method on the right hand side of equation 8 we derive

$$\operatorname{var}\left(\log\left(p_i^*\right)\right) \approx \left(\frac{1}{p_i^*}\right)\frac{\left(1-p_i^*\right)}{r_i} \quad \ldots\ldots \quad (9)$$

By putting equation 9 into equation 8 we derive

$$\operatorname{var}\left(\log\left(S^*(t)\right)\right) \approx \sum\left(\frac{1}{p_i^*}\right)\frac{1-p_i^*}{r_i} \quad (10)$$

By applying delta method on $\operatorname{var}\left(\log\left(S_i^*(t)\right)\right)$ we get

$$\operatorname{var}\left(\log\left(S^*(t)\right)\right) \approx \left(\frac{1}{S^*(t)}\right)^2 \operatorname{var}\left(S^*(t)\right) \quad \ldots(11)$$

By simplifying we derive our modified variance estimator

$$\operatorname{var}\left(S^*(t)\right) \approx \left[S^*(t)\right]^2 \sum_{i=1}^{n} \frac{e_i}{r_i\left(r_i' c_i + \left(r_i' - e_i\right)\right)} \quad (12)$$

Again, if events and censored observations do not occur at the same time then $c_i=0$ and $r_i' = r_i$, the modified variance estimator reduces to conventional Greenwood variance estimator. If the data is free from censoring, the modified variance estimator equals to the binomial variance estimator.

2.3 Modified Jump Size

We now can write the modified Kaplan-Meier Survival function as

$$S^*(t) = \prod_{i=1}^{j}\left(\frac{\left(\frac{r_i' - e_i}{r_i'}\right) + c_i}{\left(c_i + 1\right)}\right) \quad \ldots\ldots \quad (13)$$

and so the distribution function is

$$F^*(t) = 1 - \prod_{i=1}^{j}\left(\frac{\left(\frac{r_i' - e_i}{r_i'}\right) + c_i}{\left(c_i + 1\right)}\right) \quad \ldots\ldots \quad (14)$$

Just like the conventional Kaplan-Meier estimator, the modified Kaplan-Meier estimator is constant at point T_i; at this point its jump is of magnitude

$$\Delta F\left(t_{(i)}\right) = F\left(t_{(i)}\right) - F\left(t_{(i)}-\right) \quad \ldots\ldots \quad (15)$$

By substituting the relevant distribution functions in the above equation we derive

$$\Delta F\left(t_{(i)}\right) = \prod_{i=1}^{j-1}\left(\frac{\left(r_i' - e_i\right) + r_i' * c_i}{r_i'\left(c_i + 1\right)}\right) - \prod_{i=1}^{j}\left(\frac{\left(r_i' - e_i\right) + r_i' * c_i}{r_i'\left(c_i + 1\right)}\right) \quad (16)$$

By simplifying we obtain the modified Jump Size estimator

$$\Delta F\left(t_{(i)}\right) = \left(\frac{e_i}{r_i'\left(c_i + 1\right)}\right)\prod_{i=1}^{j-1}\left(\frac{\left(r_i' - e_i\right) + r_i' * c_i}{r_i'\left(c_i + 1\right)}\right) \quad \ldots\ldots \quad (17)$$

Again, in the absence of ties, our modified Jump Size estimator reduces to the conventional Jump Size estimator.

3 Simulation Study

Although we considered survival times to be discrete, we choose commonly used continuous survival distributions to draw survival and censoring times in R as compared to discrete distributions, continuous survival distributions are easy to handle and understand (R Development Core Team, 2004).

In order to incorporate tie-cases into the distributions, we now converted the continuous times into discrete times by choosing the 0 decimal point for the comparison of Kaplan-Meier and modified Kaplan-Meier estimator in terms of bias and the 1 decimal point for the comparison of Greenwood's variance estimator with modified variance estimator. Both decimal points allow for the likewise inclusion of tie-cases and there is no difference of considering either 0 or 1 regarding the results of our simulation approach. We obtained the same results by considering the reverse ordering of decimal points (results not shown).

3.1 Comparison Of Kaplan-Meier And Modified Kaplan-Meier Survival Functions

The performance of our modified Kaplan-Meier estimator was compared with conventional Kaplan-Meier estimator by simulating 1000 samples of different sizes (n=30, 50, 70 and 100) from the following survival distributions of survival times:

1. W (λ, α): Weibull with $S(t) = \exp\left(-\lambda t^{\alpha}\right)$
2. logL (λ, α): Log-logistic with $S(t) = 1/\left(1 + \lambda t^{\alpha}\right)$

Censoring was performed using the uniform distribution ranging from 0 to b. In order to obtain ties we chose the zero decimal point with different values of λ, α and b, both in the survival and censoring distributions. As in common practice, we choose three points (q_1=0.25, q_2=0.50 and q_3=0.75) at the theoretical survival distributions. In order to check whether our estimator leads to better results in different settings, we considered small sample size (n=30) as well as large sample size (n=100).

Table 1 summarizes the results of the study by means of bias of conventional Kaplan-Meier estimator and of our modified Kaplan-Meier estimator along with the corresponding sum of ties and percentage of censoring in n*1000 samples. The results show that bias of the modified estimator does not only depend on censoring but on sample size too. Column 5, containing the number of ties per simulation, shows that the performance of our new estimator in terms of bias, increases with the increase of ties per simulation, as well as with the increase of sample size. Both survival distributions (Weibull and log-logistic) have the same behaviour regarding the censoring distribution. The results strongly indicate that with increasing sample size, bias of our new method decreases, as compared to bias of conventional Kaplan-Meier Survival estimator (Figure 1).

3.2 Comparison Of Greenwood Variance Estimator And Modified Variance Estimator

As with an increasing sample size the chances of the presence of ties increases, the second part of our simulation study comprised 1000 simulations of comparatively large samples (n=50, 75, 100, 125 and 150),. We selected the exponential distribution with one decimal point for both the censoring as well as the survival times. Instead of conventionally considering three quartiles, we preferably choose the five higher percentiles (p_{50}, p_{60}, p_{70} and p_{80} and p_{90}), as Greenwood's variance regularly underestimates the true values at the right tail of survival distribution (Peto et al., 1977). However, due to the same pattern of behaviour at the 70th, 80th and 90th percentile in our simulation study, the results of the 90th percentile are not mentioned separately.

Table 2 shows the simulated standard deviations of modified Kaplan-Meier, mean estimated standard errors of Greenwood's based on Kaplan-Meier and mean standard errors of modified variance equation at the four percentiles, using Borkowf method of simulation (Borkowf, 2005). The results show that Greenwood's standard error extremely underestimates the simulated standard deviations at all four selected points and in all samples under investigation. Although in case of very heavy censoring, modified standard error also underestimates the simulated standard deviation, it still gives better results at higher percentiles compared to Greenwood's variance for each sample size (Figure 2). The same results we obtained by using different survival times and censoring times distributions.

4 Application To Leukaemia Data Set

We compared conventional Kaplan-Meier estimator and Greenwood estimator with our modified estimators on the famous leukaemia data set by Freireich et al. consisting of the survival times of 21 clinical patients, including 9 events and 12 censored observations. The set consists of weeks in maintenance of remission for leukaemia patients treated with 6-mercaptopurine and contained two ties at week 6 and 10. The weeks in remission are: 6, 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*, where (*) denotes a censored observation.

Table 3 summarizes data and methods. Column 1 shows the time in weeks, column 2 represents the events at different time points, followed by the arrangements of censored observations. As there were two ties in the data set, column 5 contains two different values from column 3. In this column we considered tie-cases and subtracted the number of censored observations from the number of persons at risk, yielding the number of persons at risk prior to and free of censoring at that time. Column 6 and 7 give the survival function values based on conventional Kaplan-Meier estimator and modified Kaplan-Meier estimator. Greenwood´s standard errors and our modified standard errors are given in the following columns.

Due to the occurrence of a tie at the first observed time, our modified Kaplan-Meier estimator performs better than conventional Kaplan-Meier estimator right from the start (Figure 3). For example at time 6, the value obtained by conventional Kaplan-Meier Survival function equals 0.857 and that obtained by modified Kaplan-Meier Survival function corresponds to 0.925. As both methods are types of Product-Limit method, our modified Kaplan-Meier approach leads to better results in terms of survival functions right from the first observed survival time. However, if ties occur only in the middle or later stages, our new method improves the curve right from these points.

In respect to standard errors, our modified estimator overcomes the right-tail underestimation-problem of Greenwood variance estimator. For example, regarding the last two events at time 6 and 7, the values of Greenwood standard error are smaller than the corresponding values of our modified standard error.

5 Discussion

Although the method of Kaplan-Meier is frequently used for univariate statistical analysis of survival data, it cannot sufficiently reflect real world problems, as the occurrence of discrete survival times,

associated with the presence of ties, is not incorporated in this concept. In case of ties, the conventional Kaplan-Meier function gives high rank to censored observations, thereby considering it as a non-informative factor (Marubini, 1995). This leads to problems regarding underestimation of survival probabilities and variance. Although we also considered censoring to be non-informative, we acknowledged that in case of ties, it still carries information. To utilize this information and to overcome the problem of ties, typically associated with discrete survival data, we developed modified Kaplan-Meier and variance estimators. The simulation study performed shows that our modified estimator gives better results in terms of bias as compared to conventional Kaplan-Meier Survival function, as well as it gives an improvement in survival curve.

Just like the conventional Kaplan-Meier Survival function, also Greenwood variance estimator does not pay any attention to possible presence of ties. Concerning this problem, we developed a modified variance estimator. The comparison of our modified variance estimator with Greenwood variance estimator by means of simulation study, shows that the modified variance estimator performs better in case of small as well as heavy censoring. Our simulation approach revealed that if censoring is in-between 30 and 60%, there are higher chances of ties as compared to very small or very heavy percentages of censoring. To check the performance of our new approaches, we applied the modified and conventional methods to a small leukaemia data set. Although this set consisted of only two tie-cases, it still gave clear impression of the superior performance of our modified methods.

On basis of our simulation study and analysis results we reach to the conclusion that as sample size increases, chances of presence of ties increase and performance of our modified estimators increases as well.

Table 1. Bias of conventional Kaplan-Meier and modified Kaplan-Meier estimator.*

Dist. of survival	Dist of censoring	Per. Censoring	n	Ties per simulation	Bias					
					Q_1		Q_2		Q_3	
					MKM	KM	MKM	KM	MKM	KM
W(0.5, 1)	U(2, 25)	4	30	0.959	-0.047	-0.044	0.029	0.034	0.233	0.195
			50	1.467	-0.040	-0.039	0.046	0.030	0.250	0.198
			70	1.893	-0.039	-0.040	0.049	0.034	0.277	0.199
			100	2.396	-0.033	-0.036	0.054	0.027	0.283	0.199
W(0.5, 25)	U(2, 25)	50	30	5.038	0.100	0.129	0.390	0.433	-0.178	-0.174
			50	9.040	0.102	0.159	0.404	0.456	-0.172	-0.167
			70	12.225	0.105	0.177	0.405	0.464	-0.169	-0.163
			100	15.868	0.090	0.192	0.394	0.469	-0.169	-0.162
W(0.5, 50)	U(2, 25)	60	30	4.845	0.063	0.079	0.316	0.388	-0.245	-0.223
			50	8.993	0.057	0.105	0.323	0.429	-0.238	-0.215
			70	12.735	0.055	0.121	0.320	0.450	-0.230	-0.208
			100	16.742	0.046	0.135	0.304	0.463	-0.232	-0.205
W(0.5, 100)	U(2, 25)	70	30	4.275	0.034	0.038	0.223	0.292	-0.349	-0.286
			50	8.285	0.025	0.049	0.217	0.351	-0.324	-0.251
			70	11.831	0.019	0.059	0.208	0.384	-0.320	-0.241
			100	16.011	0.012	0.066	0.177	0.419	-0.342	-0.236
W(0.5, 400)	U(2, 25)	83	30	2.806	0.012	0.007	0.093	0.106	-0.628	-0.497
			50	5.624	0.007	0.007	0.099	0.139	-0.624	-0.415
			70	8.352	0.005	0.009	0.089	0.152	-0.643	-0.371
			100	11.574	0.002	0.010	0.068	0.160	-0.656	-0.323
W(1.5, 0.5)	U(0, 3)	6	30	1.183	0.027	0.046	0.163	0.309	0.393	0.328
			50	1.545	0.014	0.037	0.151	0.306	0.392	0.327
			70	1.667	0.009	0.034	0.114	0.292	0.401	0.328
			100	1.862	0.003	0.027	0.050	0.264	0.473	0.323
W(1.5, 1.5)	U(0, 10)	9	30	1.855	-0.055	-0.028	0.131	0.190	0.189	0.074
			50	2.512	-0.059	-0.031	0.133	0.198	0.216	0.063
			70	2.899	-0.060	-0.028	0.114	0.204	0.215	0.053
			100	3.350	-0.072	-0.033	0.111	0.205	0.231	0.047
W(1.5, 5)	U(2, 25)	10	30	2.306	0.185	0.195	0.420	0.423	-0.203	-0.206

			50	3.616	0.193	0.202	0.422	0.427	-0.198	-0.202
			70	4.665	0.193	0.206	0.428	0.429	-0.188	-0.197
			100	5.982	0.197	0.205	0.427	0.428	-0.183	-0.184
W(1.5, 0.5)	U(0, 1.5)	34	30	1.992	0.000	0.114	-0.255	-0.134	0.082	0.364
			50	2.000	0.000	0.124	-0.367	-0.141	-0.198	0.400
			70	2.000	0.000	0.135	-0.368	-0.143	-0.222	0.417
			100	2.000	0.000	0.149	-0.368	-0.145	-0.223	0.427
W(1.5, 25)	U(2, 25)	66	30	4.681	0.044	0.055	0.245	0.332	-0.349	-0.292
			50	8.699	0.044	0.074	0.275	0.388	-0.330	-0.273
			70	12.282	0.036	0.086	0.252	0.426	-0.331	-0.267
			100	15.903	0.022	0.096	0.236	0.440	-0.347	-0.264
W(1.5, 50)	U(2, 25)	86	30	2.535	0.014	0.003	0.092	0.097	-0.710	-0.584
			50	5.133	0.007	0.005	0.073	0.100	-0.719	-0.488
			70	7.445	0.005	0.006	0.071	0.113	-0.726	-0.435
			100	10.406	0.002	0.005	0.054	0.124	-0.730	-0.374
W(1.5, 1.5)	E(1)	9	30	1.812	-0.058	-0.026	0.124	0.191	0.195	0.085
			50	2.484	-0.057	-0.031	0.139	0.200	0.213	0.058
			70	2.905	-0.063	-0.030	0.115	0.206	0.211	0.046
			100	3.298	-0.067	-0.034	0.105	0.205	0.222	0.040
W(15, 1.5)	E(0.8)	47	30	2.624	0.000	0.074	0.130	0.350	-0.104	0.013
			50	2.878	0.000	0.095	0.085	0.377	0.025	0.008
			70	2.954	0.000	0.117	0.041	0.398	-0.089	0.004
			100	2.989	0.000	0.125	0.008	0.412	-0.309	0.002
Llog(1.5, 2)	U(2, 15)	2	30	0.491	0.041	0.042	0.073	0.070	0.326	0.336
			50	0.752	0.049	0.050	0.068	0.061	0.324	0.337
			70	1.039	0.058	0.053	0.067	0.057	0.325	0.340
			100	1.415	0.064	0.056	0.077	0.059	0.334	0.338
Llog(1.5, 1)	U(2, 2.5)	20	30	2.586	0.007	0.097	-0.331	-0.193	0.121	0.087
			50	2.895	0.001	0.096	-0.433	-0.178	0.090	0.072
			70	2.974	0.000	0.092	-0.477	-0.163	-0.020	0.061
			100	2.997	0.000	0.095	-0.495	-0.158	-0.274	0.064
Llog(1.5, 0.01)	U(2, 5)	99	30	0.248	0.000	0.000	0.000	0.000	0.007	0.009
			50	0.375	0.000	0.000	0.000	0.000	0.001	0.002
			70	0.533	0.000	0.000	0.000	0.000	0.000	0.001
			100	0.723	0.000	0.000	0.000	0.000	0.000	0.000
Llog(0.5, 1)	U(2, 5)	34	30	4.102	0.052	0.141	-0.204	-0.099	0.016	0.045
			50	5.084	0.034	0.161	-0.237	-0.083	0.027	0.062
			70	5.451	0.014	0.163	-0.273	-0.086	0.000	0.068
			100	5.743	0.006	0.168	-0.315	-0.081	-0.016	0.080
Llog(0.5,0.1)	U(2, 5)	62	30	4.676	0.002	0.049	0.115	0.312	-0.285	-0.152
			50	5.528	0.000	0.052	0.060	0.351	-0.331	-0.146
			70	5.820	0.000	0.053	0.022	0.368	-0.427	-0.141
			100	5.941	0.000	0.053	0.006	0.387	-0.635	-0.139
Llog(1,1)	U(2, 50)	6	30	1.187	-0.037	-0.037	-0.094	-0.100	0.175	0.191
			50	1.903	-0.034	-0.035	-0.076	-0.094	0.187	0.208
			70	2.440	-0.031	-0.033	-0.073	-0.089	0.198	0.222
			100	3.166	-0.027	-0.032	-0.062	-0.084	0.210	0.231
Llog(1, 0.1)	U(2, 50)	33	30	3.871	-0.040	-0.021	-0.030	-0.020	-0.034	-0.039
			50	7.327	-0.034	-0.016	-0.025	-0.020	-0.035	-0.039
			70	10.479	-0.033	-0.013	-0.024	-0.018	-0.028	-0.033
			100	14.502	-0.033	-0.012	-0.023	-0.017	-0.029	-0.035

* KM=Kaplan-Meier Survival function, MKM=Modified Kaplan-Meier.

Table 2. Simulated standard deviations of modified Kaplan-Meier (MKM), mean standard errors of Greenwood´s variance estimator and modified variance estimator at four percentiles.

Dist of survival	Dist of censoring	p	t	n	Simulated SDs of MKM				Mean SEs of Greenwood				Mean SEs of modified variance			
					$p_{.5}$	$p_{.6}$	$p_{.7}$	$p_{.8}$	$p_{.5}$	$p_{.6}$	$p_{.7}$	$p_{.8}$	$p_{.5}$	$p_{.6}$	$p_{.7}$	$p_{.8}$
E(2)	E(18)	83	2277	50	0.161	0.174	0.235	0.288	0.049	0.067	0.082	0.089	0.063	0.074	0.130	0.189
		83	2491	75	0.144	0.171	0.265	0.334	0.047	0.076	0.090	0.079	0.049	0.080	0.163	0.288
		83	2651	100	0.136	0.162	0.250	0.346	0.051	0.084	0.093	0.072	0.057	0.090	0.138	0.313
		83	2754	125	0.125	0.150	0.231	0.346	0.046	0.075	0.089	0.064	0.048	0.082	0.112	0.286
		83	2823	150	0.127	0.147	0.218	0.328	0.043	0.071	0.086	0.057	0.046	0.080	0.100	0.229
E(22)	E(10)	14	1539	50	0.207	0.225	0.242	0.257	0.066	0.065	0.064	0.064	0.174	0.224	0.278	0.304
		14	1696	75	0.207	0.231	0.257	0.258	0.054	0.053	0.053	0.052	0.151	0.206	0.300	0.303
		14	1792	100	0.201	0.229	0.268	0.257	0.047	0.045	0.047	0.047	0.135	0.189	0.313	0.307
		14	1872	125	0.187	0.228	0.280	0.245	0.042	0.043	0.044	0.045	0.106	0.169	0.315	0.309
		14	1938	150	0.173	0.216	0.289	0.247	0.038	0.038	0.038	0.038	0.101	0.143	0.317	0.310
E(20)	E(10)	16	1671	50	0.208	0.231	0.247	0.253	0.067	0.069	0.069	0.070	0.172	0.237	0.338	0.374
		16	1821	75	0.196	0.233	0.263	0.246	0.055	0.056	0.056	0.056	0.129	0.195	0.340	0.360
		16	1907	100	0.180	0.228	0.280	0.243	0.048	0.048	0.048	0.048	0.101	0.170	0.347	0.355
		16	1969	125	0.165	0.216	0.287	0.252	0.043	0.045	0.043	0.043	0.100	0.143	0.285	0.343
		16	2047	150	0.158	0.206	0.288	0.276	0.039	0.039	0.040	0.040	0.102	0.133	0.275	0.326
E(15)	E(7)	18	2001	50	0.176	0.230	0.254	0.215	0.064	0.070	0.071	0.069	0.134	0.227	0.300	0.312
		18	2186	75	0.175	0.220	0.270	0.234	0.052	0.057	0.057	0.057	0.140	0.206	0.311	0.342
		18	2316	100	0.176	0.217	0.263	0.274	0.047	0.050	0.050	0.050	0.146	0.219	0.318	0.345
		18	2410	125	0.184	0.218	0.248	0.293	0.041	0.045	0.046	0.046	0.163	0.231	0.290	0.301
		18	2466	150	0.184	0.218	0.240	0.285	0.039	0.041	0.042	0.043	0.159	0.238	0.289	0.306
E(20)	E(15)	21	1631	50	0.211	0.232	0.260	0.264	0.067	0.068	0.069	0.069	0.157	0.202	0.316	0.350
		21	1777	75	0.210	0.233	0.267	0.255	0.055	0.055	0.055	0.055	0.144	0.193	0.345	0.352
		21	1850	100	0.187	0.229	0.281	0.255	0.047	0.048	0.048	0.049	0.101	0.163	0.326	0.346
		21	1919	125	0.159	0.211	0.296	0.259	0.043	0.045	0.046	0.046	0.100	0.123	0.309	0.320
		21	1985	150	0.164	0.203	0.294	0.276	0.039	0.039	0.039	0.039	0.100	0.116	0.273	0.316
E(15)	E(10)	23	1982	50	0.181	0.233	0.272	0.217	0.066	0.070	0.072	0.073	0.115	0.198	0.320	0.300
		23	2135	75	0.170	0.217	0.285	0.267	0.053	0.057	0.058	0.059	0.103	0.162	0.296	0.330
		23	2227	100	0.174	0.211	0.271	0.312	0.047	0.050	0.052	0.053	0.114	0.159	0.253	0.345
		23	2295	125	0.178	0.217	0.254	0.317	0.042	0.045	0.047	0.045	0.122	0.178	0.236	0.385
		23	2369	150	0.183	0.221	0.250	0.303	0.039	0.040	0.040	0.040	0.129	0.188	0.235	0.341

Table 3. Estimated Survival functions and standard errors using conventional and modified Kaplan-Meier estimators, Greenwood and modified standard error.*

Time	r_n	e_n	c_n	r_n'	Survival functions		Standard errors	
					S_{KM}	S_{MKM}	S_{EGW}	S_{EMKM}
0			0	21				
6			1	**20**				
7	21	0	0	17	1.000	1.000	0.000	0.000
9	**21**	3	1	16	0.857	0.925	0.076	0.062
10	17	1	1	**14**	0.807	0.871	0.087	0.087
11			1	13				
13	**15**	1	0	12	0.753	0.840	0.096	0.100
16			0	11				
17	12	1	1	10	0.690	0.770	0.107	0.132
19	11	1	1	9	0.628	0.700	0.114	0.163
20			1	8				
22			0	7				
23			0	6				
25	7	1	1	5	0.538	0.600	0.128	0..225
32	6	1	2	4	0.448	0.500	0.135	0..290
34			1	2				
35			1	1				

* S_{EGW}=Greenwood Standard error, S_{EMKM}=Modified Standard error

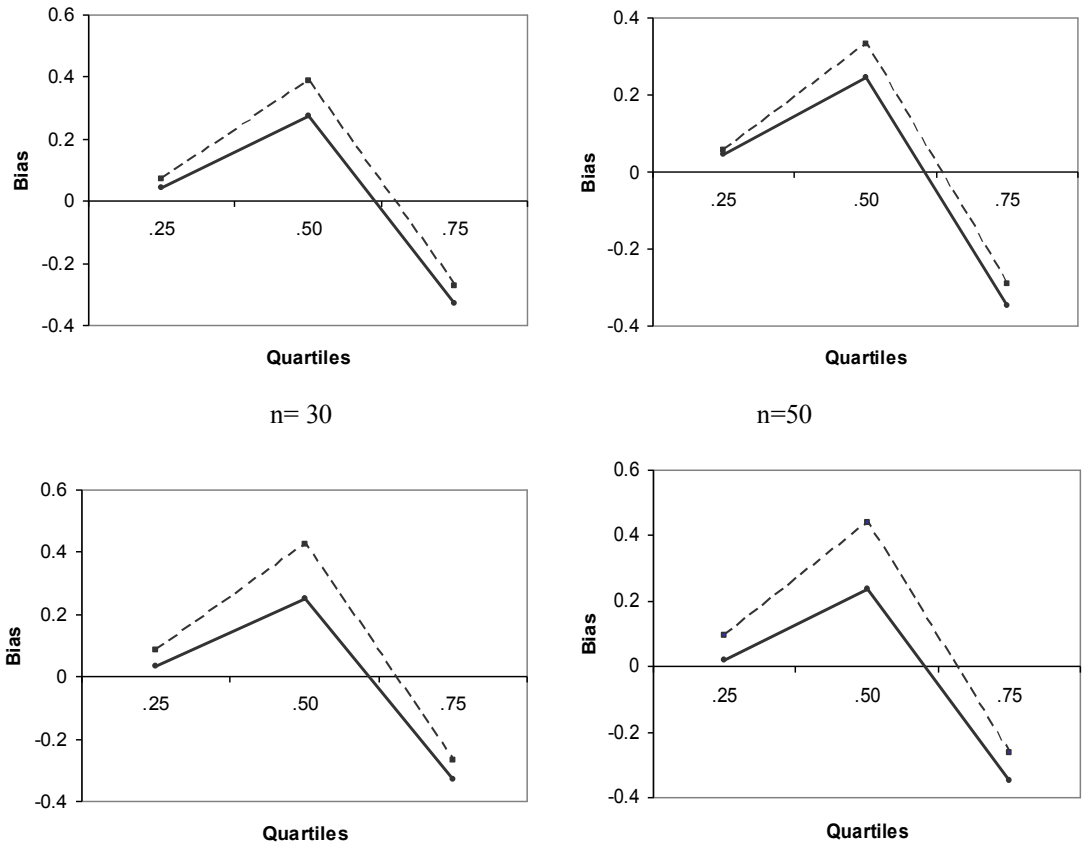
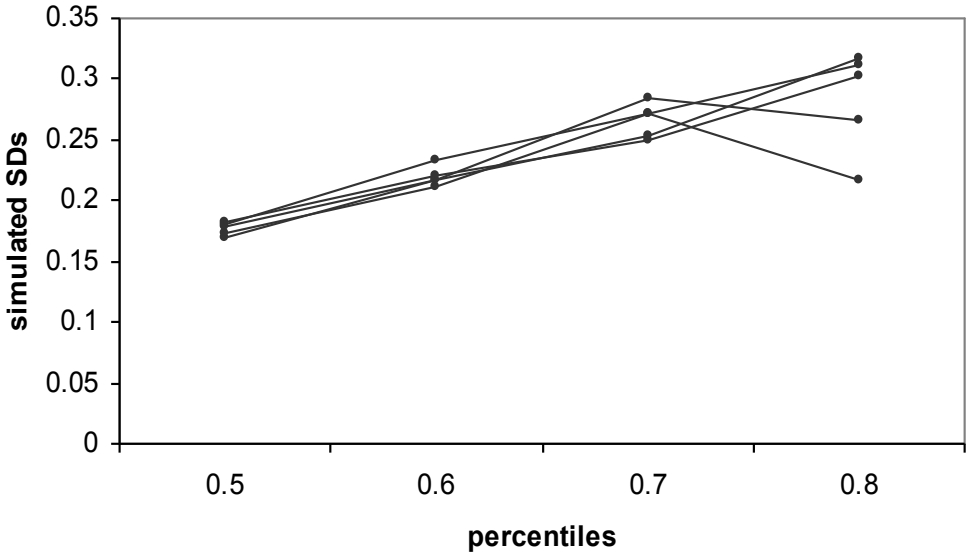
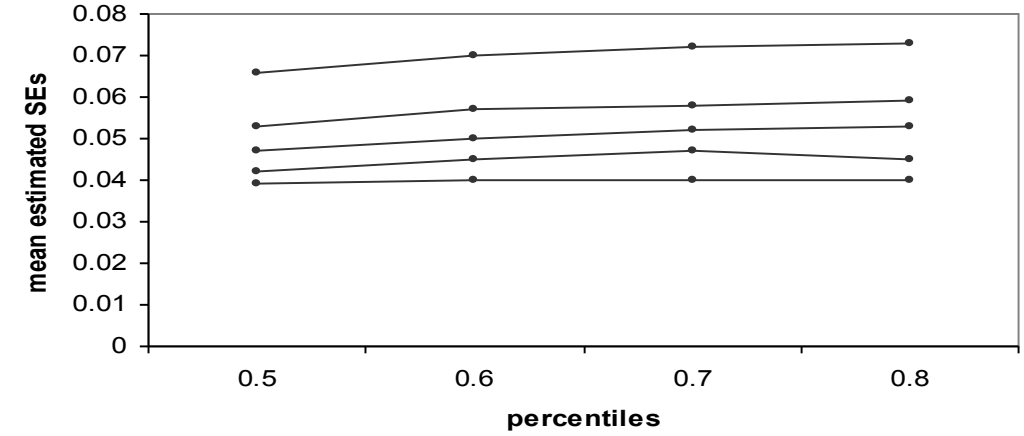


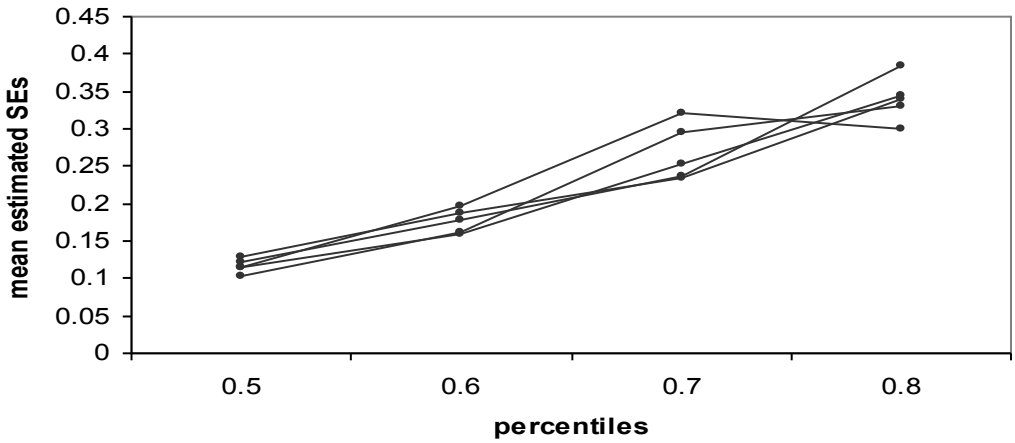
Figure 1. Simulated bias of conventional and modified Kaplan-Meier survival function (dotted curves represent bias of conventional Kaplan-Meier survival function and solid curves represent bias of modified Kaplan-Meier survival function from w(1.5, 25)).



(a)



(b)



(c)

Fig 2. Set of curves of (a) simulated SDs of modified Kaplan-Meier survival function, (b) mean estimated Greenwood standard error and (c) mean estimated standard error of modified Kaplan-Meier survival function from E(15). Each curve was estimated using 1000 simulated sets of data for a given sample size (n=50, 75, 100, 125, 150).

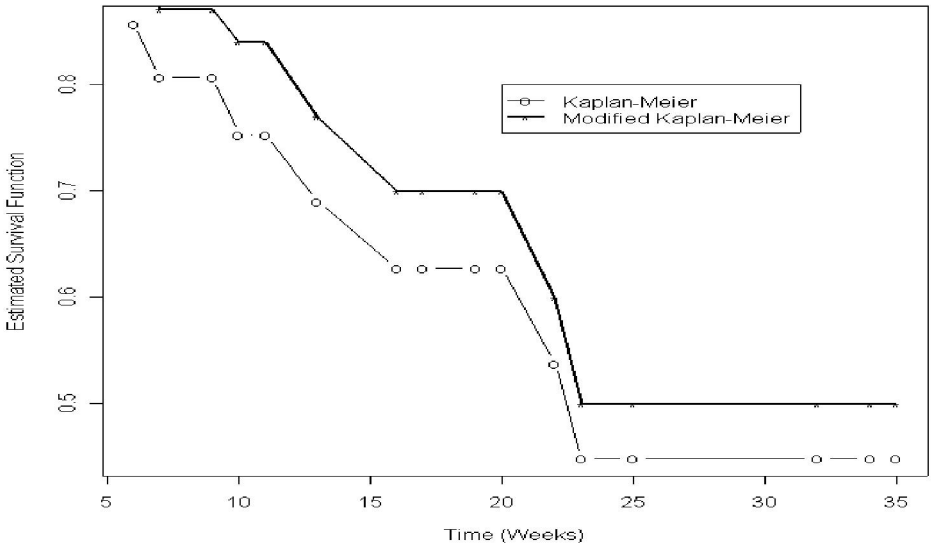


Figure 3. Kaplan-Meier Survival curve and modified Kaplan-Meier Survival curve for the leukaemia data.

References

1. Horton NJ, Switzer SS. Statistical Methods in the Journal. New England Journal of Medicine 2005;353: 1977-1979.
2. Strasak AM, Zaman Q, Marinell G, Pfeiffer KP, Ulmer H. The Use of Statistics in Medical Research: A Comparison of The New England Journal of Medicine and Nature Medicine. The American Statistician 2006; in press.
3. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. New England Journal of Medicine 1983;309: 709-713.
4. Fleming TR, Harrington DP. Nonparametric estimation of the survival distribution in censored data. Communication in Statistics – Simulation and Computation 1984;13: 1-26.
5. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 1958;53: 457-481.
6. Collet D. Modelling Survival Data in Medical Research (1st edn.). London: Chapman & Hall/CRC, 1994.
7. Greenwood Major. A report on the natural duration of cancer. In Reports on Public Health and Medical Subjects,. His Majesty's Stationery Office: London 1926; 33: 1-26.
8. Maller R, Zhou X. Survival Analysis with Long-Term Survivors. West Sussex: John Wiley & Sons Ltd, 1996.
9. Freireich EJ, Gehan E, Frei III E, Schroeder LR, Wolman IJ, Anbari R, Burgert EO, Mills SD, Pinkel D, Selwry OS, Moon JH, Gendel BR, Spurr CL, Storrs R, Haurani F, Hoogstraten B, Lee S. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukaemia: a model for evaluation of other potentially useful therapy. Blood 1963;21: 699-716.
10. Hogan JW, Laird NM. Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. Statistical Methods in Medical Research 1998;7: 28-48.
11. Brookmeyer R, Curriero FC. Survival curve estimation with partial non-random exposure information. Statistics in Medicine 2002;21: 2671-2683.
12. Lam FK, David IP. REML and ML estimation for clustered grouped survival data. Statistics in Medicine 2003;22: 2025-2034.
13. Adebayo SB, Fahrmeir L. Analysing child mortality in Nigeria with Geoadditive Discrete-time survival models. Statistics in Medicine 2005;24: 709-728.
14. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing, 2004.
15. Peto, R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. British Journal of Cancer 1977;35: 1-39.
16. Borkowf CB. A simple hybrid variance estimator for the Kaplan-Meier survival function. Statistics in Medicine 2005;24: 827-851.
17. Marubini E, Valsecchi MG. Analysing Survival Data from Clinical Trials and Observational Studies. Chichester: John Wiley & Sons, 1995.

6/13/2014