RESEARCH ARTICLE

Cancer Survival Estimates Due to Non-Uniform Loss to Follow-Up and Non-Proportional Hazards

Jagathnath Krishna K M*, Aleyamma Mathew, Preethi Sara George

Abstract

Background: Cancer survival depends on loss to follow-up (LFU) and non-proportional hazards (non-PH). If LFU is high, survival will be over-estimated. If hazard is non-PH, rank tests will provide biased inference and Cox-model will provide biased hazard-ratio. We assessed the bias due to LFU and non-PH factor in cancer survival and provided alternate methods for unbiased inference and hazard-ratio. **Materials and Methods:** Kaplan-Meier survival were plotted using a realistic breast cancer (BC) data-set, with >40%, 5-year LFU and compared it using another BC data-set with <15%, 5-year LFU to assess the bias in survival due to high LFU. Age at diagnosis of the latter data set was used to illustrate the bias due to a non-PH factor. Log-rank test was employed to assess the bias in p-value and Cox-model was used to assess the bias in hazard-ratio for the non-PH factor. Schoenfeld statistic was used to test the non-PH of age. For the non-PH factor, we employed Renyi statistic for inference and time dependent Cox-model for hazard-ratio. **Results:** Five-year BC survival was 69% (SE: 1.1%) vs. 90% (SE: 0.7%) for data with low vs. high LFU respectively. Age (<45, 46-54 & >54 years) was a non-PH factor (p-value: 0.036). However, survival by age was significant (log-rank p-value: 0.026), but not significant using Renyi statistic (p=0.067). Hazard ratio (HR) for age using Cox-model was 1.012 (95%CI: 1.004 -1.019) and the same using time-dependent Cox-model was in the other direction (HR: 0.997; 95% CI: 0.997- 0.998). **Conclusion:** Over-estimated survival was observed for cancer with high LFU. Log-rank statistic and Cox-model provided biased results for non-PH factor. For data with non-PH factors, Renyi statistic and time dependent Cox-model can be used as alternate methods to obtain unbiased inference and estimates.

Keywords: Time-dependent Cox-model- non-proportional hazards- Schoenfeld global statistic- Renyi Statistic

Introduction

Survival analyses (Breslow, 1975; Altman, 1992; Klein and Moeschberger, 2003; Kleinbaum and Klein, 2005; Taib et.al. 2008) deals with the application of various methods in order to assess the prognosis of a patient suffering from a particular disease over a time period. With the increased use of computers powered with automated tools, storage and retrieval of large volumes of cancer data are possible for evaluating the effectiveness of a given treatment. Survival estimates largely depend on the extent of follow-up information of individual patients. A typical problem encountered in survival studies is the subjects who are lost during follow-up. Survival estimates lack reliability due to loss to follow-up (LFU). Incomplete follow-up information was available on LFU patients. Standard technique such as the Kaplan-Meier (K-M) (Kaplan and Meier, 1958) method assumes LFU as random loss and having the same survival probability as those on the remaining follow-up. LFU is considered as non-random if associated factors for loss influence the study outcome. The probability of survival of these patients will be different from those who remain on follow-up. Hence, the bias in survival estimates largely depends on the degree of LFU. If the proportion of LFU is high, survival estimates may be over-estimated (Ganesh, 1995; Mathew, 1996).

Log-rank test (Mantel, 1966) is the commonly used test to assess significant difference between groups of survival data containing censored observations. Its test statistic is based on comparing expected number of deaths under the null hypothesis of no difference between groups with the observed number of deaths at each of the successive distinct death times under proportional hazards assumption. Cox-proportional hazards regression model (does not assume a specific mathematical distribution for observed survival time) (Cox, 1972) is widely used to assess the effect of various prognostic factors on cancer survival (conversely failure) after adjusting each for the other factors. The model measures hazard ratio (effect of an explanatory variable on the hazard or risk of an event) of the outcome under the assumption that the hazard is constant over the follow-up period.

Before using log-rank test and Cox-model, proportional hazard (PH) assumption needs to be checked as these methods will lose power if PH assumption violates.

Division of Cancer Epidemiology and Bio-statistics, Regional Cancer Centre, Medical College Campus, Thiruvananthapuram, India. *For Correspondence: jagath.krishna@gmail.com

Statistical tests such as Schoenfeld global test (Schoenfeld, 1980) as well as K-M plots can be used for checking PH assumption. Renyi or Supremum test (Renyi, 1953; Gill, 1980; Wei, 1984; Davis and Xie, 2011), modified Kolmogorov-Smirnov test (Fleming et al. 1980), modified log-rank test (Lin and Wang, 2004; Liu et al., 2007; Lin and Xu, 2010), two stage test procedure by Qiu and Sheng (2008) are a few among the alternate test to log-rank in testing the significance of survival probability in presence of non-PH. Li et al., (2015) compared various tests for checking the significance of survival probabilities and showed that for large sample data with more than 60% censoring of non-PH variables, Renyi test is one of the best alternatives to log-rank test. Similarly, Cox-model (Cox, 1972) provides biased inference when survival probabilities are non-proportional or the covariates depends on time. Time-dependent Cox-model (Fisher and Lin, 1999), an extension of Cox-model can be used in such situation.

In the present paper, the bias in survival estimates due to a high proportion of LFU and presence of a factor with non-proportional hazards were illustrated using realistic data and provided alternate methods to derive unbiased estimates and inferences.

Materials and Methods

Two sets of realistic data, one with a low 5-year LFU (<15%) and the other with a high LFU (>40%), of breast cancer (BC) patients reported in the Regional Cancer Centre (RCC), Trivandrum for the years 2006-2008 and 2009-2011 respectively were used for illustration. Date of diagnosis of BC was considered as the starting date of the study or the entry time of patients in to study. Age, stage, follow-up time and status at last follow-up were obtained from the cancer registry database. Age (a non-PH factor) was grouped into <45 years, 45-54 years and >54 years and stage (a non-PH factor) was grouped into I, II, III and IV based on TNM (Sobin et al., 2009; Gospodarowicz and Wittekind, 2009). Subjects who died during the course of their follow-up were accounted as 'event'. All other patients were considered as 'censored'. Five-year follow-up time was calculated by subtracting the date of last follow-up from the date of diagnosis. Based on the two data sets of high vs. low LFU, the bias in survival estimates were assessed using K-M plots.

The data set with low LFU was used for assessing the bias in inference for the non-PH factor (age) using log-rank test. The log-rank test statistic is based on comparing expected number of deaths (E_i) under the null hypothesis of no difference between groups with the observed number of deaths (O_i) at each of the successive distinct death times $t_1 < t_2 < ... < t_n$ and is calculated as:

$$\chi^2_{log-rank} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Non-proportionality by age was assessed using Schoenfeld statistic, in which residuals on functions of time provides non-zero slope for non-proportional hazards (Schoenfeld, 1980; Abeysekara and Sooriyarachchi, 2009; Xue et al., 2013). Renyi statistic was used (Renyi, 1953)

for assessing statistical significance in survival by age. Its test statistic is derived from a collection of data points (X_{11}, C_{11}), (X_{12}, C_{12}), ..., (X_{1N}, C_{1N}); ...; (X_{k1}, C_{k1}), (X_{k2}, C_{k2}), ..., (X_{kN}, C_{kN}) from the set (X_{ki}, C_{ki}) where X is the survival time and C is the right censoring time from the 3th individual in k^{th} group. Associated with each (k,i) will be the indicator variable $\Delta_{ji} = I_{(X<C)}$ which equals one if event occurred and zero otherwise. Y_k denotes the number of people at risk in group k. The survival time for each group k at time t can be represented as:

$$\widehat{S_k(t)} = \prod_{i:x_{k(i)}<t} \left(\frac{Y_k(X_{k(i)})-1}{Y_k(X_{k(i)})} \right)^{\Delta_{k(i)}}$$

Instead of taking the sum of the overall time points, the numerator of the Renyi statistic is simply

$M = Max(Z_k(t))$,

$$Z_k(t) = \sum_{j=1}^{\sum \Delta_i} W(t_i) \left[\Delta_{ji} - Y_{ji} \left(\frac{\Delta_i}{Y_j} \right) \right], \quad j = 1, ..., k$$

where W is a weight function and the variance of this estimate is

$$\widehat{\sigma_{kk}} = \sum_{j=1}^{\sum \Delta_i} W(t_i)^2 \frac{Y_{ji}}{Y_j} \left(\frac{Y_j - \Delta_j}{Y_j - 1} \right), j = 1, ..., k.$$

For a single time point t, the corresponding test statistic is $B(t) = \frac{M}{\sqrt{\sigma_{kk}}}$, where the distribution of B can be approximated by a standard Brownian motion process. The corresponding p-value (Billingsly, 1968) can be computed as:

$$P[Max|B(t)| > y] = 1 - \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} e^{\frac{-\pi^2(2k+1)^2}{8y^2}}$$

Further, hazard-ratio corresponding to age (considered as continuous variable) was assessed using Cox-regression model (PH assumption) and time dependent Cox-regression model (need not be satisfied PH assumption), where time-dependent variables are those that can change value over the course of time. The time dependent Cox model (Ata and Sozer, 2007) is given by:

$$\lambda\{t|X(t)\} = \lambda_0(t) \exp\{b^* X(t)\}, b^* = (0, 1, ... k^*)$$

where k* is the product of the levels of the variables that do not satisfy the proportional hazards assumption.

The survival analysis using K-M plot, log-rank test, Cox- model and time-dependent Cox-model were employed using SPSS 11 software. Suitable modification in the SAS program (Abeysekera and Sooriyarachchi, 2009 and Borucka, 2013) for Renyi test and Schoenfeld global test were made in the present analysis.

Results

Over-estimated 5-year BC survival of 90% was observed for 2009-11 data "with 5-year LFU>40%" compared to 67% for 2006-2008 data (with 5-year LFU <15%). Not much difference in 5-year survival for early stage BC (stage I & II) when the above two data sets were used, whereas over-estimated survival probability was observed for late stage BC corresponding to 2009-11 data with higher proportion of LFU. Five-year survival

Table 1. Breast Cancer Follow up Proportion (In %) By Stage 2006-2008 Vs. 2009-2011

Year	I		II		III		IV		Overall	
	06-08	09-11	06-08	09-11	06-08	09-11	06-08	09-11	06-08	09-11
1	100.0	100	99.6	99.2	100	97.8	100.0	87.2	99.8	98.4
2	100.0	95.2	97.2	96.1	97.6	89.0	98.6	73.6	97.8	93.4
3	98.7	90.3	94.9	92.0	95.6	81.3	97.6	64.0	95.8	87.7
4	96.1	51.0	91.9	55.2	92.7	49.4	96.3	45.6	92.9	48.5
5	87.8	29.0	86.7	31.2	90.2	30.0	95.0	37.6	88.3	26.0

Table 2. Breast Cancer Survival (%) [Standard Error: SE (%)] by Stage 2006-2008 vs. 2009-2011

Year	Stage							
	I		II		III		IV	
	06-08	09-11	06-08	09-11	06-08	09-11	06-08	09-11
1	98.7	100	98.4	99.1	94.8	97.9	73.1	88.3
	(1.3)	0	(0.5)	(0.3)	(0.9)	(0.6)	(3.2)	(0.3)
2	97.4	99.3	94.5	97.4	80.9	92.5	42.2	72.4
	(1.8)	(0.7)	(0.8)	(0.5)	(1.7)	(1.1)	(3.8)	(4.4)
3	96.1	98.6	91.1	96.1	70.2	88.6	27.9	65.1
	(2.2)	(1)	(1)	(0.6)	(2)	(1.3)	(3.3)	(5.1)
4	94.8	97.2	88	94.3	62.9	84.5	21.1	63.1
	(2.5)	(1.7)	(1.2)	(0.8)	(2.1)	(1.6)	(3)	(5.1)
5	93.5	97.2	86.1	93.2	58.3	81.9	15.3	59.2
	(2.8)	(1.7)	(1.3)	(0.9)	(2.1)	(1.9)	(2.6)	(6.1)

for stage III was 58% using 2006-2008 data (5-year LFU: 10%) and 82% using 2009-2011 data (5-year LFU: 70%) and the same for stage IV was 15% (5-year LFU: 5%) using 2006-2008 data and 59% (5-year LFU: 62%) using 2009-2011 data (Tables 1 and 2).

K-M plot by stage was found to be proportional to each other (log-rank p-value <0.001) (Figure 1). Survival by age was non-proportional (Figure 2). Survival curves for age groups 45-54 years and above 54 years crossed at different time points (Table 3) and correspondingly, Schoenfeld global test was statistically significant (p-value: 0.036). Even though PH assumption was not satisfied for survival by age, log-rank test (p-value: 0.026) showed statistical significance. Renyi statistic with respect to age (censoring for age<45 years: 74%,

45-54 years: 68% and >=55 years: 66%) showed only borderline significance (p-value: 0.069). Hazard ratios (HR) for age (continuous variable) were significant when Cox-model and time-dependent Cox-model were used, however in different directions. HR was 1.012 (95% CI: 1.004 -1.019) when Cox-model was used and 0.997 (95% CI: 0.997- 0.998) when time-dependent Cox-model was used (Table 4).

Discussion

In the present study, we illustrated the bias in survival estimates using two large data sets of BC patients treated in the RCC, Trivandrum with varying degree of LFU. Data with lower follow-up showed over-estimated BC
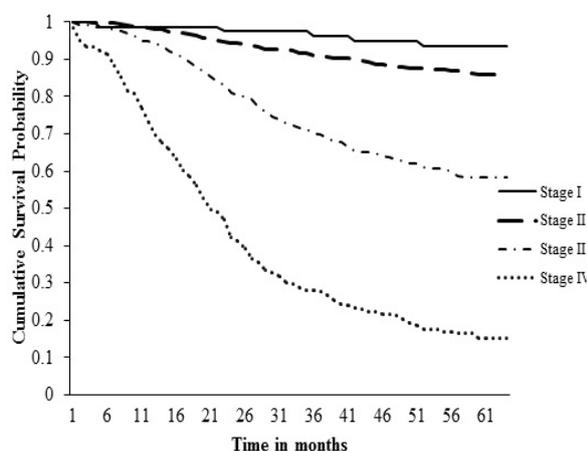

Figure 1. Survival Curve for Breast Cancer Survival with respect to Stage (2006-2008)
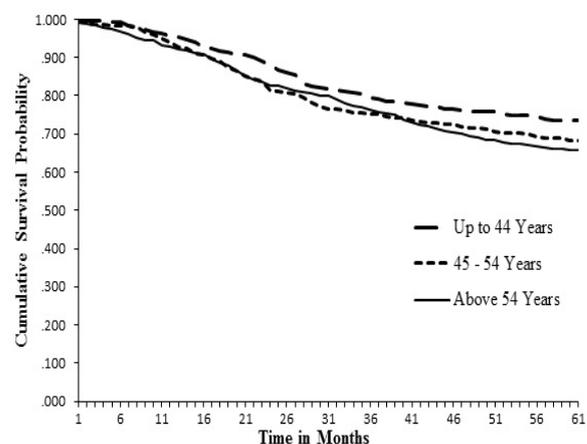

Figure 2. Survival Curve for Breast Cancer Survival with Respect to Age (2006-2008)

Table 3. Time Points at which Survival Probability (in %) for Age Cross (2006-2008)

Time in Months	< 45 Years (SE %)	45 - 54 Years (SE %)	> 55 Years (SE %)	P-Value Schoenfeld Global Test	Log-Rank Test	Renyi Test
1	99.8 (0.2)	99.5 (0.3)	99.7 (0.2)			
2	99.6 (0.3)	99.1 (0.4)	98.7 (0.4)			
9	97.3 (0.7)	96.7 (0.7)	94.7 (0.9)			
15	93.9 (1.1)	90.9 (1.2)	91.2 (1.1)	#*0.036	*0.026	0.069
17	92.4 (1.2)	89.9 (1.3)	89.6 (1.2)			
24	88.2 (1.5)	81.6 (1.6)	82.7 (1.5)			
40	78.2 (1.9)	74.2 (1.8)	73.7 (1.8)			

#Slope is significantly different from zero (Schoenfeld, 1980); *Statistically Significant

Table 4. Cox Proportional Hazards with Respect to Age

Model	Hazard Rate (HR)	95.0% CI for HR Lower	Upper
Cox Model	1.012	1.004	1.019
Time Dependent Cox Model	0.997	0.997	0.998

*Statistically significant

survival. Several studies have been reported the bias in survival estimates when the proportion of LFU was comparatively high (Ganesh, 1995; Mathew, 1996; Kristman et., 2004, Manno and Côté, 2004). In the present study, it was observed that the bias in survival estimates was minimal for early stage BC patients even though the LFU proportion was very high in one data set. However, overestimated values were observed for late stage BC patient survival for the dataset with higher proportion of LFU. Several studies have been reported that the prognosis of early stage (I and II) BC is better than the late stage (III and IV) (SEER). The present study highlighted the importance of obtaining adequate follow-up in survival analysis when the disease prognosis is poor.

In order to reduce the bias in survival estimates due to LFU, Ganesh (1995) computed breast cancer survival using actuarial method by assuming the extreme assumptions for all LFU patients as dead or alive. Mathew (1996) provided another method for estimating loss-adjusted cancer survival using logistic regression model and prognostic factors associated with the disease. Kristman et al. (2004) reported another method using logistic regression model by including a binary exposure and confounders in the cohort and a simulation was done to identify the optimum LFU proportion which gives least bias. Even though several statistical methods have been provided to reduce the bias in survival estimates, it was suggested to obtain a minimum of 80% follow-up for diseases with poor prognosis.

In the present illustration, hazards by stage were found to be proportional during the entire five-year follow-up period and hazards by age were non-proportional using K-M plot. To support this graphical evidence, Schoenfeld global tests also showed significant p-values. Abeysekera and Sooriyarachchi., (2009) compared Schoenfeld global test with classical methods including Cox-Snell residuals and recommended Schoenfeld global test as the most reliable method in validating the PH model.

Li et al., (2015) compared the power of different

alternative test for log-rank. It was showed that for large survival data with more than 60% censoring, Renyi test is one of the best alternatives to log-rank test. In the present study about 70% of the cases were censored. Hence Renyi statistic was useful in estimating the survival in presence of non-PH estimates and higher proportion of censored observations. In the present illustration, it was observed that while PH assumption failed, log-rank test has given a significant p-value and Renyi test showed an insignificant p-value. Hence when survival data are non-PH in nature, appropriate test needs to be employed.

Without checking PH assumption, Cox-model is widely used in survival analysis. Xue et al., (2013) extended the Schoenfeld residual test for testing the PH assumption of Cox model in a case-cohort analysis. Ata and Sozer., (2007) discussed time dependent Cox-model as an alternative to Cox model in presence of non-PH using lung cancer data. In the present study, we obtained hazard ratio for the non-PH factor age higher than one (null value) using Cox-model whereas the same using time-dependent Cox-model was less than one. The hazard ratios were significant in both the models. Hence the evaluation of PH assumption is essential since its violation raises the question regarding the validity of Cox model and could result in erroneous results.

In conclusion, the present illustration showed the importance of adequate follow-up to carry out survival analysis with a specific degree of precision particularly for diseases with poor prognosis and associated risk factors are non-proportional. Appropriate methods in survival analysis are to be employed to obtain accurate estimates and inference and proper checking of assumptions before using a statistical test needs to be employed. For survival data with non-PH factors, Renyi statistic and time dependent Cox model are some alternate methods to obtain unbiased inference and estimates.

Cancer Registry, which is under the network of the National Cancer Registry program of National Centre for Disease Informatics, Indian Council of Medical Research, Government of India. Their support is greatly acknowledged.

References

Abeysekara WWM, Sooriyarachchi MR (2009). Use of Schoenfeld's global test to test the proportional hazards assumption in the Cox proportional hazards model: an application to a clinical study. *J NatlSci Found Sri*, **37**, 41-51.

Altman DG (1992). Analysis of survival times. In: Practical statistics for medical research, Chapman and Hall, London (UK), pp 365–93.

Ata N, Sozer TM (2007). Cox regression models with non-proportional hazards applied to lung cancer survival data. *Hacet J Math Stat*, **36**, 157 - 67.

Borucka J (2013). Extensions of Cox model for non-proportional hazards purpose, PhUSE2013-Paper SP07.

Breslow NE (1975). Analysis of survival data under the proportional hazards model. *Int Stat Rev*, **43**, 45–57.

Cox DR (1972). Regression models and life-tables. *J R Stat Soc Series B*, **34**, 187–20.

Davis M, Xie SX (2011). Caution: Hazards crossing! Using the Renyi Test statistic in survival analysis. Pharma AUG2011-Paper SP06.

Fisher LD, Lin DY (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health*, **20**, 145-57.

Fleming TR, O'Fallon JR, O'Brien PC, Harrington DP (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right–censored data. *Biometrics*, **36**, 607-25.

Ganesh B (1995). Effect of loss to follow-up in estimating survival rates. Acta UniversitatisTamperensis, Series A, Vol. 440, University of Tampere, Tampere.

Gill RD (1980). Censoring and stochastic integrals. *Stat Neerl*, **34**, DOI: 10.1111/j.1467-9574.1980.tb00692.x.

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. *J Am Stat Assoc*, **53**, 457– 81.

Klein JP, Moeschberger ML (2003). Survival analysis: Techniques for censored and truncated data. Second Edition, Springer-Verlang, New York.

Kleinbaum DG, Klein M (2005). Survival Analysis: A self-learning text, Springer-Verlang, New York, pp 55-258.

Kristman V, Michael M, Pierre Côté (2004). Loss to follow-up in cohort studies: How much is too much. *Eur J Epidemiol*, **19**, 751-60.

Li H, Han D, Hou Y, Chen H, Chen Z (2015). Statistical inference methods for two crossing survival curves: A comparison of methods. *PLoS One*, **10**, e0116774.

Lin X, Wang H (2004). A new testing approach for comparing the overall homogeneity of survival curves. *Biom J*, **46**, 489-96.

Lin X, Xu Q (2010). A new method for the comparison of survival distributions. *Pharm Stat*, **9**, 67–76.

Liu K, Qiu P, Sheng J (2007). Comparing two crossing hazard rates by Cox proportional hazards modelling. *Stat Med*, **6**, 375–91.

Mantel N (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, **50**, 163–70.

Mathew A (1996). Removing bias in cancer survival estimates by active follow-up and information on determinants of loss to follow-up, ActaUniversitatisTamperensis, Series A, Vol. 525, University of Tampere, Tampere, pp 41-2.

Qiu P, Sheng J (2008). A two-stage procedure for comparing hazard rate functions. *J R Stat Soc Series B*, **70**, 191-08.

Renyi A (1953). On the theory of order statistics. *Acta Math Hungar*, **4**, 191- 31.

Schoenfeld D (1980). Chi-Squared goodness of fit tests for the proportional hazards regression model. *Biometrika*, **67**, 145-53.

Sobin L, Gospodarowicz M, Wittekind C (2009). TNM Classification of malignant tumors. 7th ed. Hoboken, NJ: John Wiley and Sons, Inc, UK, pp 181-93.

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch.

Taib NAM, Yip CH, Mohamed I (2008). Survival analysis of Malaysian women with breast cancer: results from the University of Malaya Medical Centre. *Asian Pac J Cancer Prev*, **9**, 197- 02.

Wei LJ (1984). Testing goodness-of-fit for proportional hazards model with censored observations. *J Am Stat Assoc*, **79**, 649-52.

Xue X, Xie X, Gunter M, et al (2013). Testing the proportional hazards assumption in case-cohort analysis. *BMC Med Res Methodol*, **13**, 88.