SciencePG
Science Publishing Group

Non-parametric Estimation of Survival Function from Grouped Observations Under Random Censorship

Job Isaac Mukangai

Department of Statistics and Actuarial Science, Kenyatta University, Nairobi, Kenya

Email address:
mkangai.jm@gmail.com

Abstract: Censoring is inevitable in survival analysis. The motivating factor for this article concerns the way censored subjects are incorporated in estimation of survival function for grouped data. In practice, the Actuarial estimator of a survival function may be biased due to unevenly distribution of censored subjects within intervals. This article presents a nonparametric estimation of a survival function using the adjusted Product Limit estimator based on grouped observations that are under random censorship. Simulation studies are carried out to assess the performance of the adjusted Product Limit estimator in comparison to the performance of Actuarial (life table) estimator to ascertain the one that is better and real data is used to show applicability of the method in real life. The results strongly indicate that adjusted Product Limit estimator of the survival function outperforms the Actuarial estimator.

Keywords: Life Table, Interval Censoring, Product Limit Estimates, Survival Analysis, Actuarial Estimator

1. Introduction

The adjusted Product Limit estimator (APLE), proposed in [1], is a flexible model for calculating survival probabilities in the presence of ties. It is closely related to Kaplan-Meier estimator discussed in [2] and in the presence of ties it gives asymptotically correct results. The main line of argument in [1] is through a series of examples that show APLE doing a commendable work under a variety of situations.

Problem with estimation of survivor function from grouped censored observations has been discussed before by several authors both parametrically and non-parametrically. Berkson and Gage [3], Culter and Ederer [4], Kaplan and Meier [2] proposed life table (actuarial) method which has been in use for decades as the standard nonparametric method for estimating the survival function when the data is grouped. The major limitation of actuarial method is about the way censored individuals are handled: the method assumes that censored individuals are evenly distributed within the interval such that half are censored before and half censored after the midpoint of the interval. With this assumption only half of the censored individuals are taken to be at risk, and as a result actuarial method tends to underestimate the true survival function. Though the

assumption made when using actuarial method might be true at one point; however, most survival distributions are often skewed or far from normal, it would be a contradiction now to assume that censored individuals are evenly distributed within any given interval. Breslow and Crowley [5] showed that actuarial estimator is consistent if and only if all individual, r_j, in the jth interval are at risk, and also showed that there is a slight overestimation in the variation in the estimated survival probability when actuarial estimator is used.

Several models that specify a parametric form for the survival distribution in which all censoring occur at the midpoint of each interval have been proposed, among others, by Elveback [6] and Chiang [7] while other non-parametric survival function estimators for interval censored data have been proposed by Peto [8], Klein and Moeschberger [9], Sun [10], among others. The fact that life table method is the most commonly used in practice in estimating survival function, see for example in [11, 12], it is an indication that there is no any other method among all models proposed in literature that can outperform it regardless of it being inconsistent and underestimating the true survival function, thus there is a need for new statistical techniques in this particular area of study.

The aim of this paper is to estimate survival function using

APLE from grouped censored observations so as to minimize, if not eliminating, the error incurred when actuarial method is used in estimating the survival function. Though APLE was developed particularly to estimate survival function for ungrouped censored data in the presence of ties, it is justifiable to use it on grouped data since it incorporates both censored and uncensored individuals in calculation of survival probabilities unlike the Kaplan-Meier estimator which does not incorporate censored individuals in case of ties; An interval in a grouped data set can be taken to be equivalent to a tie in ungrouped data set where in both cases censored and uncensored individuals are considered.

In practical situations, the researcher ought to use the most accurate estimator available so as to obtain reliable results that reflect the real life problems. This paper demonstrates that using APLE in estimating survival function from grouped censored observations leads to obtaining results with high precision than using actuarial estimate and it is anticipated that the material presented here will be of greatest interest to researchers concerned with life testing and medical follow-up studies and also of some interest to demographers and actuaries.

2. Models Description

This section presents a review of APLE and actuarial estimates of the survival function and also a brief description of random censorship model. Let d_j and c_j to be the number of individuals that fail and those that were censored, respectively, in the jth interval while r_j to be the number of individuals at risk at the start of the jth interval, then actuarial and adjusted product limit estimators are as described below.

2.1. Actuarial Estimate

Actuarial estimate (AE) also known as life table estimate of the survivor function is one of the oldest method for measuring mortality and survivorship of subjects in a population. It has been used by actuaries, demographers, and medical researchers in studies of survival, length of married life and length of working life, among others. It is obtained by multiplying together a sequence of estimates of conditional probabilities of surviving through intervals, and it depends on the selection of the intervals.

Product Limit (PL) estimate, discussed in [2], is used if all failed individuals, d_j, are known to precede all censored individuals, c_j, but if the reverse is true, then reduced sample (RS) estimate is used. When the arrangement between event and censoring times within the intervals is not known, adjustment is made on PL estimate to obtain the actuarial estimate on assumption that censored individuals are evenly spread in the interval and estimation is done at the midpoint of the interval such that only half of the censored individuals are at risk of failing. In this way the average number of individuals at risk in the jth interval is $r_j - \frac{1}{2}c_j$ and the corresponding probability of surviving in that interval $\dot{p}_j$ is then given as $\dot{p}_j = 1 - \frac{d_j}{r_j - \frac{1}{2}c_j}$ while PL and RS estimates are

respectively given as $\ddot{p}_j = 1 - \frac{d_j}{r_j}$ and $\dddot{p}_j = 1 - \frac{d_j}{r_j - c_j}$ for details on these estimates see [2]. If no individual fail in j^{th} interval, $d_j = 0$, then $\dot{p}_j = \ddot{p}_j = \dddot{p}_j = 1$ and in case no individual is censored in j^{th} interval, $\dot{p}_j = \ddot{p}_j = \dddot{p}_j = 1 - \frac{d_j}{r_j}$

As stated before, Life table method tends to underestimate the survival function due to the assumption made concerning the distribution of censored individuals within the interval. Little [13] suggested a modification of the constant $\frac{1}{2}$ in the estimate $\dot{p}_j$ in order to improve its approximation in certain circumstances. In the next subsection adjusted Product Limit estimate is discussed which also improves the estimation of the survival function.

2.2. Adjusted Product Limit Estimate

Adjusted Product Limit estimate (APLE) is a form of Product Limit estimates that is formed by multiplying together a sequence of estimates of conditional probabilities of surviving through intervals, thus the method incorporates all survival information accumulated up to the termination of the study. In forming APLE, estimated probability of failing in uncensored set of the data is used to estimate the expected number of failures out of the censored set then the overall probability of failing in the jth interval is obtained by summing the probability of an individual failing when censored or failing when uncensored.

To calculate the probability of an event occurring, it is necessary to consider all the ways in which that event can happen. The model in Figure 1 is used to illustration all possible outcomes for a subject in the jth interval can experience and from it the probability of an individual failing can be obtained as follows:

P(failing) = P(uncensored and failing) or P(censored and failing)

Therefore, P(failing) = $\lambda_1 * \lambda_3 + \lambda_2 * \lambda_5$; this is by using probability theorem for independent events.

Where $\lambda_1 = \frac{r_j - c_j}{r_j}$ is the probability of an individual being uncensored, $\lambda_2 = \frac{c_j}{r_j}$ probability of an individual being censored, $\lambda_3 = \frac{d_j}{r_j - c_j}$ probability of an individual failing when uncensored and $\lambda_5 = \frac{d_j * c_j}{(r_j - c_j)(r_j - d_j)}$ is the probability of failing when censored, see [1].

Thereafter the result is subtracted from 1 to obtain the corresponding probability of surviving in the same interval.

Probability of surviving in the jth interval, p_j according to the above procedure then is estimated as

$$p_j = 1 - \frac{d_j[(r_j - c_j)(r_j - d) + c_j^2]}{r_j(r_j - c_j)(r_j - d_j)}$$

And the adjusted Product Limit estimator as in [1] is

$$\hat{S}_*(t) = \prod_{j \leq t}\left\{1 - \frac{d_j[(r_j - d_j)(r_j - c_j) + c_j^2]}{r_j(r_j - d_j)(r_j - c_j)}\right\}$$
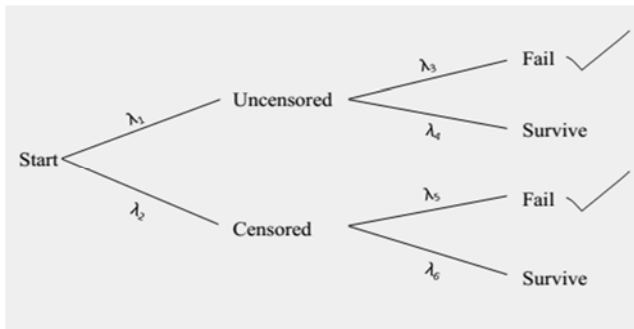
Figure 1. Probability model showing all possible outcomes.

If no individual happen to die in interval j then $p_j = 1$ and the survival probability stays constant for a whole run of such intervals just like the case with life table method, and if no individual is censored in interval j then APLE and life table method are same.

2.3. Random Censorship

Any study on survival analysis should discuss the type, causes, and treatment of censoring. Random censoring may arise in animal study or medical applications where subjects under the study may relocate/migrate to other places or may drop out from the study or the study may be terminated before some of them experience the event of interest, thus in random censoring the researcher can't tell with certainty when the subject in the study might be censored and the number of subjects that will be censored.

Let t_1 t_2 … t_N denotes the true survival times for N individuals included in the study assumed to be independent and random with a common distribution Function F(t) = P[t_i ≤ t] such that F(0)=0 and let c_1 c_2 … c_N be random variables also assumed to be independent with a common distribution Function G(c), t_i are said to be randomly censored on the right by c_i when one only observe T_i = min(t_i, c_i) and ∂_i =I$_{[ti \leq ci]}$ where ∂_i indicates whether the t_i is censored (∂_i=0) or not (∂_i=1) while t_i are said to be randomly censored on the left when one only observe T_i = max(t_i, c_i), See [14] for details. Since both t_i's and c_i's are assumed to be random samples drawn independent of each other, the observed t_i's constitute a random sample.

3. Simulation Study and Discussion

In this section, to see the performance and compare the efficiency of the aforementioned methodologies, simulation study is conducted to demonstrate the methods for several parameters in different random sample sizes with different censoring percentages. Survival times are generated from exponential E[δ] and lognormal LN[α,δ] distributions using R statistical package [15] while censoring times are generated according to uniform distribution U[0, b]. Parameter "b" is adjusted to obtain different percentage of censoring while parameters "α" and "δ" are adjusted to provide different situation in which the mentioned estimators are assessed. For each data set, the survival probabilities and

standard errors for APLE and actuarial method were computed and the results are reported in Tables 1-14 and in Figures 2-11. The range of the data sets is 25 to 3000 and the range of percentage censoring is 34% to 92.3%; though several sample sizes were considered in the study, only some results are presented for illustration because they are similar and since small samples do not necessitate grouping, the study based majorly on large sample sizes.

Results for data generated using exponential distribution are reported in Tables 1-8 while that from lognormal distribution are reported in Tables 9-14 and in both cases the results are similar. That is, adjusted product limit estimates are higher than actuarial estimates and standard errors due to adjusted product method are smaller than corresponding standard error due to actuarial method. Likewise, data generated using same distribution but different parameters, sample sizes, or different percentage of censoring give similar results. For example, Tables 10, 13 and 14 shows results for data generated from lognormal distribution with different possibilities in terms of size, censoring percentage and parameters used, in all the cases adjusted product limit estimates have smaller standard errors than actuarial estimates, see columns 7 and 8 in the stated Tables for details.

In assessing the goodness of the methods, accuracy in estimating survival probabilities was considered and the method which provides smaller standard errors was taken to be more accurate. Basing on the results in the Tables and Figures it can clearly be seen that actuarial method is less accurate both for small as well as large sample sizes and also for light, moderate as well as heavy censoring. For instance in Table 6 a large sample size of 400 is used where 86.8% of the subjects are censored (heavy censoring), at time interval [0, 1) actuarial method gives a survival probability of 0.992415 with corresponding standard error of 0.004363 while APLE gives a survival probability of 0.992496 with corresponding standard error of 0.004315 which is an improvement both in terms of estimation and accuracy. Throughout the study actuarial method gives smaller survival probabilities with corresponding larger standard errors as compared to APLE, this confirms that actuarial method is less precise, thus APLE is a better estimator than actuarial method in terms of efficiency. See Tables 1-14 in columns 7 and 8.

In case none of the subjects in the jth interval is censored the two methods give same results, see for instance Table 1 at time interval [0, 1), and if no event occurs in the jth interval then the survival probability remain constant, see for example Table 2 at intervals [14, 15) and [15, 16). These similarities in performance for the two methods shows their close relationship but only differs at some points due to the way censored subjects are treated by each method; Actuarial method can simply be taken to be another form of reduced sample estimator, discussed in [2], but for this case it reduces the number of subjects at risk in the jth interval by a half of the censored subjects only and since reduced sample estimator results to an underestimation of the true survival function, it is trivial that actuarial method will also result to

an underestimation, though not as much as reduced sample estimator. For instance, using values in Table 12 at time interval [0, 1) 250 individuals entered the interval and out of them 84 failed while 8 were censored within the interval; reduced sample estimator gives a survival probability of 0.652893, actuarial method gives 0.658537, while APLE gives a survival probability of 0.663465. Using probability theorem, probability of failing in the stated interval equals to $\frac{number\ of\ failed\ individuals}{total\ number\ of\ individuals} = \frac{84}{250} = 0.336$ with corresponding probability of surviving 0.664. Reducing the denominator by 8 (number of censored individuals) causes an increase in the probability of failing to $\frac{84}{250-8} = 0.347107$ which leads to a decrease in the corresponding probability of surviving in the same interval. Similarly, a decrease in the corresponding probability of surviving is also obtained when the denominator is reduced by half the number of censored individuals, $\frac{84}{250-4} = 0.341463$, as it is the case with actuarial method. Thus, the results reported in this study provide empiric evidence of the magnitude of underestimation of the actuarial method compared with adjusted product limit estimator. It is true the estimated survival probability 0.664 ought to reduce because some censored individuals might fail, but the question is: what magnitude should it reduce by? Reducing the denominator by half of the censored individuals

is not appropriate when the data is under random censorship. Consequently, assuming that only half of the censored individuals were at risk is unjustified; this is so because all censored individuals were observed at the start of the interval, thus being at risk of failing and if estimation is done at the midpoint of the interval, as it is done by actuarial method, then not all failed individuals were at risk since some might have failed before midpoint. Also no researcher can tell with certainty that exactly half of the censored individuals were censored before the midpoint and thus not being at risk of failing. From the above example, it is clear that actuarial method is unreliable due to its assumptions.

If the sum of the number of events and the number of censored subjects at the largest observational time equals to the number of subjects that are at risk at that time then APLE estimates goes to zero, see for example Tables 1 and 2; it is justifiable to obtain such results since no one is expected to live for undefined period of time. Likewise, such results might be obtained in situations where subjects that have not yet experienced the event of interest at the termination of study are killed, like in animal study. In such cases, the probability of surviving at the largest observation time and beyond ought to be zero. This proves that APLE is generally a better estimator than actuarial estimator.

Table 1. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=100 with 45% censoring; E [.08].*

Time interval	r_j	d_j	c_j	APLE	AE	se. APLE	se. AE
[0, 1)	100	6	0	0.940000	0.940000	0.023749	0.023749
[1, 2)	94	6	3	0.879933	0.879027	0.032504	0.032751
[2, 3)	85	9	4	0.786521	0.783711	0.041334	0.041864
[3, 4)	72	8	3	0.698952	0.694779	0.046897	0.047475
[4, 5)	61	7	3	0.618514	0.613040	0.050377	0.050960
[5, 6)	51	5	5	0.557159	0.549840	0.052245	0.052967
[6, 7)	41	1	2	0.543535	0.536094	0.052710	0.053397
[7, 8)	38	1	3	0.529132	0.521407	0.053232	0.053916
[8, 9)	34	0	4	0.529132	0.521407	0.053232	0.053916
[9, 10)	30	2	1	0.493813	0.486057	0.055221	0.055755
[10, 11)	27	4	1	0.420533	0.412690	0.057903	0.058168
[11, 12)	22	2	1	0.382212	0.374300	0.058612	0.058751
[12, 13)	19	2	2	0.341422	0.332711	0.058942	0.059127
[13, 16)	15	0	3	0.341422	0.332711	0.058942	0.059127
[16, 17)	12	1	3	0.310384	0.301025	0.060614	0.061402
[17, 19)	8	0	5	0.310384	0.301025	0.060614	0.061402
[19, 20)	3	1	2	0.000000	0.150512	NA	0.110768

NA means not applicable, se.APLE standard error for APLE, se.AE standard error for Actuarial method and E[0.08] exponential distribution with rate parameter 0.08.

Table 2. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=300 with 44.3% censoring; E [0.08].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	300	11	7	0.963312	0.962901	0.010854	0.010976
[1, 2)	282	16	12	0.908547	0.907080	0.016770	0.017040
[2, 3)	254	17	15	0.847497	0.844523	0.021176	0.021588
[3, 4)	222	19	8	0.774856	0.770918	0.025067	0.025468
[4, 5)	195	13	8	0.723102	0.718447	0.027187	0.027581
[5, 6)	174	9	13	0.685462	0.679844	0.028504	0.028945
[6, 7)	152	15	9	0.617538	0.610707	0.030584	0.031022
[7, 8)	128	11	6	0.564335	0.556965	0.031871	0.032247
[8, 9)	111	12	2	0.503303	0.496205	0.032934	0.033158
[9, 10)	97	8	6	0.461609	0.453975	0.033329	0.033530
[10, 11)	83	8	3	0.417050	0.409413	0.033624	0.033737
[11, 12)	72	9	4	0.364724	0.356774	0.033612	0.033655

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[12, 13)	59	1	4	0.358511	0.350515	0.033606	0.033641
[13, 14)	54	2	5	0.345103	0.336903	0.033647	0.033684
[14, 15)	47	3	1	0.323064	0.315167	0.033818	0.033767
[15, 16)	43	0	2	0.323064	0.315167	0.033818	0.033767
[16, 17)	41	1	4	0.315099	0.307086	0.033899	0.033855
[17, 18)	36	2	5	0.297178	0.288752	0.034207	0.034226
[18, 19)	29	2	3	0.276421	0.267752	0.034788	0.034809
[19, 20)	24	1	2	0.264812	0.256111	0.035196	0.035198
[20, 21)	21	2	3	0.238928	0.229843	0.036096	0.036151
[21, 22)	16	0	3	0.238928	0.229843	0.036096	0.036151
[22, 23)	13	3	3	0.178829	0.169884	0.039454	0.039996
[23, 24)	7	2	3	0.104742	0.108108	0.040529	0.043152
[24, 25)	2	0	2	0.000000	0.108108	NA	0.043152

Table 3. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=3000 with 44.4% censoring; E [0.1].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	3000	120	61	0.959982	0.959589	0.003578	0.003614
[1, 2)	2819	191	102	0.894844	0.893375	0.005639	0.005717
[2, 3)	2526	185	119	0.829143	0.826367	0.006991	0.007101
[3, 4)	2222	160	115	0.769257	0.765282	0.007925	0.008052
[4, 5)	1947	135	82	0.715812	0.711078	0.008604	0.008729
[5, 6)	1730	132	85	0.661045	0.655456	0.009168	0.009293
[6, 7)	1513	124	69	0.606740	0.600483	0.009622	0.009737
[7, 8)	1320	108	62	0.556972	0.550171	0.009951	0.010053
[8, 9)	1150	91	64	0.512742	0.505390	0.010180	0.010272
[9, 10)	995	74	66	0.474414	0.466514	0.010344	0.010429
[10, 11)	855	60	55	0.440964	0.432688	0.010474	0.010547
[11, 12)	740	60	63	0.404902	0.396046	0.010593	0.010662
[12, 13)	617	52	44	0.370573	0.361433	0.010706	0.010757
[13, 14)	521	37	51	0.343955	0.334444	0.010785	0.010830
[14, 15)	433	38	43	0.313407	0.303560	0.010894	0.010928
[15, 16)	352	25	35	0.290885	0.280872	0.010993	0.011013
[16, 17)	292	20	35	0.270612	0.260408	0.011108	0.011120
[17, 18)	237	15	23	0.253294	0.243086	0.011252	0.011244
[18, 19)	199	16	34	0.232149	0.221716	0.011446	0.011455
[19, 20)	149	6	27	0.222410	0.211898	0.011610	0.011628
[20, 21)	116	6	21	0.210421	0.199847	0.011933	0.011962
[21, 22)	89	5	13	0.198287	0.187735	0.012389	0.012403
[22, 23)	71	6	21	0.179256	0.169117	0.013171	0.013300
[23, 24)	44	1	19	0.173814	0.164215	0.013587	0.013788

Table 4. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=500 with 92.2% censoring; E [0.01].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	500	2	10	0.995998	0.995960	0.002823	0.002851
[1, 2)	488	3	19	0.989866	0.989715	0.004507	0.004576
[2, 3)	466	4	19	0.981354	0.981043	0.006155	0.006262
[3, 4)	443	2	22	0.976912	0.976501	0.006881	0.007009
[4, 5)	419	1	20	0.974575	0.974114	0.007249	0.007387
[5, 6)	398	3	17	0.967215	0.966611	0.008346	0.008506
[6, 7)	378	2	17	0.962086	0.961379	0.009053	0.009229
[7, 8)	359	3	20	0.954020	0.953115	0.010101	0.010310
[8, 9)	336	3	24	0.945455	0.944290	0.011149	0.011404
[9, 10)	309	2	18	0.939313	0.937995	0.011890	0.012166
[10, 11)	289	1	16	0.936052	0.934656	0.012286	0.012572
[11, 12)	272	3	27	0.925614	0.923809	0.013532	0.013899
[12, 13)	242	2	24	0.917880	0.915776	0.014471	0.014893
[13, 14)	216	1	22	0.913581	0.911309	0.015021	0.015476
[14, 15)	193	3	19	0.899226	0.896410	0.016896	0.017451
[15, 16)	171	1	18	0.893901	0.890877	0.017605	0.018199
[17, 18)	129	1	8	0.886943	0.883750	0.018788	0.019399
[18, 19)	120	1	16	0.879399	0.875859	0.020057	0.020769
[24, 25)	25	1	13	0.823582	0.828516	0.046815	0.050062

Table 5. Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=500 with 56% censoring; E [0.05].

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	500	17	11	0.965983	0.965622	0.008107	0.008193
[1, 2)	472	18	19	0.929080	0.928041	0.011551	0.011722
[2, 3)	435	20	17	0.886292	0.884522	0.014443	0.014666
[3, 4)	398	22	21	0.837149	0.834304	0.017014	0.017305
[4, 5)	355	14	15	0.804070	0.800692	0.018493	0.018796
[5, 6)	326	13	11	0.771967	0.768214	0.019780	0.020076
[6, 7)	302	13	18	0.738605	0.734130	0.020970	0.021295
[7, 8)	271	18	12	0.689439	0.684264	0.022544	0.022863
[8, 9)	241	10	16	0.660690	0.654897	0.023357	0.023693
[9, 10)	215	13	8	0.620681	0.614548	0.024433	0.024735
[10, 11)	194	12	15	0.582023	0.575006	0.025316	0.025643
[11, 12)	167	6	7	0.561072	0.553905	0.025807	0.026109
[12, 13)	154	8	6	0.531877	0.524559	0.026444	0.026707
[13, 14)	140	6	4	0.509062	0.501752	0.026899	0.027121
[14, 15)	130	4	14	0.493189	0.485435	0.027191	0.027439
[15, 16)	112	9	10	0.453180	0.444604	0.028038	0.028306
[16, 17)	93	5	8	0.428607	0.419626	0.028574	0.028835
[17, 18)	80	4	7	0.406988	0.397685	0.029089	0.029340
[18, 19)	69	2	8	0.395006	0.385448	0.029422	0.029686
[19, 20)	59	1	11	0.388020	0.378244	0.029686	0.029993
[20, 21)	47	1	9	0.379382	0.369344	0.030202	0.030579
[21, 22)	37	1	8	0.368500	0.358152	0.031128	0.031634
[22, 23)	28	1	4	0.355014	0.344376	0.032716	0.033282
[23, 24)	23	1	8	0.336585	0.326251	0.035097	0.036130
[24, 25)	14	1	9	0.282584	0.291909	0.044252	0.045829

Table 6. Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=400 with 86.8% censoring; E [0.009].

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	400	3	9	0.992496	0.992415	0.004315	0.004363
[1, 2)	388	6	14	0.977127	0.976786	0.007533	0.007649
[2, 3)	368	4	19	0.966476	0.965887	0.009138	0.009304
[3, 4)	345	7	15	0.946827	0.945854	0.011578	0.011797
[4, 5)	323	4	17	0.935067	0.933824	0.012837	0.013091
[5, 6)	302	2	12	0.928864	0.927515	0.013479	0.013741
[6, 7)	288	4	20	0.915895	0.914169	0.014761	0.015077
[7, 8)	264	6	11	0.895041	0.892951	0.016697	0.017035
[8, 9)	247	1	13	0.891407	0.889238	0.017019	0.017364
[9, 10)	233	3	15	0.879878	0.877407	0.018048	0.018427
[10, 11)	215	2	20	0.871614	0.868847	0.018792	0.019216
[12, 13)	184	2	18	0.862038	0.858918	0.019756	0.020238
[13, 14)	164	3	11	0.846192	0.842661	0.021397	0.021924
[15, 16)	133	1	10	0.839790	0.836077	0.022167	0.022720
[16, 17)	122	2	12	0.825873	0.821662	0.023863	0.024508
[22, 23)	49	3	11	0.771809	0.764996	0.036727	0.038951

Table 7. Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=800 with 92.3% censoring; E [0.005].

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	800	1	15	0.998750	0.998738	0.001249	0.001261
[1, 2)	784	3	33	0.994921	0.994834	0.002531	0.002576
[2, 3)	748	9	22	0.982939	0.982686	0.004690	0.004762
[3, 4)	717	7	20	0.973335	0.972956	0.005883	0.005968
[4, 5)	690	3	28	0.969096	0.968638	0.006345	0.006441
[5, 6)	659	2	27	0.966149	0.965637	0.006659	0.006762
[6, 7)	630	2	18	0.963080	0.962527	0.006982	0.007089
[7, 8)	610	3	20	0.958338	0.957715	0.007465	0.007578
[8, 9)	587	5	17	0.950168	0.949437	0.008246	0.008368
[9, 10)	565	2	20	0.946800	0.946016	0.008553	0.008681
[10, 11)	543	3	27	0.941555	0.940656	0.009025	0.009166
[11, 12)	513	1	22	0.939716	0.938782	0.009192	0.009338
[12, 13)	490	3	26	0.933946	0.932878	0.009719	0.009882
[13, 14)	461	3	22	0.927853	0.926658	0.010271	0.010448
[14, 15)	436	2	22	0.923586	0.922298	0.010657	0.010844
[15, 16)	412	2	24	0.919086	0.917686	0.011068	0.011270
[17, 18)	367	2	16	0.914067	0.912574	0.011561	0.011772

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[18, 19)	349	1	22	0.911437	0.909874	0.011822	0.012043
[20, 21)	308	1	18	0.908467	0.906831	0.012150	0.012381
[22, 23)	274	3	18	0.898474	0.896565	0.013310	0.013586
[24, 25)	231	1	21	0.894549	0.892499	0.013813	0.014120
[26, 27)	184	1	23	0.889600	0.887325	0.014582	0.014956
[29, 30)	117	1	21	0.881696	0.878993	0.016384	0.016978
[31, 32)	69	1	17	0.867873	0.864464	0.020831	0.022055

Table 8. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=200 with 92% censoring; E [0.003].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[3, 4)	193	1	4	0.994816	0.994764	0.005169	0.005222
[4, 5)	188	2	2	0.984232	0.984125	0.009031	0.009093
[5, 6)	184	1	6	0.978877	0.978688	0.010448	0.010544
[7, 8)	175	1	8	0.973271	0.972965	0.011794	0.011935
[9, 10)	161	1	2	0.967225	0.966884	0.013180	0.013319
[10, 11)	158	1	5	0.961097	0.960666	0.014449	0.014613
[11, 12)	152	1	6	0.954764	0.954218	0.015679	0.015874
[12, 13)	145	2	3	0.941589	0.940919	0.018018	0.018227
[13, 14)	140	1	6	0.934850	0.934051	0.019105	0.019344
[15, 16)	132	1	11	0.927714	0.926667	0.020239	0.020552
[16, 18)	120	1	3	0.919978	0.918847	0.021497	0.021816
[20, 22)	97	1	3	0.910484	0.909226	0.023275	0.023614
[27, 31)	68	1	3	0.897067	0.895553	0.026512	0.026928
[40, 42)	26	1	5	0.860921	0.857445	0.042945	0.045334

Table 9. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=100 with 42% censoring; LN [1.5, 25].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1]	100	55	1	0.449877	0.447236	0.049748	0.049846
[2, 5]	44	1	0	0.439652	0.437072	0.049657	0.049738
[6, 12]	43	0	6	0.439652	0.437072	0.049657	0.049738
[13, 16]	37	1	1	0.427760	0.425097	0.049716	0.049796
[17, 21]	35	0	4	0.427760	0.425097	0.049716	0.049796
[22, 24]	31	1	1	0.413946	0.411160	0.049991	0.050076

Table 10. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=50 with 52% censoring; LN [1.5, 15].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1]	50	19	1	0.619750	0.616162	0.068653	0.069122
[2, 3]	30	1	1	0.599067	0.595275	0.069404	0.069864
[4, 5]	28	1	0	0.577672	0.574015	0.070145	0.070529
[6, 7]	27	0	1	0.577672	0.574015	0.070145	0.070529
[8, 9]	26	1	0	0.555454	0.551937	0.070879	0.071188
[10, 34]	25	0	8	0.555454	0.551937	0.070879	0.071188
[35, 40]	17	1	0	0.522780	0.519471	0.073858	0.074035
[41, 59]	16	0	4	0.522780	0.519471	0.073858	0.074035
[60, 62]	12	1	0	0.479215	0.476181	0.079520	0.079520

Table 11. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=25 with 40% censoring; LN [1.5, 5].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	25	10	1	0.598889	0.591837	0.098025	0.099297
[1, 2)	14	2	0	0.513333	0.507289	0.100978	0.101526
[2, 3)	12	2	1	0.427000	0.419065	0.100634	0.101237
[3, 4)	9	0	0	0.427000	0.419065	0.100634	0.101237
[4, 5)	9	1	1	0.378814	0.369763	0.099994	0.100618

Table 12. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=250 with 51.6% censoring; LN [1.5, 5].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	250	84	8	0.663465	0.658537	0.029885	0.030234
[1, 2)	158	16	18	0.595183	0.587821	0.031240	0.031738
[2, 3)	124	8	17	0.555890	0.547106	0.032055	0.032641
[3, 4)	99	1	20	0.549985	0.540959	0.032227	0.032848
[4, 5)	78	5	12	0.513676	0.503392	0.033839	0.034598
[5, 6)	61	5	14	0.468436	0.456782	0.036051	0.037147
[6, 7)	42	1	7	0.456902	0.444918	0.036905	0.038029
[7, 8)	34	0	10	0.456902	0.444918	0.036905	0.038029
[8, 9)	24	1	10	0.431952	0.421501	0.040821	0.042632

Table 13. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=250 with 43% censoring; LN [1.2, 5].*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	250	69	2	0.723975	0.722892	0.028273	0.028364
[1, 2)	179	27	3	0.614736	0.612931	0.030846	0.030953
[2, 3)	149	9	4	0.577575	0.575404	0.031368	0.031484
[3, 4)	136	5	6	0.556295	0.553772	0.031620	0.031752
[4, 5)	125	4	7	0.538433	0.535541	0.031837	0.031988
[5, 6)	114	2	4	0.528974	0.525978	0.031972	0.032124
[6, 7)	108	4	1	0.509381	0.506407	0.032253	0.032385
[7, 8)	103	5	3	0.484631	0.481461	0.032528	0.032655
[8, 9)	95	1	3	0.479524	0.476311	0.032584	0.032709
[9, 10)	91	2	1	0.468984	0.465785	0.032709	0.032822
[10, 11)	88	1	5	0.463636	0.460337	0.032768	0.032887
[11, 12)	82	2	3	0.452312	0.448900	0.032931	0.033049
[12, 13)	77	3	6	0.434569	0.430702	0.033184	0.033338
[13, 14)	68	2	5	0.421710	0.417551	0.033417	0.033592
[14, 15)	61	1	4	0.414765	0.410473	0.033577	0.033760
[16, 17)	53	1	5	0.406861	0.402345	0.033846	0.034056
[18, 19)	46	2	6	0.388809	0.383631	0.034623	0.034948
[19, 20)	38	1	4	0.378447	0.372975	0.035198	0.035565
[27, 28)	7	1	2	0.317175	0.310813	0.060386	0.064020

Table 14. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for n=800 with 34% censoring; LN [0.5, 5].*

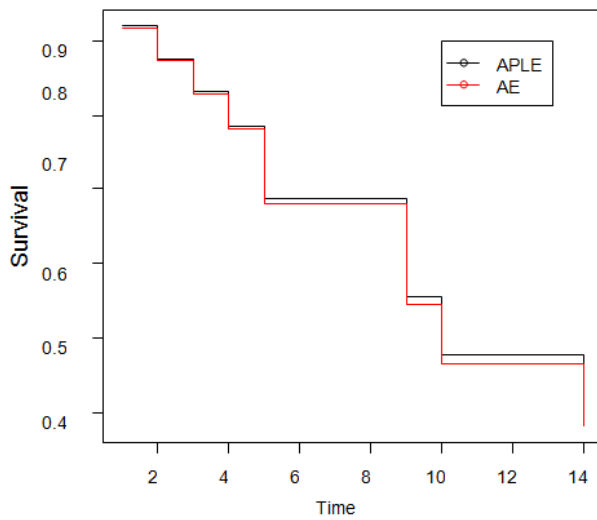| Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	800	328	8	0.589930	0.587940	0.017389	0.017446
[1, 2)	464	72	17	0.498238	0.495005	0.017724	0.017797
[2, 3)	375	26	10	0.463666	0.460221	0.017743	0.017806
[3, 4)	339	22	13	0.433527	0.429770	0.017713	0.017772
[4, 5)	304	14	8	0.413547	0.409714	0.017683	0.017733
[5, 6)	282	17	12	0.388566	0.384478	0.017620	0.017665
[6, 7)	253	5	13	0.380866	0.376680	0.017603	0.017648
[7, 8)	235	7	8	0.369507	0.365265	0.017594	0.017632
[8, 9)	220	5	14	0.361072	0.356691	0.017590	0.017631
[9, 10)	201	2	10	0.357469	0.353051	0.017597	0.017637
[10, 11)	189	8	8	0.342309	0.337784	0.017647	0.017681
[11, 12)	173	1	4	0.340329	0.335809	0.017656	0.017688
[12, 13)	168	2	7	0.336270	0.331726	0.017676	0.017707
[13, 14)	159	1	10	0.334146	0.329572	0.017691	0.017723
[14, 15)	148	2	11	0.329603	0.324946	0.017738	0.017773
[15, 16)	135	4	10	0.319778	0.314948	0.017873	0.017915
[16, 17)	121	1	14	0.317095	0.312185	0.017920	0.017970
[17, 18)	106	4	5	0.305100	0.300120	0.018216	0.018260
[18, 19)	97	3	13	0.295462	0.290171	0.018454	0.018536
[19, 20)	81	1	6	0.291792	0.286451	0.018584	0.018668
[20, 21)	74	1	6	0.287820	0.282417	0.018748	0.018836
[22, 23)	59	1	8	0.282837	0.277282	0.019060	0.019181
[24, 25)	43	1	6	0.276107	0.270350	0.019734	0.019915

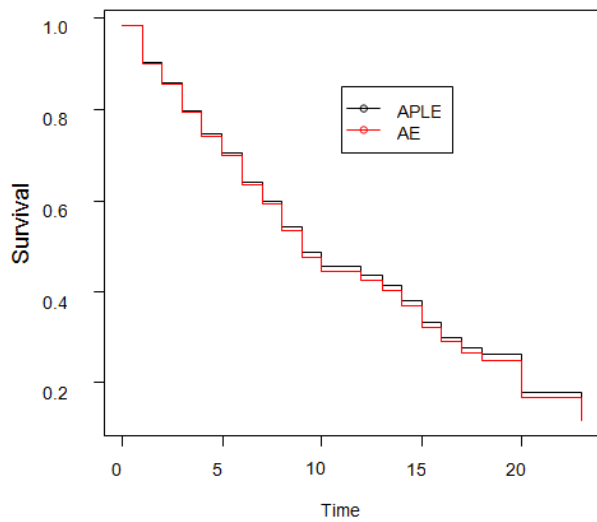**Figure 2.** *Survival curves for APLE and AE for n=25 with 52% censoring; E[0.03].*

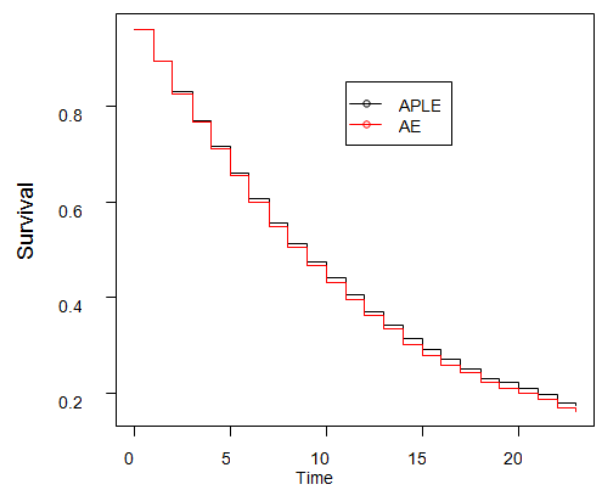**Figure 3.** *Survival curves for APLE and AE, for n=300 with 48% censoring; E[0.08].*

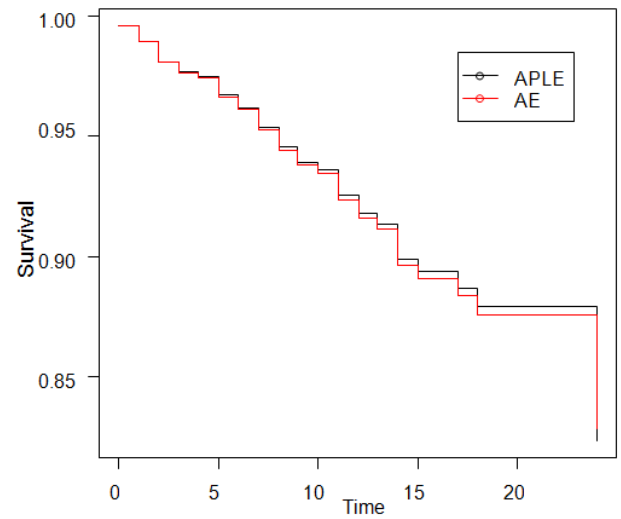**Figure 4.** *Survival curves for APLE and AE, for n=3000 with 44.4% censoring; E[0.1].*

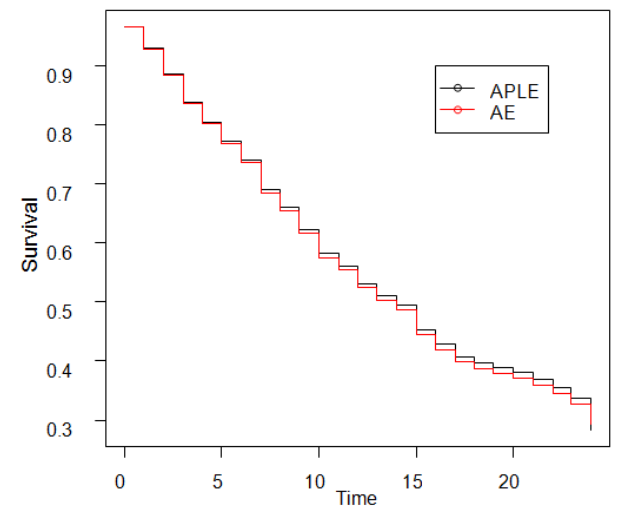**Figure 5.** *Survival curves for APLE and AE, for n=500 with 92.2% censoring; E[0.01].*

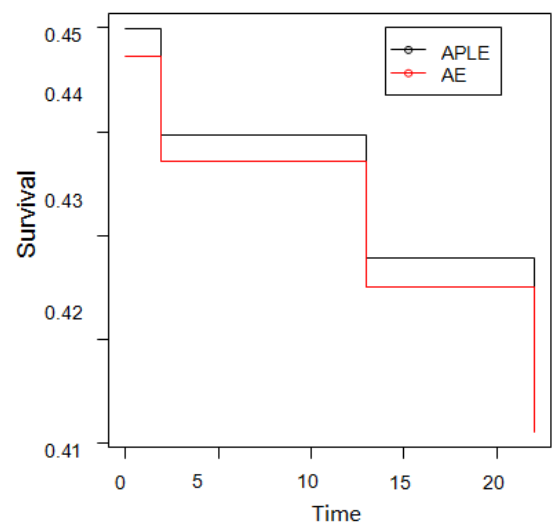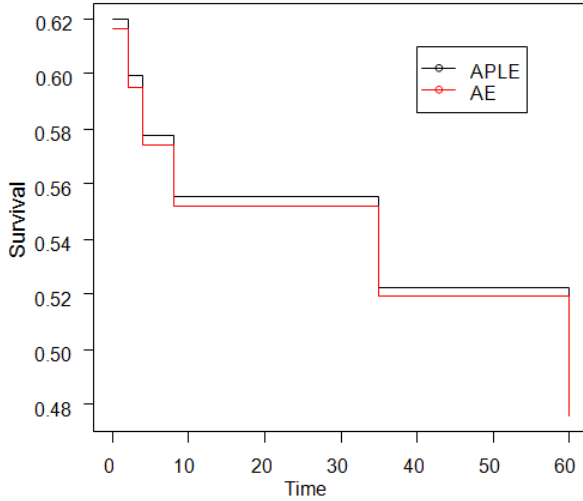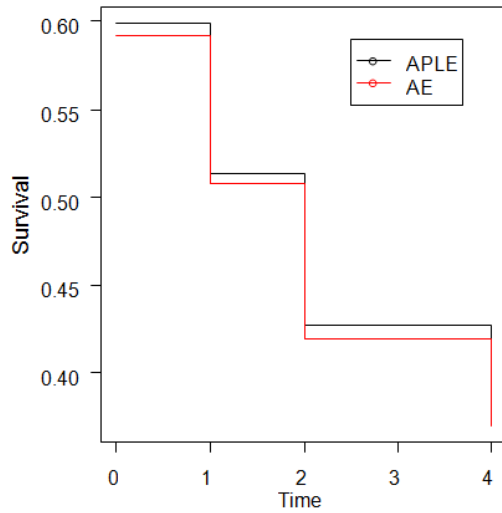**Figure 6.** *Survival curves for APLE and AE, for n=500 with 56% censoring; E[0.05].*

Figure 7. *Survival curves for APLE and AE, for n=100 with 42% censoring; LN[1.5, 25].*

Figure 8. *Survival curves for APLE and AE, for n=50 with 52% censoring; LN[1.5,15].*

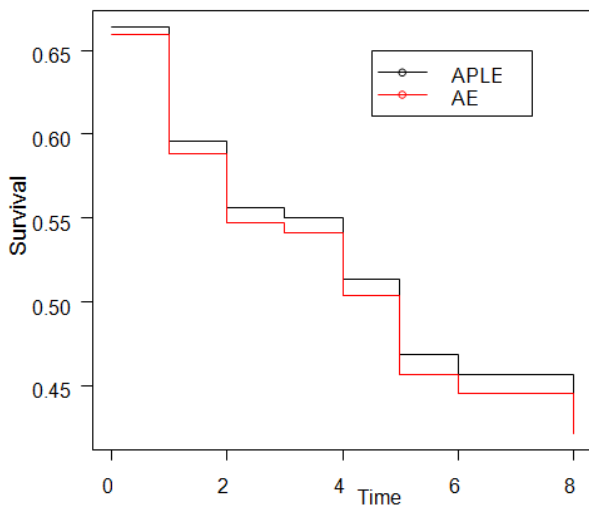*Figure 9.* *Survival curves for APLE and AE, for n=25 with 40% censoring; LN[1.5,5].*

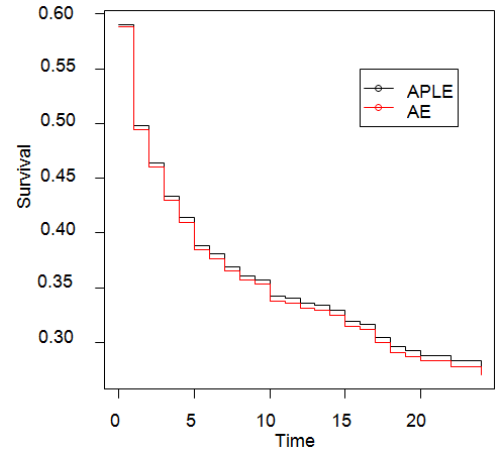*Figure 10.* *Survival curves for APLE and AE, for n=250 with 51.6% censoring; LN[1.5, 5].*

Figure 11. *Survival curves for APLE and AE, for n=800 with 34% censoring; LN[0.5, 5].*

4. Application to Real Data

Leukaemia data given in [16] and data set of 2418 Males with Angina Pectoris originally reported by Parker et al. [17] and reused in [18], are used for illustration. Table 15 and Figure 12 give the results for Leukaemia data while results for males with Angina Pectoris are reported in Table 16 and in Figure 13. These results for real data lead to similar conclusion as in the case of simulated data such that actuarial method underestimates the survival function and its results are less precise as compared to the results of adjusted product limit estimator. For instance, in Table 15 at time interval [6, 7) actuarial method give a survival probability of 0.853659 with corresponding standard error of 0.078064 while APLE give a survival probability of 0.856746 with corresponding standard error of 0.76449 and as mentioned before actuarial method assume that only half of the censored individuals are at risk and also estimation is done at the midpoint of each interval: if this assumption hold then also not all failed individuals in interval [6, 7) are at risk because some might have failed before the midpoint. From this example, it is clear that actuarial method is unreliable and APLE increases the accuracy of estimation of the survivor function by giving smaller standard errors as compared to actuarial method.

For large data set like the one reported in Table 16, the results still support the use of APLE over actuarial method in estimating survival function; the improvement in estimation of survival probabilities is significant and will have a significant impact on policy/decision making in real life problems.

Table 15. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for Leukaemia data set with n=21 with 57% censoring.*

Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[6, 7)	21	3	1	0.856746	0.853659	0.076449	0.078064
[7, 8)	17	1	0	0.806349	0.803443	0.086991	0.088155
[10, 11)	15	1	1	0.752318	0.748033	0.096422	0.097954
[13, 14)	12	1	0	0.689625	0.685697	0.106842	0.107816
[16, 18)	11	1	0	0.626932	0.623361	0.114049	0.114627
[22, 23)	7	1	0	0.537370	0.534310	0.128186	0.128261
[23, 24)	6	1	0	0.447809	0.445258	0.134519	0.135286

Table 16. *Survival Probabilities and Standard Errors for APLE and Actuarial Method for 2418 Males with Angina Pectoris with 32.8% censoring.*

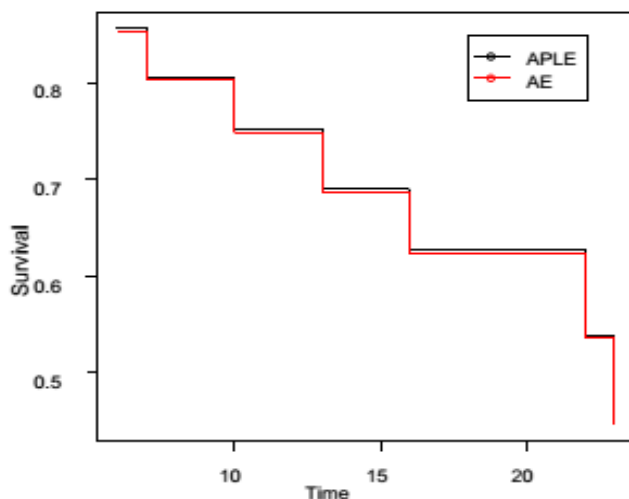| Time interval	r_j	d_j	c_j	APLE	AE	se.APLE	se.AE
[0, 1)	2418	456	0	0.8114	0.8114	0.0080	0.0080
[1, 2)	1962	226	39	0.7179	0.7170	0.0092	0.0092
[2, 3)	1697	152	22	0.6536	0.6524	0.0097	0.0097
[3, 4)	1523	171	23	0.5802	0.5786	0.0101	0.0101
[4, 5)	1329	135	24	0.5212	0.5193	0.0103	0.0103
[5, 6)	1170	125	107	0.4650	0.4611	0.0103	0.0104
[6, 7)	938	83	133	0.4228	0.4172	0.0103	0.0105
[7, 8)	722	74	102	0.3783	0.3712	0.0104	0.0106
[8, 9)	546	51	68	0.3423	0.3342	0.0106	0.0107
[9, 10)	427	42	64	0.3076	0.2987	0.0107	0.0109
[10, 11)	321	43	45	0.2653	0.2557	0.0110	0.0111
[11, 12)	233	34	53	0.2236	0.2136	0.0112	0.0114
[12, 13)	146	18	33	0.1939	0.1839	0.0116	0.0118
[13, 14)	95	9	27	0.1733	0.1636	0.0120	0.0123
[14, 15)	59	6	23	0.1508	0.1429	0.0129	0.0133
[15, 16)	30	0	30	0.0000	0.1429	NA	0.0133

Figure 12. *Survival curves for APLE and AE for Leukaemia data set with n=21 & 57% censoring.*
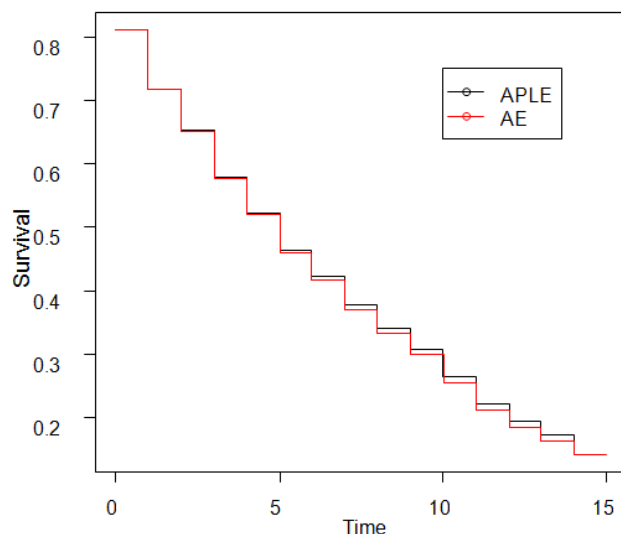

Figure 13. *Survival curves for APLE and AE for 2418 Males with Angina Pectoris & 32.8% censoring.*

5. Conclusion

This article considered the problem of estimating survival function from grouped observations that are under random censorship using adjusted product limit estimator. The performance of the stated estimator was compared with the performance of the actuarial method using simulated data as well as using real data and in assessing the results, the accuracy of the methods in estimating survival probabilities was considered whereby the method that provided smaller standard errors was taken to be more accurate. In all sample sizes considered in the study, adjusted product limit estimates were seen to have higher precision (smaller standard errors) than actuarial estimates and thus were preferred over actuarial estimates. Basing on the study findings, I urge researchers in this field of study and other related fields to use the adjusted product limit estimator so as to obtain results that are more accurate and also extend the method to observations that are under fixed censorship or data generated from other life distributions like Weibull among others.

References

[1] Job I. M. and Leo O. O. (2016). Estimating Survivor Function Using Adjusted Product Limit Estimator in the Presence of Ties. *American Journal of Theoretical and Applied Statistics*; 5(5), 290-296.

[2] Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

[3] Berkson, J. and Gage, R. R. (1950). Calculation of Survival Rates for Cancer. Proceedings of Staff Meetings, Mayo Clinic, 25, 250

[4] Cutler, S. J. and Ederer, F. (1958). Maximum Utilization of the Life Table Method in Analyzing Survival. *Journal of Chronic Diseases*, 8, 699—712.

[5] Breslow, N. and Crowley, J. (1974). A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship, *The Annals of Statistics*; 2 (3), 437-453.

[6] Elveback, L. (1958). Estimation of survivorship in chronic disease: the "actuarial" method. *J. Amer. Statist. Assoc*. 53, 420-440.

[7] Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. Wiley, New York.

[8] Peto, R. (1973). Experimental Survival Curves for Interval Censored Data, *Applied Statistics*, 22, 86-91.

[9] Klein, J. P. and Moeschberger, M. L. (1977). *Survival Analysis*. Springer-Verlag, New York.

[10] Sun, J. (1996). A non-parametric test for interval-censored failure time data with application to AIDS studies. *Statist. Medicine*, 15, 1387-1395.

[11] Gordon, A. C. (2016). *Analysis of Mortality: The Life Table and Survival*. Springer.

[12] Michael, J. S. (2015). *Survival Analysis*. John Wily & Sons Inc.

[13] Little, A. S. (1952). Estimation of the T-year survival rate from follow-up studies over a limited period of time. *Human Biol*. 24, 87-116.

[14] Gilbert, J. P. (1962). *Random censorship* Ph.D. thesis, University of Chicago.

[15] R Core Team (version 3.3.0). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org.

[16] Freireich, E. J., Gehan, E., Schroeder, L. R., Wolman, I. J., Burgert, E. O., Mills, S. D., and Lee, S. (1963). 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*; 21, 699-716.

[17] Parker, R. L., Dry, T. J., Willius, F. A., and Gage, R. P. (1946). Life Expectancy in Angina Pectoris. *Journal of the American Medical Association*, 131, 9 5—100.

[18] Lee E. T. and John W. W. (2003). *Statistical Methods for Survival Data Analysis (3rd Edn.)*. John Wiley & Sons, Inc., Hoboken, New Jersey.