Impact of Non-Random Right Censoring on Kaplan Meier Estimates and Log-Rank Test Results: A Simulation Study

Rasgele Olmayan Sağdan Sansürlü Gözlemlerin Kaplan Meier Tahminleri ve Log-Rank Test Sonuçlarına Etkisi: Bir Simulasyon Çalışması

Arzu KANIK,[a]
Seval KUL[b]

[a]Department of Biostatistics,
Mersin University Faculty of Medicine,
Mersin
[b]Department of Biostatistics,
Gaziantep University Faculty of Medicine,
Gaziantep

Geliş Tarihi/*Received:* 30.03.2010
Kabul Tarihi/*Accepted:* 12.05.2010

This study was presented in international congresses mentioned below:
Kul S, Kanik A. Does the location of right censored observations affect Kaplan-Meier estimates?", 28th Annual Conference of the International Society for Clinical Biostatistics, Alexandroupolis, Greece, 2007.
Kul S., Kanık A. "Log-Rank Test In The Absence of Random Right Censoring: A Simulation Study", 5th Conference of The Eastern Mediterranean Region of The International Biometric Society, Istanbul, Turkey, 2009.

Yazışma Adresi/*Correspondence:*
Arzu KANIK
Mersin University Faculty of Medicine,
Department of Biostatistics, Mersin,
TÜRKİYE/TURKEY
arzukanik@mersin.edu.tr

ABSTRACT Kaplan Meier method is one of the oldest and popular method to estimate the survival function from life-time data. And Log–rank test is used to compare Kaplan Meier curves to evaluate whether or not two or more groups are statistically equivalent. In survival analysis it is assumed that censored observations are randomly distributed in data and don't have any pattern. But in real life this isn't the case always. In medical studies because of dissatisfactory/ inefficient therapy, remedication, negative news about remedy or interventions and high risk of hospital infection most of the right censoring can occur within very short time interval. In the absence of censoring randomization Kaplan Meier estimations and Log-rank test, which are the most commonly used test, results may be biased due to the unbalanced pattern of the censored observations. But there is no study to show the effect of the censored observation pattern on estimation and comparison of the survival curves. In this study we aimed to show statistical properties and performance of Kaplan Meier estimates and Log-rank test under violation of random censoring assumption. A simulation performed to show impact of non-random right censoring on Kaplan Meier estimates and Log-rank test results. Our simulations cover 3 distributions, 4 levels of censoring, and two samples size for Kaplan Meier and Log-rank test. In the all simulated data the pattern of the censored data were changed and compared to data including random censoring. As a result of our simulations, we found location pattern of right censored observation in data set has no significant effect on Kaplan Meier estimate. There were two main points which have effect on Kaplan Meier estimates were sample size and censoring rate. When comparing the curves nonrandom censoring resulted in inflation of Type I error.

Key Words: Kaplan-Meiers Estimator, Logrank test, right censored data, censoring pattern

ÖZET Kaplan Meier yöntemi ve Log-rank testi ile survival analizlerinde yaşam eğrilerinin tahmininde ve eğrilerin karşılaştırılmasında en yaygın olarak kullanılan yöntemlerdir. Hayatta kalma analizlerinde sansürlü gözlemlerin rasgele olarak dağıldığı ve içinde herhangi bir yapı barındırmadığı varsayılır. Fakat bu önşartın klinik çalışmlarda her zaman karşılanması mümkün değildir. Klinik çalışmalarda bir tedavinin tatmin edici olmaması, yeni bir ilaç kullanımı, bir tedavi yöntemi hakkındaki negatif haberlerin yayılması nedeni ile kısa bir zaman aralığı içinde bir çok sansürlü gözlemin oluşması mümkündür. Sansürlü gözlemlerin bu dengesiz dağılımı ve önşartın ihlalinden Kaplan Meier tahminlerinin ve Logrank test sonuçları etkileneceği beklenmesine karşın sansürlü gözlemlerin yapısının ve oluşma zamanının yaşam eğrilerinin tahminine ve Logrank test sonuçlarına nasıl bir etkisi olacağına dair literatürde yapılmış bir çalışmaya rastlanmamaktadır. Bu çalışmada rasgele sansür varsayımının ihlali durumunda Kaplan meier ve Logrank yöntemlerinin istatistiksel özellikleri ve performansı gösterilmesi amacı ile bir simülasyon çalışması düzenlenmiştir. Simülasyon çalışmamızda 3 farklı istatistiksel dağılım, 4 farklı sansürlü gözlem düzeyi ve 2 farklı örnek genişliği kullanılmıştır. Tüm veri setlerinde sansürlü gözlemlerin yerleri değiştirilmiş ve aynı miktarda rasgele sansürlü gözlem içeren veri setleri ile karşılaştırılmıştır. Sonuç olarak sağdan sansürlü gözlemlerin yerleşim yerinin Kaplan Meier tahminlerinin üzerine istatistiksel olarak anlamlı bir etkisinin olmadığı, fakat örnek genişliğinden ve sansürlü gözlemlerin yerleşim yerinden etkilendiği ortaya konmuştur. Rasgele sansürlü gözlem içeren veri ile rasgele olmayan sansürlü yapı Logrank testi ile karşılaştırıldığında Tip I hatanın şişdiği gözlenmiştir.

Anahtar Kelimeler: Kaplan-meier tahmincisi; logrank testi, sağdan sansürlü veri, sansürlü gözlemleri yerleşimi

Turkiye Klinikleri J Biostat 2010;2(2):82-8

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs.[1] Survival analysis has its own terminology. In survival analysis the event, can be death, disease incidence, relapse from remission, recovery, discharge from hospital or any event of interest that may happen to an individual.[2] The time can refer to years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs or age of an individual when an event occurs.[3]

Sometimes patients who join the study are still alive at the terminating of the study or patient can withdraw from the study by his desire (adverse drug reaction, dissatisfaction of remedy), patient can die because of the other reason (competing risk) or patient can move during the study. It is not possible to observe these patients any more. Data for which the survival time cannot be always evaluated because of the all above mentioned reasons are called "censored data".[4] Because of censored observations, to analyse survival data special methods are used.[5]

Kaplan Meier estimate is proposed by Kaplan E.L., and Meier P. to estimate survival function of life times in consider of censored observations. Kaplan Meier estimate is proposed by Kaplan E.L., and Meier P. to estimate survival function of life times in consider of censored observations.[6] The Kaplan-Meier estimator estimates the survival function from life-time data,

Y is a survival time, C is a censoring time,

Let, (Y,C), $(Y_1,C_1),\ldots,(Y_n,C_n)$ independent and identically distributed random variables, where Y and C take values in *R, R,* respectively .

Set $Z_i = \min\{Y_i, C_i\}$, $\delta_i = I_{\{Y_i < C_i\}}$ is the censoring indicator respectively for all $i = 1,2,\ldots\ldots,n$.

Let, Z are ordered distinct times, if $Z_i \le t_{[i]}$ survival time of in this i. point can be estimated by Kaplan Meier estimation.

$$\hat{S}(t) = \prod_{Z[i]<=t,\, i=1} \frac{(\mathrm{n-i})}{(\mathrm{n-i+1})^{\delta[\mathrm{i}]}}$$

Hazard rate represents the risk that an individual fails immediately after time t given survival at time t. Hazard rates can be calculated by the help of the hazard functions and we can achieve the survival function through the hazard function.[2]

$$H(t) = \int_0^t h(u)du = -\log S(t)$$

Kaplan Meier method is one of the oldest and popular method to estimate the survival function from life-time data. And Log–rank test is used to compare Kaplan Meier curves to evaluate whether or not two or more groups are statistically equivalent.[7] Both of the methods have the same way of handling censored observations; if a survival time is censored, that individual is considered to be at risk of dying in the week of the censoring but not in subsequent weeks.

Although these methods have the considerable advantage that these methods do not require anything about the shape of the survival curve or the distribution of survival times and both methods are called non-parametric, they have several assumptions about both complete and incomplete data.[8] To apply Kaplan Meier and Log-rank methods, it is assumed that subjects are driven from population of interest randomly, survival times are independent and identically disturb, censored and uncensored observations come from the same survival distribution and distributions of censoring times are independent of the survival times.[7,8] Important assumption about censoring is non-informative which means that the absence of competing risks and subjects have the same values for covariate predictors.[9] The Log-rank test is based on the same assumptions as the Kaplan Meier survival curve.

The violations of these assumptions may have effect on the analysis survival data and results in incorrect or biased and unreliable estimates. In particular, censoring pattern of the data may increase the effect of assumption violations. In medical studies because of dissatisfactory or inefficient therapy, remedication, negative news about remedy or interventions and high risk of hospital infection most of the right censoring can occur within very short time interval.[10] Deviations from random censoring assumption matter most censoring pattern are different in the groups being compared. For example In a study two drugs effects on pain relief after surgery is compared and patients pain is persist so remedication is necessary.[10] If the times to remedication are earlier for treatment 1 than for treatment 2 non-random censoring is violet. Does censoring time affect comparison results is a very important problem question here. Another practical example for non-random censoring is that parents are very sensitive about their babies care. When any suspected infection breaks out in new born unit of a hospital, many families change their hospital because of infection risk. So these patients are censored in the same period. The comparison of two hospitals average time of discharge in neonatal intensive care unit may be biased due to the presence of non-random censored observations.

Although it is known that increasing of censored observations proportions in the data makes the Kaplan Meier estimates worse, and there are many simulation studies in the literature,[11-13] which show effect on censored data and Kaplan Meier estimates and Logrank test, there is no study to show effect of censored observations pattern on Kaplan Meier estimation and Logrank test result. In this study we aimed to show statistical properties and performance of Kaplan Meier estimates and Log-rank test under violation of random censoring assumption.

SIMULATION STUDY

In the simulation part, to be able to compare our simulation results the distributions defined in the article Hess et al[14] is used. We simulated data from 3 different distributions; survival functions derived from constant(exponential), linear increasing and linear decreasing hazard functions including censored observations with uniform distributions and censoring located at three different parts of the data with two sample sizes (n=100, n=250) for Kaplan Meier and Log-rank test. Our simulations cover four proportions of random censoring: 20 percent; 30 percent, 40 percent and 50 percent. Survival functions and parameters used in the simulations are given in the (Table 1).

In (Figure 1), an example of the simulated data used in Kaplan Meier estimates is given for constant hazard rate. Censored cases are showed with "C" and uncensored (observed) cases are with "O". For the distribution, we fixed censored data rate at 20%. And locations of the censored observations were moved from left to right in the data sets which are called random, I., II. and III. Interval, respectively.

We applied Kaplan Meier estimation for these four data sets and real survival and estimated survival functions are compared with same approach 96 different data are simulated. For Kaplan Meier estimates L_2 errors were calculated to evaluate the differences between real density and the estimated functions. A total of 1000 independent samples were simulated for each experiment.

$$L_2 = \frac{(Y - f(x))^2}{n}$$

In the case of many censored observations larger than the largest observed failure time, Kaplan

TABLE 1: Survival functions and parameters used in the simulation.

	S(t)	*n*= 100	*n*= 250
Constant Hazard	$S(t) = \exp(-\lambda_0 t)$	$\lambda_0 = 0.0256$	$\lambda_0 = 0.0357$
Linear increasing	$S(t) = \exp(\frac{-\lambda_0 t^2}{200})$	$\lambda_0 = 0.0568$	$\lambda_0 = 0.0794$
Linear decreasing	$S(t) = \exp(-\lambda_0 (t - \frac{t^2}{200}))$	$\lambda_0 = 0.04651$	$\lambda_0 = 0.0650$
		$S(t = 90) = 0.1$	$S(t = 90) = 0.04$

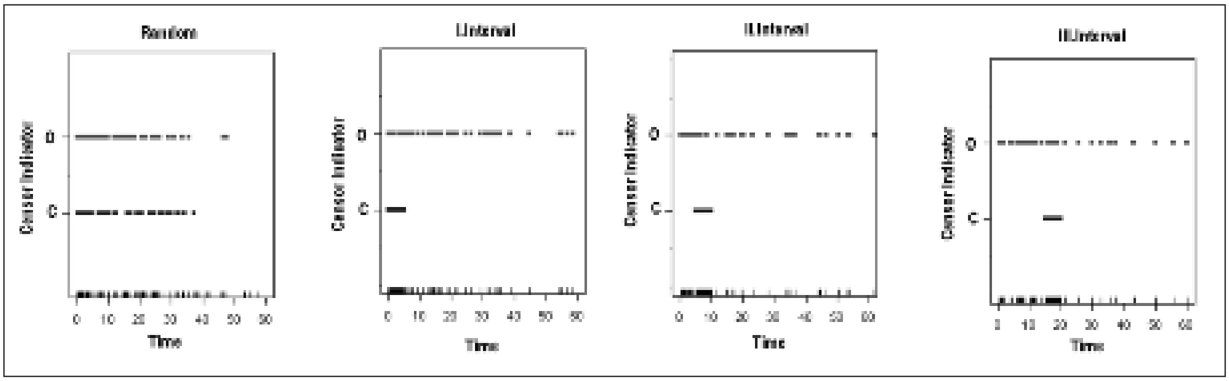

FIGURE 1: Pattern of the censored observations for constant hazard n=100, censoring rate=0.20.

Meier estimator is biased.[15] To avoid this bias in our simulation study the last observation was always observed observation.

In the Log-rank simulation part random censoring is compared versus nonrandom censoring according to Type I error rates. Functions used in Kaplan Meier estimates are used to validate the distributions. In this case for each experiment 5000 independent samples were simulated and compared with Logrank test. Simulation programs were implemented in R language.

CONCLUSION

Impact of non-random right censoring on Kaplan Meier estimates and Log-rank test results are shown with a detailed simulation study. The results of the simulation studies for Kaplan Meier estimates for N=100 and for N=250 are shown in (Figure 2) and (Figure 3) respectively. When we look at (Figure 2) and (Figure 3), we see Kaplan Meier estimate works well for all situations. Compared to random censoring, nonrandom censoring resulted

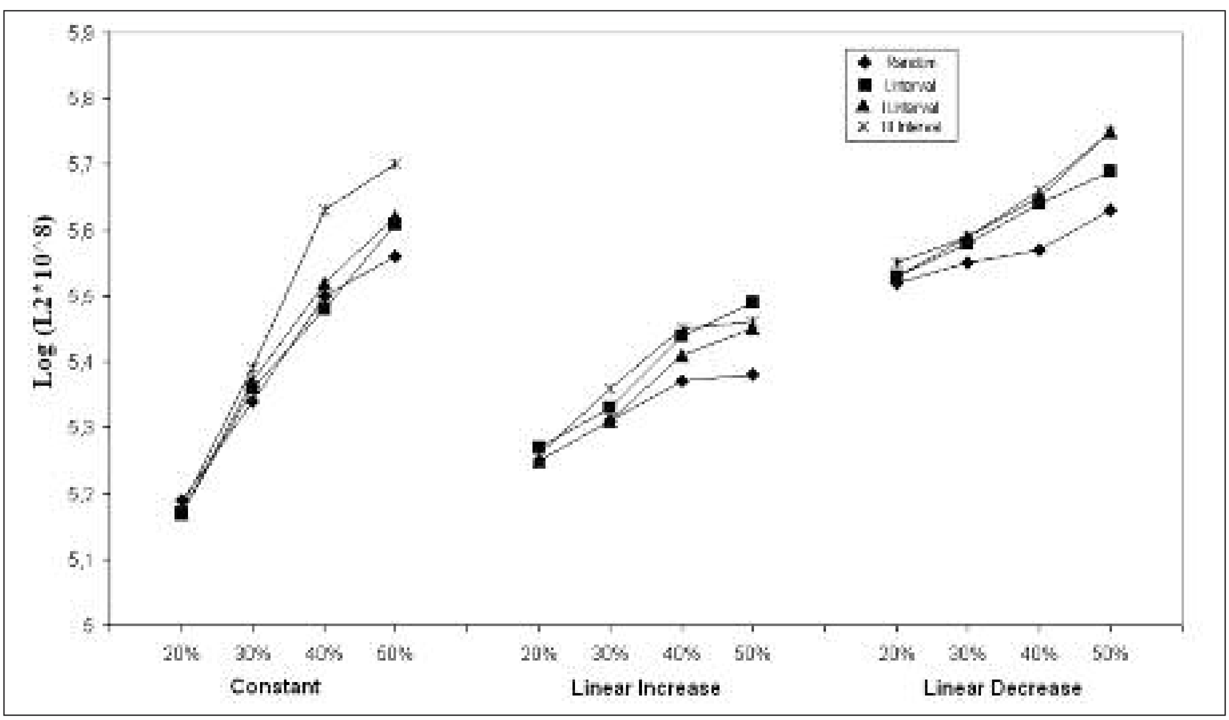

FIGURE 2: Simulation results for N=100 in terms of a L_2 (plotted as log (L_2x10^8) for four different censoring locations and with four levels of censoring; 20 percent, 30 percent, 40 percent, 50 percent.

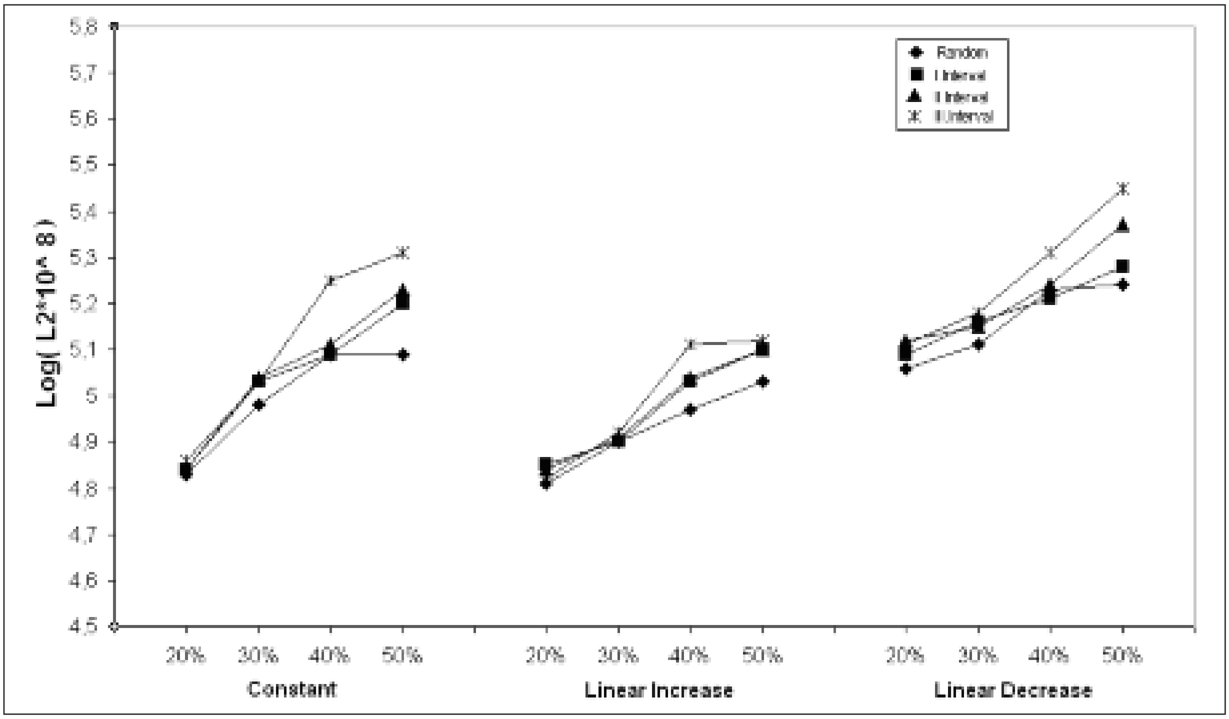

FIGURE 3: Simulation results for N=250 in terms of a L_2 (plotted as log (L_2x10^8) for four different censoring locations and with four levels of censoring; 20 percent, 30 percent, 40 percent, 50 percent.

in no statistically significant changes in a median L_2 error for 20% censoring rate, there was a modest increase for 30% censoring rate, but 40 and 50 percent censoring, nonrandom censoring resulted in a relative increase in a median L_2 error. For 50 percent censoring rate, this increase is much clearer.

In survival analysis it is thought that censored observations which occur at the beginning of the study result in more information loss and it is seen as a major problem by researchers. But our simulation study showed that censored observations occurred at the end of the study has much more negative effect on Kaplan Meier estimates. In survival studies at the end of the study, estimations are based on less observation. So this result is related with the number of the observation at the estimation point.

Another important point which has effect on Kaplan Meier estimate is censoring rate. When censoring rate is increased, estimation error increases. Because of sharp increase in after 30% censoring rate, we can conclude that 30% censoring rate is the reasonable limit which can be tolerate by Kaplan Meier estimator. Pattern of right censored observation in data set has little effect on Kaplan Meier estimate which doesn't make any effect on clinical interpretation of the results. Our results were very similar for whole distributions used in simulations which mean that Kaplan Meier estimator performance is independent from the distribution of survival times.

Although censoring pattern doesn't have any clinically significant effect on survival distribution estimates, it doesn't mean that comparison of the curves isn't biased. In the second part of our simulation we tested the impact of non-random right censoring on Log-Rank Test performance. To compare survival curves, same distributions are used with Kaplan Meier estimates. Distribution of time survival times and the censoring rates are fixed in all of the groups only censoring pattern was varied. In (Figure 4) as an example of simulated data, Cumulative survival distributions and pattern of the censored observations for Linear increasing hazard, n=100, censored rate=40% to compare random censoring versus non-random censoring are given in terms of pattern of the censoring.

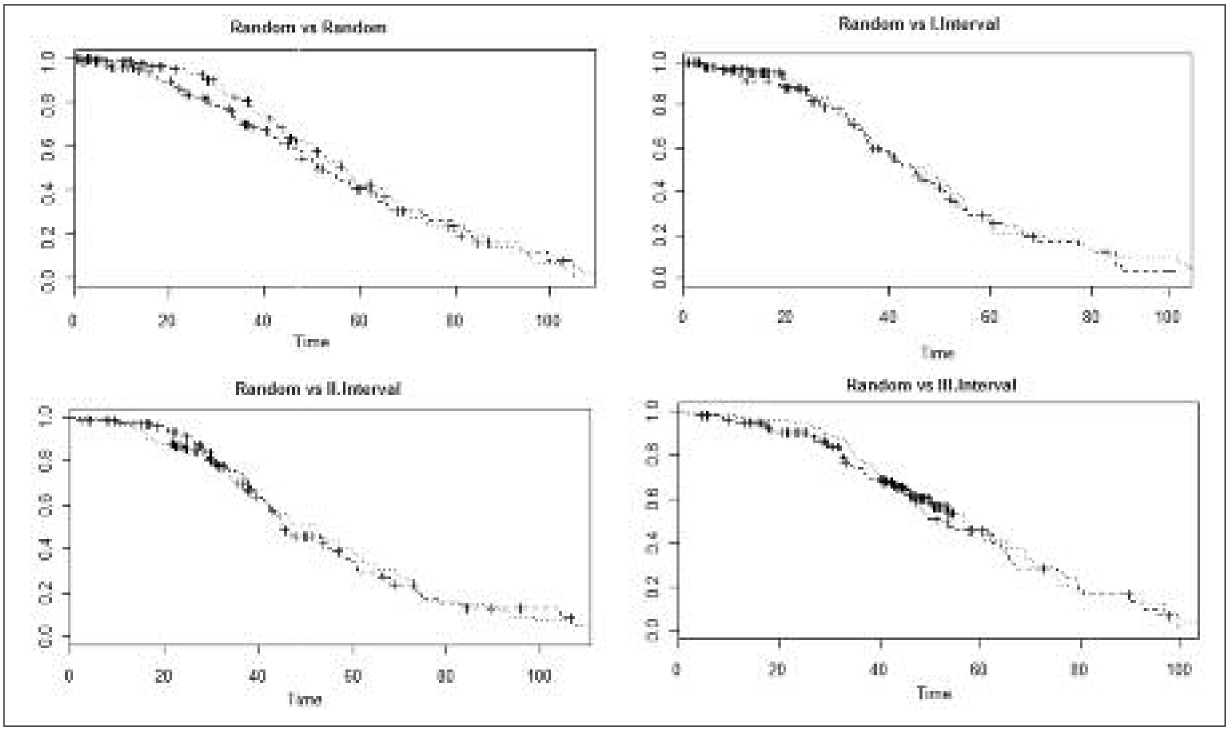

FIGURE 4: Cumulative survival distributions and pattern of the censored observations, for Linear increasing hazard, n=100, censored rate=40%.

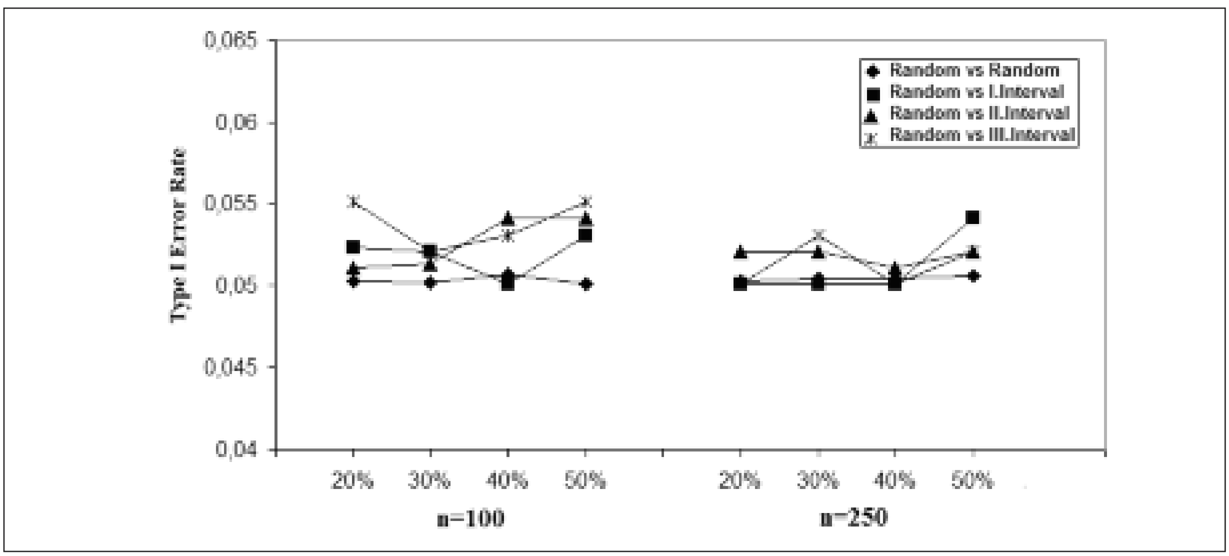

FIGURE 5: Log rank test simulation results for linear increasing function, for N=100 and 250 in terms of a Type I error for four different censoring locations and with four levels of censoring; 20 percent, 30 percent, 40 percent, 50 percent.

The simulation results for Log-rank test performance for N=100 and 250 are given in (Figure 5) in terms of a Type I error for four different censoring locations and with four levels of censoring; 20 percent, 30 percent, 40 percent, 50 percent for Linear increasing function. The results were very similar for whole distributions. As can be seen in (Figure 5), When two survival curve include randomly censored data was compared the type I error level was at 0.05 but non-random censoring pattern inflated the Type I error rate. Although it seems that type I errors are reasonable in (Figure 5), it should be considered that the Log-rank test is most likely to detect a difference between groups when the risk of an event is consistently greater for one group than another.[5] But it is unlikely to detect a difference when survival curves cross,[6] which has high possibility when two survival distributions are the same.

As a result of our simulations, we found location place of right censored observation in data set has no significant effect on Kaplan Meier estimate. There are two main points which have effect on Kaplan Meier estimates are sample size and censoring rate. This result will encourage the physicians to follow up their patients; even most of the censored observations are occurred at the beginning of the study. When comparing the curves nonrandom censoring results in Inflation of Type I error for Log Rank which is probably higher than detected. We advise to use Kaplan Meier method for non-random censoring but non random censoring has negative effect of Logrank test and we don't advise to use Logrank test under this special case.

REFERENCES

1. Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York: Springer; 1997. p.21-22.
2. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. 2nd ed. New York: Springer; 2005. p.8-14.
3. Vittinghoff E, Glidden D, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models. 1st ed. New York: Springer; 2004. p. 53-54.
4. Györfi L, Kohler M, Krzyzak A, Walk H. A Distribution-Free Theory of Nonparametric Regression. 1st ed. New York: Springer; 2002. p.540-1.
5. Dawson B, Trap GR. Basic & Clinical Biostatistics. 4th ed. Lange Medical Books-McGraw-Hill, Medical Pub. Division; 2004. p.296-297.
6. Kaplan EI, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc 1958;53(282):457-81.
7. Bland JM, Altman DG. The logrank test. BMJ 2004;328(7447):1073.
8. Altman DG, Bland JM. Time to event (survival) data. BMJ 1998;317(7156):468-9.
9. Lagakos SW. General right censoring and its impact on the analysis of survival data. Biometrics 1979;35(1):139-56.
10. Ganju J, Lakatos E, Rothe E. A simple way to estimate the median time and compare survival distributions in analgesic trials under informative censoring. J Biopharm Stat 1999;9(4):683-93.
11. Wellek S. A log-rank test for equivalence of two survivor functions. Biometrics 1993;49(3): 877-81.
12. Kim J. Kang DR, Nam CM. Logrank-type tests for comparing survival curves with interval-censored data. Computational Statistics & Data Analysis 2006;50(11):3165-78.
13. Akbar A, Pahsa GR, Naqvi SF. Properties of Kaplan-Meier Estimator: Group Comparison of Survival Curves. Eur J Sci Res 2009;32(3): 391-7.
14. Hess KR, Serachitopol DM, Brown BW. Hazard function estimators: a simulation study. Stat Med 1999;18(22):3075-88.
15. Moeschberger ML, Klein JP. A comparison of several methods of estimating the survival function when there is extreme right censoring. Biometrics 1985;41(1):253-9.